

Monitoreo remoto de cambios en el vigor de la vegetación y su relación con variables ambientales

Raúl Octavio Martínez-Rincón, César Augusto Salinas-Zavala, María Verónica Morales-Zárate y Federico Salinas-Zavala

Comparación de métodos de detección de cambio forestal en el marco del proyecto *Forests 2020*

Miguel Muñoz, Miguel Castillo, Felicia Line, Edith Mondragón, Oswaldo Carrillo, Richard Tipper y Catherine Navarro

La informalidad laboral en las entidades de México en el siglo XXI: posibles factores explicativos

Mario Camberos Castro y Joaquín Bracamontes Nevárez

Predicción automática del nivel educativo en usuarios de *Twitter* en México

Juan Carlos Gomez, Luis Miguel López Santamaría, Mario Alberto Ibarra Manzano y Dora Luz Almanza Ojeda

The Dynamics of Disability due to Mobility Limitations and Ramp-Access Coverage

Tonatiuh Suárez-Meaney, Israel Albarrán Martínez, Verónica Pérez Barrón, Ioannis Chatziioannou, Alejandro Martínez López y Luis Chías Becerril

Estimación del subregistro de las tasas de mortalidad infantil en México, 1990-2013

Lina Sofía Palacio-Mejía, Diana Molina-Vélez, Concepción García-Morales, Amado David Quezada-Sánchez, Leonel González-González, Maylen Rojas-Botero y Juan Eugenio Hernández-Ávila

Inclusión del calendario mexicano en el proceso de desestacionalización de series de tiempo: el caso del Indicador Global de la Actividad Económica

Francisco de Jesús Corona Villavicencio y Jesús López-Pérez

Selección de transformación potencia en los indicadores trimestrales de actividad económica estatal: implicaciones en el análisis de factores

Francisco de Jesús Corona Villavicencio y Jesús López-Pérez

Impacto de la crisis económica del 2008 en el empleo y salarios de las industrias manufacturera y automotriz de la región sureste de Coahuila de Zaragoza

Reyna Elizabeth Rodríguez Pérez, Liliana Meza González y David Castro Lugo



PERSONAS CON DISCAPACIDAD, LIMITACIÓN, O CON ALGÚN PROBLEMA O CONDICIÓN MENTAL



Una de cada **seis personas** tiene alguna discapacidad, limitación, problema o condición mental.

Eso significa **20.8 millones de habitantes.**

EDAD MEDIANA

Nacional:
29 años.

Hombres: **28 años.** Mujeres: **30 años.**



ENTIDADES CON MÁS ADULTOS(A) MAYORES EN MÉXICO

Ciudad de México:	11.1 %
Veracruz de I. de la Ll.:	10.0 %
Morelos:	9.6 %
Oaxaca:	9.6 %
San Luis Potosí:	9.0 %



El promedio de ocupantes por vivienda en México es de **3.6** personas.

Conociendo México

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

 **INEGI Informa**



Nota: la edad mediana divide la distribución por edades de una población determinada en dos grupos numéricamente iguales; la mitad de los casos queda por abajo de la mediana y la otra mitad, por encima; para su cálculo se excluye a la población con edad no especificada.

Fuente: INEGI. Censo de Población y Vivienda 2020. Cuestionario básico.

Contenido

Monitoreo remoto de cambios en el vigor de la vegetación y su relación con variables ambientales	5
<i>Remote Monitoring of Vegetation Vigor Change and its Relationship with Environmental Variable</i>	
Raúl Octavio Martínez-Rincón, César Augusto Salinas-Zavala, María Verónica Morales-Zárate y Federico Salinas-Zavala	
Comparación de métodos de detección de cambio forestal en el marco del proyecto <i>Forests 2020</i>	16
<i>Comparison of Forest Change Detection Methods in the Framework of the Forests 2020 Project</i>	
Miguel Muñoz, Miguel Castillo, Felicia Line, Edith Mondragón, Oswaldo Carrillo, Richard Tipper y Catherine Navarro	
La informalidad laboral en las entidades de México en el siglo XXI: posibles factores explicativos	30
<i>Labor Informality in Mexican Entities during the 21st Century: Possible Explanatory Factors</i>	
Mario Camberos Castro y Joaquín Bracamontes Nevárez	
Predicción automática del nivel educativo en usuarios de <i>Twitter</i> en México	48
<i>Automatic Prediction of the Educational Level of Twitter Users in Mexico</i>	
Juan Carlos Gomez, Luis Miguel López Santamaría, Mario Alberto Ibarra Manzano y Dora Luz Almanza Ojeda	
The Dynamics of Disability due to Mobility Limitations and Ramp-Access Coverage	62
<i>La dinámica de la discapacidad debida a las limitaciones de la movilidad y cobertura de rampas de accesibilidad</i>	
Tonatiuh Suárez-Meaney, Israel Albarrán Martínez, Verónica Pérez Barrón, Ioannis Chatzioannou, Alejandro Martínez López y Luis Chias Becerril	
Estimación del subregistro de las tasas de mortalidad infantil en México, 1990-2013	80
<i>Estimation of Underreporting in Infant Mortality Rates in Mexico, 1990-2013</i>	
Lina Sofía Palacio-Mejía, Diana Molina-Vélez, Concepción García-Morales, Amado David Quezada-Sánchez, Leonel González-González, Maylen Rojas-Botero y Juan Eugenio Hernández-Ávila	
Inclusión del calendario mexicano en el proceso de desestacionalización de series de tiempo: el caso del Indicador Global de la Actividad Económica	94
<i>Inclusion of the Mexican Calendar in the Process of Seasonal Adjustment of Time Series: the Case of the Global Indicator of Economic Activity</i>	
Francisco de Jesús Corona Villavicencio y Jesús López-Pérez	
Selección de transformación potencia en los indicadores trimestrales de actividad económica estatal: implicaciones en el análisis de factores	108
<i>Power Transformation Selection in Quarterly Indicators of State Economic Activity: Implications for Factor Analysis</i>	
Francisco de Jesús Corona Villavicencio y Jesús López-Pérez	
Impacto de la crisis económica del 2008 en el empleo y salarios de las industrias manufacturera y automotriz de la región sureste de Coahuila de Zaragoza	124
<i>Impact of the 2008 Economic Crisis on Employment and Wages in the Manufacturing and Automotive Industries in the Southeast Region of Coahuila de Zaragoza</i>	
Reyna Elizabeth Rodríguez Pérez, Liliana Meza González y David Castro Lugo	
Colaboran en este número	149

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Presidente del Instituto

Julio Alfonso Santaella Castell

Vicepresidentes

Enrique de Alba Guerra

Paloma Merodio Gómez

Graciela Márquez Colín

Adrián Franco Barrios

Dirección General de Estadísticas Sociodemográficas

Edgar Vielma Orozco

Dirección General de Estadísticas de Gobierno, Seguridad Pública y Justicia

Óscar Jaimes Bello

Dirección General de Estadísticas Económicas

José Arturo Blancas Espejo

Dirección General de Geografía y Medio Ambiente

María del Carmen Reyes Guerrero

Dirección General de Integración, Análisis e Investigación

Sergio Carrera Riva Palacio

Dirección General de Coordinación del Sistema Nacional de Información Estadística y Geográfica

María Isabel Monterrubio Gómez

Dirección General de Comunicación, Servicio Público de Información y Relaciones Institucionales

Eduardo Javier Gracida Campos

Dirección General de Administración

Luis María Zapata Ferrer

Contraloría Interna

Manuel Rodríguez Murillo

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Editor responsable

Sergio Carrera Riva Palacio

Editor técnico

Gerardo Leyva Parra

Coordinación editorial

Virginia Abrin Batule y Mercedes Pedrosa Islas

Corrección de estilo

José Pablo Covarrubias Ordiales y Laura Elena López Ortiz

Corrección de textos en inglés

Gerardo Piña

Diseño y formación edición impresa

Juan Carlos Martínez Méndez y Eduardo Javier Ramírez Espino

Indizada en: Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal *Latindex Catálogo*; Citas Latinoamericanas en Ciencias Sociales y Humanidades (*CLASE*) y en la Red Iberoamericana de Innovación y Conocimiento (*REDIB*).

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, Vol. 12, Núm. 1, enero-abril, 2021, es una publicación cuatrimestral editada por el Instituto Nacional de Estadística y Geografía, 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, México. Teléfono 55 52781069. Toda correspondencia deberá dirigirse al correo: rde@inegi.org.mx

Editor responsable: Sergio Carrera Riva Palacio. Reserva de Derechos al Uso Exclusivo del Título Núm. 04-2012-121909394300-102, ISSN Núm. 2007-2961, ambos otorgados por el Instituto Nacional del Derecho de Autor. Certificado de Licitud de Título y Contenido Núm. 15099, otorgado por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación. Domicilio de la publicación, imprenta y distribución: 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, México.

El contenido de los artículos, así como sus títulos y, en su caso, fotografías y gráficos utilizados son responsabilidad del autor, lo cual no refleja necesariamente el criterio editorial institucional. Asimismo, la Revista se reserva el derecho de modificar los títulos de los artículos, previo acuerdo con los autores. La mención de empresas o productos específicos en las páginas de la Revista no implica el respaldo por el Instituto Nacional de Estadística y Geografía.

Se permite la reproducción total o parcial del material incluido en la Revista, sujeto a citar la fuente.

Versión electrónica: <http://rde.inegi.org.mx>

ISSN 2395-8537



Offline Work

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

CONSEJO EDITORIAL

Dr. Enrique de Alba Guerra

Presidente del Consejo

Mtra. Claudia Aburto Rancaño

Instituto Tecnológico Autónomo de México
México

Dr. Clemente Ruiz Durán

Universidad Nacional Autónoma de México
México

Dr. Emilio Cunjamá López

Instituto Nacional de Ciencias Penales
México

Dr. Fernando Cortés Cáceres

Profesor emérito de FLACSO PUEDE de la UNAM
México

Dra. Graciela Teruel Belismelis

Universidad Iberoamericana
Ciudad de México
México

Dra. Landy Sánchez Peña

El Colegio de México
México

Dra. María Martha Téllez Rojo Solís

Instituto Nacional de Salud Pública
México

Dr. Víctor Manuel Guerrero Guzmán

Instituto Tecnológico Autónomo de México
México

Editorial

El presente es el número 33, y representa un poco más de 10 años de publicar ininterrumpidamente *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*; esto, gracias a un equipo formado por autores(as), dictaminadores(as), correctores(as) de estilo, diseñadores(as), revisores(as), impresores(as), entre otros(as) más que, de manera comprometida, han hecho posible la permanencia de la *Revista*; a todos ellos(as), nuestro reconocimiento. Durante este tiempo, hemos fortalecido su contenido y cumplido con su propósito inicial de servir de enlace entre la generación de la información estadística y geográfica oficial y la investigación académica para compartir el conocimiento entre especialistas e instituciones con propósitos similares.

A continuación, se describe el contenido de los artículos que forman parte de este número.

Monitoreo remoto de cambios en el vigor de la vegetación y su relación con variables ambientales (Remote Monitoring of Vegetation Vigor Change and its Relationship with Environmental Variable) evalúa el desempeño de una estación de monitoreo que detecta modificaciones en el estado de salud de las plantas y el efecto que tienen factores medioambientales sobre estas. Para ello, se implementó una base de estudio de bajo costo que integra la medición de algunos de estos mediante un sensor radiométrico dentro de la reserva ecológica Laura Arriaga Cabrera. Lo novedoso de este trabajo es el uso de un instrumental que permite correlacionar su impacto en la flora.

Comparación de métodos de detección de cambio forestal en el marco del proyecto Forests 2020 (Comparison of Forest Change Detection Methods in the Framework of the Forests 2020 Project) realiza un análisis comparativo entre distintas técnicas con la finalidad de detectar modificaciones en la cobertura vegetal en una región de la Selva Lacandona en Chiapas.

El objetivo del artículo *La informalidad laboral en las entidades de México en el siglo XXI: posibles factores explicativos (Labor Informality in Mexican Entities during the 21st Century: Possible Explanatory Factors)* es identificar factores de dicha condición de trabajo; para ello, se aplicó un modelo econométrico de mínimos cuadrados ordinario multivariado en tres etapas para probar la relación entre esta y las variables de mercado, institucionales y de otro tipo utilizando los datos de la Encuesta Nacional de Ocupación y Empleo (ediciones 2015-2019) e información proveniente de otras fuentes.

El trabajo *Predicción automática del nivel educativo en usuarios de Twitter en México (Automatic Prediction of the Educational Level of Twitter Users in Mexico)* presenta el extracto de una serie de características del contenido textual de los tuits publicados que se utilizaron para construir modelos basados en aprendizaje automático, los cuales predicen si un usuario tiene estudios universitarios o no.

The Dynamics of Disability due to Mobility Limitations and Ramp-Access Coverage (La dinámica de la discapacidad debida a las limitaciones de la movilidad y cobertura de rampas de accesibilidad) tiene como finalidad reconocer si la oferta de esta infraestructura crece de forma proporcional a la población que la requiere; para ello, se compara su crecimiento por motivos de movilidad en relación con el de las rampas en términos de cobertura dentro de zonas urbanas entre 2010 y 2014 en México.

Estimación del subregistro de las tasas de mortalidad infantil en México, 1990-2013 (Estimation of Underreporting in Infant Mortality Rates in Mexico, 1990-2013) tiene como objetivo compararlas a través de las estadísticas vitales con las estimadas a partir de la Encuesta Nacional de la Dinámica Demográfica 2014, según sexo y ámbitos urbano o rural, para calcular el nivel de subregistro que hay, cuyos resultados permiten concluir que este ha disminuido, pero todavía se considera alto para el país, sobre todo en áreas rurales.

En el trabajo *Inclusión del calendario mexicano en el proceso de desestacionalización de series de tiempo: el caso del Indicador Global de la Actividad Económica (Inclusion of the Mexican Calendar in the Process of Seasonal Adjustment of Time Series: the Case of the Global Indicator of Economic Activity)* se presenta una alternativa para usar, en la etapa de ajustes previos del paquete *X-13ARIMA-SEATS*, fechas movibles (como el Día del Trabajo, la Independencia de México y la interacción que existe entre Semana Santa con el natalicio de Benito Juárez), y se recomienda al INEGI evaluar su utilización para el desarrollo de sus indicadores más relevantes.

En *Selección de transformación potencia en los indicadores trimestrales de actividad económica estatal: implicaciones en el análisis de factores (Power Transformation Selection in Quarterly Indicators of State Economic Activity: Implications for Factor Analysis)* se presenta un análisis empírico de las implicaciones que tiene la elección de esta metodología en el proceso de ajuste estacional para el Indicador Trimestral de la Actividad Económica Estatal (ITAEE).

Impacto de la crisis económica del 2008 en el empleo y salarios de las industrias manufacturera y automotriz de la región sureste de Coahuila de Zaragoza (Impact of the 2008 Economic Crisis on Employment and Wages in the Manufacturing and Automotive Industries in the Southeast Region of Coahuila de Zaragoza) tiene como objetivo analizar las consecuencias de esta perturbación en empresas altamente integradas con la economía mundial, como es el caso de la automotriz por la fuerte integración de esta actividad de la zona metropolitana de Saltillo con el mercado de Estados Unidos de América, donde las remuneraciones (sin importar su nivel de calificación) de los trabajadores de este subsector resultaron más afectadas que las de los del sector manufacturero en general de la misma área.

<http://rde.inegi.org.mx>

Monitoreo remoto de cambios en el **vigor de la vegetación** y su relación con variables ambientales

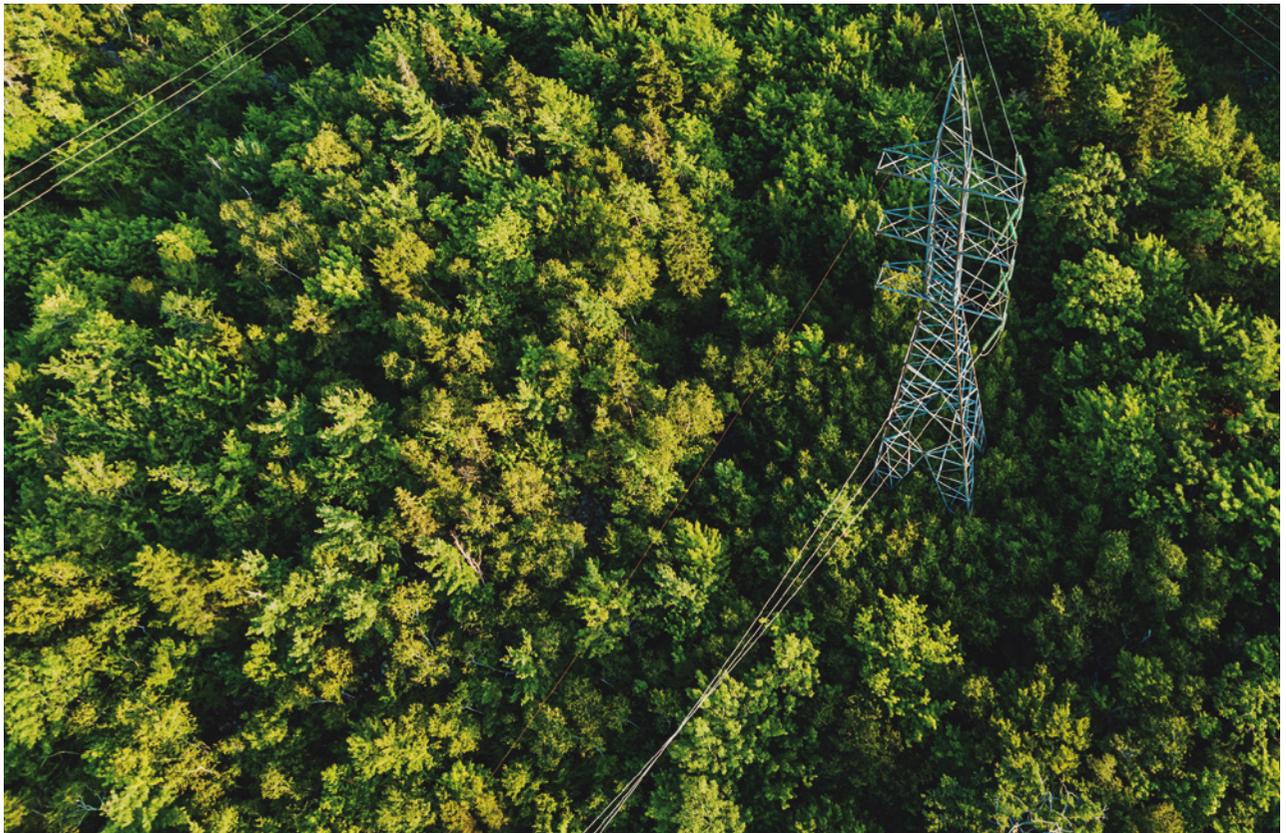
*Remote Monitoring of **Vegetation Vigor** Change and its Relationship with Environmental Variable*

Raúl Octavio Martínez-Rincón,* César Augusto Salinas-Zavala,
María Verónica Morales-Zárate** y Federico Salinas-Zavala****

* CONACYT-Centro de Investigaciones Biológicas del Noroeste (CIBNOR), SC, rrincon@cibnor.mx

** CIBNOR, SC; csalinas@cibnor.mx, mzarate04@cibnor.mx y fsalinas04@cibnor.mx, respectivamente.

Nota: este trabajo fue financiado por el fondo sectorial CONACYT-INEGI (proyecto: 278789).



Aerial Drone View: Power Pylon/Shami/iStock

El monitoreo de los cambios en la flora se realiza de manera tradicional con visitas a campo de forma periódica o en tiempos específicos; sin embargo, este método requiere de un gran esfuerzo de muestreo y usualmente se limita en tiempo y espacio. Como alternativa, los sensores remotos han permitido inferir modificaciones en la misma de manera sistemática en diferentes resoluciones tanto temporales como espaciales; no obstante, a pesar de que es una herramienta muy utilizada para ello, presenta ciertas limitaciones, ya que solo se pueden describir algunos aspectos generales, como el vigor de la vegetación, su variación estacional y cambios espacio-temporales. En el presente trabajo se describen las modificaciones en el vigor de la vegetación y su relación con las variables ambientales de manera remota. Para esto, se implementó una estación de monitoreo de bajo costo que integra la medición de algunas de las variables con un sensor radiométrico dentro de la reserva ecológica Laura Arriaga Cabrera.

Palabras clave: estación de monitoreo remoto; índice de vegetación; correlación cruzada.

Recibido: 3 de diciembre de 2019.
Aceptado: 24 de junio de 2020.

Introducción

El monitoreo del vigor de la vegetación (estado de salud) se ha realizado en las últimas décadas de forma tradicional por medio de la observación *in situ* o de la percepción remota (imágenes satelitales o aéreas). La primera permite evaluar de manera detallada los cambios estacionales de las plantas (fenología), como floración, reproducción y pérdida de área foliar, entre otros; sin embargo, este método es costoso y tiene una cobertura espacio-temporal muy limitada, impidiendo, así, evaluar los efectos climáticos de eventos de largo plazo en la fenología de la vegetación de un área en particu-

Monitoring vegetation changes is traditionally done by periodic field survey on specific times. However, this method requires a great deal of sampling effort and is usually time-space limited. Alternatively, remote sensing has allowed to infer vegetation changes systematically at different time and space resolutions. Although it is a widely used tool for this purpose, it presents certain limitations, since only some general aspects can be described, such as vegetation's health, its seasonal variability and spatiotemporal changes. In this work, modifications in vegetation's health variability and its relationship with environmental factors are described remotely. For this, a low-cost monitoring station was implemented, which integrates the measurement of some of the variables with a radiometric sensor within the Laura Arriaga Cabrera Ecological Reserve.

Key words: Remote Monitoring Station; Vegetation Index; Cross Correlation.

lar. La segunda los evalúa en una escala espacial y temporal mucho mayor y a menor costo debido a que existen imágenes satelitales disponibles para el público en general de manera gratuita; no obstante, en la escala fina, este método no permite distinguir las modificaciones estructurales o fisiológicas de las plantas, además de tener poca capacidad para identificar patrones contrastantes entre diferentes especies de ecosistemas silvestres; otra limitación radica en que los índices de vegetación que se infieren a partir de las imágenes de satélite necesitan un proceso de supervisión, ya que pueden carecer de interpretabilidad si no se conocen las características de los tipos de vegetación del área que se está analizando.

NDVI

En ecosistemas terrestres, un método indirecto utilizado para describir la fenología de las plantas es a través de la percepción remota mediante el uso de índices de vegetación (Chen *et al.*, 2000), los cuales se emplean para medir la densidad relativa y la salud de la flora; en la actualidad, existen varios: uno de los más utilizados alrededor del mundo es el Índice de Vegetación de Diferencia Normalizada (NDVI, por sus siglas en inglés), que se calcula a partir de la reflectancia de la luz en las longitudes de onda del color rojo y el infrarrojo cercano.

$$NDVI = \frac{IRC - Rojo}{IRC + Rojo} \quad (1)$$

donde:

IRC = reflectancia en la longitud de onda del infrarrojo cercano (750-900 nm).

Rojo = reflectancia de la luz en la longitud de onda del rojo (620-750 nm).

Mediante el uso de datos del Índice es posible describir la fenología de la vegetación. A partir de sus valores, se pueden obtener parámetros biométricos que reflejan cambios fenológicos de ella; por ejemplo, el valor máximo del NDVI indica la época del año de mayor desarrollo foliar de las plantas y viceversa. Con la amplitud del Índice (diferencia entre los mínimos y máximos) se puede inferir el inicio de la estación de crecimiento —que es la época del año donde el NDVI se incrementa de manera sostenida— y el fin de la misma —cuando se manifiesta un decremento de forma sostenida del Índice— (Vázquez, Adema y Fernández, 2013).

Por otra parte, el NDVI ha sido muy utilizado para evaluar cambios espacio-temporales en la cobertura vegetal de diferentes ecosistemas alrededor del mundo (Jiang *et al.*, 2015; Li *et al.*, 2015; Vandendorj, Gantsetseg y Boldgiv, 2015). En investigaciones recientes, algunos autores (Meng, Ni y Zong, 2011; Revadekar, Tiwari y Kumar, 2012; Wang

et al., 2015) han relacionado las modificaciones espacio-temporales del Índice al efecto del cambio climático y/o al de las actividades humanas; sin embargo, otros investigadores (Sohoulande Djebou, Singh y Frauenfeld, 2015; Wang *et al.*, 2014) mencionan que modificaciones en la precipitación y la temperatura atmosférica tienen una relación más estrecha con los valores registrados del NDVI. El vínculo entre la lluvia y la cobertura vegetal es más estrecho aún en ecosistemas áridos debido a que las precipitaciones estacionales son la principal fuente de agua en estos (Zhu *et al.*, 2015).

Jiang *et al.* (2014) utilizaron los valores del NDVI para caracterizar la cobertura vegetal en una región de China; en adición, usaron herramientas estadísticas (como modelos de regresión) para describir la tendencia espacial de este índice y, con ello, describir cuáles son los tipos de vegetación que experimentaron cambios negativos o positivos en la última década. De manera similar, Salinas-Zavala *et al.* (2017) emplearon el Índice para describir la tendencia de la cobertura vegetal en Baja California Sur, sugiriendo que en la parte sur de esa entidad se presentan valores negativos de tendencia del NDVI debido al efecto negativo causado por el crecimiento urbano en esta región.

Efecto de la variabilidad ambiental

El que impacta en las plantas ha sido muy estudiado y descrito en diferentes ecosistemas terrestres (Yepes y Silveira Buckeridge, 2011); las investigaciones abordan el uso de percepción remota para describir la respuesta del vigor de la vegetación y variables ambientales y han demostrado que la precipitación y la temperatura tienen un fuerte efecto en la producción de flora durante su fase de crecimiento (Wang, Rich y Price, 2003); sin embargo, de acuerdo con Zhao *et al.* (2015), en ecosistemas semiáridos, la temperatura parece no tener una relación tan determinante como la tiene la precipitación o la circulación atmosférica asociada a dicha producción (Salinas-Zavala, Douglas y Diaz, 2002). Por el contrario, otros autores (Deng, Su y Liu, 2007) sugieren que, en bosques subtropi-

cales, la radiación fotosintéticamente activa (PAR, por sus siglas en inglés) y la temperatura tienen un mayor efecto en la producción vegetal en este tipo de ecosistemas debido a que la precipitación no es una limitante.

Por todo lo anterior, se considera que el monitoreo de la vegetación debe realizarse en conjunto con el ambiental para poder explicar algunos de los factores más importantes que promueven cambios en el vigor de esta de algún sitio de interés. Así, el objetivo principal de este trabajo es evaluar el desempeño de una estación de monitoreo remota para valorar modificaciones en el estado de salud de las plantas y el efecto que tienen algunas variables ambientales sobre estas.

Metodología

Área de estudio

La zona de interés se localiza en la reserva ecológica Laura Arriaga Cabrera del Centro de Investigaciones Biológicas del Noroeste (CIBNOR), SC. En esta, la vegetación es de tipo matorral sarcocaulé, la cual se caracteriza por la dominancia de plantas con troncos grandes y carnosos, como el ciruelo (*Cyrtocarpa edulis* Brandegees), el lomboy (*Jatropha cinérea* Ortega), la matacora (*J. cuneata* Wiggins &

Rollins), el mezquite (*Prosopis articulata* S. Watson), el palo adán (*Fouquieria diguetii* Tiegh) y el torote (*Bursera microphylla* A. Gray), además de cactáceas, como el cardón (*Pachycereus pringlei* S. Watson) y la pitaya agria (*Stenocereus gummosus* Engelm) y la dulce (*S. thurberi* Engelm).

Equipos y recolecta de datos

El registrador electrónico es el aparato que se encarga de anotar las lecturas de los sensores montados en la estación de monitoreo remoto y fue programado para recolectar datos con una frecuencia de 15 minutos; con esta configuración, puede almacenar hasta un año de registros de cada variable (ver cuadro 1). Cuenta con panel solar para carga de las baterías y con tarjeta SIM para el envío de información. Las variables que capta este equipo son: dirección del viento (en grados), velocidad del viento ($m s^{-1}$), precipitación (mm), temperatura del aire ($^{\circ}C$), humedad relativa (proporción) y PAR ($\mu mol m^{-2} s^{-1}$), así como reflectancias en las bandas del rojo (630 nm) y del infrarrojo cercano (800 nm); a partir de los datos de estas dos últimas se calcula el NDVI con la ecuación 1.

Los sensores fueron colocados en una estructura tipo antena a diferentes alturas de la superficie del suelo: los de temperatura, humedad, viento y PAR,

Cuadro 1

Continúa

Equipos montados en la estación de monitoreo remoto ubicado en la reserva Laura Arriaga Cabrera

Nombre	Marca/modelo	Parámetro/unidades	Precio aproximado (dólares estadounidenses)
Registrador electrónico	Decagon/EM50G	No aplica	1 980.00
Sensor de reflectancia espectral	Decagon/SRS-Nr	<ul style="list-style-type: none"> Reflectancia en la banda del rojo (630 nm)/Watts por metro cuadrado ($W m^{-2}$). Reflectancia en la banda del infrarrojo cercano (800 nm)/Watts por metro cuadrado ($W m^{-2}$). NDVI/sin unidades. 	535.00
Sensor de presión de vapor, humedad y temperatura	Decagon/VP3	<ul style="list-style-type: none"> Humedad relativa/porcentaje (%). Temperatura/grados centígrados ($^{\circ}C$). 	470.00

Equipos montados en la estación de monitoreo remoto ubicado en la reserva Laura Arriaga Cabrera

Nombre	Marca/modelo	Parámetro/unidades	Precio aproximado (dólares estadounidenses)
Anemómetro	Decagon/Copas David	· Dirección del viento/grados (°). · Velocidad de viento/metros por segundo ($m s^{-1}$). · Rachas de viento/metros por segundo ($m s^{-1}$).	300.00
Sensor PAR	Decagon/QSO-S	· PAR/micromoles por metro cuadrado por segundo ($\mu mol m^{-2} s^{-1}$).	420.00
Pluviómetro	Decagon/ECRN-50	· Precipitación/milímetros.	300.00

a 5 m; el pluviómetro y el registrador de datos, a 1 m y el de reflectancia espectral, a 7 m con una inclinación de 75° con respecto a la superficie de la zona de estudio. Este sensor tiene un diámetro de muestreo de 52.2 m, equivalente a un área de 2 141 metros cuadrados.

Todos los datos pueden ser recolectados manualmente en cualquier momento mediante el uso de un puerto de comunicación y un *software* especializado (ECH₂O *utility*); sin embargo, el registrador EM50G cuenta con una antena para el envío de información por medio de telefonía celular, lo que permite acceder a ella de manera remota.

Para este trabajo, se programaron cinco horarios diferentes (6:00, 9:00, 12:00, 15:00 y 18:00 horas) para el envío remoto de datos diarios a un servidor (https://zentracloud.com/#/dashboard_map), en el cual se puede visualizar y descargar la información de los últimos 15 días. Gracias a que se puede registrar más de una estación de monitoreo, el sistema permite la posibilidad de crear una red de monitoreo en la que todos los usuarios acceden a los datos de todas las estaciones en el momento que lo deseen, si cuentan con conexión a internet.

Si bien para la presente investigación se utilizó la información de 18 meses, la estación de monitoreo

ha funcionado ininterrumpidamente desde su colocación en diciembre del 2017, censando de forma eficiente durante todo el periodo. En este sentido, es importante mencionar que registró datos a la intemperie de manera continua sin sufrir ningún desperfecto o desgaste evidente; los equipos se revisaron de manera periódica y se les dio mantenimiento en seis ocasiones durante el periodo de estudio, el cual consistió en el reemplazo de los sujetadores plásticos de algunos sensores (debido a que se vuelven frágiles por el efecto de la radiación solar). Cabe hacer mención que el costo asociado a este procedimiento de mantenimiento está directamente relacionado con la distancia a la cual se encuentren los sensores y el observador, y que el precio total de la estación es de alrededor de 4 mil dólares estadounidenses.

Análisis de información

Los datos ambientales y del Índice analizados en este trabajo corresponden al periodo de diciembre del 2017 a junio del 2019; son valores promedio por hora de cada día, con excepción de la precipitación, a la cual se le calculó la suma por hora y fecha. Los de humedad relativa fueron transformados a porcentajes, multiplicando el valor original (en proporción) por 100. Debido a la gran variabilidad de la información y para describir los patrones

temporales de las variables analizadas, se aplicaron promedios móviles con una ventana de 15 días, es decir, se utilizaron todos los promedios por hora registrados en esos días.

A partir de los valores promedio por hora, se calcularon los coeficientes de correlación de Pearson mediante la ecuación 2 con la finalidad de describir la relación lineal entre las variables ambientales y el Índice:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2) \sum (y_i - \bar{y})^2}} \quad (2)$$

donde:

r_{xy} = coeficiente de correlación de Pearson.

x_i y_i = observaciones de las variables.

\bar{x} \bar{y} = promedios de las variables.

Se aplicaron análisis de componentes principales para reducir dimensiones (capturar la señal común en solo dos ejes) y para describir las correlaciones de todas las variables ambientales mediante el uso de los *eigen*-vectores de los componentes 1 y 2.

Resultados y discusión

Variabilidad temporal

En el cuadro 2 se muestra el resumen estadístico (mínimo, promedio y máximo) de los valores registrados de cada uno de los parámetros en el periodo de estudio. Como se aprecia, los datos presentan una alta variabilidad, con excepción de la precipitación, cuyos valores se mantienen en cero en la mayoría de los registros.

En las figuras 1 y 2 se muestran los valores promedio por hora y día de cada una de las variables ambientales captadas durante el periodo de estudio, con excepción de la precipitación, a la cual se le calculó la suma por hora y fecha. Como se aprecia en la figura 1A, el sensor de esta registró valores de 0 mm la mayor parte del periodo; sin embargo, se observa que en septiembre captó los más altos y que entre diciembre del 2018 y enero del 2019 se registraron eventos de lluvia. Este comportamiento coincide con lo reportado por Ayala Niño (2018), quien describe que la precipitación registrada en el sitio durante 2006-2007 es mayor en septiembre y, de manera ocasional, se registran lluvias en diciembre y enero; también reporta que, de febrero a junio, la precipitación en la zona es casi nula. Estos resultados con-

Cuadro 2

Resumen estadístico de las variables ambientales registradas con la estación de monitoreo remoto ubicado en la reserva Laura Arriaga Cabrera

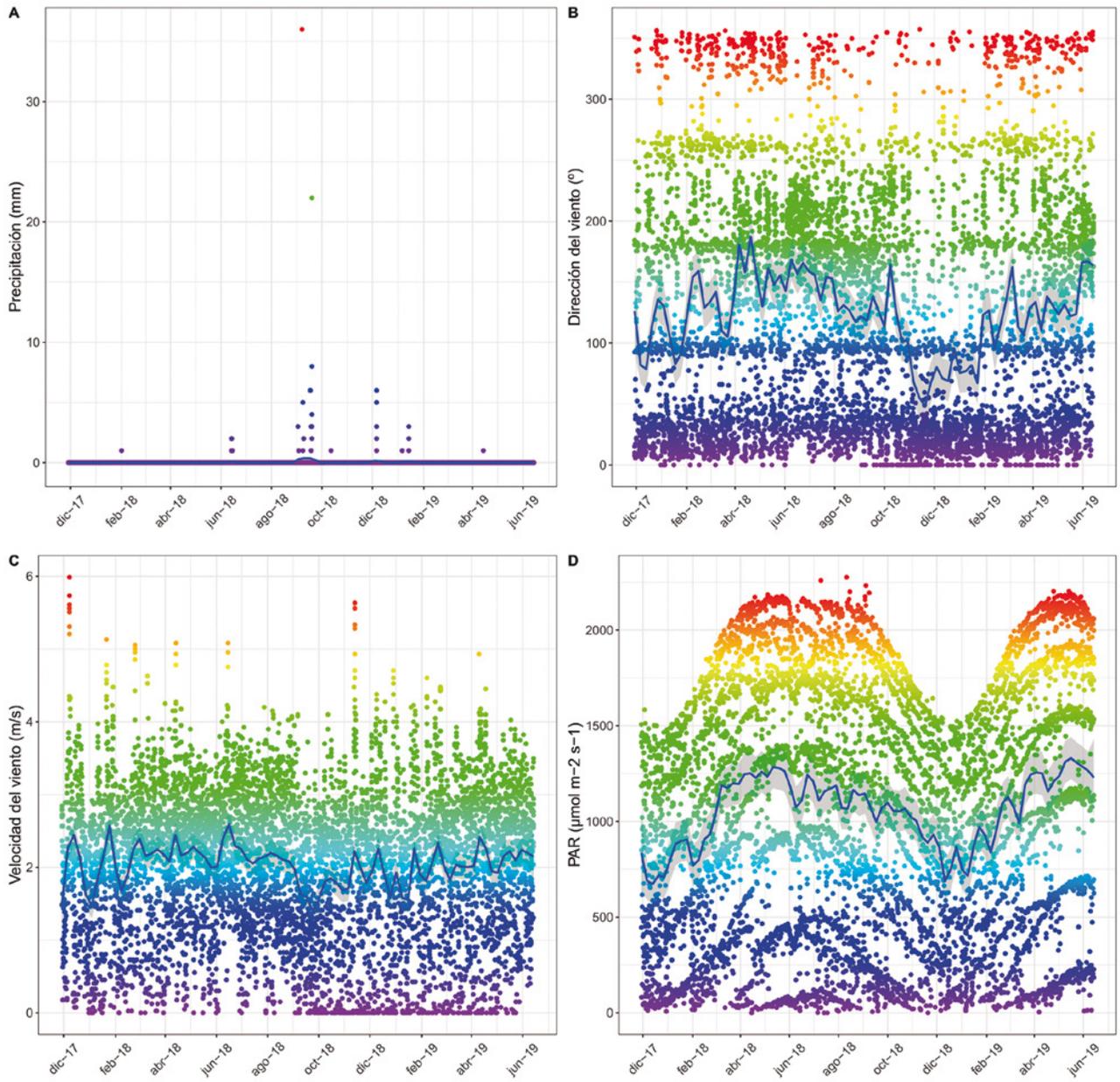
Variable ambiental	Mínimo	Promedio	Máximo	Fecha mín.	Fecha máx.
Humedad relativa (%)	18.4	52.5	88.2	12/04/19	20/09/18
Temperatura (°C)	7.8	26.8	41.2	31/12/18	23/07/18
Precipitación (mm)	0.0	0.0	36.0	*	07/09/18
Dirección del viento (°)	0.0	125.3	357.3	*	13/11/18
Velocidad del viento (m s ⁻¹)	0.0	2.0	6.0	*	08/12/17
PAR (μmol m ⁻² s ⁻¹)	0.0	1 068.1	2 276.0	*	11/08/18
NDVI	0.1	0.5	0.8	18/07/18	22/12/18

Fecha mín. = fecha con la observación mínima.

Fecha máx. = fecha con la observación máxima.

* Varias fechas presentan valores cero, por lo tanto, no se especificó ninguna como la mínima.

Valores promedio por hora de las variables



Nota: la línea azul representa un promedio móvil de aproximadamente 15 días; el área en gris, su intervalo de confianza a 95 por ciento.

firman la importancia de las lluvias estacionales en el área, que se registran sobre todo en septiembre.

La dirección del viento (ver figura 1B) muestra que los valores dominantes de octubre a mayo son de 0 a 50 °, es decir, vientos del norte y noroeste, mientras que, de junio a septiembre, los valores dominantes son de 150 a 250 °, correspondientes a las direcciones suroeste, sur y sureste. Su velocidad

(ver figura 1C) es mayor ($> 5 \text{ m s}^{-1}$) en diciembre y febrero, mientras que en septiembre se registran los menores valores. Al respecto, Herrera-Cervantes (2019) demostró que los vientos más fuertes de la región se presentan en el invierno y que provienen del noroeste.

PAR es una variable ambiental que se utiliza para estimar la cantidad de radiación disponible para fo-

tosíntesis; sin embargo, también ha sido utilizada para la creación de índices de aridez en conjunto con la precipitación y la temperatura. Los valores de la PAR (ver figura 1D) sugieren una mayor radiación de abril a octubre; por el contrario, de noviembre a enero es cuando se registran los valores más bajos de esta variable. Estos resultados coinciden con los reportados por Troyo-Diéquez *et al.* (1990), quienes mencionan que los valores más altos de radiación en La Paz se presentan de abril a agosto. Al respecto, se ha demostrado que la poca lluvia de la región y los altos valores de radiación causan la aridez de la zona (Ayala Niño, 2018; Troyo-Diéquez *et al.*, 1990).

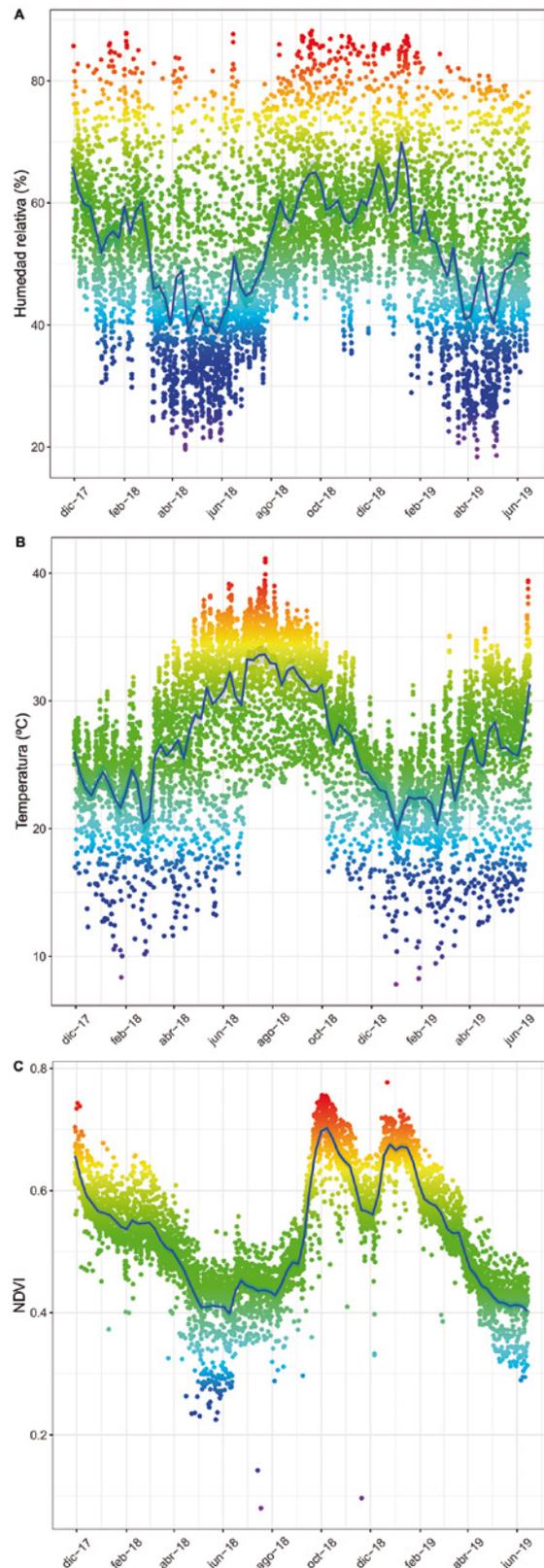
Como se aprecia en la figura 2A, los valores más bajos de humedad relativa se presentan en abril y mayo, mientras que los más altos, en diciembre y enero. Según Bell *et al.* (2012), la cantidad de humedad en el suelo en este sitio está relacionada de forma directa con la precipitación y humedad relativa debido a que la del suelo se mantiene relativamente alta en meses sin lluvias.

Las temperaturas (ver figura 2B) mínimas se presentan en febrero y marzo, mientras que las más altas, en junio y julio. Estos resultados difieren con respecto a los presentados por Ayala (2018), quien menciona que las menores son de diciembre a marzo. Al respecto, dichas diferencias podrían estar relacionadas con la variación interanual de las variables ambientales, dado que algunos años pueden ser más frescos/cálidos que otros y que las lluvias invernales inciden de manera directa en la temperatura de la región (Rodríguez-Moreno *et al.*, 2014).

Los valores del NDVI (ver figura 2C) muestran que el vigor de la vegetación es mayor en septiembre, mientras que de abril a junio es cuando se registran los más bajos. A pesar de este componente estacional marcado, en diciembre del 2018 y enero del 2019 se presentaron valores altos, reflejo de las lluvias que hubo en estos meses. Esto sugiere que el estado de salud de las plantas responde muy rápido a la presencia de la precipitación fuera de la temporada de verano. Salinas-Zavala *et al.* (2017) demostraron, mediante el uso del Índice, que el vi-

Figuras 2

Valores promedio por hora de las variables



Nota: la línea azul representa un promedio móvil de aproximadamente 15 días; el área en gris, su intervalo de confianza a 95 por ciento.

gor máximo de la vegetación en la parte sur de la península de Baja California se presenta de agosto a octubre; sin embargo, en algunos años se pueden presentar valores altos del NDVI en los meses de invierno debido a la presencia de lluvias de esa temporada. Esta variabilidad interanual ha sido relacionada de manera directa con eventos de gran escala como El Niño, mismos que generan una mayor precipitación en la región debido a las anomalías positivas de temperatura en el mar (Hawinkel *et al.*, 2016).

Relación entre el NDVI y las variables ambientales

Se calcularon los coeficientes de correlación de Pearson y el valor de probabilidad asociado a estos de todas las variables medidas con la estación de monitoreo para describir la relación lineal que existe entre estas y el Índice. Los primeros muestran que la humedad relativa se correlaciona de forma negativa con la temperatura, la velocidad del viento y la PAR; además, observan una correlación lineal casi nula entre la precipitación y las otras variables ambientales (ver cuadro 3).

Como se aprecia en el cuadro 3, el NDVI se correlaciona de forma positiva con la humedad relativa y de manera negativa con la temperatura. La que hay entre la precipitación y el Índice fue la más baja de todas, sugiriendo una correlación lineal casi

nula entre estas variables. Al respecto, en ecosistemas áridos o semiáridos, Agam y Berliner (2006) muestran que la formación de rocío puede exceder la cantidad de agua que aportan las lluvias; por lo tanto, la temperatura y la humedad relativa juegan un papel más importante que la precipitación en el desarrollo de la vegetación en este tipo de ecosistemas. Debido a esto, se sugiere que el crecimiento de especies que se encuentran en la vegetación de matorral sarcocaula tiene una estrecha dependencia con la humedad relativa y la temperatura.

El análisis de componentes principales aplicado a la información registrada con la estación de monitoreo sugiere que en el 1 y el 2 se explica 56 % de la variación de los datos (ver figura 3A). Además, que las variables temperatura y humedad relativa están correlacionadas de manera negativa en el primero, que la humedad y el NDVI guardan una positiva y que entre la temperatura y el Índice es inversa. La PAR y la velocidad del viento están muy relacionadas y la precipitación tiene una variación muy baja (ver figura 3B). Las variables que más aportan en el primer componente principal son la humedad relativa, temperatura, PAR y velocidad del viento (ver figura 3C), mientras que, en el segundo componente, las más importantes son la dirección del viento y el NDVI (ver figura 3D). Estos resultados resumen y confirman los valores de correlación descritos con anterioridad, de tal manera que se enfatiza que el vigor de la vegetación está muy relacionado con la humedad y la temperatura.

Cuadro 3

Coefficientes de correlación de Pearson (diagonal inferior) y su probabilidad (diagonal superior) categorizada en $p < 0.05^*$ (significativos) o $p > 0.05$ (no significativos)

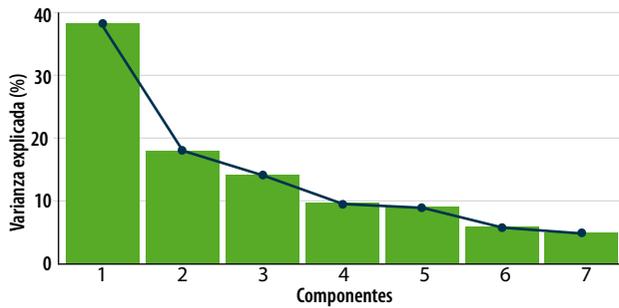
	Humedad	Temperatura	Precipitación	Dir. viento	Vel. viento	PAR	NDVI
Humedad	---	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$
Temperatura	-0.638	---	$p < 0.05$	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$
Precipitación	0.078	-0.014	---	$p < 0.05$	$p < 0.05$	$p < 0.05^*$	$p < 0.05^*$
Dir. viento	-0.048	0.081	0.018	---	$p < 0.05^*$	$p < 0.05^*$	$p < 0.05^*$
Vel. viento	-0.406	0.355	-0.007	-0.102	---	$p < 0.05^*$	$p < 0.05^*$
PAR	-0.549	0.5	-0.053	-0.176	0.422	---	$p < 0.05^*$
NDVI	0.444	-0.407	-0.026	-0.214	-0.199	-0.177	---

Nota: en negritas están los valores de correlación más altos entre el NDVI y las variables ambientales

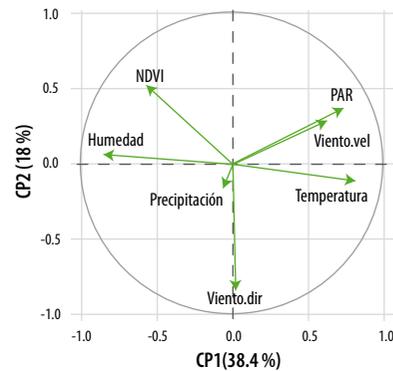
Figuras 3

Análisis de componentes principales aplicados a las variables

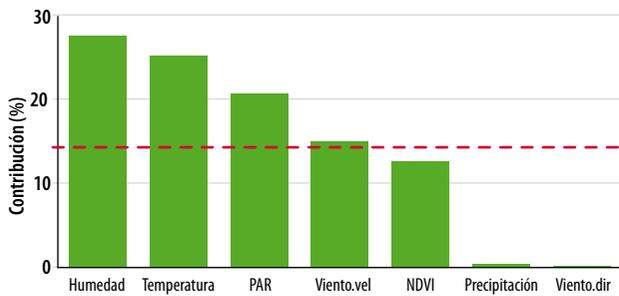
A) Porcentaje de contribución de los componentes principales



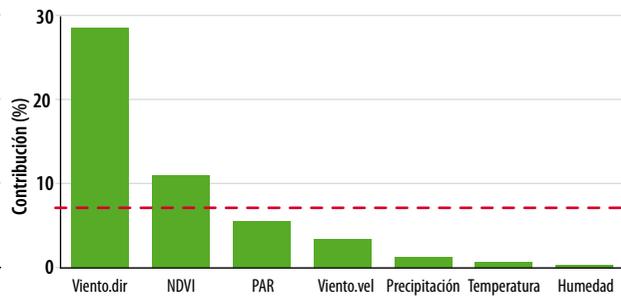
B) Eigen-vectores de los componentes 1 y 2



C) Contribución de las variables en el componente principal 1



D) Contribución de las variables en el componente principal 2



Nota: las líneas rojas representan valores de referencia a partir de los cuales se consideran significativos o no los porcentajes de contribución de cada variable.

Conclusiones

Los resultados muestran que este tipo de estaciones de bajo costo permite la recolecta de datos ambientales y del vigor de la vegetación en escalas de alta frecuencia, en tiempo casi real y de manera remota; por lo tanto, tiene un gran potencial para su implementación en casi cualquier parte del territorio nacional que tenga cobertura de telefonía celular.

Si bien la cobertura espacial no es muy amplia (≈ 0.22 ha), el bajo costo de la implementación y fácil manejo de la estación lo subsana en el análisis costo-beneficio. Debido a que se pueden incluir varias de estas en el servidor de datos, es posible concluir que estos equipos pueden trabajar en conjunto, a manera de una red de monitoreo remoto. Gracias al sensor de vegetación, esta red puede ser utilizada para evaluar cambios estacionales en el estado de salud de las plantas, en conjunto con la variabilidad ambiental de los sitios de interés.

Los resultados del análisis muestran que el vigor de la vegetación en un ecosistema terrestre con matorral sarcocaula está influenciado sobre todo por la humedad relativa y la temperatura del aire. Esto se debe a la escasa precipitación que se presenta en este tipo de ecosistema. En particular, se observó que el mayor vigor de la vegetación se presenta cuando la humedad relativa es más alta y cuando la temperatura ambiental disminuye a finales del verano.

Fuentes

Agam, N. and P. R. Berliner. "Dew Formation and Water Vapor Adsorption in Semi-Arid Environments-A Review", en: *Journal of Arid Environments*. 65(4), 2006, pp. 572-590.

Ayala Niño, Fernando. *Flujos y almacenes de carbono edáfico en un ecosistema árido del noroeste de México: un enfoque multiscalar*. Tesis de doctorado. Centro de Investigaciones Biológicas del Noroeste, SC. La Paz, Baja California Sur, 2018.

- Bell, Tom W., Olaf Menzer, Enrique Troyo-Diéquez and Walter C. Oechel. "Carbon Dioxide Exchange over Multiple Temporal Scales in an Arid Shrub Ecosystem near La Paz, Baja California Sur, Mexico", en: *Global Change Biology*. 18(8), 2012, pp. 2570-2582.
- C. A. Salinas-Zavala, R. O. Martínez-Rincón y M. V. Morales-Zárte. "Tendencia en el siglo XXI del Índice de Diferencias Normalizadas de Vegetación (NDVI) en la parte sur de la península de Baja California", en: *Investigaciones Geográficas. Boletín del Instituto de Geografía*. 94, 2017, pp. 82-90.
- Chen, X., Zhongjun Tan, D. Mark Schwartz and Chengxin Xu. "Determining the Growing Season of Land Vegetation on the Basis of Plant Phenology and Satellite Data in Northern China", en: *International Journal of Biometeorology*. 44(2), 2000, pp. 97-101.
- Deng, F., G. Su and C. Liu. "Seasonal Variation of MODIS Vegetation Indexes and Their Statistical Relationship With Climate Over the Subtropic Evergreen Forest in Zhejiang, China", en: *IEEE Geoscience and Remote Sensing Letters*. 4(2), 200, pp. 236-240.
- Hawinkel, P., W. Thiery, S. Lhermitte, E. Swinnen, B. Verbist, J. Van Orshoven and B. Muys. "Vegetation response to precipitation variability in East Africa controlled by biogeographical factors", en: *J. Geophys. Res. Biogeosci.* 121, 2016, pp. 2422-2444.
- Herrera-Cervantes, Hugo. "Sea Surface Temperature, Ocean Color and Wind Forcing Patterns in the Bay of La Paz, Gulf of California: Seasonal Variability", en: *Atmósfera*. 32, 2019, pp. 25-38.
- Jiang, Weiguo, Lihua Yuan, Wenjie Wang, Ran Cao, Yunfei Zhang and Wenming Shen. "Spatio-Temporal Analysis of Vegetation Variation in the Yellow River Basin", en: *Ecological Indicators*. 51, 2014, pp. 117-126.
- _____ "Spatio-Temporal Analysis of Vegetation Variation in the Yellow River Basin", en: *Ecological Indicators*. 51, 2015, pp. 117-126.
- Li, Shuangshuang, Saini Yang, Xianfeng Liu, Yanxu Liu and Mimi Shi. "NDVI-Based Analysis on the Influence of Climate Change and Human Activities on Vegetation Restoration in the Shaanxi-Gansu-Ningxia Region, Central China", en: *Remote Sensing*. 7(9), 2015, pp. 11163-11182.
- Meng, Meng, Jian Ni and Meijuan Zong. "Impacts of Changes in Climate Variability on Regional Vegetation in China: NDVI-Based Analysis from 1982 to 2000", en: *Ecological Research*. 26(2), 2011, pp. 421-428.
- Revadekar, J. V. Yogesh K. Tiwari and K. Ravi Kumar. "Impact of Climate Variability on NDVI over the Indian Region during 1981-2010", en: *International Journal of Remote Sensing*. 33(22), 2012, pp. 7132-7150.
- Rodríguez-Moreno, Víctor Manuel, José Ariel Ruíz-Corral, Guillermo Medina-García, José Saúl Padilla-Ramírez y Thomas Gunter Kretzschmar. "Efecto de la condición ENSO en la frecuencia e intensidad de los eventos de lluvia en la península de Baja California (1998-2012)", en: *Revista Mexicana de Ciencias Agrícolas*. 5(spe10), 2014, pp. 1923-1937.
- Salinas-Zavala, C. A., A. V. Douglas and H. F. Diaz. "Interannual Variability of NDVI in Northwest Mexico. Associated Climatic Mechanisms and Ecological Implications", en: *Remote Sensing of Environment*. 82(2), 2002, pp. 417-430.
- Sohoulande Djebou, Dagbegnon C., Vijay P. Singh and Oliver W. Frauenfeld. "Vegetation Response to Precipitation Across the Aridity Gradient of the Southwestern United States", en: *Journal of Arid Environments*. 115, 2015, pp. 35-43.
- Troyo-Diéquez, Enrique, Francisco de Lachica-Bonilla and José L. Fernández-Zayas. "A Simple Aridity Equation for Agricultural Purposes in Marginal Zones", en: *Journal of Arid Environments*. 19(3), 1990, pp. 353-362.
- Vandandorj, Sumiya, Batdelger Gantsetseg and Bazartseren Boldgiv. "Spatial and Temporal Variability in Vegetation Cover of Mongolia and Its Implications", en: *Journal of Arid Land*. 7(4), 2015, pp. 450-461.
- Vázquez, Pablo, Eduardo Adema y Beatriz Fernández. "Dinámica de la fenología de la vegetación a partir de series temporales de NDVI de largo plazo en la provincia de La Pampa", en: *Ecología Austral*. 23(2), 2013, pp. 77-86.
- Wang, Fei, Xuan Wang, Ying Zhao and Zhifeng Yang. "Temporal Variations of NDVI and Correlations between NDVI and Hydro-Climatological Variables at Lake Baiyangdian, China", en: *International Journal of Biometeorology*. 58(7), 2014, pp. 1531-1543.
- Wang, J., P. M. Rich and K. P. Price. "Temporal Responses of NDVI to Precipitation and Temperature in the Central Great Plains, USA", en: *International Journal of Remote Sensing*. 24(11), 2003, pp. 2345-2364.
- Wang, Jing, Kelin Wang, Mingyang Zhang and Chunhua Zhang. "Impacts of Climate Change and Human Activities on Vegetation Cover in Hilly Southern China", en: *Ecological Engineering*. 81, 2015, pp. 451-461.
- Yepes, Adriana y Marcos Silveira Buckeridge. "Respuestas de las plantas ante los factores ambientales del cambio climático global (revisión)" en: *Colombia Forestal*. 14(2), 2011, pp. 213-232.
- Zhao, Zhiqiang, Jiangbo Gao, Yanglin Wang, Jianguo Liu and Shuangcheng Li. "Exploring Spatially Variable Relationships between NDVI and Climatic Factors in a Transition Zone Using Geographically Weighted Regression", en: *Theoretical and Applied Climatology*. 120(3), 2015, pp. 507-519.
- Zhu, Lin, Huili Gong, Zhenxue Dai, Tingbao Xu and Xiaosi Su. "An Integrated Assessment of the Impact of Precipitation and Groundwater on Vegetation Growth in Arid and Semiarid Areas", en: *Environmental Earth Sciences*. 74(6), 2015, pp. 5009-5021.

Comparación de métodos de detección de cambio forestal en el marco del proyecto **Forests 2020**

*Comparison of Forest Change Detection Methods in the Framework of the **Forests 2020 Project***

Miguel Muñoz,* Miguel Castillo,** Felicia Line,*** Edith Mondragón,** Oswaldo Carrillo,* Richard Tipper*** y Catherine Navarro*

El monitoreo y medición precisa del cambio forestal son esenciales para evaluar el progreso hacia la reducción de emisiones por deforestación y degradación exigido por el *Acuerdo de París* bajo la Convención Marco de las Naciones Unidas ante el Cambio Climático. No obstante, persisten amplias diferencias en las estimaciones de la extensión del bosque remanente y la tasa a la que se está perdiendo. Dichas variaciones se deben a una serie de factores entre los que destacan: las disparidades en la escala de los mapas, los métodos y los insumos utilizados e, incluso, en las propias definiciones de los conceptos *bosque* y *deforestación*. En este trabajo, en el marco del proyecto *Forests 2020*, se realizó un análisis comparativo entre distintos enfoques y métodos para detectar cambios en la cobertura forestal en una región de la Selva Lacandona en Chiapas con la finalidad de poner en perspectiva los distintos alcances y limitaciones de cada uno.

Palabras clave: monitoreo; deforestación; imágenes satelitales; evaluación de la incertidumbre; México.

Recibido: 1 de abril de 2020.

Aceptado: 26 de junio de 2020.

* Comisión Nacional Forestal; miguelmrv.munoz@gmail.com, ocarrillo.ute@conafor.gob.mx y oswaldisma@gmail.com, catherine.navarro@conafor.gob.mx, respectivamente.

** El Colegio de la Frontera Sur, mcastill@ecosur.mx y edith.mondragonvazquez@gmail.com, respectivamente.

*** Ecometrica, felicia@ecosistemasconsultora.org y flicline@gmail.com, Richard.tipper@ecometrica.com y richard.tipper@resilienceconstellation.com, respectivamente.

Nota: este trabajo es parte del proyecto *Forests 2020*, el cual fue desarrollado del 2016 al 2020, financiado por la Agencia Espacial del Reino Unido a través de su programa de Socios Internacionales y coordinado por Ecometrica.

Accurate measurement and monitoring of forest cover change are essential for assessing progress towards reducing emissions from deforestation and forest degradation as agreed in the UNFCCC Paris Agreement. Unfortunately, wide differences persist in estimates of the extent of remaining forest area and the rate at which it is being lost. The variations in the estimations of these parameters arise from several factors, among which are the scale of the maps, the methods used, inputs used and even in the very definitions of forest and deforestation. In this work, within the framework of the *Forests 2020* project, a comparative analysis was carried out between different approaches and methods to detect forest change in the Lacandon forest region in Chiapas, in order to put into perspective the different scopes and limitations of each one.

Key words: Monitoring; Deforestation; Satellite Imagery; Accuracy Assessment; Mexico.



Una triste realidad/Carlos Miguel Herrera Tapia

Introducción

México es considerado como líder en políticas de cambio climático (CC) entre los países de economías emergentes. Fue una de las primeras naciones en regular las acciones de mitigación y adaptación al CC mediante la publicación de la *Ley General de Cambio Climático* en el 2012; en el 2015, presentó sus *Contribuciones nacionalmente determinadas*, que fueron decretadas en el 2018 para su alineación con el *Acuerdo de París* (Averchenkova y Guzman Luna, 2018; CONAFOR, 2020).

En el 2017, nuestro país publicó la *Estrategia para la reducción de emisiones por la deforestación y degradación (REDD+)*, en la que se reconoce que el sector forestal contribuirá con casi 22 % de la meta de mitigación de México en el 2030 (CONAFOR, 2020).

Por lo tanto, existe una creciente necesidad de monitorear el progreso en la reducción de las emisiones relacionadas con cambios de usos de suelo, sobre todo asociados a la deforestación, la cual se define en la *Ley General de Desarrollo Forestal Sustentable (LGDFS)* como la "... pérdida de la vegetación forestal en forma permanente, por causas inducidas o naturales..." (art. 7, fracción XVIII) o, en términos operativos, como la reducción del dosel por debajo del umbral mínimo (10 %) con cambio de uso de suelo forestal a no forestal, de manera permanente (CONAFOR, 2020).

Si bien a nivel mundial se han creado iniciativas para el monitoreo de los bosques —como el caso de *Global Forest Change (GFC)* desarrollado por la Universidad de Maryland (Hansen *et al.*, 2013), que en la actualidad se encuentra disponible en la plataforma *Global Forest Watch*—, estas

no cuentan con procesos de estimación de la exactitud temática o medición de error asociado a los cambios detectados; por ejemplo, los mapas que se generan a partir de GFC reportan solo la pérdida de cobertura forestal, que no necesariamente se relaciona con superficies de deforestación; además, en algunos casos no se consideran las definiciones locales de *bosque*, lo que puede provocar sobreestimaciones o bien, sesgos estadísticos significativos.

Considerando lo anterior, de acuerdo con los insumos disponibles y las capacidades técnicas en cada país, se han diseñado y puesto en operación sistemas de monitoreo forestal tratando de consolidar y adecuar métodos precisos en las estimaciones tanto estadísticas como geográficas (Argotty *et al.*, 2019).

En el caso de México, desde el 2014, la Comisión Nacional Forestal (CONAFOR) ha operado el Sistema Nacional de Monitoreo Forestal (SNMF), establecido en el artículo 34 de la *LGDFS*, mediante el cual se han elaborado y sometido los Niveles de Referencia de Emisiones Forestales (NREF) en la Convención Marco de las Naciones Unidas ante el Cambio Climático (CONAFOR, 2020). De manera histórica, la Comisión ha utilizado los mapas de uso de suelo y vegetación de las series I a la VI del Instituto Nacional de Estadística y Geografía (INEGI) para el monitoreo de los cambios en esta y el cálculo de tasas de deforestación (TD).

En el primer NREF para México, donde se estimaron TD brutas (CONAFOR, 2015), estas fueron sustancialmente mayores que las reportadas con anterioridad a la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés) debido, sobre todo, a modificaciones en las metodologías y en las definiciones de los conceptos de *bosque* y *deforestación* (CONAFOR, 2020). Además, las series del INEGI no son insumos adecuados para el propósito de detectar la deforestación, ya que son elaboradas como mapas de referencia de la vegetación existente en el país para su respectivo año base en una escala de 1:250 000.

Por lo tanto, en el marco del proyecto *Forests 2020*,¹ la CONAFOR y El Colegio de la Frontera Sur (ECOSUR), en colaboración con la Universidad de Leicester y el Centro de Geografía Ambiental (CIGA) de la Universidad Nacional Autónoma de México, se han desarrollado metodologías robustas para la estimación y mapeo de las coberturas y la detección de cambios de cobertura del suelo mediante el uso de herramientas satelitales, empleando definiciones locales de bosque, así como métodos adaptados a las condiciones específicas del país, para generar TD con alta certidumbre.

Bajo este contexto es que se llevó a cabo un análisis comparativo de enfoques de detección de cambios en la cobertura forestal y deforestación para evaluar sus alcances y limitaciones.

Se eligió como área de estudio a la región de Marqués de Comillas, en Chiapas, ya que se ha caracterizado por presentar una alta frecuencia de cambio en el uso del suelo, lo cual se consideró idóneo para realizar este análisis y evaluar la efectividad de los diferentes métodos de detección y mapeo de la deforestación.

A partir de un enfoque espacialmente explícito, se elaboraron cuatro mapas de deforestación de la zona de estudio y se estimó la superficie dañada de acuerdo con los siguientes métodos: 1) sobreposición de cartas de uso de suelo y vegetación (INEGI), 2) Sistema Satelital de Monitoreo Forestal (SAMOF) de la CONAFOR, 3) detección de cambios desarrollado en el Laboratorio de Análisis Geográfico de ECOSUR (LAIGE-ECOSUR) y 4) el algoritmo *Breaks For Additive Season and Trend (BFAST)*.

Además de los métodos espacialmente explícitos, se calculó la superficie de deforestación usando el enfoque de muestreo propuesto en el capítulo 3 del volumen 4 de las *Directrices del Panel Intergubernamental de Cambio Climático* (IPCC, por sus siglas en inglés) del 2006. Para evaluar y comparar de forma objetiva la efectividad de los diferentes métodos de mapeo, fue necesario obtener

¹ Para más información, ver <http://ecometrica.com/space/forests2020>.

estimaciones insesgadas de las superficies siguiendo la metodología de Olofsson *et al.* (2014). Por último, se realizó un análisis comparativo entre los métodos propuestos y se discutieron los alcances y limitaciones de cada uno.

Área de estudio

El análisis comparativo sobre los enfoques de estimación de la deforestación se hizo en la región de la Selva Lacandona, en Chiapas, denominada Marqués de Comillas (ver mapa 1), que abarca una extensión de 200 mil hectáreas. De acuerdo con la *Carta de uso de suelo y vegetación* correspondiente a la serie

VI (INEGI, 2014), en la zona se encuentran presentes de forma natural las selvas alta perennifolia (SAP) y mediana subperennifolia (SMQ), en sus diferentes fases vegetativas (ver imagen 1).

En esta zona hay una gran dinámica de cambio en la flora natural derivada de actividades como la ganadería y el cultivo de la palma de aceite, como se aprecia en las imágenes 2 (Covaleda *et al.*, 2014). Fue elegida como una de las áreas piloto del proyecto *Forests 2020* y ahí se realizaron las experimentaciones metodológicas iniciales, antes de escalar a nivel estatal pues, además, cuenta con amplia información de campo por parte de ECOSUR para la validación de los métodos.

Mapa 1

Ubicación del área de estudio, zona Marqués de Comillas, Chiapas, México

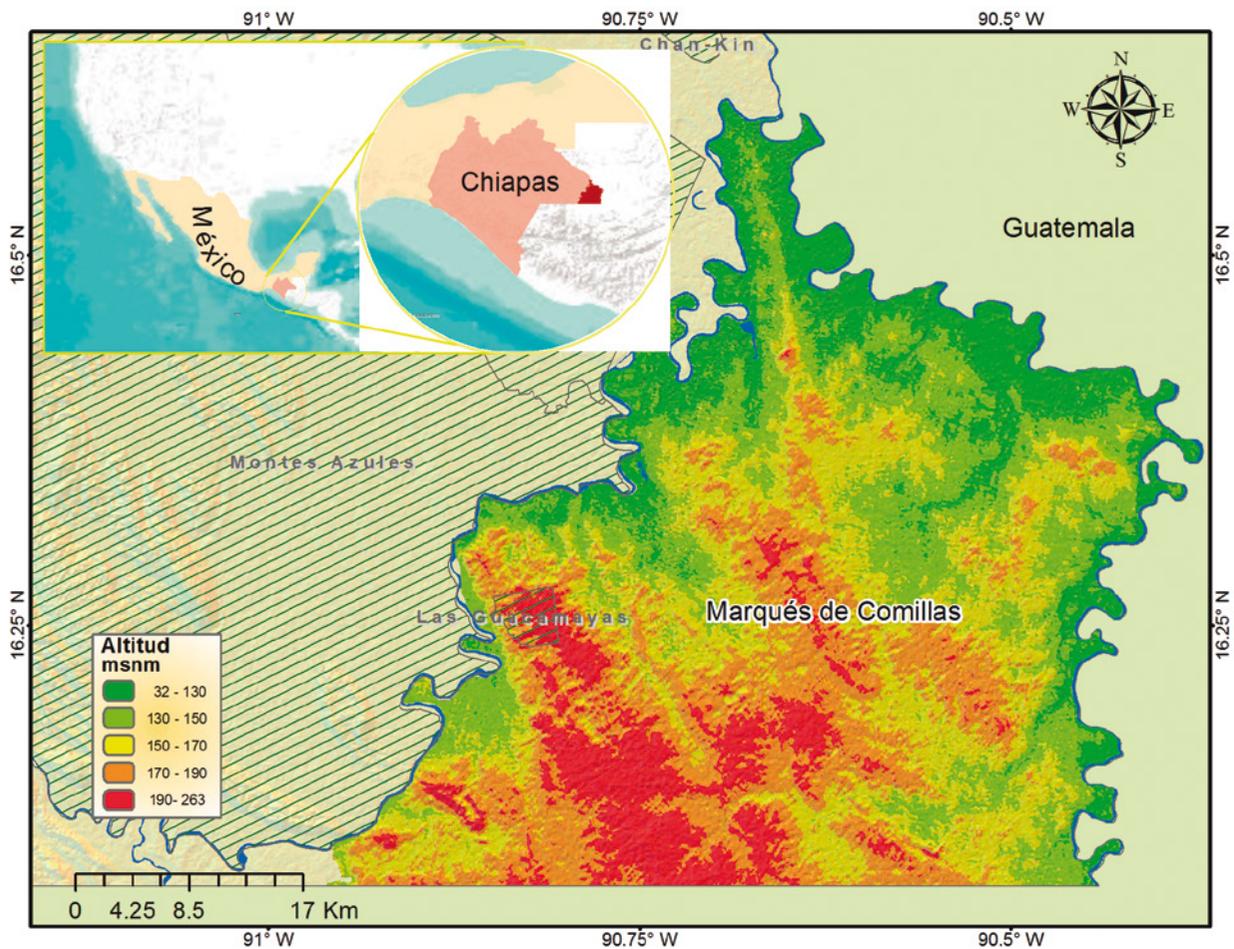


Imagen 1



Agricultura de temporal (maíz), con fragmento de selva mediana perennifolia en el fondo, San Isidro, Marqués de Comillas; coordenadas: UTM 15. X: 735699 Y: 1799272 Z: 169.
Fuente: LAIGE-ECOSUR.

Imágenes 2



Ganadería en San Isidro, Marqués de Comillas, Chiapas; coordenadas UTM 15. X: 728810 Y: 1801748 Z: 171.
Fuente: LAIGE-ECOSUR.



Plantación de palma de aceite, en la Victoria, Marqués de Comillas, Chiapas; coordenadas: UTM 15 X: 743323 Y: 1798520 Z: 181.

Fuente: LAIGE-ECOSUR.

Metodología

De manera general, son tres las etapas en las que se divide y que abonarán después al análisis de los resultados y la discusión. La primera es la construcción de mapas para detección de cambios en la cobertura forestal empleando distintos métodos de mapeo, así como la estimación de la deforestación bruta. La segunda es la descripción del método empleado para el cálculo de superficies insesgadas realizada a las estimadas de manera previa. Por último, la tercera es una propuesta de valoración de deforestación no espacialmente explícita basada en un enfoque de muestreo sistemático.

Métodos para la detección y mapeo de cambios

Su análisis comparativo se llevó a cabo para el periodo 2011-2014 o en fechas muy cercanas. Los mapas

obtenidos se evaluaron solo en la clase de deforestación bruta con el objetivo de que todas las cartas fueran compatibles y comparables para poder determinar la efectividad de cada método, ya que la definición de bosque varía en cada trabajo.

Sobreposición de mapas de uso de suelo y vegetación de las series del INEGI, 2011-2014

Con este método se compararon y detectaron las modificaciones que ocurren en el tiempo 2 (T2) con respecto al 1 (T1) para, después, estimar la superficie de cambio. Es importante señalar que tiene algunas limitantes debido a las características propias de las cartas, *i. e.* la Unidad Mínima de Mapeo (UMM) de las áreas forestales, que es de 50 y 25 ha para áreas agrícolas. Además, su escala es de 1:250 000, lo que la hace un insumo poco adecuado para el monitoreo de cambios de la cobertura del suelo en áreas con poca extensión.

Enfoque pared a pared del SAMOF, 2011-2014

Este método se enmarca en el Sistema Nacional de Monitoreo Forestal de la CONAFOR y su objetivo principal es la generación de mapas de cobertura y cambios de cobertura del suelo (usando imágenes *Landsat*). El SAMOF genera esta información geoespacial a través de una cadena articulada de seis pasos que incluyen el establecimiento de definiciones operativas de bosque y deforestación, la calibración de algoritmos, la selección de insumos, el procesamiento de imágenes satelitales,² el posprocesamiento manual y la evaluación de exactitud temática (Gebhardt *et al.*, 2014). En particular, la elaboración de los mapas y la detección de los cambios se realizó mediante el análisis de dos compuestos multitemporales (2011-2014) *Landsat* con los algoritmos *iterative Multivariate Alteration Detection (iMAD)* y *Maximum Autocorrelation Factor (MAF)* (Nielsen *et al.*, 1998), así como la aplicación de un proceso manual de fotointerpretación para corregir y validar las modificaciones.

Método de mapeo de cambios de ECOSUR (LAIGE-ECOSUR), 2010-2015

En primer lugar, se elaboró un mapa de cobertura del suelo con la mayor cantidad de datos de campo posibles con el fin de obtener una carta con altos niveles de precisión; a este producto se le denominó *mapa de referencia*.

En la fase siguiente, se identificaron las áreas que sufrieron cambios con respecto al mapa de referencia. Para lograrlo, se empleó la siguiente secuencia de métodos: se calculó la diferencia algebraica entre las imágenes satelitales de los años de interés y al resultado de esa operación se le aplicó una segmentación, es decir, se agruparon los píxeles con valores similares; después, se aplicó el algoritmo *Isolation Forest* (Pedregosa *et al.*, 2011) para detectar segmentos sospechosos de haber sufrido modificaciones y se clasificaron los segmen-

² Se implementa con la versión de la CONAFOR del *software* Monitoreo de Datos de Actividad de México (MAD-Mex).

tos usando el método de *Random Forest* (Breiman, 2001); al final se aplicó un proceso de fotointerpretación para corregir y validar los cambios. Se emplearon imágenes *SPOT 5* y *7* (tamaño de píxel 10 y 6 m, respectivamente) correspondientes al 2010 y 2015; para la aplicación de los procesos automatizados, se utilizó el *software Python* y las librerías para el procesamiento de imágenes satelitales.

Detección de cambios con el algoritmo BFAST, 2011-2014

Este es un algoritmo basado en el análisis de series de tiempo de imágenes *Landsat* y *MODIS*, principalmente; integra la descomposición de las series en componentes de tendencia y temporalidad con métodos para detectar y caracterizar el cambio dentro de estas (Verbesselt *et al.*, 2010); calcula, de forma iterativa, el tiempo y el número de cambios bruscos dentro de series de tiempo y caracteriza la modificación por su magnitud y dirección. Para la detección de estos en el área de estudio, se utilizó la plataforma de computación en la nube del *System for Earth Observations, Data Access, Processing & Analysis for Land Monitoring (SEPAL)*, desarrollado por la FAO (SEPAL-FAO, 2018), donde se incluye la herramienta de *software* estadístico *R* (R Core Team, 2020) que permite la implementación del paquete de procesamiento de imágenes BFAST.

Enfoque de estimación de superficies insesgadas

Para evaluar y comparar de manera objetiva la efectividad de los diferentes métodos de mapeo espacialmente explícitos descritos antes, es necesario obtener estimaciones insesgadas de las superficies analizadas. Por lo tanto, a continuación, se presenta el método utilizado para estimarlas tomando como insumos las superficies mapeadas.

Para obtener estimaciones insesgadas de las TD brutas en el área ventana de estudio empleando los productos espaciales de los diferentes métodos de mapeo, se llevó a cabo una evaluación de exac-

titud siguiendo el método propuesto por Olofsson *et al.* (2014), el cual está basado, por un lado, en el conteo de muestras que coinciden con los mapas de cambios de cobertura generados y el llenado de una matriz de confusión con la cual se reportan las exactitudes global, del usuario y del productor y, por el otro, se integran estimadores estadísticos insesgados de las superficies (y sus incertidumbres) de las clases mapeadas a partir de los resultados de la matriz. A continuación, se explica cada uno de los pasos.

Estimación del tamaño de muestra y distribución de las muestras

Para efectuar este análisis, se empleó un muestreo aleatorio estratificado (MAE), el cual es flexible en la distribución de las muestras en zonas de cambio (Olofsson *et al.*, 2013),

En el MAE utilizado, los estratos a evaluar fueron las clases mapeadas: deforestación, bosque estable (BE) y no bosque estable (NBE). Sin embargo, para este ejercicio solo se presentan los resultados para la clase deforestación bruta.

Las muestras se diseñaron considerando *a priori* una exactitud del usuario (U_i) a nivel de estrato con

valores conservadores de 0.5. El nivel de confianza utilizado fue de 95 % ($\alpha = 0.95$) y un error de 0.02 % ($S(O)$) en la precisión total de omisión siguiendo a Olofsson *et al.* (2013). El tamaño de la muestra se calculó según Cochran (1977) y su distribución por estrato se realizó mediante la asignación óptima de Neyman. Para aquellos estratos pequeños, el tamaño de muestra asignado fue de 50 puntos con la finalidad de alcanzar la suficiente representación estadística.

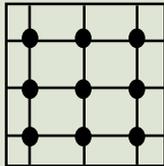
Exactitud temática de los mapas de cambio

Una vez diseñada la muestra, especialistas en interpretación visual de imágenes satelitales analizaron los puntos de muestreo empleando aquellas de mayor resolución disponibles, como *RapidEye* o *SPOT*, para los años de referencia de cada método. En particular, en estos puntos de muestreo se identificó el estrato (bosque y no bosque) al 2011 y 2014 y, después, su estrato de cambio o permanencia (deforestación, BE y NBE).

A partir de la información de referencia generada por los intérpretes, fue posible construir una matriz de confusión y obtener la exactitud total de los mapas de cambio espacialmente explícitos, y las exactitudes de usuario y del productor,

Cuadro 1

Proceso de estimación de superficies por medio de proporciones. Modificado de IPCC, 2006

Ejemplo de la estimación de superficies por medio de proporciones			
Procedimiento de muestreo	Estimación de proporciones	Superficies estimadas de categorías de uso de la tierra	Error estándar
	$p_i = n_i/n$ $p_1 = 3/9 \approx 0.333$ $p_2 = 2/9 \approx 0.222$ $p_3 = 4/9 \approx 0.444$ $\text{Suma} = 1.0$	$A_i = p_i * A$ $A_1 = 300 \text{ ha}$ $A_2 = 200 \text{ ha}$ $A_3 = 400 \text{ ha}$ $\text{Total} = 900 \text{ ha}$	$s(A_i)$ $s(A_1) = 150.0 \text{ ha}$ $s(A_2) = 132.2 \text{ ha}$ $s(A_3) = 158.1 \text{ ha}$

donde:

A = superficie total (= 900 ha en el ejemplo).

A_i = superficie estimada de la categoría de uso de la tierra i .

n_i = cantidad de puntos ubicados en la categoría de uso de la tierra i .

n = cantidad total de puntos.

además de los estimadores insesgados de áreas y sus respectivas incertidumbres, siguiendo la metodología de Olofsson *et al.* (2014).

Metodología para la estimación de superficies de deforestación basada en el enfoque de muestreo

En el capítulo 3 del volumen 4 de las *Directrices del IPCC* (2006) se indica que es posible estimar superficies y cambios mediante un enfoque de proporciones basado en un muestreo sistemático de parcelas. El IPCC sintetiza este enfoque en el cuadro 1.

Siguiendo este enfoque, se procedió al análisis y fotointerpretación manual de parcelas de 1 ha sobre imágenes de satélite, anidadas en el diseño muestral del Inventario Nacional Forestal y de

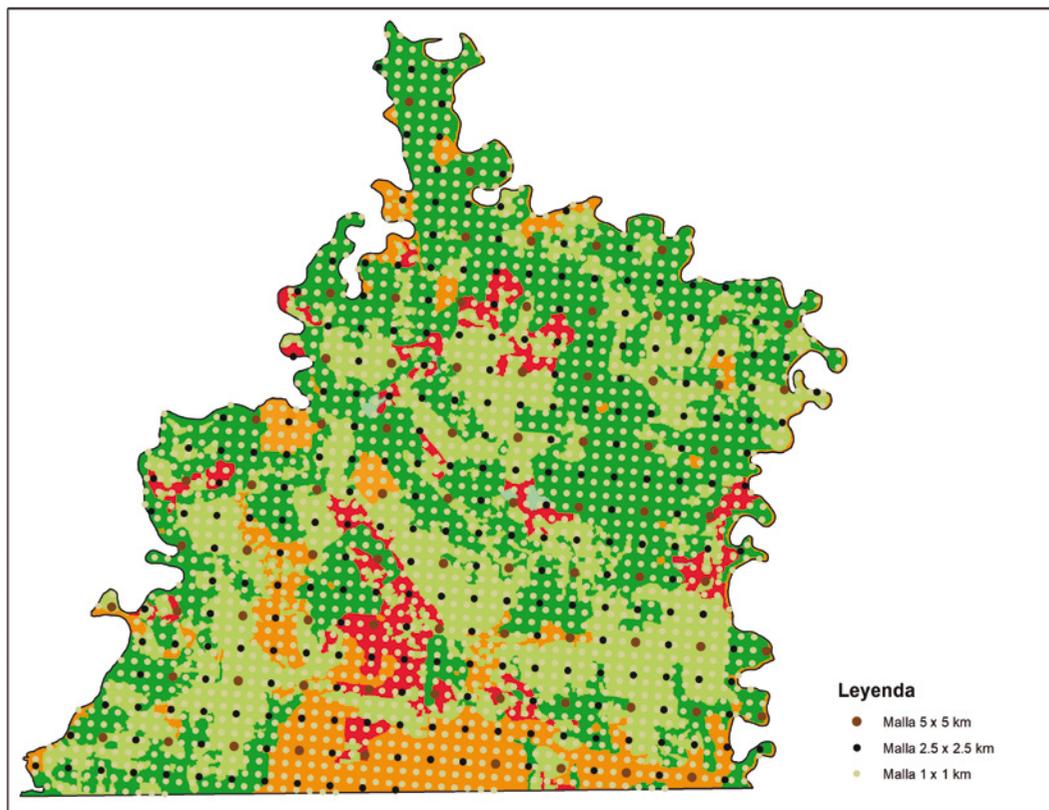
Suelos (INFYS)³ y con una equidistancia de 1 km x 1 km, lo que arrojó en total 1 968 parcelas de muestreo (ver imagen 3). De manera adicional, mediante un protocolo de fotointerpretación con criterios claros, se estableció el tipo de cobertura presente en la parcela en el 2011 y 2014, así como los cambios que ocurrieron en el periodo de análisis (2011-2014).

Una vez que se contó con el análisis de las coberturas de todas las parcelas para el periodo de análisis en el área de estudio, con el enfoque de estimación de superficies por medio de proporciones se calcularon las estadísticas de las pérdidas de la cubierta forestal y las superficies de cada clase que se hubieran determinado al inicio del muestreo.

³ Para la ubicación exacta de los puntos de muestreo, visite la página <https://marquesdecomillas.ecosur.ourecosystem.com/interface/>

Imagen 3

Distribución de las parcelas de la malla de muestreo anidadas en el diseño muestral del INFYS



Las áreas de cambio se obtuvieron de acuerdo con la ecuación:

$$\widehat{ATF}_i = \frac{n_i}{N_i} \times AT$$

donde:

\widehat{ATF}_i = área estimada en el i -ésimo estrato.

n_i = número de parcelas del estrato i .

N_{ij} = número total de las parcelas muestreadas en el estrato i .

AT_{ij} = área total de la zona de estudio.

La incertidumbre del \widehat{ATF}_i se estimó de acuerdo con la ecuación:

$$U(\widehat{ATF}_i) = \frac{Z_{\alpha/2} * s(\widehat{ATF}_i)}{\widehat{ATF}_i} * 100$$

donde:

$Z_{\alpha/2}$ = percentil 95 % del modelo de distribución empírico que ajusta los datos,

$s(\widehat{ATF}_i) = AT * \sqrt{\frac{p_i(1-p_i)}{N_i-1}}$ y $s(ATF_i)$ = desviación estándar del área estimada en el

i -ésimo estrato obtenida anteriormente con los insumos ya definidos y $p_i = \frac{n_i}{N_i}$.

Resultados

El mapa 2 muestra el resultado de mapeo de los cuatro métodos explicados;⁴ en ellos se pueden apreciar las diferencias en la distribución geográfica y tamaño de las áreas de deforestación. Los que utilizaron insumos de mayor resolución arrojaron zonas de deforestación más pulverizada y, en el caso de BFAST, se detectó mayor superficie total de esta. El patrón espacial de la deforestación obtenido con los métodos LAIGE-ECOSUR, BFAST y SAMOF presentó cierto nivel de similitud, pero el que se tuvo con los mapas del INEGI no se asemeja a ninguno de los otros tres; por ejemplo, en la parte central-inferior del segmento a del mapa se observa un área grande de cambios, que no aparece en ninguno de los otros.

⁴ Para ver y comparar los mapas visite la plataforma interactiva en línea de EO Labs: <https://marquesdecomillas.ecosur.ourecosystem.com/interface/>

Mapas

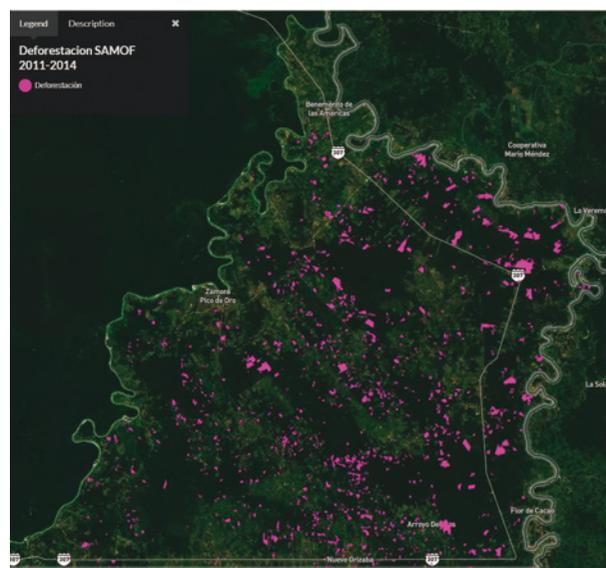
Continúa

Cartas de deforestación para el área de Marqués de Comillas utilizando los cuatro métodos

a) INEGI 2011-2014



b) SAMOF (2011-2014)

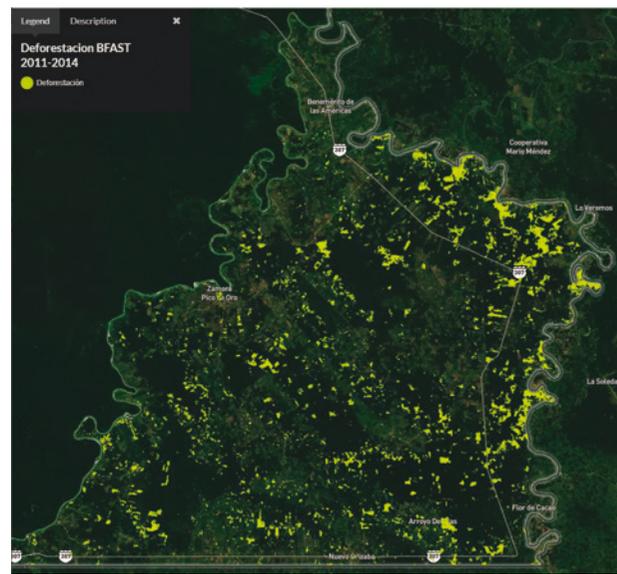


Cartas de deforestación para el área de Marqués de Comillas utilizando los cuatro métodos

c) LAIGE-ECOSUR (2010-2015)



d) BFAST (2011-2014)



En el cuadro 2 se muestran las estimaciones de deforestación bruta obtenidas por cada método, así como la gráfica asociada. Cabe señalar que, en el caso de la metodología LAIGE-ECOSUR, el periodo de análisis fue más amplio, contemplando cinco años (2010-2015). Sin embargo, para que estas fueran comparables, se anualizaron los datos. De

forma adicional, y como referencia, se pone la estimación de GFC, al ser considerada como una de las cifras que se utilizan con mayor frecuencia para referir datos de deforestación, aunque se define como la pérdida de cobertura forestal (en lugar de deforestación); entonces, no es estrictamente comparable con las demás estimaciones.

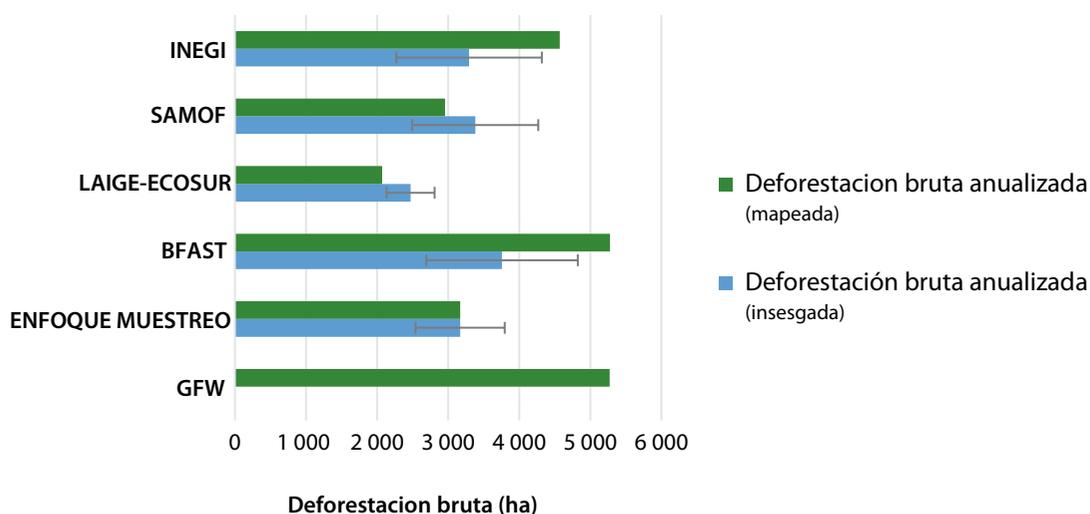
Cuadro 2

Resultados en la estimación de deforestación bruta para los métodos analizados (superficies mapeada y ajustada)

Clases	Método de estimación					
	INEGI	SAMOF-CONAFOR	LAIGE-ECOSUR	BFAST	Enfoque de muestreo	GFC
Deforestación bruta para todo el periodo (superficie mapeada-estimada) en hectáreas	13 708	8 868	10 361	15 838	9 507	15 821
Deforestación bruta para todo el periodo (superficie insesgada) en hectáreas	9 887	10 146	12 360	11 277	N/A	-
Deforestación bruta anualizada (superficie mapeada) en hectáreas	4 569	2 956	2 072	5 279	3 169	5 273
Deforestación bruta anualizada (superficie insesgada e intervalo de confianza) en hectáreas	3 295 ±1 024	3 382 ±887	2 472.09 ±338	3 759 ±1 066	3 169 ±628	-
Exactitud del usuario (%)	7.25	48.65	97	30.77	N/A	-

N/A: no aplica.

Superficies insesgadas y mapeadas de deforestación bruta anual 2010-2014 de diferentes métodos con niveles de incertidumbre asociados



Discusión y recomendaciones

Como se puede observar en los resultados, existe una amplia variabilidad en las estimaciones de las superficies mapeadas de deforestación e, incluso, en los patrones espaciales del cambio. En este último caso, la falta de concordancia entre los obtenidos con los mapas del INEGI y el resto de los métodos se debe, en apariencia, a su bajo nivel de detalle espacial lo cual confirma, en cierta medida, lo que se anunciaba de forma previa: la escala espacial gruesa puede ser una seria limitante para obtener estimadores precisos de deforestación.

Los cambios de uso del suelo con frecuencia se presentan en áreas pequeñas (alrededor de 1 ha), por lo que mapas como las series del INEGI (cuya superficie mínima cartografiada es de 50 ha) podrían subestimar o sobreestimar dichos cambios. Es necesario recalcar que estas cartas fueron elaboradas para representar el uso de suelo, por lo que su propósito no es detectar modificaciones.

Por otra parte, dentro de los métodos que presentan un patrón espacial de cambios similar,

también existen variaciones considerables en las estimaciones de la cantidad de bosque perdido; ejemplo de esto es el de BFAST, que calcula más del doble de lo obtenido con el de LAIGE-ECOSUR. Por lo anterior, cobra sentido y relevancia el criterio de exactitud temática pues, ante un escenario en el que se cuenta con dos estimaciones de un mismo fenómeno, es recomendable, entonces, elegir el método o la estimación con el menor nivel de error. La exactitud del usuario y del productor puede ser una medida adecuada de la confiabilidad de una clase en un determinado mapa. Valores altos indican resultados más confiables, o menor error.

Otro criterio importante por mencionar al elegir un método de estimación u otro es la definición del concepto de bosque. En el caso de los mapas de GFC, los cambios detectados corresponden a la pérdida de dosel en áreas que de inicio tenían una cobertura arbórea por lo menos de 30 % y esto no necesariamente puede considerarse como deforestación, sino una posible perturbación más relacionada con la degradación. Otro fenómeno que puede estar siendo detectado por los mapas de GFC es la pérdida de vegetación secundaria, si-

tuación muy común en el área de estudio, donde se practican métodos agrícolas de roza-tumba y quema; es por ello que, con frecuencia, las tasas de deforestación provenientes de este método son altas.

Por otra parte, el método basado en muestreo estadístico produce estimadores de la cantidad de deforestación cercanos a los que se obtienen con los de mapeo de menor error (SAMOF y LAIGE-ECOSUR), por lo que resulta una alternativa atractiva y confiable cuando se desean estimadores de la cantidad y no de la distribución espacial del proceso. Si bien no aplica el criterio de exactitud temática del usuario, cuenta con una incertidumbre asociada baja, lo cual, vuelve a este método una opción costo-eficiente para replicarlo a niveles tanto regionales como nacionales.

En perspectiva, los métodos del SAMOF y de LAIGE-ECOSUR resultan ser muy efectivos en cuanto a la precisión de los cálculos de deforestación, teniendo exactitudes del usuario de 48.65 y 97 %, respectivamente. Ambos involucran en su implementación la estimación de superficies insesgadas. Sin embargo, el costo de posprocesamiento es alto, ya que requiere tiempo y esfuerzos adicionales por parte de los intérpretes visuales que validan una a una las etiquetas de cambio asignadas de una muestra representativa para eliminar falsas modificaciones.

Se puede decir que la implementación tanto del SAMOF (cuyos productos se enfocan a la estimación de coberturas y sus cambios a niveles estatal y nacional), así como la iniciativa local por parte de LAIGE-ECOSUR parten de enfoques que generan información con una mayor fiabilidad, ya que incorporan a sus análisis una fase de posprocesamiento, la cual requiere del conocimiento de expertos locales y tiempo para asegurar niveles mayores de certidumbre y, si es posible, información auxiliar de campo (como fue en el caso de LAIGE-ECOSUR), lo cual implica una elevación en los costos de mapeo. Sin embargo, es una etapa necesaria para asegurar fiabilidad de los productos y eliminar los falsos cambios.

A nivel de recomendaciones, para aquellos usuarios que generan sus propias estimaciones de la deforestación, se sugiere que, además del procesamiento automatizado de imágenes para detección de cambios, se deberá incluir un proceso final de análisis, revisión y ajuste de los resultados denominado *fase de posprocesamiento*.

Otra buena práctica para el monitoreo de la deforestación es incluir dentro del proceso de producción de mapas la estimación de parámetros de exactitud (Strahler *et al.*, 2006). Para implementar el análisis de exactitud temática a los mapas generados, es necesario llevar a cabo un diseño de muestreo donde se determinan las clases a evaluar, el método y el tamaño de muestras. Es deseable involucrar campañas intensivas de validación en campo; sin embargo, debido a que las verificaciones en el terreno suelen ser costosas, estas se pueden sustituir por un método más costo-eficiente, como es el caso del análisis y fotointerpretación de las muestras sobre imágenes de alta resolución, lo que facilita la generación de información de referencia con una mayor fiabilidad. Tener un análisis de este tipo ayudará a los usuarios finales a conocer sus alcances y limitaciones, además de conocer qué tan cercanas a la realidad son las estadísticas de deforestación reportadas.

Por último, se recomienda seguir invirtiendo en el fortalecimiento de las capacidades de monitoreo del país, ya que, evidentemente, las metodologías SAMOF y LAIGE-ECOSUR demuestran ser más eficientes y robustas para calcular deforestación, comparadas con métodos o fuentes internacionales. De esta manera, será posible escalar este tipo de análisis del nivel local o regional al nacional, disminuir los tiempos de procesamiento y mejorar la exactitud de los resultados, utilizando la mejor información disponible como las imágenes satelitales de alta resolución, la experiencia de especialistas locales, además de los datos de campo para la calibración o validación final de los mapas. Esto, se verá reflejado en estimaciones más precisas y mejores de la deforestación, poniendo a México a la vanguardia a nivel internacional en materia de su monitoreo.

Fuentes

- Averchenkova A. & S. Guzman Luna. *Mexico's General Law on Climate Change: Key achievements and challenges ahead*. London, Grantham Research Institute on Climate Change and the Environment and Centre for Climate Change Economics and Policy, London School of Economics and Political Science, 2018.
- Argotty, F., E. Fung, P. Imbach, C. Medellín, O. Carrillo, E. Larios, G. Islas, D. Lopez & M. Muñoz. *Estado del monitoreo forestal en Latinoamérica y el Caribe: tipos de iniciativas y uso de tecnologías. Nota técnica núm. IDB-TN-01690*. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE)-Banco Interamericano de Desarrollo (BID), 2019, 82 p.
- Breiman, L. "Random forests", en: *Machine Learning*. 45(1), 2001, pp. 5-32.
- Comisión Nacional Forestal (CONAFOR). *Nivel de Referencia de Emisiones Forestales de México (2007-2016)*. En rev. 2020 (DE) https://redd.unfccc.int/files/nref_2007-2016_mexico.pdf
- _____. *Inventario Nacional Forestal y de Suelos Informe 2004-2009*. Zapopan, México, 2012.
- Covalada, S., S. Aguilar, A. Ranero, I. Marín & F. Paz. *Diagnóstico sobre determinantes de deforestación en Chiapas*. México, Alianza México para la Reducción de Emisiones por Deforestación y Degradación (Alianza México-REDD+)-Agencia de los Estados Unidos para el Desarrollo Internacional (USAID), 2014.
- Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). *Directrices voluntarias sobre monitoreo forestal nacional*. Roma, Italia, 2017, 90 p.
- Gebhardt, S., T. Wehrmann, M. A. Muñoz-Ruiz, P. Maeda, J. Bishop, M. Schramm, R. Kopeinig, O. Cartus, J. Kellendorfer, R. Ressler, M. Schmidt & L. A. Santos. "MAD-MEX: automatic wall-to-wall land cover monitoring for the Mexican REDD-MRV program using all Landsat data", en: *Remote Sensing*. 6(5), 2014, pp. 3923-3943.
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau & R. Moore. "Google Earth Engine: Planetary-scale geospatial analysis for everyone", en: *Remote Sensing of Environment*. 202, 2017, pp. 18-27.
- Instituto Nacional de Estadística y Geografía (INEGI). *Conjunto de datos vectoriales de uso del suelo y vegetación. Serie VI. (Capa Unión). Escala 1: 250 000*. México, INEGI, 2016.
- Intergovernmental Panel on Climate Change, Suiza (IPCC). *IPCC guidelines for national greenhouse gas inventories. Prepared by the National Greenhouse Gas Inventories Programme*. Kanagawa, Japan, IGES, 2006.
- "Ley Nacional de Desarrollo Forestal Sustentable (LGDFS) 2003", en: *Diario Oficial de la Federación*. Última reforma publicada el 24 de enero de 2017 (DE) <http://extwprlegs1.fao.org/docs/pdf/mex36056.pdf>
- Nielsen, A., K. Conradsen & J. Simpson. "Multivariate alteration detection (MAD) and maf postprocessing in multispectral bitemporal image data: New approaches to change detection studies", en: *Remote Sensing of the Environment*. 64(1), 1998.
- Olofsson, P., G. M. Foody, S. V. Stehman & C. E. Woodcock. "Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation", en: *Remote Sensing of Environment*. 129, 2013, pp. 122-131.
- Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock & M. A. Wulder. "Good practices for estimating area and assessing accuracy of land change", en: *Remote Sensing of Environment*. 148, 2014, pp. 42-57.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, P. Perrot & J. Vanderplas. "Scikit-learn: Machine learning in Python", en: *Journal of Machine Learning Research*. 12, octubre de 2011, pp. 2825-2830.
- R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing, 2020 (DE) URL <https://www.R-project.org/>
- Romijn, E., C. B. Lantican, M. Herold, E. Lindquist, R. Ochieng, A. Wijaya, D. Murdiryaso & L. Verchot. "Assessing change in national forest monitoring capacities of 99 tropical countries", en: *Forest Ecology and Management*. 352, 2015, pp. 109-123.
- Schrödle, B. & L. Held. "Spatio-temporal disease mapping using INLA", en: *Environmetrics*. 22 (6), 2011, pp. 725-734.
- Sepal-FAO. *System for earth observations, data access, processing & analysis for land monitoring*. 2018 (DE) <https://sepal.io/>, consultado en marzo de 2018.
- Strahler, A. H., L. Boschetti, G. M. Foody, M. A. Friedl, M. C. Hansen, M. Herold & C. E. Woodcock. "Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps", en: *European Communities, Luxembourg*. 51(4), 2006.
- Verbesselt, J., R. Hyndman, G. Newnham & D. Culvenor. "Detecting trend and seasonal changes in satellite image time series", en: *Remote Sensing of Environment*. 114(1), 2010, pp. 106-115.

La informalidad laboral en las entidades de México en el siglo XXI: posibles factores explicativos

Labor Informality in Mexican Entities during the 21st Century: Possible Explanatory Factors

Mario Camberos Castro y Joaquín Bracamontes Nevárez*



Mexican musicians on the streets/Scharfsm86/istock

* Centro de Investigación en Alimentación y Desarrollo (CIAD) A. C., mcamberos@ciad.mx y joaco@ciad.mx, respectivamente.

En este artículo, el objetivo es analizar el problema de la informalidad laboral con el fin de identificar sus factores determinantes en México y en las entidades que forman el país. Para ello, se aplica un modelo econométrico de mínimos cuadrados ordinario multivariado en tres etapas para probar la relación entre esta y las variables de mercado, institucionales y de otro tipo utilizando los datos de la Encuesta Nacional de Ocupación y Empleo (ediciones 2015-2019) e información proveniente de otras fuentes. Los resultados principales llevan a concluir que la informalidad no es del todo voluntaria, más bien resulta de una mezcla de factores de mercado e institucionales, por lo que se desprende que las medidas para reducirla deben estar orientadas en ambos sentidos.

Palabras clave: informalidad; mercado; instituciones.

Recibido: 11 de abril de 2020.
Aceptado: 5 de julio de 2020.

Introducción

En un principio, la informalidad era referida como el sector informal, un segmento del mercado laboral al lado del formal caracterizado por no cumplir con las regulaciones establecidas de manera legal, sobre todo de las leyes hacendaria y del trabajo. Además de observar condiciones de precariedad en los espacios donde esta se desarrollaba, cuyas consecuencias eran baja productividad, falta de competitividad y salarios de bajo monto (ILO, 2012). Esta definición inicial ignoraba tanto a millones de trabajadores que laboraban en empresas registradas como formales, pero los mantenían en la informalidad —sea por no pagar prestaciones de salud, vivienda o fondo de pensiones—, como a los cientos de miles de empleados domésticos (mujeres y hombres) que no trabajaban en negocios formales ni en informales, sino que lo hacían en hogares (Negrete, 2010).

Es en este marco que se propone como objetivo investigar sobre la informalidad en las entidades de México y conocer los posibles factores que la explican diferenciando entre factores de mercado [a) tasa de crecimiento del Producto Interno Bruto

In this article, the objective is to analyze the problem of labor informality in order to identify its determining factors in Mexico and in the entities that make up the country. To this end, 3-step, multivariate Ordinary Minimum Squares (MCO) econometric model is applied to test the relationship between this and market, institutional and other variables using data from the National Survey of Occupation and Employment (2015-2019 editions) as well as information from other sources. The main results lead to the conclusion that informal labor is not entirely voluntary, but rather results from a mix of market and institutional factors, and therefore it follows that measures to reduce it should be oriented in both directions.

Key words: informality; market; institutions.

(PIB), b) nivel de salarios y c) tasa de desempleo] e institucionales [d) no pago de impuesto sobre la renta, e) no pago de servicios de salud y f) no contribución al pago de pensiones, además de g) corrupción y h) impunidad].

En este trabajo discutimos, primero, sobre la conceptualización de la informalidad y la forma de estimarla a la luz de trabajos de autores e instituciones reconocidos, como Tokman, quien mucho ha contribuido su estudio en América Latina (2001a, b; 1993), la Organización Internacional del Trabajo (OIT) y el Instituto Nacional de Estadística y Geografía (INEGI); luego, hacemos un análisis retrospectivo para encontrar sus orígenes, evolución y causas en México; después, mostramos la metodología utilizada para estimarla, incluido el modelo econométrico aplicado para probar los factores explicativos; enseguida, presentamos una descripción de los distintos tipos de informalidad en las entidades destacando, en cada caso, aquellas que la encabezan; y, por último, analizamos los resultados de la aplicación del modelo econométrico y las pruebas que lo avalan, para terminar con una discusión y las conclusiones.

Conceptualización como economía informal y su medición

El origen del concepto de informalidad como un sector de trabajo data de 1972 (OIT) y se identificó en un estudio para Kenia, África (Hart, 1973; Portes y Haller, 2004; Tokman, 2001a; Rivera y Sánchez, 2018), en el que se refería a un empleo por cuenta propia (Portes y Haller, 2004) y que, en América Latina, la Comisión Económica para América Latina y el Caribe (CEPAL) lo había reconocido en la década de los 50 del siglo pasado como parte de la marginalidad (Cortés, 2000) en el mercado laboral capitalista, en tanto que Tokman lo consideró, 30 años después, como el velo que ocultaba el problema del desempleo en América Latina (1993).

En un inicio, la OIT la definió, con base en Hart (1973), como una "... modalidad urbana..."; un sector de la economía formada, sobre todo, por trabajadores por cuenta propia y microempresas que operaban fuera de la formalidad, es decir, el sector informal. Después, la misma OIT, en su Conferencia Internacional del Trabajo realizada en el 2002 (Rei y Bhattacharya, 2008; Negrete, 2010), la cuestionó por no considerar a empleados en empresas formales que carecían de prestaciones sociales a pesar de que la ley lo exigía y excluía, así como también —como señala Tokman (2001b), autoridad en la materia— a trabajadores domésticos (mujeres y hombres), a quienes han laborado por cuenta propia, y a microempresas de menos de cinco personas; con ello, dio un giro a la conceptualización al incluir tanto a empresas como trabajadores que formarían ya no un sector informal, sino una economía informal cualitativa y cuantitativamente diferente, definiéndola como aquella en la que se desempeñan empleados y negocios que están fuera de la formalidad (ILO, 2012) por no cumplir una o varias de las siguientes condiciones: a) no pagar impuestos, b) no regirse por contratos de trabajos individuales o colectivos y c) no otorgar prestaciones contempladas en las leyes generales de seguridad social, como salud, pensión para el retiro por invalidez o vejez, seguro de desempleo, vivienda, entre otras. En este punto hay que hacer notar la preocupación de autores reconocidos en el tema

por conceptualizar la informalidad —Maloney (2004), Rei y Bhattacharya (2008), por ejemplo— como economía sombra, en alusión a que —agregamos nosotros— constituye un reflejo de la real, que es la formal.

La estimación de la economía informal puede hacerse con base en tres criterios: a) la proporción del producto de esta respecto al Producto Interno Bruto (PIB) total, como lo trabajaron Rei y Bhattacharya (2008) de la Organización Internacional del Trabajo (OIT) en un estudio para países de bajos y medianos ingresos; y b) la que considera a las empresas de menos de cinco trabajadores, como lo realizaron Rodríguez (2016) para Colombia y Jiménez (2012) para varias naciones de América Latina. Estas dos medidas dejan fuera una parte de los informales en el sector formal que, de ser adoptadas, sesgaría los resultados; por lo tanto, acudimos a una tercera, c), que se refiere a la suma de personas empleadas en el sector informal, más las que se encuentran en el formal que no cumple con las obligaciones legales consideradas por la OIT arriba citadas, como proporción del empleo total, y es utilizada por instituciones como la propia OIT (ILO, por sus siglas en inglés, 2011) y el INEGI (2013), así como por diferentes autores (Maloney, 2004; Lehman y Muravyev, 2012; Fernández y Almagro, 2008; y Camberos y Bracamontes, 2015b), razón por la cual es la medida de la informalidad que adoptaremos en este trabajo.

Considerando el último criterio, su estimación, de acuerdo con la OIT, se puede realizar a partir de su definición no en un opuesto a la de formalidad, como si se tratase de mercados segmentados, sino como mercados que se entrelazan, tal cual lo concibe Tokman (2001a) que tiene su equivalente a la informalidad que existe también en un mercado formal. Por ello, la definición de trabajo informal, base de la definición de economía informal, está compuesto por tres tipos de empleos: a) trabajadores informales en empresas formales, b) trabajadores informales en negocios informales y c) trabajadores familiares pagados que producen bienes para autoconsumo o empleados domésticos (ILO, 2012), los

cuales se adoptan en México y que ha captado cada trimestre la Encuesta Nacional de Ocupación y Empleo (ENOE) del INEGI (2015a, 2017 y 2019).

Origen de la informalidad en México

En México, si bien es cierto que esta condición del empleo, tal como se define en algunas de las modalidades actuales, ha existido sobre todo en el campo y en los negocios familiares a lo largo de prácticamente el siglo XX, el concepto de informalidad como sector informal data de 1971 (Rivera y Sánchez, 2018), los registros estadísticos del INEGI y de diferentes autores como Alba (1993), Lustig (1993) y Boswarth (1998) de esta se remontan apenas a la década de los 80, en la modalidad de sector informal. Antes de esta fecha no existían estadísticas oficiales sobre la informalidad en México. No obstante, es en el 2005 que en el país se adopta de manera oficial el concepto de sector informal con fines estadísticos en la ENOE 2005 (Negrete, 2010).

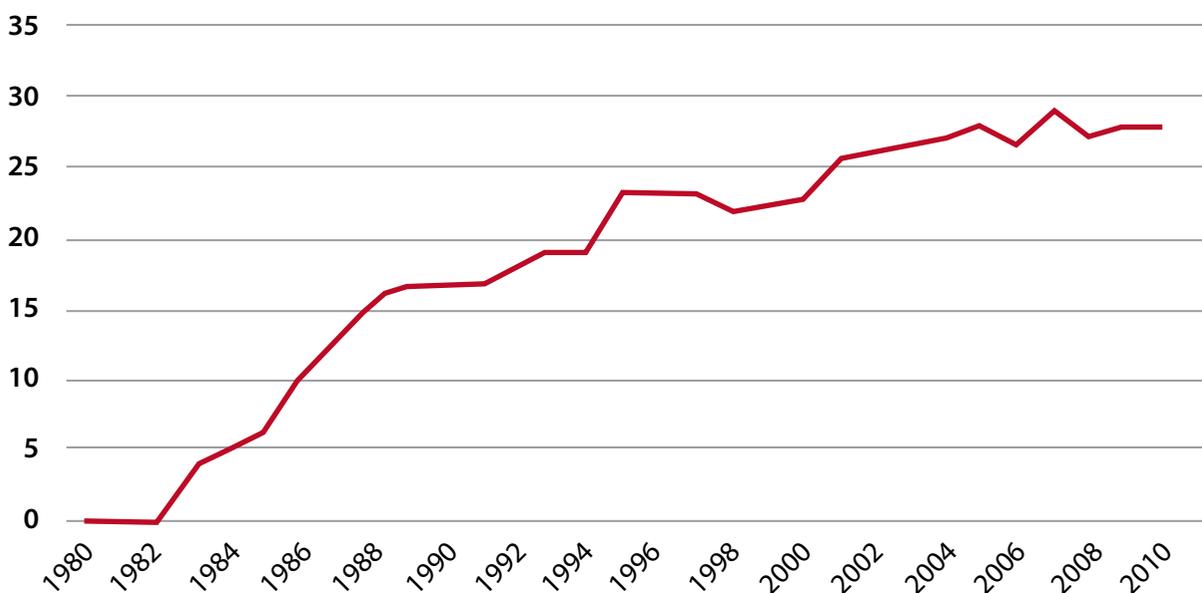
Su evolución se muestra en la gráfica 1, en la que se observan claramente los cambios en las pendientes positivas de la curva a lo largo de tres

décadas, las cuales se intensifican con las crisis recurrentes de la década de los 80 y la globalización, pero que no se detienen con el ingreso de México al Tratado de Libre Comercio (TLC) ni con la década del cambio; más bien, parece haber tomado carta de naturalización en el mercado laboral mexicano hasta alcanzar al sector formal (INEGI, 2017).

Sobre este proceso de informalización, diferentes autores han dado cuenta de esta situación en México para la década de los 80, como los arriba citados, además de Soria (2001) para la de los 90; tanto Fernández y Almagro (2008) como Tello (2010) agregan la informalidad creciente en la primera década del siglo XXI y, por último, Camberos y Bracamontes (2015a) revisan la serie más amplia hasta la segunda década del siglo XXI. Los datos de la informalidad que recogen el efecto de la crisis 2008-2009 y consideran también a los informales en el sector formal del mercado laboral son, por un lado, de Tello (2010), quien la estima para el 2008 en 55 % de trabajadores urbanos en empresas formales, pero al margen de toda prestación de ley y, por el otro —en años más recientes—, de la OIT, la cual estimó que llegaba hasta 60 % (ILO, 2015 y 2011) y del INEGI (2013), si bien se ha reducido a 57 % en la actualidad (INEGI, 2019; ILO, 2019).

Gráfica 1

Informalidad en México, 1980-2010



Fuentes: Camberos y Bracamontes, 2015b; INEGI, 2013.

Una consecuencia del crecimiento de la informalidad en México a lo largo de las cuatro últimas décadas es su contribución a la pobreza de millones de trabajadores y sus familias, porque los ha dejado al margen de los beneficios de un sistema de salud y los ha privado de una pensión para el retiro, que los ha condenado a un futuro incierto.

Factores que la explican

Es importante diferenciar entre las causas que originaron la informalidad y los factores que ahora explican el porqué prefieren los trabajadores y las empresas mantenerse en ella. Las primeras son de dos tipos, aquellas que la consideran voluntaria, como Maloney (1999), para quien los mercados de trabajo formal e informal no son segmentados, sino integrados. Él se apoya en los resultados de una encuesta de autoempleados, realizada en México, que contestaron a la pregunta de las razones del porqué dejaron su último empleo: 37 % los hombres para ser independientes y 46 % de las mujeres por matrimonio, como los porcentajes mayores de ocho tipos de respuesta, de donde se puede desprender que son informales voluntariamente (Maloney, 2004), si bien en el caso de las mujeres no es muy claro. El agente considera los atributos de los empleos informales como los horarios flexibles, el trabajo independiente, ser su propio jefe como microempresario y el no valorar los beneficios de la seguridad social (Maloney, 1999) para decidir emigrar de manera voluntaria a la informalidad. A estas ventajas agrega el autor la oportunidad de convertirse en microempresario en los países en desarrollo, como América Latina (Maloney, 2004), o de acuerdo con otro autor, haber encontrado ventajas importantes en el sector informal para combatir el desempleo e incrementar el ingreso mediante pequeñas empresas en países de África, como muestra un estudio para Ghana (Debrah, 2007).

El segundo tipo corresponde a la corriente de pensamiento que argumenta que esta condición laboral es involuntaria, originada por los ciclos y las crisis de la economía (Shapiro, 2017; Camberos y Bracamontes, 2015a; Lustig, 1993 y Alba, 1993) que

en su fase descendente reduce la generación de empleos y cuando entra en crisis hace perder millones de puestos de trabajo, situación que obliga a los trabajadores y microempresarios a buscar la informalidad como una vía de subsistencia mientras se reactiva el mercado laboral, pero cuando esto sucede, ante la competencia de quienes se incorporan por primera vez a este con mayor calificación y menor salario, prefieren permanecer en la informalidad.

En este tema, en un estudio comparativo de Lehman y Muravyev (2012) para países de América Latina y Europa Central se utilizaron los impuestos, la legislación salarial, el seguro del desempleo, la duración del seguro y el grado de sindicalización como factores explicativos. En otro trabajo, Rei y Bhattacharya (2008) consideraron una gran cantidad y diversidad de variables *proxy* para acercarse a estos, que se dividen en tres grupos: a) institucionales para medir la calidad, como índices de regulación del mercado de trabajo y de los negocios, además de derechos de los trabajadores y de los sociales de las mujeres, así como un indicador del nivel de democracia política y un conjunto de estos de gobernanza, entre otros; b) macroeconómicos, como el PIB per cápita de 111 países, la relación entre la formación bruta de capital y el gasto del gobierno como proporción del PIB, el porcentaje de contribución de la agricultura al PIB y la proporción de matrícula en primaria del total de niños en edad de asistir a la escuela, por ejemplo; y c) de política, como índices de nivel global tarifario, libertad de comercio y de la calidad del uso de la política monetaria; así como la relación del comercio total y de la inversión extranjera directa respecto al PIB; solo por mencionar algunas variables.

En este marco, Maloney (2004) pone especial énfasis como factores explicativos de la informalidad a lo que califica como rigideces laborales, como los impuestos y los costos laborales por el pago de prestaciones; considera, también, el grado en educación, y encuentra una relación inversa entre este y el índice de informalidad del empleo. Cabe decir que la posición acerca de las rigideces salariales ha sido la dominante actualmente, la comparten los

autores aquí citados de acuerdo con sus hallazgos, argumentos que a nuestro juicio alimentaron el sistema de contratación *outsourcing*, que ha resultado beneficioso para un grupo de empresarios; sin embargo, ha sido perjudicial para la mayoría de los trabajadores e inhibido el crecimiento económico.

En esta investigación se utilizan tres tipos de factores: 1) de mercado: a) PIB per cápita, b) tasa de desempleo, c) nivel del salario y d) tipo de desempleo voluntario (1) o involuntario (0); 2) institucionales: e) impuestos federales y estatales, f) aportaciones a la seguridad social, g) contribución al fondo de pensiones y h) condición de contratación; y 3) otros: i) grado de sindicalización, j) corrupción y h) impunidad. Si bien es cierto que los factores anotados no agotan la explicación de los asociados a la informalidad, toda vez que pudieran agregarse más (como en el caso de los de mercado, el grado de exposición o apertura de la economía al exterior, el nivel de inversión, la productividad o el grado de competitividad; entre los institucionales, desglosar la variedad de impuestos existentes, además del impuesto sobre la renta (ISR) y el de nóminas estatales, como también las prestaciones sociales, así como las aportaciones a la salud o, también, incorporar la gran variedad de otros factores que contienen los trabajos arriba citados), el problema estriba en que no existen series estadísticas disponibles para todas y cada una de las entidades que permitan probar la relación entre informalidad y los factores explicativos a nivel de país, como los casos de las naciones de la Unión Europea y algunas de América Latina (como México) y el de la Organización para la Cooperación y el Desarrollo Económicos (OCDE).

Metodología y datos

El modelo utilizado es del tipo recomendado por la gran mayoría de los autores reconocidos, como Cameron y Trivedi (2010) y Greene (2008), el de mínimos cuadrados ordinario (MCO), aunque cabe aclarar que el tamaño de la muestra se reduce a 31 observaciones cuando se prueba con las variables de mercado e institucionales porque Guanajuato

no reporta el salario, mientras que cuando se corre con todas estas se excluye a Michoacán de Ocampo por carecer del índice de sindicalización, y se probó en tres etapas con dos hipótesis por medio de las variables *proxy*, como se indica a continuación.

En la etapa 1, para comprobar la primera hipótesis de que existe una relación inversa entre la informalidad y las variables de mercado (ecuación 1), tomamos los logaritmos naturales (\ln) del PIB per cápita y el salario, más la tasa de desempleo y una variable cualitativa, expresada como relación desocupación voluntaria/involuntaria, con signo positivo, la cual confirmaría que la informalidad es voluntaria.

En la etapa 2 (ecuación 2), que corresponde a la segunda hipótesis, la cual supone que las variables institucionales aumenten la informalidad, de acuerdo con la literatura revisada, agregamos al modelo los impuestos, junto con el resto de las variables institucionales, como el no tener contrato de trabajo por escrito, la falta de atención a la salud y la no contribución al sistema de pensiones, que deberían estar relacionadas positivamente con esta condición laboral. En este ejercicio, probaremos también si las variables de mercado mantienen su consistencia estadística o si serán las institucionales las que determinen la incidencia de la informalidad en las entidades de México, la que, por lo tanto, confirmaría que es voluntaria.

En la tercera etapa (ecuación 3) utilizaremos otras variables, como el grado de sindicalización, que se supone contribuye a la informalidad empresarial porque aumenta el costo de contratación, de la mano de la corrupción y la impunidad; por lo tanto, se espera una relación positiva con la informalidad.

Enseguida mostramos las tres ecuaciones:

$$Inf_{ij} = \alpha_0 + \sum_{ij} \beta_i X_{ij} + Div + \varepsilon_{ij} \quad \text{con } \alpha, \beta > 0 \quad (1)$$

$$Inf_i = \alpha_0 + \sum_j \beta_j X_{ij} + Div + \sum_n \lambda_n linkin + \varepsilon_{ij} \\ \text{con } \alpha, \beta, \lambda > 0 \quad (2)$$

$$Infi = \alpha_0 + \sum_j \beta_j X_{ij} + Div + \sum_n \lambda_n kin + \sum_n \delta_n \theta_{in} + \varepsilon_{ij} \quad \text{con } \alpha, \beta, \lambda, \delta > 0 \quad (3)$$

definidas las variables como:

Infi: índice de informalidad de la entidad $i = 1, \dots, 32$ para un solo año.

Xij: conjunto de variables de mercado $j = 1, 2, 3$; a) PIB per cápita de la entidad, b) salario por día y c) tasa de desempleo.

Div: variable *dummy* para tipo de empleo voluntario (1)/involuntario (0) en cada entidad.

Kin: conjunto de variables institucionales $n = 1, 2, 3, 4$; a) ISR, porcentaje de la población que no paga impuesto sobre la renta; b) porcentaje de la población ocupada en cada entidad sin servicio de salud; c) porcentaje de la población sin fondo de pensiones; y d) porcentaje de población ocupada sin contrato.

δin: otras variables $n = 1, 2, 3$; a) porcentaje de la población ocupada sindicalizada en cada entidad, b) índice de corrupción en cada entidad y c) índice de impunidad de cada entidad.

εij: error estocástico.

Por último, es importante agregar que aplicamos el método de corte transversal, que consiste en comparar la situación de los 32 estados en un mismo punto del tiempo (junio del 2015), por la dificultad de encontrar más de una observación para todas y cada una de las entidades que pudiera hacerlas comparables en diferentes años.

Fuentes de información

Para completar la investigación, fue necesario acudir a una diversidad de estas para las 32 entidades, cuyos datos se registran en el cuadro 1 y el Anexo, las cuales a continuación señalamos:

1. Los índices de empleo formal e informal se obtuvieron de la ENOE (INEGI, 2015a).
2. Variables de mercado: a) el PIB de las entidades se obtuvo de las estadísticas del INEGI

(2015b); b) los salarios mínimos base determinados por el Instituto Mexicano del Seguro Social (IMSS) se tomaron de la ENOE (INEGI, 2015a); c) las tasas de desempleo de las entidades, de la ENOE (INEGI, 2015a); d) la *dummy* desempleo voluntario/involuntario fue construida con base en su conceptualización y adopta valores 1 desempleo voluntario y 0 desempleo involuntario.

3. Variables institucionales: a) impuestos (a.1) ISR, porcentaje de población que no paga; se consideró como tal el índice de trabajadores en el sector informal que registra la ENOE (INEGI, 2015a); b) porcentaje de la población ocupada sin contrato de trabajo por escrito tomado de la ENOE (2015a); c) porcentaje de población sin servicio de salud; d) porcentaje de población sin fondo de pensión. Estas dos últimas son consideradas como los dos factores institucionales en los que se ha centrado el programa de formalización que anunció la Secretaría del Trabajo y Previsión Social (STPS, 2016), ambos tomados del *Anuario estadístico y geográfico por entidad federativa*, cuadro 5.1 (INEGI, 2015c).
4. Otras variables: a) porcentaje de sindicalización, considerado como un factor de rigidez del mercado que contribuye a la informalidad, tomado de INEGI (2015a); b) Índice Nacional de Corrupción y Buen Gobierno por entidad, tomado de Transparencia Internacional (2010) y c) impunidad, del Índice Global de Impunidad México (Leclerq y Rodríguez, 2016).

Informalidad en las entidades

Un primer factor para calificar la informalidad institucional es el no pago de impuestos. Originalmente, esta se refería a la población que ha laborado en el sector informal y que lo ha hecho de manera voluntaria para no pagar el ISR, el cual es importante para el Estado mexicano por cuestiones recaudatorias, su principal fuente de ingresos que, de acuerdo con el cuadro 1 (columna 3), en el país llegaron a 27.2 %, lo cual significaría que un porcentaje mayor de trabajadores

son involuntariamente informales, pues trabajan en empresas o negocios propios formales, pero buena parte de estos (30 %), que son los que trabajan en el sector formal, pero no reciben alguna o varias prestaciones, son informales contra su voluntad.

En las entidades, la mayor parte de las empresas y trabajadores paga ISR, pero en estados con altos índices de informalidad, como el de México y la Ciudad de México (antes Distrito Federal), la mayor parte lo son, por no pagar el ISR, 30 y 35 %, respectivamente. Es probable que estos altos índices de informalidad por el no pago se deba a la reforma impositiva vigente, que parece indicar poco impacto en la reducción de esta y que, más bien, persiga resultados recaudatorios.

La no pertenencia a un sistema de salud es el factor con menor porcentaje de informalidad, representó 22.73 % a nivel nacional; tal vez, su bajo nivel se explica por la creación del sistema del Seguro Popular, que incorporó a millones de informales, quienes carecían de atención médica. En este rubro, Puebla y Michoacán de Ocampo, con poco más de 31 %, y el estado de México, con casi 28 %, presentan la mayor proporción de habitantes sin servicio médico, entidades con altas concentraciones de población que pudieran explicar esa carencia de atención en salud; cabe decir que la población de la Ciudad de México registra niveles de 26 % sin este servicio, en tanto que estados de baja densidad (como Colima, Campeche y San Luis Potosí) muestran los menores porcentajes de carencia.

La no formalización de la relación trabajador-empleador, que pone en desventaja al primero, se refiere a la ausencia de un contrato; esta situación registra que poco más de 45 % de los trabajadores a nivel nacional carecen de contrato por escrito, siendo Puebla, Michoacán de Ocampo, Oaxaca y Chiapas los más altos con cifras mayores a 60 %, en tanto que Baja California (31), Baja California Sur (37), Chihuahua (28), Sonora (36), Nuevo León (24), Coahuila de Zaragoza (28) y Querétaro (37) observan los menores porcentajes del personal ocupado sin contrato.

El no pertenecer a un sistema de pensiones es otro tipo de informalidad que está relacionado, más bien, con el futuro del trabajador, el cual corre el riesgo de no contar con un ingreso que le permita subsistir o llevar una vida cómoda cuando ya no le sea posible trabajar. Este indicador se registra en la columna 6 del mismo cuadro y es el factor que contabiliza la mayor informalidad por separado, siendo Chiapas (63.10 %), Oaxaca (58.31%) y Guerrero (50.30%) las entidades con la mayor proporción de empleados en esa situación y donde existe el mayor riesgo de los trabajadores y sus familias de caer en pobreza extrema al final de su vida laboral; por el contrario, Coahuila de Zaragoza, Nuevo León y la Ciudad de México, con 13.85, 14.99 y 19.13 % ofrecen el panorama más previsor de todas.

La STPS (2016) considera prioritarios para la reducción de la informalidad en México el pertenecer a un sistema de pensiones y tener acceso a una institución de salud pública, posición apoyada por la Secretaría de Hacienda y Crédito Público (SHCP, 2013), el IMSS (2015) y la OIT (2014). Los resultados conseguidos al respecto en ambos rubros en México de 57 % se pueden dimensionar comparándolos con los registrados por Colombia, donde los informales alcanzaron 55.4 % en el 2014 (Rodríguez y Andrés, 2016) y un porcentaje menor en varios países de América Latina, como Brasil, Chile y Argentina (Jiménez (2012), quizá producto del impulso a las micro y pequeñas industrias en estos (Ferraro y Rojo, 2018).

Para completar el análisis de la informalidad, agregamos tres factores que hasta ahora no han sido considerados como importantes en esta, pero que a nuestro juicio han constituido formas coercitivas no económicas ni institucionales, que inciden en el clima laboral que ha permeado las relaciones entre el Estado y la sociedad desde hace décadas y que se ha agudizado recientemente. En primer lugar, la sindicalización, que en México ha sido utilizado por las centrales de trabajadores de un partido como presión para el empleador y pudiera abonar a la informalidad, porque aumenta los costos de la formalidad y orilla a quienes generan empleos o se autoemplean para migrar hacia esta condición laboral.

Cuadro 1

Tipos de factores de la informalidad en las entidades de México*

Entidad	Tasalnf** (%)	SecInfor* (%)	No pagan ISR**(%)	Sin contrato	inf x salud (%)***	inf x pen (%)****	Sindicalización (%) (+)	Índice de Corrupción	Índice de Impunidad
Nacional	57.81	27.20	27.20	45.34	22.73	34.69	23.86		
Aguascalientes	43.67	22.64	22.64	36.02	16.03	26.48	19.49	4.7	68.37
Baja California	40.15	19.89	19.89	30.60	22.34	25.44	10.42	7.1	74.42
Baja California Sur	40.40	16.97	16.97	37.26	15.95	24.13	25.92	1.8	70.74
Campeche	61.02	26.65	26.65	48.48	11.79	46.15	26.94	8.3	47.22
Coahuila de Zaragoza	35.07	20.22	20.22	28.58	18.13	13.85	29.76	8.3	72.93
Colima	54.40	19.22	19.22	50.95	13.43	32.14	31.26	10.4	67.10
Chiapas	78.84	23.49	23.49	61.43	20.24	63.10	25.13	7.6	59.96
Chihuahua	36.04	15.36	15.36	28.00	18.26	23.56	12.85	7.1	59.94
Ciudad de México	51.28	30.03	30.03	35.33	26	19.13	19.35	17.9	56.97
Durango	55.98	23.92	23.92	43.66	18.95	29.00	22.23	3.9	73.97
Guanajuato	58.30	28.07	28.07	52.45	18.39	40.60	24.04	7.6	65.15
Guerrero	79.50	33.42	33.42	58.98	27.16	50.30	35.84	16.0	73.19
Hidalgo	71.60	35.25	35.25	59.39	19.78	49.98	23.78	11.6	67.90
Jalisco	50.13	23.29	23.29	43.07	24.26	26.93	19.61	10.3	67.45
México	59.28	35.98	35.98	46.41	27.68	29.40	18.69	16.4	76.48
Michoacán de Ocampo	71.13	31.94	31.94	62.68	31.12	43.76	33.29	8.4	67.71
Morelos	66.37	34.70	34.70	57.27	22.1	41.69	30.69	6.9	69.19
Nayarit	62.08	26.41	26.41	52.95	21.17	35.14	34.66	4.4	50.42
Nuevo León	36.89	20.42	20.42	24.42	16.16	14.99	21.37	9.1	71.63
Oaxaca	80.00	32.41	32.41	61.98	22.61	58.31	38.54	13.4	71.63
Puebla	72.41	27.86	27.86	64.21	31.53	42.18	27.85	7.6	69.19
Querétaro	45.37	19.88	19.88	37.32	18.07	36.04	18.44	6.7	72.24
Quintana Roo	47.27	20.68	20.68	37.30	20.59	26.09	26.97	5.9	76.71
San Luis Potosí	58.25	24.05	24.05	39.74	12.38	45.12	34.17	5.6	56.86
Sinaloa	50.90	21.33	21.33	48.04	15.94	27.59	24.07	9.9	69.34
Sonora	45.37	22.06	22.06	34.62	17.75	21.49	26.27	8.4	59.94
Tabasco	63.66	28.71	28.71	52.16	20.79	46.83	29.46	10.8	69.10
Tamaulipas	47.32	21.04	21.04	35.49	19.07	26.08	41.69	6.5	73.38
Tlaxcala	73.22	39.46	39.46	57.17	23.21	48.19	39.07	10.2	67.69
Veracruz	68.10	25.28	25.28	59.71	26.92	41.47	37.29	6.9	72.38
Yucatán	62.33	32.25	32.25	48.27	16.11	38.80	19.31	5.0	72.80
Zacatecas	66.02	25.20	25.20	55.45	17.21	48.94	37.79	6.1	67.10

* Datos del segundo trimestre del 2015.

** Estimación de acuerdo con la definición del INEGI (2016): trabajadores en el sector informal más trabajadores del sector formal sin prestaciones, como servicio de salud o pertenecer a un sistema de pensión de acuerdo con STPS (2016).

*** Es el (%) de población que no tiene atención de salud de ningún tipo. Se refiere a los que no cuentan con ningún sistema de pensión, aunque tengan atención de salud del Seguro Popular y/o de la SSA.

(+) INEGI, 2015a.

Fuente: estimaciones propias con base en 1) INEGI, 2015 a, b y c, cuadro 5.1, 2) Transparencia Internacional, 2010 y 3) Leclercq y Rodríguez, 2016.

Al respecto, pudiera esperarse un elevado índice de correlación entre entidades con mayor porcentaje de sindicalización y mayor informalidad en general, pero menor informalidad por pensión, porque es una de las prestaciones que exige el sindicato; sin embargo, los resultados del cuadro 1 apoyan la primera suposición porque Oaxaca, Veracruz de Ignacio de la Llave, Zacatecas, Tlaxcala y Tamaulipas registran los más elevados índices de sindicalización, alrededor de 40 %, como también las tasas más altas de informalidad —con excepción de Tamaulipas (45 %)—, superiores a 60 % (en el extremo se encuentra Oaxaca con 80 %). En cuanto a la suposición de que a una mayor sindicalización habrá menor informalidad por pensión, los resultados del cuadro 1 parecen decir lo contrario, que a menor sindicalización menor es la informalidad por pensión, como sucede con Aguascalientes, Baja California, Baja California Sur, Nuevo León y Chihuahua, la excepción es Coahuila de Zaragoza.

Los dos siguientes factores son corrupción e impunidad; su relación con la informalidad estriba en que ambos elevan los costos de mantenerse en la formalidad que imponen las instituciones y se considera que a mayor corrupción e impunidad será más elevada la informalidad. En esta noción, la calificación que ha recibido México en el Índice Internacional de Corrupción (Transparencia Internacional, 2010) es alta: de los más corruptos en el *ranking* mundial, mientras que en el de Impunidad se encuentra en el segundo lugar, solo por detrás de Colombia (Leclerq y Rodríguez, 2016).

El análisis de las entidades basado en las dos últimas columnas del cuadro 1 de los índices correspondientes revela que los de corrupción más altos, en ese orden, son de la Ciudad de México, México, Guerrero y Oaxaca. De los cuatro estados mencionados con esa situación, en tres (Oaxaca, Guerrero y México) se observan altas tasas de informalidad, por encima de la media nacional, lo cual puede ser indicativo de la corrupción como factor que contribuye a la informalidad; por el contrario, los valores más bajos se registran en Yucatán, Nayarit, Durango y Aguascalientes, pero con excepción de este último, los tres restantes registran elevadas tasas de

informalidad, por lo cual no se puede concluir que bajos índices de corrupción se corresponden con baja informalidad. En impunidad, destacan con el mayor valor Quintana Roo, México, Baja California y Durango, y con los menores niveles, Campeche, Nayarit, Ciudad de México y San Luis Potosí, donde no se encontró correspondencia entre informalidad e impunidad.

Una última forma de mostrar la informalidad y probar cuáles son los factores determinantes —si los de mercado, los institucionales u otros— es acudiendo a la estadística y la econometría, lo que haremos enseguida.

Un modelo econométrico para probar los factores explicativos

En principio, analizamos la correlación entre las variables explicativas que componen el modelo para probar que no presenta el problema clásico de multicolinealidad, tal como lo recomiendan Cameron y Trivedi (2010) y Greene (2008); luego, aplicamos el de MCO en tres etapas para probar las hipótesis con la información de las variables para el 2015, excepto para los índices de Corrupción e Impunidad, los cuales se tomaron de los años más cercanos (2016 y 2010) con la información disponible.

Correlación entre las 12 variables

Esta se analiza en el cuadro 2: las cuatro de mercado, cinco institucionales y otras tres como se definieron antes en el apartado metodológico. Se observa que entre las institucionales no existe alta colinealidad, excepto sin pensión y sin contrato, donde alcanza 0.85, y la *dummy* desempleo voluntario/involuntario y tasa de desempleo 0.82 y signo negativo es síntoma de que el desempleo es involuntario; sin embargo, la correlación entre los factores institucionales y las variables de mercado son bajas y, en su mayoría, negativas, como ocurre con las otras de sindicalismo, corrupción e impunidad; pero tenemos un primer hallazgo importante, la

Cuadro 2

Matriz de correlaciones entre variables explicativas

	Inpibcap	Insal	tasades	vol1inv0	nopaganisr	sincontrato	salud	pension	sind7	corrup6	impunida5
Inpibcap	1.0000										
Insal	0.7086	1.0000									
tasades	0.2440	0.1964	1.0000								
vol1inv0	-0.2559	-0.1307	-0.8627	1.0000							
nopaganisr	-0.3508	-0.0889	-0.2340	0.2631	1.0000						
sincontrato	-0.5642	-0.4043	-0.4710	0.5002	0.6464	1.0000					
salud	-0.4368	0.0111	0.0773	0.0867	0.4729	0.3947	1.0000				
pension	-0.4706	-0.3085	-0.6072	0.6269	0.5507	0.8516	0.1431	1.0000			
sind7	0.3380	0.4885	0.2148	-0.2152	0.1157	-0.1772	0.1513	-0.2666	1.0000		
corrup6	-0.0491	0.2922	0.0815	-0.0441	0.5196	0.2015	0.4724	0.1187	0.4060	1.0000	
impunida5	-0.3234	-0.2199	0.0601	-0.1294	0.0372	-0.0491	0.3014	-0.1643	-0.2959	0.0349	1.0000

Fuente: estimaciones propias con base en cuadros 1 y del Anexo, y programa estadístico Stata 13.

relación inversa entre PIB y factores institucionales se puede interpretar en el sentido de que la forma en que contribuyen a la informalidad inhiben el crecimiento económico; por lo tanto, podemos esperar estimaciones más robustas con las regresiones que realizaremos a continuación y confirmar el hallazgo mencionado.

Aplicación del modelo con las variables de mercado

Al inicio se aplicó la prueba de la normalidad para la variable dependiente informalidad (infinv), cuyos resultados completos (y gráfica) se presentan en el Anexo; anotamos solo el valor que sirve para calificar la distribución de la variable $Prob > chi2 = 0.167$.

El criterio de Skewness/Kurtosis, Shapiro Wilks y Shapiro Francia (Taype, 2015) para calificar una distribución normal es que el valor sea $Prob > chi2 = 0.05$; por lo tanto, la variable dependiente presenta una distribución normal. Habría que agregar que los resultados se obtuvieron utilizando MCO y con

vce robustos (ver cuadros 3 y 3a, 4 y 4a, 5 y 5a y cuadro en Anexo); en general, muestran que, para las variables de mercado, los errores se redujeron en 2.97 %, pero para las institucionales aumentaron en 6.39 % y se redujeron en 3.56 % cuando se incluyen todas las variables.

Para el modelo con las de mercado (ver cuadros 3 y 3a), podemos comprobar que los coeficientes son diferentes a cero (prueba de Wald), el estadístico $F = 9.48$ y $Prob > F = 0.0002, < 0.05$. Cuando se corre el modelo heterocedástico con mínimos errores estándar (vce) robustos, aumentan para las variables logaritmo natural del PIB (Inpib) y la tasa de desempleo (tasades) y se reducen para el logaritmo natural de los salarios (Insal), el valor del estadístico $F = 8.60$ se reduce y $Prob > F = 0.0004$. Los signos son los esperados, negativo para el PIB y el desempleo; esto es, el incremento en el PIB reduce la informalidad, en tanto que aumenta cuando se reduce el desempleo y ambos son estadísticamente significativos a 5 %, si bien el salario tiene baja significación y signo positivo, diferente al esperado. La $R^2 = 0.4589$ ajustada por los grados de libertad es aceptable.

Cuadro 3

Relación informalidad y variables de mercado

Source	SS	df	MS	
Model	2815.4714	3	938.49045	Number of obs. = 31
Residual	2672.5792	27	98.984414	$F(3, 27) = 9.8$
Total	5488.0505	30	182.93502	$Prob > F = 0.0002$

$R\text{-squared} = 0.513$
 $Adj\ R\text{-squared} = 0.4589$
 Root MSE 9.9491

Fuente: estimaciones propias con base en cuadros 1 y del Anexo, y programa estadístico Stata 13.

Cuadro 3a

Relación informalidad y variables de mercado, vce robustos

tasainf15	Coef.	Std. Err.	t	P > t	[95 % Conf. Interval]	
lnpibcap	-14.39205	5.040817	-2.86	0.008	-24.73496	-4.049752
lnsal	12.72541	18.34279	0.69	0.494	-24.91089	50.36171
tasades	-5.259486	1.691043	-3.11	0.004	-8.729219	-1.789753
_cons	175.6343	72.85282	2.41	0.023	26.15268	325.116

Núm. obs. = 31, $F(3, 27) = 8.60$, $Prob > F = 0.0004$, $R^2 = 0.5130$.

* Los errores estándar de los coeficientes se redujeron en promedio 2.97 % utilizando vce robusto.

Fuente: estimaciones propias con base a cuadros 1 y del Anexo, y programa estadístico Stata 13.

Otras pruebas, como la de Ramsey, de variables omitidas arroja el resultado:

$$F(3, 24) = 6.23$$

$$Prob > F = 0.0028$$

como $Prob > F < 0.005$, significa que las variables de mercado no explican la informalidad actual. La prueba Breusch-Pagan/Cook-Weisberg para detectar si el modelo es heterocedástico registra los siguientes valores:

$$chi2(1) = 7.20$$

$$Prob > chi2 = 0.0073$$

cuya $Prob > F < 0.05$ implica que el modelo de mercado presenta heteroscedasticidad.

Inclusión de las variables institucionales

La incorporación de estas en el modelo (ver cuadros 4 y 4a) también pasa la prueba de Wald:

la $F = 73.24$ y la $Prob > F = 0.0000$ dan cuenta de ello e individualmente. De acuerdo con los resultados registrados en el cuadro 4, las variables se relacionan de manera positiva con la informalidad, como se esperaba, y con elevada significación estadística, un error menor a 5 % y aumenta la capacidad explicativa, la $R^2 = 0.9353$ ajustada por los grados de libertad; sin embargo, en presencia de los factores institucionales, todas las variables de mercado pierden la consistencia que tuvieron en la determinación de la informalidad.

La prueba de Ramsey estima que el modelo no presenta problemas de variables omitidas importantes:

$$F(3, 21) = 0.51$$

$$Prob > F = 0.6809$$

como $Prob > F > 0.05$, se desprende que las variables de mercado e institucionales en conjunto explican la informalidad actual. En lo que se refiere a la prueba para detectar si el modelo es hetero-

Cuadro 4

Relación informalidad y variables de mercado en institucionales

Source	SS	df	MS
Model	5203.8601	6	867.31001
Residual	284.19049	24	11.84127
Total	5488.0505	30	182.93502

Number of obs. = 31
 $F(3, 24) = 73.24$
 $Prob > F = 0.0000$
 $R\text{-squared} = 0.9482$
 $Adj\ R\text{-squared} = 0.9353$
Root MSE 3.4411

Fuente: estimaciones propias con base en cuadros 1, 2 y del Anexo, y programa estadístico Stata 13.

Cuadro 4a

Relación informalidad y variables de mercado e institucionales, vce robustos

tasainf15	Coef.	Std. Err.	t	P > t	[95 % Conf. Interval]
lnpibcap	1.473971	2.394658	0.62	0.544	-3.468361 6.416303
lnsal	-13.34429	7.372719	-1.81	0.083	-28.56083 1.872258
tasades	-0.749707	0.7208225	-1.04	0.309	-2.237412 0.7379975
nopaganisr	0.5831343	0.1420587	4.1	0.000	0.2899397 0.876329
salud	0.581583	0.1799465	3.23	0.004	0.2101917 0.9529743
pension	0.7103892	0.0794704	8.94	0.000	0.5463704 0.8744081
_cons	64.30547	26.94967	2.39	0.025	8.684082 119.9269

Núm. obs. = 31, $F(3, 24) = 104.40$, $Prob > F = 0.0000$, $R^2 = 0.9482$.

* Los errores estándar de los coeficientes aumentan 6.39 % cuando se aplica vce robusto.

Fuente: estimaciones propias con base en cuadros 1, 2 y del Anexo, y programa estadístico Stata 13.

dástico, la Breusch-Pagan/Cook-Weisberg estima los siguientes valores:

$$chi2(1) = 2.46$$

$$Prob > chi2 = 0.1170$$

cuya $Prob > chi2 > 0.05$, resultados que dan cuenta que el modelo no presenta heterocedasticidad.

Inclusión de otras variables explicativas

Cuando se hace esto, la prueba de Wald conjunta de coeficientes = 0, muestra valores:

$$F(9, 20) = 97.79 \quad Prob > F = 0.0000,$$

con lo que se rechaza la hipótesis; por lo tanto, las variables explican los diferentes valores que presenta la informalidad de las entidades.

También, cabe destacar que los resultados que se muestran en el cuadro 5 señalan que la inclusión de las otras variables reduce la capacidad explicativa del modelo, valores $F = 55$, $Prob > F = 0.0000$ y $R^2 = 0.9437$ ajustada por los grados de libertad. Individualmente, sindicalización y corrupción se relacionan de manera positiva con la informalidad, en tanto que la impunidad lo hace de forma negativa, si bien las tres variables con baja significación estadística.

Es de señalar que el poder explicativo de las variables de mercado se mantiene muy bajo, opacado por las institucionales. La prueba de Ramsey, de acuerdo con los siguientes resultados, señala que no hay el problema de variables omitidas:

$$F(3, 17) = 2.11; \quad Prob > F = 0.1371$$

Cuadro 5

Relación informalidad y todas las variables

Source	SS	df	MS
Model	5082.84693	9	564.76077
Residual	205.303081	20	10.265154
Total	5288.15001	29	182.35

Number of obs. = 30
 $F(9, 20) = 55.02$
 $Prob > F = 0$
 $R\text{-squared} = 0.9612$
 $Adj\ R\text{-squared} = 0.9437$
Root MSE 3.2039

Fuente: estimaciones propias con base en cuadros 1, 2 y del Anexo, y programa estadístico Stata 13.

Cuadro 5a

Relación informalidad y todas las variables, vce robustos

tasainf15	Coef.	Std. Err.	t	P > t	[95 % Conf. Interval]
lnpibcap	1.211139	2.302109	0.53	0.605	-3.590976 6.013255
lnsal	-19.62137	7.372511	-2.66	0.015	-35.00016 -4.242583
tasades	-1.077613	0.7051132	-1.53	0.142	-2.548453 0.3932276
nopaganisr	0.4728454	0.1482209	3.19	0.005	0.163662 0.7820289
salud	0.6264766	0.1882771	3.33	0.003	0.2337375 1.019216
pension	0.7101473	0.08433	8.42	0.000	0.534238 0.8860566
sind7	0.1636611	0.1030194	1.59	0.128	-0.051234 0.3785558
corrup5	0.1716739	0.2224786	0.77	0.449	-0.292408 0.6357561
impunida6	-0.0940378	0.1029988	-0.91	0.372	-0.30889 0.1208139
_cons	108.7233	31.22054	3.48	0.002	43.5984 173.8482

Núm. obs. = 30, $F(3, 24) = 97.79$, $Prob > F = 0.0002$, $R^2 = 0.9612$.

* Los errores estándar se redujeron en promedio para todas las variables 3.56 % con vce robustos.

Fuente: estimaciones propias con base en cuadros 1, 2 y del Anexo, y programa estadístico Stata 13.

Respecto al problema de heterocedasticidad, los valores $chi2(1) = 4.53$ y $Prob > chi2 = 0.0333 < 0.05$ indican que esta existe, que estadísticamente podrían corregirse aumentando el tamaño de la muestra, porque la prueba de Ramsey mostró que no hay variables omitidas relevantes, pero lo que no se puede eliminar en la realidad es justo esa heterogeneidad en la informalidad entre las entidades, que dan origen a la heterocedasticidad.

Discusión de los resultados del modelo

Los factores económicos muestran que, por el lado del PIB, su incremento contribuye a reducir la informalidad, que corrobora estadísticamente lo que ocurre en la realidad: con el crecimiento economi-

co aumenta la demanda de empleos formales y se reduce el desempleo; por su parte, la reducción del desempleo coincide con un aumento en la informalidad, porque justamente se oculta en ella, de acuerdo con el velo que refería Tokman (1993); sin embargo, en presencia de los factores institucionales, todas las variables de mercado pierden la consistencia que tuvieron en la determinación de la informalidad y son las institucionales las que adquieren mayor significancia.

A la luz de estos resultados, podemos comprobar que los esfuerzos de formalización del mercado laboral mexicano con el programa lanzado en el 2016 por la Secretaría del Trabajo y Previsión Social, centrado en la atención a la salud y garantizar una pensión, ha sido poco fructífero, si nos

guiamos por el elevado porcentaje que mantiene, mayor al registrado por Tokman (2001b), menos de 50 % hace 20 años para México y otros países de América Latina, como Argentina, Brasil y Colombia, naciones en las que claramente se explica que la causa principal de la informalidad han sido las crisis económicas y que ahora se complicará aún más con la de salud del 2020, máxime si la recuperación se da de la misma forma en la que ha ocurrido a lo largo de décadas, el *goteo* al que se refería Tokman (2001b) con base en bajos salarios, razón por la cual la población preferirá autoemplearse, es decir, mantenerse en la informalidad.

Es indudable que la reducción de la informalidad requiere de una política que combine de manera deliberada los factores aquí investigados, que establezca, por un lado, el crecimiento del PIB como condición necesaria —así lo indica la relación negativa que guarda con las variables institucionales en el modelo— con base en la generación de empleos bien remunerados y con todas las prestaciones de ley, lo cual implica, de suyo, reducir la contratación mediante el sistema *outsourcing*, que ha enmascarado la falta de contratos y el no pago de prestaciones visto como una forma de flexibilización y reducción de los costos laborales, lo cual ha abonado a la informalidad dentro de la formalidad y, además, que prevalezca el respeto a la libre elección de actividades lícitas de trabajo e inversión que cumplan con las exigencias de la ley, como el pago de impuestos y de prestaciones al trabajo; y como condición suficiente, que el Estado mexicano despeje dudas sobre actos de corrupción en el manejo de los recursos e impunidad en la aplicación de la ley que favorezcan a grupos o familias y que, en su lugar, los dedique para apoyar a emprendedores e impulsar proyectos generadores de empleos formales.

Conclusiones

La economía informal se refiere a los trabajadores que laboran tanto en el sector informal como aquellos que lo hacen en empresas formales, pero trabajan sin contrato o no perciben prestaciones

que marca la ley, lo cual representa en México poco más de 57 % de la población ocupada. Constituye una suerte de flexibilidad clásica moderna y un grave problema que ha dado lugar a la búsqueda de las causas y de soluciones alternativas, donde se inscribe esta investigación.

El presente estudio mostró que el origen de la informalidad fueron las crisis recurrentes de la década de los 80, posteriormente alimentada por los gobiernos neoliberales de los factores que explican la situación actual en las entidades de México, que fue nuestro objetivo.

Se utilizaron dos hipótesis para probar cuáles son los factores explicativos de la informalidad actual: 1) los de mercado (entre los que se incluye el PIB per cápita, el salario, la tasa de desempleo y una variable *dummy*, compuesta por la relación voluntario/involuntario del tipo de desempleo) reducen la informalidad, 2) los institucionales (no pago de ISR, sin contrato, carencia de servicio de salud y no pertenencia a un sistema de pensiones) aumentan la informalidad y 3) otros factores, que no pueden conformarse como hipótesis debido a que se esperan efectos diferenciados: sindicalización, corrupción e impunidad.

Los resultados muestran que de 57.81 % de los trabajadores en México que están en la informalidad, poco más de la mitad (27.4 %) no paga ISR; si bien en entidades con altos índices de informalidad combinados con mayor cantidad de causantes (como México y la Ciudad de México), la mayoría de los empleados, más de las dos terceras partes, lo son por no pagar impuestos.

Por medio del modelo econométrico se comprueba que la variable crecimiento del PIB reduce la informalidad, resultando no significativas el salario y la tasa de desempleo; pero cuando son incluidas las variables institucionales, son estas las que explican la informalidad y las de mercado pierden consistencia estadística.

Por último, de esta investigación se desprende que si el desempleo es involuntario en la mayo-

ría de las entidades, parte de los trabajadores se convierten a la informalidad para poder subsistir y otra parte lo hacen orillados por factores institucionales, tal como lo muestra el modelo; por lo tanto, este trabajo concluye que, contrario a lo que encontró Maloney (2004), la informalidad no es del todo voluntaria, más bien resulta una mezcla de los dos tipos de factores, de mercado e institucionales, de lo cual se desprende que las medidas para reducirla en México y sus entidades deben ir en ambos sentidos. Es claro que la reducción de la informalidad requiere de una política que vaya más allá del *goteo*, que combine de manera deliberada los factores aquí investigados, que establezca como condición necesaria el crecimiento del PIB mediante empleos bien remunerados y con todas las prestaciones de ley, contrario a la práctica del siglo XXI en México basada en el *outsourcing*; como condición suficiente, que el Estado mexicano, en lugar de combatir a los informales, como intentaron hacerlo en el pasado reciente, poner más énfasis en la lucha contra la corrupción en el manejo discrecional y desvío de los recursos hacia actividades fuera del presupuesto, eliminar la impunidad y, en su lugar, los dedique en el apoyo a emprendedores para impulsar proyectos generadores de empleos formales.

Fuentes

- Alba, F. "El mercado de trabajo: cambios en el modelo de absorción de la fuerza laboral. México auge, crisis y ajuste", en: Bazdreh, C.; N. Bucay; S. Loaeza y N. Lustig (comps). *El Trimestre Económico*. México, FCE Lecturas, 73, 1993, pp. 179-200.
- Boswarth, B. *Productivity Growth in Mexico*. Report Núm. 1732-ME, Washington, The World Bank, 1998.
- Camberos C., Mario y Joaquín Bracamontes N. "Las crisis económicas y sus efectos en el mercado de trabajo, en la desigualdad y en la pobreza de México", en: *Contaduría y Administración*. 60 (S2), 2015a, pp. 219-249 (DE) www.sciencedirect.com www.cya.unam.mx/index.php/cya
- _____. "La caída de los salarios en México 1980-2010: la productividad y la informalidad como explicación", en: *Equilibrio Económico, Revista de Economía, Política y Sociedad*. Vol. 11 (2), semestre julio-diciembre, núm. 40, 2015b, pp. 169-194, ISSN: 2007-2627; ISSN (versión electrónica): 2007-3666.
- Camberos C., Mario y Abel Castro R. "La informalidad en las entidades de México: ¿voluntaria?, ¿institucional?", en Rivera-Huerta, Rene, Nidia López Lira y Luz Marina Sánchez Rodríguez. *Economía informal y otras formas de producción y trabajo atípico. Estudios para el caso de México*. UAM-X, 2018a, pp. 47-76.
- _____. "Desempleo, salarios e informalidad: un análisis de las entidades de México y la frontera norte", en: Rodríguez Reyna y D. Castro L. (coords.). *Mercado laboral: México y frontera norte*. Ediciones Laurel-UAdeC, 2018b, pp. 73-95. ISBN Universidad Autónoma de Coahuila: 978-607-506-324-9, ISBN Ediciones de Laurel: 978-607-8477-30-2.
- Cameron, A. y Trivedi K. *Microeconometrics Using Stata*. Stata press College Station, Tx., 2010.
- Cortés, F. "La metamorfosis de los marginales: la polémica sobre el sector informal en América Latina", en: *Tratado Latinoamericano de Sociología*. 2000, pp. 592-618 (DE) <http://sgpwe.izt.uam.mx/pages/egt/publicaciones/libros/Tratadolatinoamericanodesociologiadeltabajo/Metamorfosisdelosmaginales.pdf>, consultado el 2 de julio de 2020.
- Debrah, Yaw A. "Promoting the Informal Sector as a Source of Gainful Employment in Developing Countries: Insights from Ghana", en: *The International Journal of Human Resource Management*. 18(6), 2007, pp. 1063-1084.
- Falco, P. y William Maloney. "Heterogeneity in subjective wellbeing: An application to occupational allocation in Africa", en: *Journal of Economic Behavior & Organization*. Vol. 111, 2015, issue C, 137-153.
- Fernández, O. y F. Almagro. "Caracterización de la ocupación del sector informal en México", en: *Boletín del Sistema Nacional de Estadística, Geografía e Informática*. 1(1), 2008, pp. 3-16.
- Ferraro, C. y S. Rojo. *Las MIPYMES en América Latina y el Caribe. Una agenda integrada para promover la productividad y la formalización*. OIT Cono Sur, Reportes Técnicos núm. 7, 2018 (DE) https://www.ilo.org/wcmsp5/groups/public/---americas/---ro-lima/---sro-santiago/documents/publication/wcms_654249.pdf, consultado el 1 de septiembre de 2019.
- Greene, W. *Econometric Analysis*. New York, Prentice Hall, 2008.
- Hart, Keith. "Informal Income Opportunities and Urban Employment in Ghana", en: *Journal of Modern African Studies*. 11, 1973, pp. 61-89.
- Instituto Mexicano del Seguro Social (IMSS). *Ley del Seguro Social*. 2015 (DE) <http://www.imss.gob.mx>, consultado el 15 octubre de 2015.
- Instituto Nacional de Estadística Geografía e Informática (INEGI). *Población ocupada en el sector informal, 2005-2015*. México, INEGI, 2016.
- _____. *Encuesta Nacional de Ocupación y Empleo (ENOE) 2015*. 2.º trimestre. México, INEGI, 2015a.
- _____. *Banco de Información Económica (BIE). Producto Interno Bruto de las entidades de México 2005-2013*. México, INEGI, 2015b.
- _____. *Población ocupada en el sector informal, 2005-2010*. México, INEGI, 2013.
- _____. *Anuario estadístico y geográfico por entidad federativa*. México, INEGI, 2015c.
- _____. *Encuesta Nacional de Ocupación y Empleo (ENOE)*. 4.º trimestre. México, INEGI, 2016.
- _____. *Encuesta Nacional de Ocupación y Empleo (ENOE)*. 4.º trimestre 2016, México, INEGI, 2017.

- _____. *Encuesta Nacional de Ocupación y Empleo (ENOE) 2019. 2.º trimestre*. México, INEGI, 2019.
- International Labor Organization (ILO). *Statistical update on employment in the informal economy*. Geneve, Department of Statistic, ILO, 2011 (DE) http://www.ilo.org/wcmsp5/groups/public/dgreports/stat/documents/presentation/wcms_157467.pdf, consultado el 12 de marzo de 2013.
- _____. *Global Trends: Employees*. Geneve, Department of Statistic, ILO, 2012 (DE) http://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_171571.pdf, consultado el 12 de marzo de 2013.
- _____. *Transition from the Informal to formal Economy Recommendations*. Geneve, Department of Statistic, ILO, 2015.
- _____. *Informal Economy*. Geneve, Department of Statistic, ILO, 2019 (DE) <https://www.ilo.org/global/topics/employment-promotion/informal-economy/lang--en/index.htm>, consultado el 1 de septiembre de 2019.
- Jiménez, D. "La informalidad laboral en América Latina: ¿explicación estructuralista o institucionalista?", en: *Cuadernos de Economía*. Vol. 31, núm. 58, julio-diciembre. Bogotá, 2012. Versión impresa ISSN 0121-4772.
- Leclerq, O. J. A. y G. Rodríguez S. L. *Índice Global de Impunidad México (IGI-MEX)*. UDLAP, 2016.
- Lehmann, H. y A. Muravyev. "Labor Market Institutions and Informality in Transition and Latin American Countries," en: *IZA Discussion Paper No. 7035*. Munchen, 2012.
- Lustig, N. "El efecto social del ajuste. México auge, crisis y ajuste", en: Bazdrech, C.; N. Bucay; S. Loaeza y N. Lustig (eds.). *El Trimestre Económico*. México, 1993, pp. 201-238.
- Maloney, W. "Does Informality Imply Segmentation in Urban Labor Markets? Transition Evidence from Sectoral Transitions in Mexico", en: *World Bank Economic Review*. 13, 1999, pp. 275-302.
- _____. "Informality Revisited", en: *World Development*. 32, No. 7, Washington, D. C., The World Bank, 2004, pp. 1159-1178.
- Negrete, R. *El empleo informal en México visto bajo el esquema conceptual OIT-Grupo de Delhi*. OIT-Grupo de Delhi, 2010.
- Organización Internacional del Trabajo (OIT). *El empleo informal en México: situación actual, políticas y desafíos*. México, Oficina Regional para América Latina y El Caribe, 2014.
- _____. *El sector informal; un estudio de Kenia*. OIT, 1972.
- Portes, A. y W. Haller. *La economía informal*. Santiago de Chile, ONU-CEPAL, 2004.
- Rei, D. y Bhattacharya M. *The Impact of Institutions and Policy on Informal Economy in Developing Countries an econometric exploration*. Working Paper No. 84. Policy Integration and Statistics Department, International Labor Office Geneva, 2008.
- Rivera-Huerta, René y Luz Marina Sánchez Rodríguez. "Economía informal, sector informal: concepto y estadísticas de las definiciones básicas", en: Rivera-Huerta, Rene, Nidia López Lira y Luz Marina Sánchez Rodríguez. *Economía informal y otras formas de producción y trabajo atípico. Estudios para el caso de México*. UAM-X, 2018, pp. 23-46.
- Rodríguez, P. y Andrés J. *Las instituciones y su efecto en la informalidad laboral para las 13 áreas metropolitanas de Colombia*. Tesis de maestría. Bogotá, Universidad Nacional de Colombia, 2016.
- Shapiro, A. *Institutions, Informal Labor Markets, and Business Cycle Volatility*. 2017 (DE) <http://www.jstor.org/stable/24570866>, consultado el 26 de diciembre de 2016.
- Secretaría del Trabajo y Previsión Social (STPS). *Programa de formalización del empleo*. México, 2016 (DE) <https://www.forbes.com.mx/mexico-29-millones-personas-empleos-informales/#gs.N7HHSE4>, consultado el 29 de marzo de 2017.
- Secretaría de Hacienda y Crédito Público (SHCP). *Fomentando la formalidad laboral en todo el territorio nacional*. Nota técnica. México, Gobierno de México, 2013.
- Soria, V. (2001). "El mercado de trabajo en Brasil y México a la luz de la integración internacional y la crisis financiera", en: *Comercio Exterior*. México: BNCE.
- Taype R. Álvaro. *Evaluando la normalidad en Stata*. Stata you tube, 2015 (DE) <https://www.youtube.com/watch?v=dDT-wsNwUK0>
- Tello, C. *Sobre la desigualdad en México*. México, Universidad Nacional Autónoma de México, 2010.
- Transparencia Internacional. *Índice Nacional de Corrupción y Buen Gobierno. Resultados por entidad federativa*. 2010 (DE) http://imco.org.mx/politica_buen_gobierno/mexico-anatomia-de-la-corrupcion/, consultado 4 de abril de 2017.
- Tokman, V. "Las relaciones entre los sectores formal e informal. Una exploración sobre su naturaleza", en: *Revista Economía*. Vol. 24, núm. 48, PUCP, Lima, 2001a, pp. 17-74.
- _____. "Mercados de trabajo y empleo en el pensamiento económico latinoamericano", en: Osvaldo Sunkel (comp.). *El desarrollo desde adentro*. México, FCE, 1993, pp.163-191.
- _____. "De la informalidad a la modernidad", en: *Boletín Cinterfor*. Núm. 155, 2001b, pp. 9-31. Tomado del original *De la informalidad a la modernidad*, publicado por la OIT, 2001. Nota del autor (DE) https://www.oitcinterfor.org/sites/default/files/file_articulo/tokman.pdf, consultado el 1 de julio de 2020.

Anexo

Informalidad y variables de mercado de las entidades de México

Entidad	Tasalnf (15)*	SecInfor*	pib/cap**	salarios***	tasa de des	vol/inv(+)
Aguascalientes	43.67	22.64	130 102	229.85	4.46	0.00
Baja California	40.15	19.89	122 357	257.75	4.75	0.00
Baja California Sur	40.40	16.97	139 599	242.34	4.33	0.00
Campeche	61.02	26.65	628 311	336.14	2.51	1.00
Coahuila de Zaragoza	35.07	20.22	159 651	261.18	5.30	0.00
Colima	54.40	19.22	115 578	229.44	4.46	0.00
Chiapas	78.84	23.49	43 345	210.58	3.25	1.00
Chihuahua	36.04	15.36	113 386	242.77	4.00	1.00
Ciudad de México	51.28	30.03	257 121	357.25	5.36	0.00
Durango	55.98	23.92	92 669	204.23	4.80	0.00
Guanajuato	58.30	28.07	101 980		4.37	0.00
Guerrero	79.50	33.42	56 671	222.36	1.97	1.00
Hidalgo	71.60	35.25	78 669	230.76	4.05	1.00
Jalisco	50.13	23.29	112 825	244.26	5.28	0.00
México	59.28	35.98	75 983	264.06	5.60	0.00
Michoacán de Ocampo	71.13	31.94	68 737	234.34	2.79	1.00
Morelos	66.37	34.70	83 545	264.80	3.16	1.00
Nayarit	62.08	26.41	78 047	208.80	5.61	0.00
Nuevo León	36.89	20.42	205 952	304.31	4.94	0.00
Oaxaca	80.00	32.41	54 034	221.77	2.38	1.00
Puebla	72.41	27.86	70 868	247.12	3.25	1.00
Querétaro	45.37	19.88	156 244	306.49	5.02	0.00
Quintana Roo	47.27	20.68	149 983	206.49	3.68	1.00
San Luis Potosí	58.25	24.05	99 285	241.95	2.86	1.00
Sinaloa	50.90	21.33	97 974	200.87	4.33	0.00
Sonora	45.37	22.06	139 712	227.35	4.93	0.00
Tabasco	63.66	28.71	179 044	238.44	6.00	0.00
Tamaulipas	47.32	21.04	122 206	253.04	4.65	0.00
Tlaxcala	73.22	39.46	60 477	214.64	4.53	0.00
Veracruz de Ignacio de la Llave	68.10	25.28	83 928	234.01	3.58	1.00
Yucatán	62.33	32.25	97 215	198.07	2.35	1.00
Zacatecas	66.02	25.20	83 640	233.16	2.81	1.00

* Datos del segundo trimestre del 2015. La diferencia son los trabajadores informales en el sector formal.

** A precios constantes del 2008, INEGI (2015b).

*** Establecidos por el IMSS como base 2.º trimestre (INEGI, 2015a) en pesos corrientes.

(+) vol 1 se refiere a si está desempleado voluntariamente; inv 0, si es desempleado involuntario.

Fuente: estimaciones propias con base en Encuesta Nacional de Ocupación y Empleo (INEGI, 2015 a y b).

Predicción automática del nivel educativo en usuarios de **Twitter en México**

Automatic Prediction of the Educational Level of Twitter Users in Mexico

Juan Carlos Gomez, Luis Miguel López Santamaría, Mario Alberto Ibarra Manzano y Dora Luz Almanza Ojeda*

Para resolver la situación planteada en este trabajo, se extrajo una serie de características del contenido textual de los tuits publicados por los usuarios, que se utilizaron para construir modelos basados en aprendizaje automático, los cuales predicen si un usuario tiene estudios universitarios o no. Ambos se probaron con un conjunto de datos extraído de forma directa del sitio, compuesto por más de un millón de tuits en español, correspondientes a 195 usuarios ubicados en México. Con él, se hicieron experimentos siguiendo una validación cruzada de 10 partes. La evaluación se realizó utilizando las métricas macro *F1* y el área bajo la curva ROC (AUC). Los resultados indican que la tarea es compleja, siendo las mejores características las abreviaturas, que alcanzaron valores arriba de 60 % para ambas métricas, mientras que los modelos de máquinas de vectores de soporte y árboles de decisión presentaron un desempeño similar.

Palabras clave: analítica de datos; *Twitter*; perfilado de autor; aprendizaje automático.

Recibido: 14 de noviembre de 2019.

Aceptado: 9 de julio de 2020.

* Universidad de Guanajuato, Campus Irapuato-Salamanca, jc.gomez@ugto.mx (autor de correspondencia), lm.lopezsantamaria@ugto.mx, ibarram@ugto.mx, dora.almanza@ugto.mx, respectivamente.

Nota: esta investigación fue apoyada por el Fondo Sectorial CONACYT-INEGI, proyecto número 290910.

To solve the task raised in this work, a series of characteristics were extracted from the textual content of the tweets published by users, which were used to build models based on machine learning, which predict whether a user has a university degree or not. Both were tested with a data set extracted directly from the site, composed of more than one million tweets in Spanish, corresponding to 195 users located in Mexico. With it, experiments were made following a 10-fold cross-validation. The evaluation was performed using the *F1* macro metrics and the area under the ROC (AUC) curve. The results indicate that the task is complex, the best characteristics being the abbreviations, which reached values above 60% for both metrics, while the support vector and decision tree machine models showed similar performance.

Key words: data analytics; *Twitter*; author profiling; machine learning.



Futuristic cyborgs learning about humans/gremlin/iStock

Introducción

El análisis de las redes sociales visualiza al entorno social como un conjunto de patrones y conexiones entre unidades de interacción (usuarios) a partir de los cuales es posible estudiar relaciones económicas, políticas y afectivas existentes entre los usuarios (Wasserman & Faust, 1994). En este contexto, surge la tarea denominada *perfilado de autor* (*Author Profiling*), la cual se refiere al análisis del contenido generado o compartido en estos medios (p. ej. *Facebook* y *Twitter*, entre otros) para predecir diferentes atributos de quien los realiza, como género, edad, personalidad u orientación política.

El perfilado de autor tiene una amplia gama de aplicaciones en áreas como mercadotecnia, política, seguridad y educación; por ejemplo, en esta puede apoyar a las plataformas digitales para

adecuar las estrategias de enseñanza para perfiles de estudiantes específicos. Por otro lado, muchas compañías y organizaciones lo utilizan para segmentar a sus clientes/usuarios para determinar qué contenido proveerles con fines de publicidad, campañas políticas, programas sociales o entretenimiento, entre otros.

Uno de los enfoques más populares para perfilar los atributos de interés (edad, género, etc.) de una persona consiste en entrenar modelos basados en aprendizaje automático (*Machine Learning*) utilizando el texto publicado por un grupo de personas de las cuales ya se tiene conocimiento de los atributos de interés. Una vez entrenados los modelos, se pueden usar para predecir esas particularidades en individuos no conocidos de forma previa.

Dentro de las redes sociales, *Twitter* es una de las más populares en diversos países; cuenta con,

al menos, 330 millones de usuarios activos,¹ de los cuales se conectan diario cerca de 126 millones.² En México, es la tercera más popular con cerca de 5 millones de usuarios (solo detrás de *Facebook* y *YouTube*), quienes publican mensajes (llamados tuits), con una longitud máxima de 280 caracteres para compartir lo que piensan, hacen o lo que está sucediendo a su alrededor y, aunque pueden incluir contenido multimedia (imágenes, audio o video), en su mayoría es texto debido a la facilidad de su escritura y envío. De manera adicional, todos los usuarios pueden interactuar entre sí, respondiendo o volviendo a publicar los tuits de otros.

En este trabajo se presenta un estudio sobre la predicción automática del nivel educativo en usuarios de *Twitter* en México a través del análisis del texto de las publicaciones que han hecho en la red social. Su hipótesis central es que algunas de las características textuales extraídas de las publicaciones son adecuadas para identificar si un usuario de *Twitter* en México tiene estudios superiores o no, partiendo del supuesto de que un entorno universitario favorece el dominio de conocimientos y usos lingüísticos específicos (Ramos, 2012).

Para conducir la investigación presentada aquí, se hizo una recopilación de 1 101 241 textos en español correspondientes a 195 personas, sin distinción de género, ubicados en México, quienes fueron seleccionados haciendo un muestreo aleatorio de usuarios activos utilizando la *Application Program Interface (API)* nativa de *Twitter* y fueron cuestionados de forma directa sobre su género y nivel de estudios; los que respondieron, se colocaron en dos clases: nivel superior (positivo) y no superior (negativo).

Después, de los tuits se extrajo una serie de características textuales que representan distintos aspectos del contenido publicado: palabras, menciones (@ o *ats*), etiquetas (# o *hashtags*), ligas (*http links*), abreviaturas, los vectores de palabras *GloVe* y los promedios de uso de cada característica.

1 <https://bit.ly/3f4a29K>

2 <https://bit.ly/3eXA18W>

Utilizando los distintivos anteriores, se construyeron modelos basados en aprendizaje automático para la predicción. En específico, se entrenaron y probaron modelos de máquinas de vectores de soporte lineales (SVM, por sus siglas en inglés), clasificadores Naïve Bayes (MNB, por sus siglas en inglés) y árboles de decisión (DT, por sus siglas en inglés). Se experimentó con ellos siguiendo una validación cruzada de 10 partes para tener resultados más consistentes. Su desempeño se midió usando dos métricas populares en la clasificación de textos: macro F1 y el área bajo la curva ROC (AUC).

Las contribuciones de este trabajo son dos: la primera es la presentación de la tarea de la predicción automática del nivel educativo en usuarios de *Twitter* a partir del análisis del contenido textual que publican; la segunda, el estudio del desempeño de diferentes características textuales y modelos de predicción para el análisis.

Los resultados muestran que la labor es compleja, con los mejores valores apenas arriba de 60 % para ambas métricas utilizando abreviaturas y máquinas de vectores de soporte o DT. Debido a esto, se determina que es necesario continuar con la investigación para encontrar otras características o modelos que mejoren el desempeño.

El resto del artículo está organizado de la siguiente forma: primero se presenta una revisión de los trabajos relevantes en la literatura; enseguida, se describe la metodología, incluyendo la descripción del conjunto de datos utilizado y de la experimentación; después, se muestran los resultados obtenidos y, por último, se resumen las conclusiones y se establecen algunas ideas para trabajos futuros.

Revisión de la literatura

En el estudio del problema de perfilado de autor se han desarrollado varios trabajos a lo largo de los años. Entre los más reconocidos se encuentran los presentados en los eventos PAN labs (*Plagiarism Analysis, Authorship Identification and Near Duplication*).

te Detection), que se han realizado como parte de las conferencias CLEF desde el 2013 (Rangel *et al.*, 2013; Rangel *et al.*, 2014, 2015, 2016, 2017 y 2018). Las tareas de perfilado que se han propuesto en estos eventos han tratado sobre la predicción de edad, género, idioma nativo y rasgos de personalidad. De entre ellos, las de la edad y el género han sido las tareas más populares. En estos eventos se ha utilizado una serie de conjuntos de datos extraídos de tuits en varios idiomas, siendo el inglés el dominante. De manera adicional, hasta el 2018, estos contenían de forma exclusiva texto; a partir de ese año, se ha incluido el uso de imágenes para reforzar la tarea.

En los trabajos presentados en los eventos de PAN@CLEF se han empleado muchos enfoques para resolver la tarea de perfilado de autor. Lo anterior, incluye el uso de varias características extraídas del texto, como diccionarios de frases; palabras en general, de sentimiento, de opinión o específicas asociadas con un grupo de edad o género; lemas; categorías gramaticales; etiquetas HTML; emoticonos; atributos de segundo orden; vectores singulares o de palabras; entre otros. Además, se exploraron diferentes modelos de predicción basados en aprendizaje automático. Entre los más populares se encuentran las máquinas de vectores de soporte, la regresión logística, los clasificadores bayesianos y las redes neuronales profundas (*Deep Learning*).

Para el análisis de información proveniente de *Twitter*, existen varias tareas que se han estudiado; por ejemplo, el análisis de sentimientos, que consiste en determinar si un tuit, usuario o grupo tiene una opinión positiva, negativa o neutra sobre un tema o tópico o en un momento determinado (Desai, 2018; Vashishtha & Susan, 2019). En la detección de temas de tendencia se intenta clasificar el impacto de uno de discusión actual, catalogándolo como algo disruptivo, popular o rutinario (Indra *et al.*, 2019). Por otro lado, la predicción de cambios en el mercado de valores utilizando *Twitter* es un campo de investigación de interés actual. Algunos estudios han concluido que el estado de ánimo público recopilado de esta red social bien puede estar

correlacionado con el Índice Dow Jones (Pagolu *et al.*, 2016). Por último, en la predicción de resultados políticos se desea predecir al ganador de una contienda electoral con base en la popularidad de los contendientes en *Twitter* (Wang & Gan, 2018).

Para todas estas tareas, se han estudiado varias características textuales, como palabras, *n*-gramas, emoticonos, *hashtags* (Jianqiang *et al.*, 2018), *BN*-gramas (Indra *et al.*, 2019), diccionarios de sentimientos (Desai, 2018), vectores *GloVe*, patrones de difusión de sentimientos (Wang & Yu, 2018) y estadísticas, por ejemplo, el número de menciones a un individuo o su número de seguidores (Serban *et al.*, 2014). De igual forma, se han utilizado diversos modelos basados en aprendizaje automático con enfoques supervisados y no supervisados.

En cuanto a la predicción del nivel educativo de usuarios, los trabajos en la literatura se orientan en la detección del nivel de aprendizaje en entornos educativos virtuales (en los cuales se tiene más control sobre las etapas del proceso de aprendizaje y del vocabulario empleado en el mismo, con menor cantidad de ruido que en una red social abierta), enfocándose en tópicos específicos y siguiendo ciertas taxonomías de clasificación (Echeverría *et al.*, 2013; Pincay & Ochoa, 2013). De forma adicional, hay trabajos donde se estudia cómo los estudiantes usan las redes sociales para integrarlas en su proceso de aprendizaje (Tess, 2013).

Metodología

La utilizada en este trabajo está compuesta por tres fases: la adquisición de los datos, su procesamiento y la experimentación. Esta última incluye, a su vez, dos procesos: la construcción de los modelos para la predicción y la evaluación de los mismos.

Adquisición de datos

Para esto, primero se creó un *script* en *Python* que realizó un muestreo aleatorio de más de 500 usuarios activos ubicados en México utilizando la API na-

tiva de *Twitter*. A esta muestra inicial se le preguntó directamente a su cuenta por su género y su nivel educativo, obteniendo respuesta de 260 de ellos. Se les etiquetó en dos clases para nivel educativo: superior (positiva) y no superior (negativa).

Con otro *script* de *Python* se recopilaron los tuits de los usuarios identificados previamente, que fueron recuperados por fecha (del más reciente al más antiguo), con un límite máximo de 10 mil por cada uno. Con algunos usuarios hubo problemas en la captura de la información y se eliminaron, quedando, al final, 195. En total, se recopilaron 1 101 241 mensajes, para un promedio de 5 647.4 tuits por usuario; todos están en español, pero en varios se hacen mezclas de palabras en otros idiomas, sobre todo inglés.

En el cuadro 1 se observa la distribución de los usuarios en nuestro conjunto de datos por nivel educativo y género; se puede ver que tanto la distribución por género como por nivel educativo es similar a la encontrada en *Twitter* en su totalidad. En esta red se estima que 66 % son hombres,³ mientras que 80 % tienen un nivel educativo universitario⁴ (*college degree*). *Twitter* (junto con *LinkedIn*) es reconocida por ser popular entre las

personas con alto nivel educativo y de ingresos.⁵ Por lo tanto, se considera que nuestro conjunto de datos es representativo de la demografía de *Twitter*.

Procesamiento de los datos

De los tuits recopilados para extraer una serie de características relevantes, primero se limpiaron para remover etiquetas HTML, dejando solo el texto de cada uno; a continuación, se concatenaron todos los mensajes correspondientes a un mismo usuario en una sola cadena larga de texto y se pasó todo el contenido a minúsculas; después, se utilizaron expresiones regulares para extraer del texto seis características superficiales: palabras, *emojis*/emoticonos (*emoticons*), etiquetas (*#* o *hashtags*), menciones (*@* o *ats*), ligas (*http links*) y abreviaturas comunes.

Las palabras se consideraron como una secuencia de letras, más los caracteres *_* y *-*; de estas, se filtraron las vacías (*stopwords*), quedando una lista de 778 que se reunieron de diferentes fuentes en internet. En el caso de las abreviaturas, se recopiló de forma manual una lista de las 130 más utilizadas en *Twitter*. El cuadro 2 presenta una muestra de

3 <https://www.omnicoreagency.com/twitter-statistics>

4 <https://www.pewinternet.org/fact-sheet/social-media/>

5 <https://blog.hootsuite.com/twitter-demographics/>

Cuadro 1

Distribución de usuarios por género y nivel educativo

Nivel educativo	Hombres	Mujeres	Total
Superior	103	58	161 (82.5 %)
No superior	22	12	34 (17.5 %)
Total	125 (64 %)	70 (36 %)	195 (100 %)

Cuadro 2

Ejemplos de abreviaturas comunes encontradas en *Twitter*

Abr.	Significado	Abr.	Significado	Abr.	Significado
<i>xd</i>		<i>fb</i>	Facebook	<i>pq</i>	¿Por qué?
<i>rt</i>	<i>Retweet</i>	<i>omg</i>	<i>Oh my God</i>	<i>wtf</i>	What the fuck?
<i>lol</i>	<i>Laughing Out Loud</i>	<i>alv</i>	A la verga	<i>tqm</i>	Te quiero mucho

ellas con su significado; algunas provienen del inglés, pero su uso se ha difundido a otros idiomas, incluido el español.

Analizando la frecuencia del uso de abreviaturas y palabras para cada clase (superior y no superior), en los cuadros 3 y 4 se muestran las 15 más utilizadas. Las listas están ordenadas por frecuencia, primero por columna y luego por fila. En negritas se marcan las palabras o abreviaturas que son diferentes entre clases. Como puede observarse, los usuarios de ambas presentan ciertas diferencias en su uso.

En el cuadro 5 se muestran los promedios y las desviaciones estándar del número de características por tuit por clase. En promedio, un mensaje tiene poco menos de seis características, siendo las palabras las más abundantes y las abre-

viaturas, las menos. Los usuarios de la clase superior escriben mensajes más largos que contienen, en su mayoría, más palabras, etiquetas, menciones y ligas, mientras que los de la no superior lo hacen con más *emojis*/emoticonos y abreviaturas.

Una vez extraídas las características, se obtuvieron seis archivos (uno por cada una) con 195 líneas por archivo; cada línea representa las características de un usuario. Todos ellos se dividieron en 10 partes para generar el mismo número de conjuntos de entrenamiento y de prueba para poder aplicar una validación cruzada estratificada, manteniendo la proporción de cada clase en cada parte. En esta validación, los conjuntos de entrenamiento se encuentran formados por nueve partes (193 usuarios) y la sección restante se usa como conjunto de prueba (22 usuarios).

Cuadro 3

Listas de las 15 palabras más comunes por clase

Superior				No superior			
gracias	bendiciones	dios	mañana	gente	juego	años	espero
vida	the	amo	feliz	quiero	vida	mañana	canal
gente	años	saludos	méxico	mierda	the	alguien	veo
quiero	alguien	mundo		gracias	cosa	gusta	

Cuadro 4

Listas de las 15 abreviaturas más comunes por clase

Superior					No superior				
xd	by	ptm	alv	bb	xd	lol	md	cm	wtf
pa	omg	dm	tqm	fb	rt	pa	omg	pls	dm
rt	lol	cc	tkm	gpi	yt	tl	alv	oc	rip

Cuadro 5

Promedio del número de características por tuit por clase

Promedio por tuit							
Nivel educativo	Palabras	<i>Emojis</i>	Etiquetas	Menciones	Ligas	Abr.	Todas
Superior	4.97 (4.49)	0.48 (1.63)	0.14 (0.55)	0.17 (0.55)	0.35 (0.51)	0.04 (0.22)	6.15 (4.93)
No superior	4.68 (4.12)	0.55 (1.97)	0.11 (0.48)	0.09 (0.41)	0.24 (0.44)	0.06 (0.28)	5.73 (4.58)
Promedio	4.83 (4.31)	0.52 (1.81)	0.13 (0.52)	0.13 (0.49)	0.30 (0.48)	0.05 (0.25)	5.94 (4.76)

En el cuadro 6 se muestran ejemplos de tuits del conjunto de datos con sus clases, antes de ser procesados y de las características extraídas. Debido a la poca presencia promedio de la mayoría de estas, no hay mensajes que contengan todas al mismo tiempo.

Para las características anteriores, de cada conjunto de entrenamiento se extrajo un vocabulario, formado por el grupo de particularidades únicas en ese conjunto. En el cuadro 7 se muestran los promedios y las desviaciones estándar de los tamaños de cada vocabulario. Las ligas tienen el más extenso, seguidas de las palabras, y las abreviaturas, el menor; esto indica que hay una gran cantidad de ligas y palabras únicas.

Usando el vocabulario correspondiente, se realizó un proceso de vectorización por usuario utilizando el método *term-document-inverse-document-frequency* (*tf-idf*), el cual está definido como $tfidf(t, d) = tf(t, d) \times idf(t)$. En esta ecuación, $tf(t, d)$ es la frecuencia de la característica t (palabra, emoji, liga, etc.) en el documento d que, para nuestro caso de estudio, corresponde a un usuario. La parte $idf(t)$ está definida como $idf(t) = \log\left(\frac{1+n_d}{1+df(d, t)}\right) + 1$, donde

$df(d, t)$ representa el número de usuarios que contienen la característica t , y n_d es la cantidad total de usuarios en el conjunto de entrenamiento. Para cada característica en el conjunto de entrenamiento se calculó el *idf* y se guardó para ser utilizado en la vectorización del conjunto de prueba.

Adicional a las características individuales, también se construyeron matrices con el conjunto combinado de todas ellas. Al utilizar el método *tf-idf*, cada usuario queda representado como un vector, cuyo tamaño corresponde con el tamaño de los vocabularios mostrados en el cuadro 7, dependiendo de la característica utilizada. El método *tf-idf* es común cuando se trabaja con texto para medir la importancia de una característica respecto a un documento (usuario) y a toda la colección de documentos (usuarios).

De igual forma, se construyeron matrices usando los promedios de uso de cada característica por tuit por usuario. De esta forma, un usuario queda representado por un vector de siete promedios (uno para cada una de las seis características y uno para el uso de todas combinadas).

Cuadro 6

Ejemplos de tuits con sus características extraídas

Clase	Tuit	Palabras	Emojis	Etiquetas	Menciones	Ligas	Abr.
Superior	Cuando irán a traer Promare a Latinoamérica 😊	irán traer promare latinoamérica	😊				
Superior	No sé, me reí por 10 minutos XD https://bit.ly/3iFL4EU	reí minutos				https://bit.ly/3iFL4EU	Xd
No superior	#Bowsette al estilo megaman rollo 8bits	estilo megaman rollo bits		bowsette			
No superior	Sopa misógina como @DieBatsuDie https://bit.ly/3faA4Nx	sopa misógina			diebatsudie	https://bit.ly/3faA4Nx	

Tamaños de los vocabularios por característica

Palabras	Vocabulario					
	Emojis	Etiquetas	Menciones	Ligas	Abr.	Todas
233 607	1 222	32 122	31 356	305 730	108	594 469
(7 261)	(8)	(1 626)	(1 773)	(12 607)	(1)	(20 643)

Por último, se construyeron matrices usando como características los vectores de palabras *GloVe*, cuyo modelo mide estadísticas de coocurrencia entre palabras a partir de un conjunto de datos de entrenamiento. Para este trabajo, se utilizó un modelo preentrenado⁶ sobre el conjunto de datos *Spanish Billion Word Corpus* (Cardellino, 2016), el cual contiene un diccionario de más de 800 mil palabras, cada una representada por un vector de 300 particularidades. Para las características *GloVe*, se calculó el vector promedio de todos los vectores de las palabras encontradas en los tuits de cada usuario. De esta forma, cada usuario queda representado por un vector promedio de 300 características densas. Se aplicó el proceso de vectorización para los usuarios en cada conjunto de entrenamiento y prueba.

Una vez obtenidas las matrices de entrenamiento y prueba, se aplicó una normalización euclidiana (l_2) sobre cada vector de usuario de cada matriz para evitar problemas de escalamiento. La normalización euclidiana se define como $\bar{x}_N = \frac{\bar{x}}{\|\bar{x}\|_2} = \frac{\bar{x}}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}}$, donde \bar{x}_N representa el vector \bar{x} (un usuario) normalizado y $\|\bar{x}\|_2$ es la norma l_2 del vector.

Experimentación

Después de la vectorización, para cada una de las características utilizadas para representar a los usuarios se tenían 10 matrices de entrenamiento y 10 más de prueba normalizadas, siguiendo la validación cruzada estratificada de 10 partes. Con las primeras se construyeron y probaron modelos de predicción usando máquinas de SVM, MNB y DT.

Para optimizar el hiperparámetro de regularización C de los modelos SVM, en cada iteración de la validación cruzada se probaron los siguientes valores de C : 0.1, 1.0, 10.0, 100.0, realizando una subvalidación cruzada de tres partes con cada conjunto de entrenamiento con cada posible valor de C . Una vez ajustado el parámetro, se construyó el modelo final con el mejor valor de C y con todo el conjunto de entrenamiento.

Para medir el desempeño de los modelos, se usaron dos métricas basadas en una *matriz de confusión*, la cual está formada por cuatro celdas (ver cuadro 8) con número de verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Esta matriz relaciona las clases reales de los usuarios (columnas) con las predichas por el modelo (renglones). Se construye una matriz por cada clase, donde un usuario puede (positivo) o no (falso) pertenecer a la clase.

La primera métrica utilizada es macro $F1$, definida como $F1 = 2 \left(\frac{Precision \cdot Recall}{Precision + Recall} \right)$, donde $Precision = \frac{TP}{TP + FP}$ y $Recall = \frac{TP}{TP + FN}$. $F1$ representan la media armónica de *precision* y *recall*. En su versión macro, se calcula de forma independiente para cada matriz de confusión de cada clase y después se promedia. De esta forma, cada clase tiene el mismo peso en el resultado de la métrica, independientemente del número de usuarios en cada una.

La curva *Receiver Operating Characteristic* (ROC) es una de probabilidad que grafica la razón de verdaderos positivos contra la de falsos positivos en varios umbrales. El área bajo la curva ROC (AUC) evalúa un grado de separabilidad, midiendo la probabilidad de que un modelo clasifique más alto a un usuario en la clase superior que a uno de la no superior, elegidos ambos de forma aleatoria.

⁶ Disponible en: <https://github.com/dccuchile/spanish-word-embeddings>

Cuadro 8

Matriz de confusión

	Salida del modelo: verdadera	Salida del modelo: falsa
Valor real: verdadera	TP	FN
Valor real: falsa	FP	TN

La métrica macro $F1$ toma valores entre 0 y 1, donde 0 representa una clasificación totalmente errónea y 1 una perfecta. AUC toma valores entre 0.5 y 1, donde 0.5 significa un clasificador aleatorio y 1 representa al perfecto.

Para efectos de comparación, se consideran tres modelos base: el primero es un MNB que usa palabras con un pesado binario. Este primer modelo crea un vector para cada usuario con 1 si una palabra aparece en el texto del usuario y 0, si no aparece. El segundo es totalitario, el cual asigna a todos los usuarios en el conjunto de prueba a la clase mayoritaria (superior). El tercero es aleatorio uniforme, el cual destina un usuario a una clase con probabilidad de 0.5. Para los modelos 2 y 3, se tienen valores para las métricas de 0.45 para macro $F1$ y 0.5 para AUC.

No se considera la métrica de *accuracy* (definida con la ayuda del cuadro 8 como $\frac{TP+TN}{TP+FP+FN+TN}$), muy popular en clasificación, ya que puede no representar de manera adecuada el desempeño de los modelos para conjuntos con datos no balanceados (como es el caso del presente trabajo). En estas situaciones, por lo general, una (o más) clase(s) domina(n) el conjunto de datos. Esto propicia que el modelo tienda a clasificar a todos los usuarios en la clase dominante, obteniendo un buen resultado en esta métrica. Sin embargo, tal resultado puede ser engañoso, pues los usuarios en la clase minoritaria son ignorados.

Todos los códigos se implementaron en *Python* utilizando las librerías *NLTK*, *scikit-learn*, *Beautiful Soup* y *emoji*. Los códigos y los datos utilizados

para el procesamiento y la experimentación están disponibles en un repositorio en *GitHub*.⁷

Resultados

En el cuadro 9 se presentan los resultados de los experimentos con las características extraídas y los modelos de predicción. En los renglones 3 a 11 se indican las características: palabras, *emojis*/emojiconos, etiquetas, menciones, ligas, abreviaturas, la combinación de todas las anteriores, vectores *GloVe* y el número promedio de características en un tuit por usuario (longitud). Las columnas 2 a 7 indican las métricas separadas por modelo: SVM, MNB y DT. El renglón 12 (sombreado) muestra el promedio de la métrica de todas las características por modelo. Las columnas 8 y 9 (sombreadas) presentan los promedios de las métricas de todos los modelos por característica. Los renglones 13, 14 y 15 ofrecen los resultados de los modelos base: MNB con palabras y pesado binario (base 1), modelo totalitario (base 2) y modelo aleatorio uniforme (base 3). Se observa que el modelo base 1 tiene un desempeño similar a los otros modelos base. Los resultados mostrados son el promedio y la desviación estándar de las 10 iteraciones en la validación cruzada. En negritas se resaltan los valores más altos para cada métrica.

También, se observa que las características de etiquetas, menciones y ligas tienen un desempeño similar a los modelos base, por lo que no tienen poder discriminatorio. Solo las etiquetas con el DT o

⁷ https://github.com/jcgarranza/education_twitter

el SVM presentan resultados arriba de los modelos base. Considerando que los vocabularios de estas características son grandes (ver cuadro 7) y que su presencia en los tuits de los usuarios es baja (ver cuadro 5), se estima que ellas están distribuidas sobre los usuarios y no sobre las clases; es decir, que cada usuario introduce nuevas características de manera aislada. Por otro lado, las etiquetas pueden tener mejor desempeño porque se refieren a temas (tendencias), los cuales estarían relacionados con intereses de un determinado nivel educativo.

En el caso de las palabras, a pesar de tener una mayor presencia en los tuits y un vocabulario más alto, tienen un desempeño inferior a los *emojis*, cuya presencia y vocabulario son más limitados. Esto nos indica que hay una mayor superposición de las distribuciones de uso de palabras entre las

clases. Por otro lado, los emojis están más presentes en los tuits que las etiquetas, menciones o ligas (ver cuadro 5), por lo que su uso indica, en cierta medida, que las expresiones emocionales pictográficas, además de ser más populares, cambian entre los niveles educativos.

Más interesante es el uso de las abreviaturas, cuyo vocabulario es muy pequeño (en promedio, 108, con desviación estándar de 1) y tienen la menor presencia en los tuits (con solo 0.05 en promedio por mensaje, con desviación estándar de 0.25), pero que presentan el mejor desempeño de todas las características. Esto nos indica que las distribuciones del uso de abreviaturas entre las clases presentan diferencias, donde los usuarios de los dos niveles educativos las utilizan de forma distinta para comunicarse.

Cuadro 9

Resultados de la predicción del nivel educativo utilizando las diferentes características

Caract.	SVM		MNB		DT		Promedio	
	Macro F1	AUC	Macro F1	AUC	Macro F1	AUC	Macro F1	AUC
Palabras	0.52 (0.15)	0.54 (0.11)	0.45 (0.01)	0.50 (0.00)	0.55 (0.13)	0.57 (0.14)	0.51 (0.11)	0.54 (0.10)
<i>Emojis</i>	0.57 (0.16)	0.57 (0.15)	0.45 (0.01)	0.50 (0.00)	0.60 (0.09)	0.60 (0.09)	0.54 (0.10)	0.56 (0.10)
Etiquetas	0.45 (0.02)	0.49 (0.03)	0.45 (0.01)	0.50 (0.00)	0.53 (0.11)	0.53 (0.10)	0.48 (0.06)	0.51 (0.06)
Menciones	0.45 (0.01)	0.50 (0.00)	0.45 (0.01)	0.50 (0.00)	0.48 (0.08)	0.51 (0.07)	0.46 (0.05)	0.50 (0.04)
Ligas	0.45 (0.01)	0.50 (0.00)	0.45 (0.01)	0.50 (0.00)	0.45 (0.01)	0.50 (0.00)	0.45 (0.01)	0.50 (0.00)
Abr.	0.62 (0.18)	0.65 (0.17)	0.45 (0.01)	0.50 (0.00)	0.60 (0.15)	0.62 (0.15)	0.56 (0.13)	0.59 (0.13)
Todas	0.56 (0.15)	0.56 (0.13)	0.45 (0.01)	0.50 (0.00)	0.59 (0.11)	0.61 (0.12)	0.53 (0.11)	0.56 (0.10)
<i>GloVe</i>	0.60 (0.14)	0.61 (0.13)	--	--	0.53 (0.12)	0.53 (0.12)	0.57 (0.13)	0.57 (0.12)
Longitudes	0.51 (0.10)	0.54 (0.07)	0.45 (0.01)	0.50 (0.00)	0.57 (0.12)	0.58 (0.12)	0.51 (0.09)	0.54 (0.08)
Promedio	0.53 (0.12)	0.55 (0.10)	0.45 (0.01)	0.50 (0.00)	0.54 (0.11)	0.56 (0.11)		
Base 1			0.45 (0.01)	0.50 (0.00)				
Base 2			0.45	0.50				
Base 3			0.45	0.50				

Para ilustrar la diferencia en las distribuciones del uso de las abreviaturas, en la gráfica 1 se presenta la frecuencia relativa (dividida entre la frecuencia máxima) de cada una en cada clase. Algunas se utilizan con frecuencia similar (p. ej. *xd*, que es la más común en ambas clases, o *+ 1*, también usada con frecuencia), mientras que otras se usan de manera distinta (p. ej. *amix* se utiliza más en la clase superior y *yt*, en la no superior). Se realizó la prueba χ^2 de Pearson para homogeneidad sobre ambas distribuciones, con la hipótesis nula de que ambas son iguales. El resultado para χ^2 de la prueba fue de 6 236.93. Comparando este valor con el cuadro de referencia con una significancia de 95 % ($p = 0.05$) y 129 grados de libertad (valor en tabla de 156.508), se observa que es mayor, por lo que se rechaza la hipótesis nula y podemos decir que ambas distribuciones son distintas.

De forma adicional, se aplicó un análisis de componentes principales (PCA) a las matrices de los usuarios representados con características de palabras y abreviaturas. La gráfica 2 muestra el resultado de la proyección de los usuarios de cada clase al espacio de los dos primeros PCA. Se observa que los usuarios representados con palabras tienen una mayor superposición entre clases, lo cual

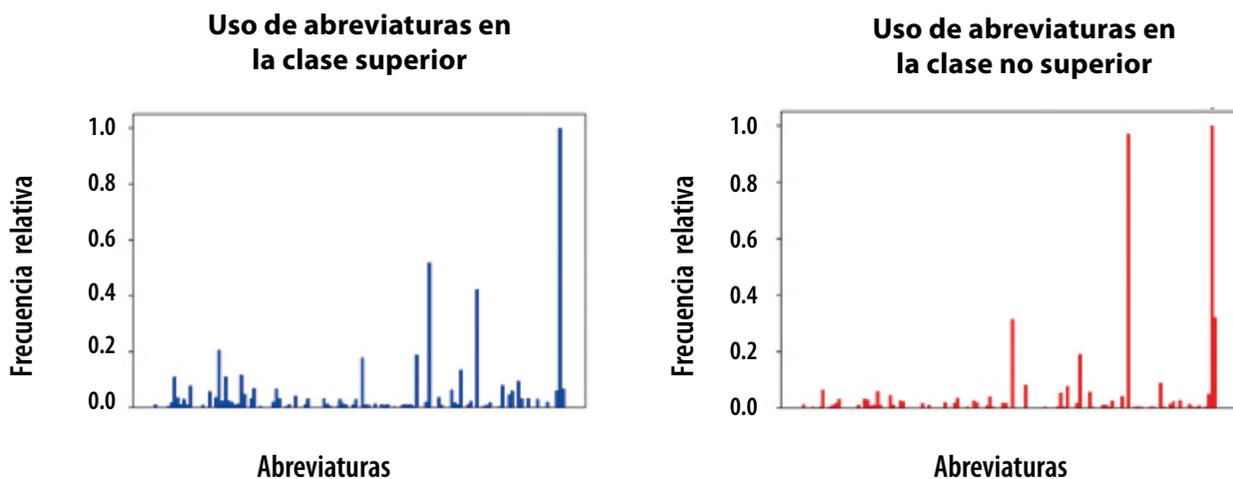
complica la clasificación, mientras que cuando se ha hecho con abreviaturas, se distingue una mayor separación. No obstante, a pesar de tener una mayor separación entre clases con las abreviaturas, los usuarios de la clase no superior están agrupados en áreas diferentes, con lo que la clasificación no es tan directa y, por ello, se observan valores aún bajos para las métricas de clasificación en esta tarea.

Por otro lado, cuando se consideran todas las características anteriores en conjunto (lo que sería equivalente a tomar todo el contenido del tuit), el desempeño es un poco mejor al uso de solo las palabras, pero no al de emplear solo los *emojis* o las abreviaturas. Esto indica que agregar el resto de las características a las abreviaturas o los *emojis* modifica la distribución de los usuarios en los espacios de las características, creando una mayor superposición y limitando la discriminación entre las clases.

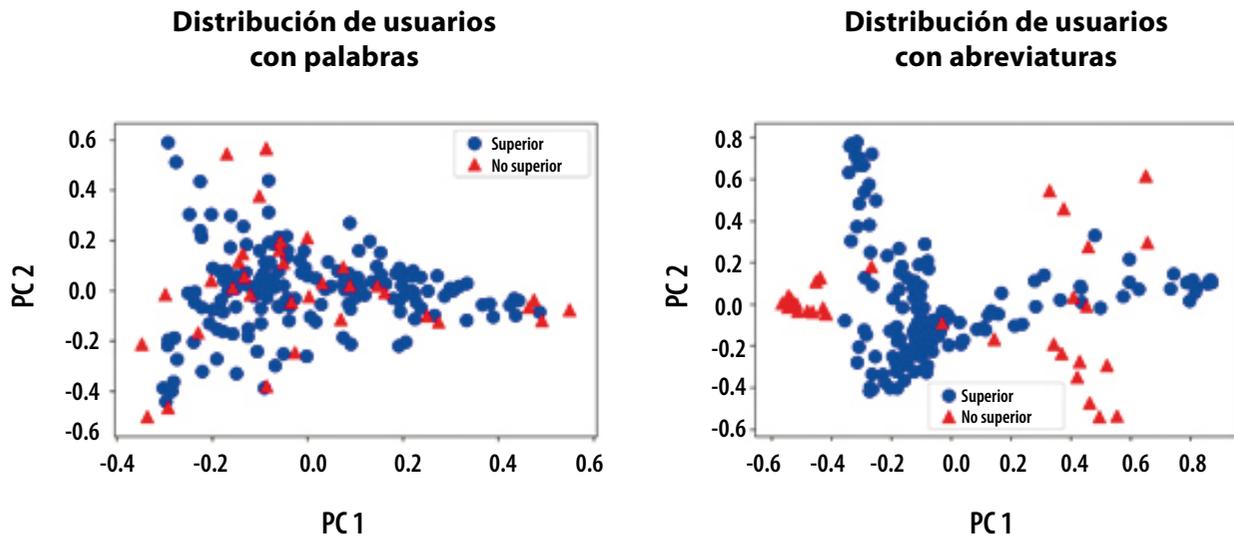
En el caso de los vectores de palabras *GloVe*, estos presentan un buen desempeño con el modelo SVM (casi similar a las abreviaturas), pero más bajo con el DT (inferior al uso de palabras). Esto significa que los usuarios representados en el espacio de

Gráfica 1

Frecuencia de uso de las abreviaturas entre clases



Proyección de los usuarios con PCA utilizando palabras y abreviaturas



características de estos vectores pueden ser mejor separados utilizando el hiperplano del modelo SVM, que con las reglas de decisión del DT.

Por último, cuando se utiliza el número promedio de características por tuit, se alcanza un desempeño similar al del uso de las palabras. Esto indica que el tamaño promedio de los mensajes que escribe un usuario refleja de manera parcial una diferencia entre las clases (ver cuadro 5), donde se muestra que los usuarios de la clase superior tienden a escribir tuits un poco más largos que los de la otra clase. No obstante, las diferencias no son suficientes para una discriminación completa.

Respecto a los modelos de predicción utilizados, se observa en el cuadro 9 que el MNB tiene un comportamiento igual al de los modelos base, por lo que las probabilidades condicionales de las características son similares entre clases y el modelo no puede discriminar entre ellas. Por otro lado, cuando se promedian sus resultados sobre todas las características, los modelos SVM y DT tienen un comportamiento similar, con ambos presentando su mejor desempeño con el uso de abreviaturas. El primero

también tiene buen comportamiento con los vectores de palabras *GloVe*, mientras que el segundo lo tiene con los *emojis*. En general, el mejor desempeño se alcanza con el uso de SVM y abreviaturas. Esto indica que los usuarios representados en el espacio de las abreviaturas pueden separarse de una mejor manera utilizando un hiperplano.

Conclusiones y trabajo futuro

En este trabajo se presentó un estudio sobre la predicción automática del nivel educativo en usuarios de *Twitter* en México. Se extrajo una serie de características del contenido textual de los tuits publicados por los usuarios y se utilizaron para construir tres tipos de modelos basados en aprendizaje automático: SVM, MNB y DT. La predicción fue de tipo binaria, donde los modelos determinaban si un usuario tenía estudios universitarios o no. Se probaron las características y los modelos con un conjunto de datos extraído de forma directa del sitio, compuesto por más de un millón de tuits en español, correspondientes a 195 usuarios ubicados en México.

De acuerdo con los experimentos para la predicción automática del nivel educativo de usuarios de *Twitter* en México, se concluye lo siguiente:

- Algunas características individuales, como las ligas, menciones y etiquetas, son demasiado específicas (haciendo referencia a sitios en la red, usuarios o temas particulares) y no aportan información suficiente para distinguir entre las clases.
- Otras, como las palabras, aportan un mejor contexto de los usuarios sobre temas de interés propios de la clase, lo cual ayuda a una mejor discriminación que las características previas.
- El uso de *emojis* es más popular que el de ligas, etiquetas o menciones. Además, estas expresiones emocionales pictográficas cambian entre los niveles educativos y pueden ayudar a separar entre clases con el empleo de reglas de decisión.
- Las abreviaturas presentan el mejor desempeño para separar entre usuarios de los dos niveles educativos. Esto indica que los usuarios de distintas clases usan las abreviaturas de formas diferentes.
- La combinación de todas las características introduce ruido, generando una mayor superposición entre las distribuciones de uso de estas y en la representación de los usuarios en el espacio de características.
- Los vectores de palabras *GloVe* capturan el contexto de cada palabra, lo que permite representar a los usuarios en un espacio donde pueden ser separados de forma parcial con un hiperplano.
- El número promedio de características en los tuits de un usuario es un indicador parcial de las clases, considerando que los de la clase superior tienden a escribir mensajes más largos.

Algunas ideas por explorar para trabajos futuros incluyen el uso de otros modelos de clasificación, como las redes neuronales, las cuales pueden funcionar de manera adecuada con características densas, como los vectores de palabras

GloVe. De igual forma, podemos considerar el uso de otros tipos de vectores de palabras, como *FastText*, que consideran de manera diferente el contexto de estas. Por último, se puede estudiar el uso de métodos de reducción de dimensión, como *Latent Dirichlet Allocation*, *Latent Semantic Indexing*, *Principal Component Analysis*, *Biased Discriminant Analysis* y *Non Negative Matrix Factorization*, entre otros, que se encargan de calcular asociaciones entre palabras para agruparlas en tópicos o temas.

Fuentes

- Cardellino, C. *Spanish Billion Words Corpus and Embeddings*. 2016.
- Desai, R. D. "Sentiment Analysis of Twitter Data", en: *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. Madurai, India, IEEE, 2018, pp. 114-117.
- Echeverría, V.; J. C. Gómez & M. F. Moens. "Automatic labeling of forums using bloom's taxonomy", en: *International Conference on Advanced Data Mining and Applications*. Springer, 2013, pp. 517-528.
- Kumar, S.; F. Morstatter & H. Liu. *Twitter Data Analytics*. New York, Springer Publishing Company, 2014.
- Özgiiven, N. & B. Mucan. "The relationship between personality traits and social media use", en: *Social Behavior and Personality: an International Journal*. 2013, vol. 41, no. 3, pp. 517-528.
- Pagolu, V. S.; K. N. Reddy; G. Panda & B. Majhi. "Sentiment analysis of Twitter data for predicting stock market movements", en: *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. IEEE, 2016, pp. 1345-1350.
- Pal, N. R.; K. Pal; J. M. Keller & J. C. Bezdek. "A possibilistic fuzzy c-means clustering algorithm" en: *IEEE Transactions on Fuzzy Systems*. 2005, vol. 13, no. 4, pp. 517-530.
- Pincay, J. & X. Ochoa. "Automatic classification of answers to discussion forums according to the cognitive domain of Bloom's taxonomy using text mining and a Bayesian classifier", en: *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 2013, pp. 626-634.
- Ramos, J. E. "Bases para la evaluación del dominio de las formas disciplinares de comunicación y de los usos lingüísticos especializados en el Espacio Europeo de Educación Superior (EEES)", en: *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*. Núm. 12, 2012, pp. 88-119.
- Rangel Pardo, F. M.; F. Celli; P. Rosso; M. Potthast; B. Stein & W. Daelemans. "Overview of the 3rd Author Profiling Task at PAN 2015", en: *CLEF 2015 Evaluation Labs and Workshop Working Notes Papers*. 2015, pp. 1-8.

- Rangel, F.; P. Rosso; I. Chugur; M. Potthast; M. Trenkmann; B. Stein (...) & W. Daelemans. "Overview of the 2nd author profiling task at PAN 2014", en: *CLEF 2014 Evaluation Labs and Workshop Working Notes Papers*. UK, Sheffield, 2014, pp. 1-30.
- Rangel, F.; P. Rosso; M. Koppel; E. Stamatatos & G. Inches. "Overview of the author profiling task at PAN 2013", en: *CLEF Conference on Multilingual and Multimodal Information Access Evaluation*. CELCT, 2013, pp. 352-365.
- Rangel, F.; P. Rosso; M. Montes-y-Gómez; M. Potthast & B. Stein. "Overview of the 6th author profiling task at PAN 2018: multimodal gender identification in Twitter", en: *Working Notes Papers of the CLEF*. 2018.
- Rangel, F.; P. Rosso; M. Potthast & B. Stein. "Overview of the 5th author profiling task at PAN 2017: Gender and language variety identification in twitter", en: *Working Notes Papers of the CLEF*. 2017.
- Rangel, F.; P. Rosso; B. Verhoeven; W. Daelemans; M. Potthast & B. Stein. "Overview of the 4th author profiling task at PAN 2016: cross-genre evaluations", en: *Working Notes Papers of the CLEF 2016 Evaluation Labs*. CEUR Workshop Proceedings/Balog, Krisztian (edit.) et al. 2016, pp. 750-784.
- Serban, I. V.; D. S. González & X. Wu. *Prediction of changes in the stock market using twitter and sentiment analysis*. 2014.
- Tess, P. A. "The role of social media in higher education classes (real and virtual)-A literature review", en: *Computers in human behavior*. Vol. 29, no. 5, 2013, pp. A60-A68.
- Vashishtha, S. & S. Susan. "Fuzzy rule based unsupervised sentiment analysis from social media posts", en: *Expert Systems with Applications*. Vol. 138, 2019, p. 112834.
- Wang, J.; H. Zhao & Z. Liu. "Topic Propagation Prediction Based on Dynamic Probability Model", en: *IEEE Access*. Vol. 7, 2019, pp. 58685-58694.
- Wang, L. & J. Q. Gan. "Prediction of the 2017 French Election Based on Twitter Data Analysis Using Term Weighting", en: *2018 10th Computer Science and Electronic Engineering (CEECE)*. IEEE, 2018, pp. 231-235.
- Wang, L.; J. Niu & S. Yu. "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis", en: *IEEE Transactions on Knowledge and Data Engineering*. 2019.
- Wasserman, S. & K. Faust. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge, Cambridge University Press, 1994.
- Winarko, E. & R. Pulungan. "Trending topics detection of Indonesian tweets using BN-grams and Doc-p", en: *Journal of King Saud University-Computer and Information Sciences*. Vol 31, no. 2, 2019, pp. 266-274.
- Wisdom, V. & R. Gupta. *An introduction to Twitter Data Analysis in Python*. 2016.
- Zhao, D. & M. B. Rosson. "How and why people Twitter: the role that micro-blogging plays in informal communication at work", en: *Proceedings of the ACM 2009 International Conference on Supporting group work*. New York, NY, USA, ACM, 2009, pp. 243-252.

The Dynamics of Disability due to Mobility Limitations and Ramp-Access Coverage

La dinámica de la discapacidad debida a las limitaciones de la movilidad y cobertura de rampas de accesibilidad

Tonatiuh Suárez-Meaney, Professor of Universidad Autónoma del Estado de México. Israel Albarrán Martínez, Special Project manager of Geoestrategias Electrónicas de Mercado, S. A. de C. V. Verónica Pérez Barrón, Researcher of Hospital Psiquiátrico Infantil Dr. Juan N. Navarro. Ioannis Chatzioannou, Instituto de Ingeniería, Universidad Nacional Autónoma de México. Alejandro Martínez López, Actuary manager of bussines intelligence of Registro Único de Vivienda. Luis Chías Becerril, Professor of Instituto de Geografía, Universidad Nacional Autónoma de México.

In this paper we seek to recognize whether the supply of disability ramps grows proportionally to the population that requires them. We compare the growth of the population with disabilities for mobility reasons in relation to the growth of ramps in terms of coverage within urban areas between 2010 and 2014 in Mexico. To carry out this comparison, it is necessary to discuss the volatility of the concept of disability due to its tendency to evolve steadily over the years, as it changed, even, in the study period making it very difficult to generate adequate measures that can provide useful geospatial comparisons. In this sense, the importance of tools such as the International Classification of Functioning, Dis-

En este trabajo buscamos reconocer si la oferta de rampas para discapacidad crece de forma proporcional a la población que las requiere. Se compara el crecimiento de la población con discapacidad por motivos de movilidad en relación con el de las rampas en términos de cobertura dentro de zonas urbanas entre el 2010 y 2014 en México. Para llevar a cabo este cotejo, es necesario discutir la volatilidad del concepto de discapacidad debido a su tendencia a evolucionar de manera constante a lo largo de los años, pues cambió, incluso, en el periodo de estudio haciendo muy difícil generar medidas adecuadas que puedan ofrecer comparaciones geoespaciales de utilidad. En este sentido, se destaca la importancia

Corresponding author: Verónica Pérez Barrón, Peten 46-304, col. Narvarte, CP 03020, CDMX, demarcación territorial Benito Juárez. Telephone: 5523008122, veronicaperezbarron@gmail.com

ability and Health (ICF) of the World Health Organization to achieve appropriate and useful results is highlighted. The provision of ramps in urban areas in Mexico was taken from the National Institute of Statistics and Geography (INEGI), as an index of local government response to the disability problem. It should be noted that the aforementioned INEGI databases are useful to describe the real situation in terms of coverage of this infrastructure. The results allow us to observe that the population with mobility limitations is growing faster than the construction of ramps and that there is a need for actions to increase the speed of their provision in order to fulfill the purposes that the new vision of disability due to mobility limitations has generated.

Key words: geography of disability; census; ramps; mobility limitation; ICF.

Recibido: 6 de mayo de 2019.
Aceptado: 10 de julio de 2020.

de herramientas como la Clasificación Internacional del Funcionamiento, la Discapacidad y la Salud (CIF) de la Organización Mundial de la Salud para lograr resultados apropiados y útiles. La provisión de rampas en áreas urbanas de México fue tomada del Instituto Nacional de Estadística y Geografía (INEGI), como índice de la respuesta del gobierno local al problema por discapacidad. Cabe destacar que las mencionadas bases de datos del INEGI son útiles para describir la situación real en términos de cobertura de esta infraestructura. Los resultados permiten observar que la población con limitaciones de movilidad crece más rápido que la construcción de rampas y que existe la necesidad de acciones para aumentar la velocidad de su provisión para cumplir con los propósitos que ha generado la nueva visión de discapacidad debido a limitaciones en la movilidad.

Palabras clave: geografía de la discapacidad; censos; rampas; limitación en la movilidad; CIF.



Una ayuda con las discapacidades de las personas/ Professional Studio/iStock

Introduction

The problem of disability is important because it is a factor of inequity against the families that face it. According to the micro-data published in ENIGH 2018 by INEGI, Mexican families are divided into two groups, those with at least one member facing disability and those without. It is observed that the former have on average less income, and while the proportion of their spending in housing and transportation it is very similar, the money spending for health by the people facing disability problems is much higher. This is just one example of how disability differentiates families. The modern concept of disability states that disability can pose serious problems in the realization of peoples everyday activities when the city's environments intensify it. The establishment of environments that are able to simplify the life of people with disabilities, is capable to reduce the differences between disabled and non-disabled households, as well as the large economic difference between them. Nowadays, in terms of percentages there are 2.1% of households with disabilities in Mexico (Table 1).

Legal framework and urban equipment for disability

This paper compares the dynamics of the population with disabilities due to mobility limitations in Mexican cities, and the advancement in the coverage of ramps for disability (the ramps are so-called

popularly, even though they have a wider use). The ramps nowadays, represent the main equipment for universal accessibility, which since the signing of the Convention about the Rights of People with Disabilities (1) are recognized in the laws of the Signatories, as mandatory implementation.

Mexican law seeks to "ensure accessibility in the basic infrastructure, equipment or urban environment and public spaces" as well as to promote the infrastructure for disability "to be universal, obligatory, progressive, signposted and adapted for all people", both in urban and rural areas¹. The Urban Development Act establishes that accessible spaces ensure free transit², and the actual regulations detail their minimum architectural characteristics in private centers³. Specifically, the legislation of Mexico City already considers the attention to mobility for people with disabilities as a top priority^{4,5}; for that reason, it obliges the architectural adaptation of the

- 1 *Ley General para la Inclusión de las Personas con Discapacidad*. Artículo 16. *Diario Oficial de la Federación* 30/05/2011. México. http://dof.gob.mx/nota_detalle.php?codigo=5191516&fecha=30/05/2011
- 2 *Ley General de Asentamientos Humanos, Ordenamiento Territorial y Desarrollo Urbano*. *Diario Oficial de la Federación*. 28/11/2016. México http://dof.gob.mx/nota_detalle.php?codigo=5462755&fecha=28/11/2016
- 3 *Norma Oficial Mexicana Nom-034-Stps-2016, Condiciones de Seguridad para el Acceso y Desarrollo de Actividades de Trabajadores con Discapacidad en los Centros de Trabajo*. 20/07/2016 México. *Diario Oficial de la Federación*. http://www.dof.gob.mx/nota_detalle.php?codigo=5445287&fecha=20/07/2016
- 4 *Ley de Movilidad del Distrito Federal* Artículo 6. Asamblea Legislativa del Distrito Federal. 15-03-2017. <http://www.aldf.gob.mx/archivo-c1649ef58afecae210136d1cc81036b5.pdf>
- 5 *Ley para la Integración al Desarrollo de las Personas con Discapacidad del Distrito Federal*, Artículo 9. Asamblea Legislativa del Distrito Federal. 03/mar./2016. <http://www.aldf.gob.mx/archivo-b28392e3c828c3108d8b0873830eb35d.pdf>

Table 1

Average quarterly income of households with members presenting or not presenting and proportion of spending of that income in housing, health, education and transportation, in Mexico, during 2018

	No	Yes
Average income	50,003	30,907
Proportion of spending on housing	6.11	5.95
Proportion of health spending	1.66	3.75
Proportion of spending on education	5.55	2.89
Proportion of spending on transportation	12.86	10.98

Source: National Survey of Households Income and Expenditure, National Institute of Statistics and Geography INEGI. 2019 Micro-data

spaces⁶, punishing those who do not respect them⁷. Thus, it is natural for the number of ramps to grow, but at the same time the user population grows as well, so it is necessary to measure the growth differential between the equipment designed for the mobility disabled population and the population with mobility disabilities, which is a complex concept.

Problem of definition of the disability concept and its evolution

Until 1900, following the juridical and global positivism of the time, the people with disabilities were not counted, they were even criminalized (2) (3), and referred with qualifiers such as blind, deaf, crazy and crippled. From 1950 to 2000, the disability “disappears” again from Mexican statistics.

Ledesma (4, P26) summarizes the modern vision of disability in four stages: a) Compassionate or Philan-

thropic stage; b) Rights recognition stage, since the 1950s disability is accepted as a “social problem”; c) Integration stage, between 1980 and 1990, people with disabilities are integrated actors in public decisions; d) Equality stage, since the present century, this stage seeks the visibility of people with disabilities, raising by that way the awareness about human rights.

On the other hand, Toboso and Arnau (5) consider the same number of stages but a different logic: Firstly, is the *stage of disregard*. Secondly, follows the *stage of medical rehabilitation or care model, focused on the individual*. Thirdly, is the *Social Model stage*, which is focused upon the surroundings of people with bases in the economic philosophy of Amartya Sen (6), or the lowest income, closely related in Mexico to disability (7). Finally, follows the *model of diversity* that is still under development. Pantano (8) since 2003 was able to understand that the available statistics were taking measurements of different disability’s concepts, while Schkolnik (10) warned of the need to include in the census data questions such as the severity of disability. Until 2010 returns in Mexico the census quantification that allows the comparisons between different census populations. The evolution of the concept of disability is presented in table 2.

6 *Ley de Salud del Distrito Federal*. Artículo 86-VI. Asamblea Legislativa del Distrito Federal. 15-03-2017. <http://www.aldf.gob.mx/archivo-2af4aa984657fe215d5fa76e7ae9cb97.pdf>
 7 *Ley de Desarrollo Urbano del Distrito Federal*. Artículo 99-III. Asamblea Legislativa del Distrito Federal. 24-03-2015. <http://www.aldf.gob.mx/archivo-3a6419f3c20189c5c79382d35f87c41f.pdf>

Table 2

Comparison of the evolution of the concept of disability

Phase	Period	General principles	Linguistics forms
Elimination	Primitive	Mystical fear of people with disabilities.	All kinds of stigmatization.
Discrimination and segregation	From antiquity until the nineteenth century	The people with disabilities represent a problem for society.	Words concerning the deficiencies: lame, one-armed, blind, deaf.
Compassionate or philanthropic	Early twentieth century	People with disabilities require care and help.	Sick, helpless.
Recognition of rights	1950-1979	The people with disabilities have rights.	Although the UN came to use the term “disabled” it began to be used the word “handicap”.
Integration	1980-2000	The people with disabilities participate.	Handicapped, disabled.
Equality	2001-2017	The person with disabilities is like any other person. In order to be able to make his/her rights valid, an equality of opportunities is needed.	Disability, but is still used the vision of disability as something intrinsic to people. Euphemisms like: people with different capacities.
Social model	21st Century	What’s important in a person is what they can do, not what they can’t. The disability is external-independent to the individual.	People with disabilities.

Proper elaboration through different texts revision (2 - 14).

The ICF. Functioning, disability, and health

After numerous international studies and consultations, in 2001, the World Health Assembly approved the use of the International Classification of Functioning, Disability and Health (ICF) (15). The ICF contains a series of classifications in order to describe the person, his or her health and the states related to each individual. Thus, it is a tool that permits to describe the disability of the people by detailing their conditions, documenting in depth the possible deficiencies in their functions and corporal structures, as well as their limitations and restrictions in the activity and participation; in interaction with environmental and personal factors and the health (15, 16).

The new concept of disability within the censuses

The quantification of the population with disabilities faces problems that have been already identified by several authors (6, 8, 9), in order to achieve similar and indispensable data so that to evaluate equal-universal access. Table 3 shows that in Mexico from 1885 to 1940, the ancient concepts of disa-

bility are used. In addition (Table 4), even though the 1990s are the time of greatest change in universal adaptation due to technology and political activism (17). It is until the beginning of the new millennium, thanks to efforts such as the ICF, when the disability is recognized as a condition and not as a problem. In 2010, in Mexico, the censuses begin to record the levels of functioning and disability. In 2000, names that are more appropriate are given to disabilities, and in 2010 is added the origin of the disabilities, that is a very useful fact considered for the elderly. Moreover, in 2010 there are discussions about different level of limitations based upon the ICF that have as a result the perfection of measurements concerning the severity levels of disabilities in 2014.

Methodology

Based on the comparative concepts, it is sought to know within different localities in Mexico, at what speed the population with disabilities due to mobility limitations increases compared to the growth of the main urban access equipment such as the ramps.

Table 3

Continues

Historical accounting of disability according to concepts of each year

Concepts	Year					
	1900	1910	1921	1930	1940	1950-1990
Total population (Thousands)	13,607.2	15,160.3	14,334.7	16,552.7	19,653.5	
Registered people with a disability (or addition of substitute concepts)	33.9	31.2	93.3	109.6	106.9	
Percentage of registered with disabilities with respect to the total population	0.2	0.2	0.7	0.7	0.5	

Table 3

Concludes

Historical accounting of disability according to concepts of each year

Concepts	Year					
	1900	1910	1921	1930	1940	1950-1990
Blinds	12.9	11.8	16.2	18.9	16.9	
Deaf-mute	9.0	7.8	3.5	6.1	5.1	
Deaf			15.0	15.5	13.3	
Mute			5.9	7.7	3.5	
Idiotism	4.1	4.2	4.8	6.8	6.0	
Cretinism	3.0	2.6				
Alienation	4.9	4.8				
Mads			5.2	8.3	9.2	
Hunchbacked			3.2	2.9	1.8	
One-armed			10.3	10.3	9.2	
Crippled			10.3	13.7	10.8	
Lame			19.0	19.3	21.9	
With physical and mental defects					9.4	

Table 4

Continues

Historical accounting of disability according to concepts of each year age via ICF approach

Concepts	Year		
	2000	2010	2014
Total population (Thousands)	97,483.4	112,336.5	119,990.0
Registered people with a disability (or addition of substitute concepts)	1,784.8	4,527.7	7,184.0
Percentage of registered with disabilities with respect to the total population	1.8	4.0	6.0
With some motor disability	809.7		
With some hearing limitation	280.0		
With some language limitation	86.7		
With some limitation			
Mental disability	287.6		
People who have difficulties in the performance and/or accomplishment of everyday tasks.		4,527.8	

Table 4

Concluces

Historical accounting of disability according to concepts of each year via ICF approach

Concepts	Year		
	2000	2010	2014
People who have difficulties in walking or moving, getting on and off the bus, sidewalks, stairs etc.		2,437.4	
People who have difficulties in bathing, dressing and/or eating.		229.0	
People who have difficulties in listening, even if they are using hearing devices.		498.6	
People who have trouble seeing, even when they use lenses.		1,292.2	
People who have difficulties maintaining a level of concentration in simple things.		209.3	
People who have some kind of difficulty or mental limitations		448.9	
People who have difficulties communicating with others or who have limitations on receiving and producing messages in order to make themselves understood through language, signs and symbols.		401.5	
People who are not able to walk, move and getting on and off the sidewalks, buses, stairs etc.			560.4
People who are not able to see even though they are using lenses			202.3
People who are not able to talk, communicate or converse			350.4
People who are not able to hear even though they are weare hearing devices			190.2
People who are not able to dress, bath or eat			481.8
People who cannot pay attention or learn simple things			251.1
People who have considerable difficulties in order to move, walk, ascend and descend			3,043.4
People who have considerable difficulties to see even when they weare glasses			2,612.9
People who have considerable difficulties to speak, communicate or converse			472.8
People who have considerable difficulties to hear even though they weare hearing devices			1,172.0
People who have considerable difficulties to bath, dress or eat			522.3
People who have considerable difficulties in order to pay attention or learn simple things			1,113.3

Material

In this paper we were able to use different databases that can be appreciated in table 5, the above mentioned databases were very helpful in order to perform the following two measurements:

a. Evolution of the population with disabilities

In order to obtain the quantity of people with disabilities at a Census Tract (CT) as well as at a Municipality level, it was necessary to consult the data of the 2010 Population and Housing census (*Censo de Población y Vivienda in Spanish*). Moreover, through the census sample we were able to identify the cause of the disability for each of the individuals. For the year 2014 it was used the National Survey of Demographic Dynamics (*Encuesta Nacional de la Dinámica Demográfica in Spanish*) available by INEGI (See table 5 in order to consult the sources) and we made use of the expansion factors offered by the above-mentioned survey.

b. Evolution of the supply of ramps

We were able to measure the coverage of ramps for every block (and its centroid) through the public data of INEGI named "Infraestructura y características del entorno urbano" (2010 y 2014) (See table 6), where each block can be evaluated and qualified according to the ramps availability that offers to the people via the surrounding roads. Every block can be qualified with the following five possible numerical indicators: 1) "all the surrounding roads have ramps" (total coverage), 2) "some of the surrounding roads have ramps" (partial coverage), 3) "none of the surrounding roads have ramps", 4) "housing complex", 5) "No specified". For this work we took under consideration only the blocks with values equal to "1", "2" or "3" because the blocks with values equal to "4" and "5" are very peculiar and they present a very small part of the sample. In figure 1 can be seen the qualification procedure, as well as the geo-informatics process that needs to be followed within a Geographic Information System. In table 6 can be appreciated the growth in ramps coverage according to the localities size growth in terms of total population.

Table 5

Continues

Data structure

	Capture instrument	Year	Primary sampling unit	Sample	Conceptual synthesis	Source	Obtained category or indicator	Web reference
1	Population and housing census	2000/2010	Population/housing	Universe	Its purpose is to count the resident population of the country, to update the information on its main demographic and socioeconomic characteristics, and to locate its distribution in the national territory.	INEGI	Variables described in tables 3-4	http://www.beta.inegi.org.mx/proyectos/ccpv/2010/
2	Census sample or expanded questionnaire	2010	Population/housing	2.9 millions of houses	The expanded questionnaire contains all the questions of the above-mentioned basic questionnaire and additional questions about the characteristics of the housing and its inhabitants.	INEGI	Variables described in tables 3-4	http://www.beta.inegi.org.mx/proyectos/ccpv/2010/

Table 5

Concludes

Data structure

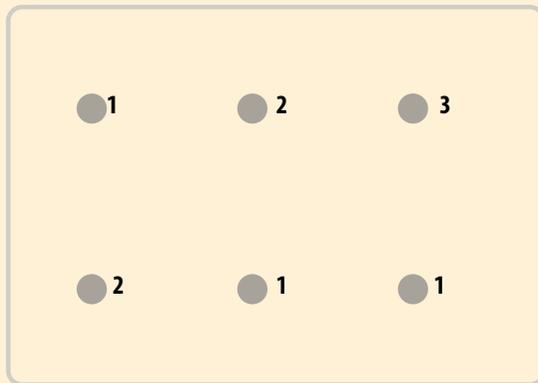
	Capture instrument	Year	Primary sampling unit	Sample	Conceptual synthesis	Source	Obtained category or indicator	Web reference
3	National survey of demographic dynamics	2014	Population/housing	101,389 houses	It was a statistical project aimed at enriching the supply of information of national interest linked to the national subsystem of demographic and Social information.	INEGI	Variables described in tables 3-4	http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/enadid/2014/
4	Characteristics of localities and the urban environment	2010	Blocks	1,806 localities	Derived from the population and housing census. This topic focuses on the roads infrastructure conditions, the availability of urban equipment, the existence of urban services and the accessibility conditions presented by the towns of 5,000 and more inhabitants.	INEGI	Blocks with wheelchair ramps availability in "All of its Roads"	http://www.inegi.org.mx/est/contenidos/Proyectos/acescomicrodatos/Otros/2010/default.aspx
5	Characteristics of localities and the urban environment	2014	Blocks		Its objective is to update the information related to the housing environment, published for the first time in 2010, whose capture instruments were applied during the operation called the update route of the national geostatistical framework.	INEGI	Blocks with wheelchair ramps availability in "All of its Roads"	http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/otras/localidades/2015/

Figure

Continues

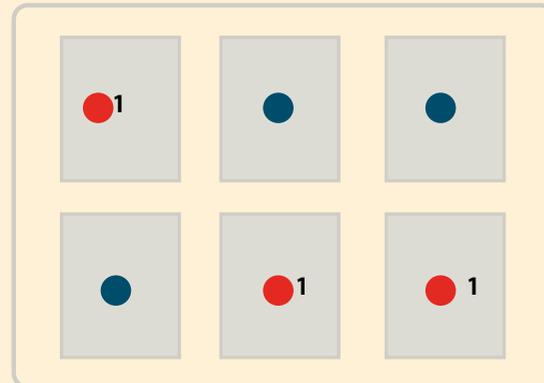
Structure of spatial data

Source Data



Points are generated from the Latitude-Longitude coordinates of the source data.

Blocks Data

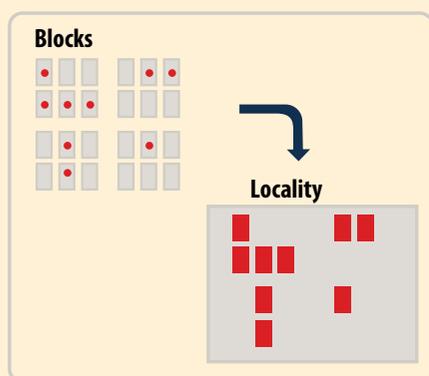


The source data is added to the polygons of blocks and the ones that do count with ramps in all roads are selected.

- 1. With ramps presence in all roads.
- 2. With ramps presence in some roads.
- 3. Without ramps presence.

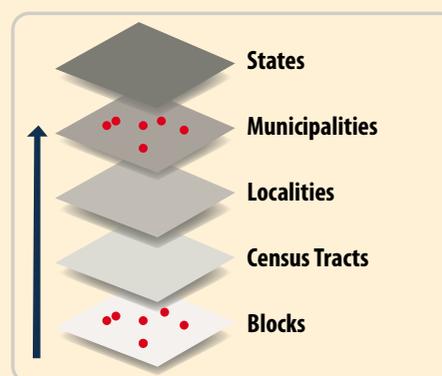
Structure of spatial data

Locality Data



The blocks contained in their corresponding Locality that do have presence of wheelchair ramps in all their roads are counted.

Geostatistical Framework



From the National Geostatistical Framework it is possible to apply geospatial data integrations and/or desintegrations depending on the level of the source data.

Table 6

Growth of ramps coverage according to locality size

Size of the locality by total population	Urban Blocks in the country		Blocks with complete ramps coverage		Blocks with partial ramps coverage		Percentage of blocks with complete ramps coverage		Percentage of blocks with partial ramps coverage		Growth in ramps coverage 2010 - 2014	
	2010	2014	2010	2014	2010	2014	2010	2014	2010	2014	Complete	Partial
3 millions and more	253,297	259,933	16,216	17,668	31,995	33,757	6.40	6.80	12.63	12.99	1.09	1.06
From 1 million to 2,999,999	148,895	154,424	6,220	7,482	15,182	17,083	4.18	4.85	10.20	11.06	1.20	1.13
From 500,000 to 999,999	245,751	254,598	10,076	12,497	25,464	29,161	4.10	4.91	10.36	11.45	1.24	1.15
From 100,000 to 499,999	228,216	234,786	8,817	10,440	23,046	28,005	3.86	4.45	10.10	11.93	1.18	1.22
From 50,000 to 99,999	56,249	58,073	972	1,193	4,427	5,758	1.73	2.05	7.87	9.92	1.23	1.30
From 15,000 to 4,999	139,319	143,939	1,163	1,492	8,401	10,256	0.83	1.04	6.03	7.13	1.28	1.22
Urban subtotal	1,071,727	1,105,753	43,464	50,772	108,515	12,020	4.06	4.59	10.13	11.22	1.17	1.14
Total national	1,377,010	1,456,594	44,278	60,904	113,913	142,664	3.22	4.18	8.27	9.79	1.38	1.25

Method

In order to obtain the previously mentioned measurements a) and b), it is necessary to analyze changes both in time and space with the following variables:

a , year. a_1 year 2010 or a_2 year 2014.

$T^a(c)$, total population in the year a within localities of size c .

$D^a(c)$, population with some kind of limitation in the year a within localities of size c .

$M^a(c)$, population with mobility's limitation in the year a within localities of size c .

$B^a(c)$, total number of blocks in the year a , within localities of size c .

$R^a(c)$, blocks with total coverage of ramps in the year a within localities of size c .

$r^a(c)$, blocks with partial coverage of ramps in the year a within localities of size c .

The growth of ramp coverage (expression 16 and 20 of table 7) and the relative population growth to mobility limitation (expression 11 and 14 of table 7) were estimated for all the localities.

We were able to calculate the population's proportion with some kind of limitation according to the size of locality via the expressions 1-9 seen in the table 7. The total national population is equal to the population of all the localities of all sizes as indicated in expression 6 of table 7. The sum of the population with disabilities due to some kind of limitation within all localities, is the population with disabilities at a national level, while the sum of population with disabilities in terms of mobility limitation within all localities is the national representation of mobility's disability. These are our national disability indicators. The principal indicators that need to be calculated can be seen in the expressions 9-20 of the table 7.

Table 7

Continues

Expressions for the comparison between population growth and ramps coverage

ID	Expression	Concept	Comparative indicator	Data
1	CPC	Population growth in localities of c size from year a_1 to a_2 .	$CPC = \frac{T^{a_2}(c)}{T^{a_1}(c)}$	
2	CPD	Growth of population with some kind of limitation in localities of size c .	$CPD = \frac{D^{a_2}(c)}{D^{a_1}(c)}$	
3	$CPDM$	Growth of the disabled population that presents mobility's limitation in localities of size c .	$CPDM = \frac{M^{a_2}(c)}{M^{a_1}(c)}$	
4	CCR	Growth of blocks with wheelchair ramps availability in all of their roads within localities of c size.	$CCR = \frac{R^{a_2}(c)}{R^{a_1}(c)}$	
5	$CMCR$	Growth of blocks with wheelchair ramps availability in some of their roads within localities of c size.	$CMCR = \frac{r^{a_2}(c)}{r^{a_1}(c)}$	
6	$PT(a)$	National population in the year a within all the localities n of size c .	$\sum_{c=1}^n T^a(c)$	

Table 7

Concludes

Expressions for the comparison between population growth and ramps coverage

ID	Expression	Concept	Comparative indicator	Data
7	$PTD(a)$	National disabled population in the year a within all localities n of size c .	$\sum_{c=1}^n D^a(c)$	
8	$PTM(a)$	National population presenting mobility's disability in year a within all localities n of size c .	$\sum_{c=1}^n M^a(c)$	
9	%PCD 2010	% of disabled population in 2010.	$PTD(2010) \times 100 / PT(2010)$	4.0
10	%PCD 2014	% of disabled population in 2014.	$PTD(2014) \times 100 / PT(2014)$	6.0
11	CPCD10_14	Growth of population with disabilities from 2010 to 2014	$\%PCD\ 2014 / \%PCD\ 2010$	1.5
12	%PCDM2010	% of population with mobility's disabilities in 2010	$PTM(2010) \times 100 / PT(2010)$	2.2
13	%PCDM2014	% of population with mobility's disabilities in 2014	$PTM(2014) \times 100 / PT(2014)$	3.0
14	CPCDM10_14	Growth of population with disabilities in mobility from 2010 to 2014	$\%PCDM2014 / \%PCDM2010$	1.4
15	%MzaT2010	% of blocks with ramps in all of their roads in 2010	$\sum_{c=1}^n M^a(c) \times 100 / R^a(c)$	3.2
16	%MzaT2014	% of blocks with ramps in all of their roads in 2014		4.2
17		Growth of %MzaT2010 to %MzaT2014	$\%MzaT2014 / \%MzaT2010$	1.3
18	%MzaA2010	% of blocks with ramps in some of their roads in 2010	$\sum_{c=1}^n M^a(c) \times 100 / r^a(c)$	8.3
19	%MzaA2014	% of blocks with ramps in some of their roads in 2014		9.8
20		Growth of %MzaA2010 to %MzaA2014	$\%MzaT2014 / \%MzaT2010$	1.2
21		Comparison		

When the population within a geographic entity grows, it is expected to grow the special groups of population as well (for example the disabled population). So if $a_2 > a_1$ (2014 > 2010), it would be expected that relations of all kinds of population for the two periods would grow, that is to say: $CPC > 1$, $CPD > 1$ y $CPDM > 1$. However, if we take into consideration proportions, and if the social and integrative model of the disability were fulfilled in accordance with the ICF canons, it would be expected that the proportion of the disabled population would decline, that is to say,

$$\frac{T^{a_2}(c)}{D^{a_2}(c)} < \frac{T^{a_1}(c)}{D^{a_1}(c)}$$

Also a social model of disability carried out in the different local governments, could suppose that the infrastructure in general for the disabled people, and in particular the ramps would increase each year in a greater proportion than that of the population with disability in order to

cover the objective of 100% universal access. That is to say, the $CMCR > 1$ and $CCR > 1$. Each local government would be initiating the urbanization of new blocks, trying to achieve complete coverage of ramps in neighborhoods that had partial coverage. In Mexico, we see that in the year 2000, 45% of the population with some kind of disabilities was reported; In 2010, this percentage raised as a 53% of the population with limitations in walking, moving, getting off and on public transportation modes was registered. Finally, in 2014, we are able to see that 50% of the population with some kind of disability was reported. As we can see the statistics is not comparable, but it oscillates between 45% and 53% of the mobility's limitation and also it oscillates between 4% and 6% of the population's total disability.

Change in time

The graph 1 shows the data of table 8. The growth rates of the disabled population ratios are higher

Graph 1

Growth of the proportion of ramps against disability



Table 8

Growth of the proportion of ramps against disability

Year	Population with disability	Population with mobility disability	Partial coverage of ramps	Total coverage of ramps
2010	4.0	2.2	3.2	8.3
2014	6.0	3.0	4.2	9.8
Growth Rate	1.5	1.4	1.3	1.2

Source from the INEGI National Demographic Dynamics Survey, 2014. INEGI Population and Housing Census, 2010.

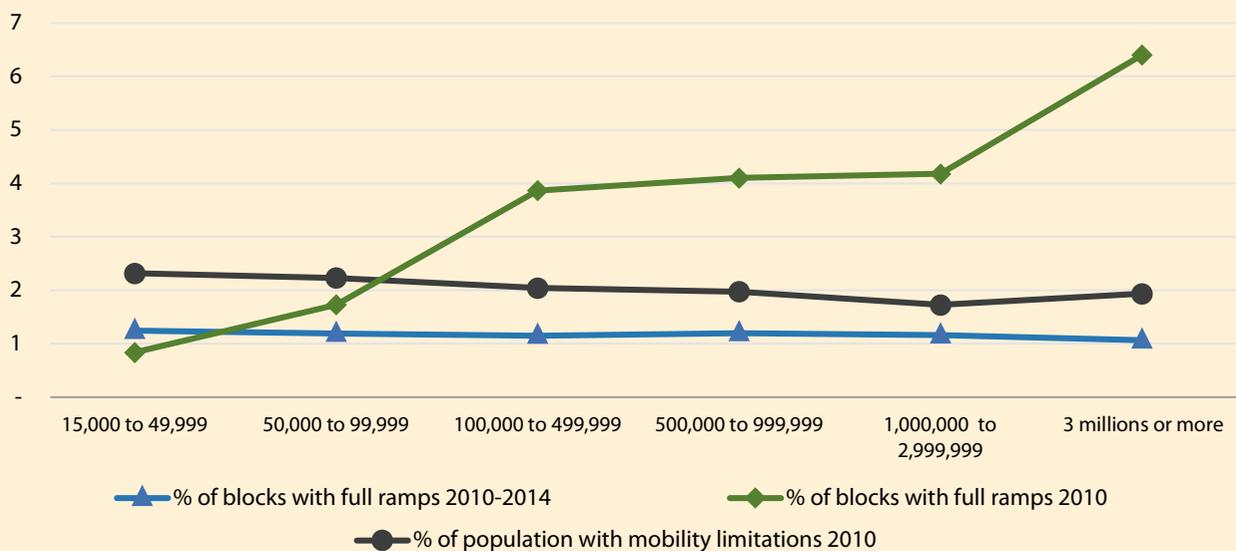
than the growth rates of ramp's coverage. In other words, the population with disabilities is growing faster than the coverage of ramps.

In general it can be noted that the coverage of ramps increase. The growth rate of partial ramps coverage though, is higher than that of total coverage. Which seems to indicate something that is usually seen on the streets, where the local governments who are in charge of establishing ramps coverage, implement their lines of action accord-

ing to their budgets, although it is known that partial coverage is not completely useful because if a wheelchair goes up at one point of the block, it is understood that it will have the opportunity to go down at another. This may involve working in parallel instead of following a sequential approach, that is, moving forward in many blocks at the same time instead of starting when only having completed the previous one. However, it is obvious, that the population with disabilities due to mobility grows faster than the coverage of ramps.

Graph 2

Full ramp coverage, disabled population due to mobility limitation and growth rate, according to locality size



Change in space

In graph 2 which shows the data of table 9, we can see how there is a tendency for the larger localities to have a lower proportion of population with disabilities due to mobility's limitation (by the year 2010). Maybe, this tendency can be justified because there are more health services or more prevention in these larger localities. It may also be due to the fact that the adjustments made by large cities allow people with different conditions to adapt to various activities. Moreover, through graph 2 we are able to see that the larger localities not only have fewer people with disabilities but more coverage of ramps infrastructure. This may be because in large cities there are more health services, greater access to education and greater possibilities for transportation.

The following graph shows how in larger cities there is more complete coverage of ramps, and at the same time, a smaller population with disabilities, even though the growth of the latter is almost constant in all city sizes.

Discussion

Urbanization-Metropolisation is currently one of the most important phenomena of the 21st century because concentrates a great number of people to urban centers in order to find better life opportunities. Nevertheless, urban environments, infrastructures, facilities and services, depending how they are planned and built, can impede or enable access, participation and inclusion of members of society (18).

Since the late 1980s it was known that a metropolis accessible to everyone did not have to be a utopia (19).

There is evidence in literature, revealing a widespread lack of accessibility in built environments, from roads and housing, to public buildings and spaces. Evidence also reveals a lack of accessibility to basic urban services such as sanitation and water, health, education, transportation, emergency and disaster response, resilience building, and access to information and commu-

Table 9

Full ramp coverage, disabled population due to mobility limitation and growth rate, according to locality size

City's size	% of blocks with full ramps 2010-2014	% of blocks with full ramps 2010	% of population with mobility limitations 2010
15,000 to 49,999	1.24	0.8	2.32
50,000 to 99,999	1.19	1.7	2.23
100,000 to 499,999	1.15	3.9	2.04
500,000 to 999,999	1.20	4.1	1.97
1,000,000 to 2,999,999	1.16	4.2	1.73
3 millions or more	1.06	6.4	1.93

nications. These accessibility limitations greatly contribute to the disadvantages and marginalization faced by persons with disabilities, leading to disproportionate rates of poverty, deprivation and exclusion (18).

However, major cities show greater attention to disability, due to their medical services and the level of accessibility to different services that they offer to their inhabitants, so there is less disability, but they also have more ramps coverage. Which would explain why people with disabilities from other localities come to them for treatment and if their income allows it, to live in them so that to find better accessibility conditions, not only due to the greater number of ramps, but also due to the possibilities offered by some public transport services, greater employment possibilities, access to health and educational facilities. Considering the current concept, the disability should not only be registered in terms of absolute numbers, but also in terms of severity, and whether it is experienced within a favorable environment or not. If the smaller cities do not offer universal access, they can expel disabled people, as well as, people belonging to priority care groups (elderly, pregnant women and people with strollers) and enter into a vicious circle: there are no ramps because there is no population that needs them.

The unfavorable disproportion that results comparing the growth speed of ramps coverage with the count of the number of people who have mobility's limitation is barely studied. Most countries do not even count their ramps, making it more difficult to diagnose barriers related to disability. Efficient ways to generate geo-spatial knowledge through well-established governmental inventories should be found that can help to assess the mobility of disabled people. Moreover, this kind of geo-spatial mechanisms should be enriched via information related to ramps quality, such as ramps slope, ramps width and ramps direction in order to secure that the ramps will be user friendly. Furthermore, it should be highlighted that ramps coverage is only one element required to improve universal accessibility and is one of the easiest and

perhaps least expensive to solve. On the other hand, generating public transport services, or having specialized health services, lifts and escalators is more complicated and expensive.

It is worth mentioning, that in the disability scenario, multiple complex relationships are registered between the limitations of people and the physical and socio-economic barriers that exist in the environmental context, including economic supports, jobs, access to education and personal situations. Hence, it is not surprising, to discover the inverse relationship between the number of ramps and the population with disabilities due to limitations in mobility.

Conclusions

Although the issue of disability is not new in the world, in Mexico it has statistically recovered since 2010, reflecting the little attention it has received officially and socially. The lack of ramp equipment and its below demand growth, precisely expresses the little attention given to the issue of disability in Mexico. The latter requires not only certain infrastructures such as ramps, but changes of perspective (social interventions and public policy) that propose and generate comprehensive and inclusive solutions. The concept of "smart cities" represents for countries like Mexico an utopia, for that reason, urban planners, sociologists, engineers, geographers and other specialists must work on the design and functionality of "inclusive cities" that allow universal accessibility for children, older adults, women, pedestrians and of course people with disabilities.

Considering these facts, the objective of this paper has been to recognize whether the supply of ramps equipment grows in a similar way as the population that requires it.

The importance of comparing the infrastructure for disability with the dynamics of the disabled population is that the programs and resources allocated and promoted by the government should

be evaluated via objective data and information, so that to recognize whether the rights of people with disabilities are being recognized.

The soundest part of this investigation is the creation of a methodology that through the application of a series of indicators allows us to recognize if the supply of ramps really improves.

The summary of our findings is very interesting because it shows that the lack of equipment such as ramps and their growth rate that is below the demand, requires actions that will permit the increase in the speed of ramps allocation in order to meet the purposes that have been generated by the new vision of disability by mobility's limitations.

Nevertheless, it is worth mentioning that the above mentioned methodology has some limitations, for instance, the fact that there is ramp coverage does not tell us anything about the appropriate quality of these ramps. In addition, there are blocks with very complex shapes that are not well surveyed by counting how many of their sides are covered by ramps. Moreover, there are very large blocks that may have a covered side by ramp but only in a small proportion, resulting by that way, insufficient the coverage of infrastructure.

In Mexico the INEGI has undertaken the task of counting and registering the characteristics related to urban environment, but not in all the countries can be seen this type of data collection, thus the methodology will not be easily applied in many other countries.

It is essential in an issue as important as disability that governments measure progress based on the infrastructure's quality and quantity destined to this condition. Therefore, measuring accessibility is an important task. Global efforts have already been made to classify disability, but we still do not know whether the city's environment is suitable. Thus, it can be measured within the urban space where and how destinations are dispersed (20).

To measure the city's environment accessibility, the local and regional scales must be considered. Within the latter it is important to take into account, the structure of travels origins and destinations, urban morphology, urban equipment (including ramps, sidewalks, etc.) and locally architecture. All these elements must form a system to measure the accessibility of the city's environment.

For the near future, a set of similar evaluations like the ones presented in this investigation can be carried out periodically in order to recognize which places have advanced more in the supply of universal spaces and which ones need to be helped.

References

1. Asamblea General-Naciones Unidas [Internet]. Convención sobre los Derechos de las Personas con Discapacidad, protocolo facultativo; 2006 [cited on 12 June 2020]. Retrieved from <https://www.ohchr.org/SP/ProfessionalInterest/Pages/OptionalProtocolRightsPersonsWithDisabilities.aspx>
2. Schweik SM. *The Ugly Laws: Disability in Public (The History of Disability)*. New York: University Press; 2010.
3. Da Silva Bampi LN, Guillhem D, Dornelles Alves E. Social model: a new approach of the disability theme. *Rev Latino-Am Enferm*. [Internet]. 2010 [cited on 5 Jun 2020]; 18(4):1–9. Retrieved from http://www.scielo.br/scielo.php?pid=S0104-11692010000400022&script=sci_arttext&lng=es
4. Ledesma JA (editor). *La imagen social de las personas con discapacidad. Estudios en homenaje a José Julián Barriga Bravo* [Internet]. España: Comité Español de Representantes de Personas con Discapacidad. Ediciones Cinca. Fundación Academia Europea de Yuste; 2009. [cited on 28 May 2020]. Retrieved from http://www.codajic.org/sites/www.codajic.org/files/Laimagensocialdelaspersonascondiscapacidad_0.pdf
5. Toboso Martín M, Arnau Ripollés MS. La discapacidad dentro del enfoque de capacidades y funcionamientos de Amartya Sen. *Araucaria, Rev Iberoam Filos Política y Humanidades* [Internet]. 2008 [cited on 6 Jun 2020]; 10(20):64-94. Retrieved from <http://www.redalyc.org/articulo.oa?id=28212043004> ISSN 1575-6823
6. Martínez Ríos B. Pobreza, discapacidad y derechos humanos. *Rev Española Discapac* [Internet]. 2013;1(1):9-32. Retrieved from <http://www.cedd.net/redis/index.php/redis/article/view/33/22>
7. Chías Becerril L, Pérez Barrón V, Suarez Meaney T, Albarrán Martínez I. Edad discapacidad y movilidad. *Revista Accesos* [Internet]. 2014;5(30):10-12. Retrieved from https://www.researchgate.net/publication/277405657_Discapacidad_Movilidad_y_edad

8. Pantano L. La medición de la discapacidad en Latinoamérica: los marcos conceptuales de preguntas censales. Lecciones de algunas experiencias. Boletín del real Patronato sobre Discapacidad [Internet]. 2003 [cited on 5 jun 2020]; 56. Retrieved from <http://sid.usual.es/idocs/FB/ART9326medicion.pdf>
9. Schkolnik S. América Latina: la medición de la discapacidad a partir de los censos y fuentes alternativas. Notas de Población CEPAL. 2011; 37(92): 275-306.
10. Stiker HJ. A history of disability. Series Corporalities: Discourses of Disability. Michigan: The University of Michigan Press; 1999.
11. Rose M. The staff of Oedipus. Transforming disability ancient Greece. Michigan: The University of Michigan Press; 2003.
12. Trompoukis C, Kourkoutas D. Greek mythology: the eye, ophthalmology, eye disease, and blindness. Can J Ophthalmol Can ... [Internet]. 2007;42(4):562–6. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0008418207801682>
13. Metzler I. A social history of disability in the middle ages. New York: Routledge. Taylor And Francis Group; 2013.
14. Pérez de Plá E. Sujeto, inclusión y diferencia: Investigación psicoanalítica y psicosocial sobre el Síndrome de Down y otros problemas del desarrollo. Ciudad de México: Universidad Autónoma Metropolitana Unidad Xochimilco; 2000.
15. Organización Mundial de la Salud. Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud. España: Publicaciones la Organ Mund la Salud. 2001;1-288.
16. Verhoef J, Toussaint PJ, Zwetsloot-Schonk JHM, Breedveld FC, Putter H, Vliet Vlieland TPM. Effectiveness of the introduction of an international classification of functioning, Disability and health-Based rehabilitation tool in multidisciplinary team care in patients with rheumatoid arthritis. Arthritis And Rheum [Internet]. 2007; 57(2):204-8.
17. Shapiro J. No Pity. People with disabilities forging a new civil rights movement. New York: Three Rivers Press; 1994.
18. United Nations [Internet]. Enhancing the accessibility of information and technology and inclusive development. Convention on the Rights of Persons with Disabilities. 29 March 2016 [cited on 12 June 2020]. Retrieved from https://www.un.org/disabilities/documents/COP/9/RT3/CRPD_CSP_2016_4-1603540E.pdf
19. Hahn H. Disability and the urban environment: a perspective on Los Angeles. Environment and Planning D: Society and Space. 1986;4(3): 273-288.
20. Church RL, Marston JR. Measuring accessibility for people with a disability. Geographical Analysis. 2003;35(1), 83-96.

Estimación del subregistro de las tasas de mortalidad infantil en México, 1990-2013

Estimation of Underreporting in Infant Mortality Rates in Mexico, 1990-2013

Lina Sofía Palacio-Mejía,* Diana Molina-Vélez,** Concepción García-Morales,** Amado David Quezada-Sánchez,**
Leonel González-González,** Maylen Rojas-Botero*** y Juan Eugenio Hernández-Ávila**

El presente trabajo tiene como objetivo comparar las tasas de mortalidad infantil observadas a través de las estadísticas vitales en México con las estimadas a partir de la Encuesta Nacional de la Dinámica Demográfica 2014, según sexo y ámbitos urbano o rural, en el periodo comprendido de 1990 al 2013 para calcular el nivel de subregistro que hay, cuyos resultados permiten concluir que este ha disminuido, pero todavía se considera alto para el país, sobre todo en áreas rurales.

Palabras clave: mortalidad infantil; omisiones de registro; estadísticas vitales; México.

The objective of this work is to compare the infant mortality rates observed through vital statistics in Mexico with those estimated from the National Survey of Demographic Dynamics 2014, according to sex and urban or rural areas, in the period from 1990 to 2013. This, in order to calculate the level of under-registration that exists, whose results allow us to conclude that this has decreased, but is still considered high for the country, especially in rural areas.

Key words: infant mortality; under-registration; vital statistics; Mexico.

Recibido: 29 de octubre de 2019.

Aceptado: 14 de septiembre de 2020.

* CONACYT-Instituto Nacional de Salud Pública de México (INSP), lpalacio@insp.mx

** INSP, cie06@insp.mx, cie03@insp.mx, amado.quezada@insp.mx, leonel.gonzalez@insp.mx y juan_eugenio@insp.mx, respectivamente.

***Universidad de Antioquia, Colombia, maylen.rojas@udea.edu.co



The child lies unconscious in the middle of a crowd of people. People look at the fallen child/ Andrii Yalanskyi / iStock

Introducción

La tasa de mortalidad infantil (TMI) relaciona el número de defunciones en niños menores de 1 año de edad y la cantidad de nacidos vivos (nv) en un mismo lugar y periodo, y expresa la probabilidad de que un recién nacido muera antes de cumplir el primer año por cada mil nv (OMS, 2005). Su magnitud, distribución y estructura se consideran indicadores sensibles de la salud general, del bienestar social, del acceso de la población a los servicios de salud y del grado de desarrollo de las poblaciones (Blaxter, M., 1981; Chin, B., L. Montana y X. Basagaña, 2011).

Pese a que en el mundo se han registrado importantes progresos en su reducción (UNICEF, 2017), la TMI continúa siendo una prioridad, sobre todo para los países en vías de desarrollo, de manera que la Organización de las Naciones Unidas (ONU) la incluyó entre las metas de los Objetivos de Desarrollo del Milenio (ODM), y en los subsiguientes Objetivos de

Desarrollo Sostenible (ODS) la sigue manteniendo (ONU, 2018; OMS, 2020 y ONU, 2005). En estos últimos, se ha considerado mejorar el registro de nacimientos y, de esta manera, proporcionar acceso a una identidad jurídica para todos (ONU, 2020), lo cual ayudará a disminuir el subregistro de la TMI. Según la Organización para la Cooperación y el Desarrollo Económicos (OCDE), México se encontraba en 2016 por debajo del promedio regional en la tasa de mortalidad infantil (11.5 por cada mil nv) con respecto al indicador de la región de América Latina y el Caribe (14.7) (OCDE, 2019).

Para evaluar el avance hacia el logro de estos objetivos, resulta fundamental tener información confiable que permita evaluar tendencias en salud. En este sentido, se recomienda a los gobiernos adoptar las medidas pertinentes para disponer de un sistema de información de estadísticas vitales (EEVV) con datos precisos, completos, oportunos y continuos sobre los sucesos vitales, como

nacimientos y defunciones (ONU, 2014); y si bien la mayoría de los países de América Latina cuentan con EEVV, algunos estudios han reportado deficiencias en la cobertura y calidad de los registros (OPS, 2007 y 2017).

Para el caso específico de México, el subregistro de nacimientos se ha estimado entre 5.8 y 33.8 % (UNICEF-INEGI, 2012; González Cervera, A. S. y R. Cárdenas, 2005), mientras que, en las defunciones totales, las estimaciones varían entre 5 y 5.3 % (Mathers, C. D. *et al.*, 2005). Estas investigaciones han evidenciado desigualdades en la magnitud del subregistro de la mortalidad en los ámbitos urbano y rural, así como en los grupos de edad, sobre todo en los niños y adultos mayores (UNICEF-INEGI, 2012 y Tomé, P. *et al.*, 1997). No obstante, las estimaciones no son consistentes entre sí debido al uso de diferentes metodologías, a las discrepancias en las fuentes de información o a los agregados territoriales analizados (Tomé, P. *et al.*, 1997; Cordero, E., 1968 y Langer, A. *et al.*, 1990).

Se espera que las EEVV registren los eventos vitales ocurridos en el país de manera universal, gratuita y oportuna; no obstante, en México se han identificado diferentes barreras en los ámbitos legal, geográfico, económico, administrativo y/o cultural que dificultan el registro efectivo en su sistema (UNICEF-INEGI, 2012); esta situación afecta la precisión y validez de las estimaciones con las que se hace seguimiento al fenómeno y comparaciones territoriales. Para subsanar de manera parcial esta situación, pueden aplicarse técnicas indirectas para la estimación del subregistro que permiten el ajuste de la tasa (ONU, 1986).

En este sentido, el propósito de este trabajo fue calcular el subregistro de las TMI en México comparando las tasas calculadas desde las estadísticas vitales con las estimadas a partir de la Encuesta Nacional de la Dinámica Demográfica (ENADID) 2014,¹ según sexo y ámbito urbano/rural, entre 1990 y el 2013.

¹ Levantada por el Instituto Nacional de Estadística y Geografía (INEGI) de México.

Material y métodos

Se llevó a cabo un estudio observacional, retrospectivo, con fuentes secundarias de información.

Fuentes de información

En México, el origen oficial de datos sobre muertes y nacimientos corresponde a los registros de estadísticas vitales recabados por el INEGI (2003), que fueron utilizados para calcular las TMI observadas (*TMI_o*) y, para realizar el análisis de comparación, las estimadas (*TMI_e*) a partir de la ENADID 2014 (INEGI, 2015).

Unidad de observación

Es la TMI nacional, por año de ocurrencia y con desglose según sexo y ámbitos urbano y rural.

TMI observada en las estadísticas vitales

Se calcularon a partir de los registros de estas, para cada año de ocurrencia en el periodo 1990-2013, así:

$$TMI_o = \frac{\text{Número de defunciones en menores de un año}}{\text{Número de nacidos vivos}} \times 1\,000$$

Para el denominador, se utilizó información de los certificados de nacimiento, ya que se tomaron en cuenta los nv, registrados por año de ocurrencia, del INEGI.

TMI estimada a partir de la ENADID

Para calcularlas, se utilizaron los datos de dos instrumentos de la ENADID (considerando que cada uno es una tabla en la base de datos de la Encuesta): el cuestionario para el hogar y el módulo para la mujer. En este caso, la población objetivo estuvo formada por los residentes habituales en las viviendas seleccionadas en el cuestionario del hogar y las mu-

eres de 15 a 54 años de edad, en el módulo mencionado (INEGI, 2015), correspondientes al 2014.

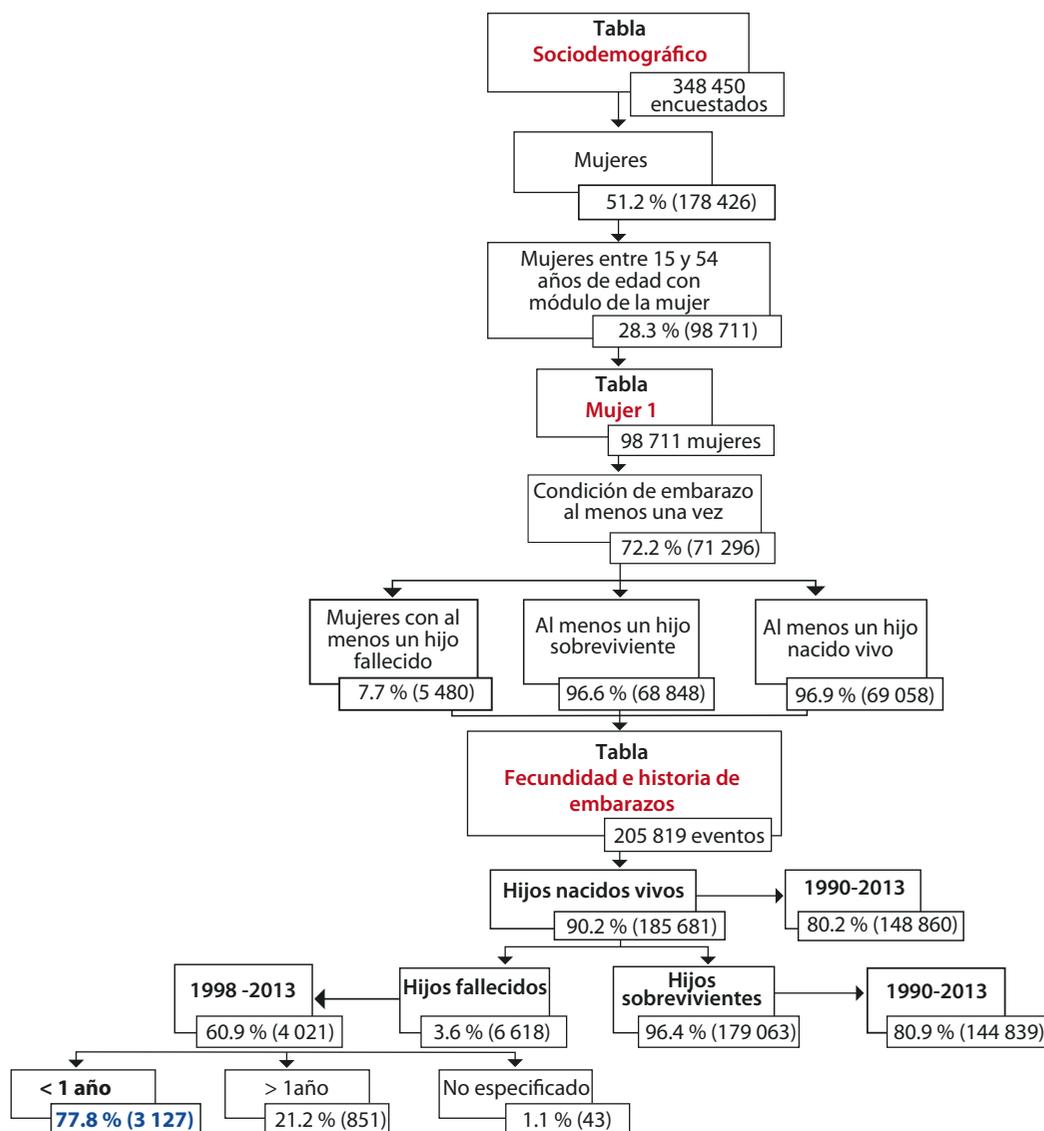
Las variables necesarias se identificaron en dos tablas: *Características de la mujer (Mujer1)* y *Fecundidad e historia de embarazos (TFec_Hemb)* (ver cuadro diagrama). En la primera, la unidad de observación (UO) fueron las *mujeres elegibles* (15 a 54 años de edad que contaban con el módulo de la mujer completo) y en la segunda lo constituían los eventos de embarazo, donde se desagrega la información de acuerdo con el tipo de desenlace

de cada observación: abortos, mortinatos, hijos sobrevivientes e hijos fallecidos, además de datos sobre la fecha de ocurrencia de los nacimientos o pérdidas, edad, sexo y ámbito (urbano y rural).

La obtención de la muestra para el análisis y estimación de las TMI en el periodo de interés se formó a partir de la información de los hijos sobrevivientes y fallecidos y el total de hijos nv; este último dato se estimó a partir de la suma de los dos primeros. Una vez identificada la población objetivo, se utilizaron los ponderadores de la muestra de la

Diagrama

Proceso de exploración de las bases de la ENADID 2014 para identificar la muestra y estimar las TMI en México, 1990-2013



ENADID para obtener el número de defunciones en menores de 1 año de edad y el de nv desagregados según sexo y ámbito. Al cociente entre ambas medidas multiplicado por mil se le denominó *TM_e*.

El porcentaje de subregistro en la TMI para México se estimó como la diferencia entre las *TM_e* (ENADID 2014) y las *TM_o* (INEGI, 2003), tomando como referencia los cálculos obtenidos a partir de la ENADID, así (INEC, 2016):

$$SBR \% = \left(\frac{(TM_e - TM_o)}{TM_e} \right) \times 100$$

donde:

SBR = porcentaje estimado de subregistro de mortalidad.

Con el fin de analizar las tendencias suavizadas de la tasa de mortalidad estimada y del subregistro, se aplicaron modelos de regresión por segmentos con especificación cúbica restringida y nudos en 1994, 1998, 2005 y 2009, los cuales corresponden, aproximadamente, a los percentiles 5, 35, 65 y 95 del periodo de observación (Harrell, F. E., 2001), para cada serie de tiempo. Los errores estándar se ajustaron por la autocorrelación

con el método de Newey-West y se especificó un proceso autorregresivo de segundo orden (ONU, 1986). Se estimaron modelos separados para cada sexo y para cada ámbito de residencia (rural o urbano). Todos los análisis se hicieron en *Stata*.

Resultados

Tasa de mortalidad infantil en México

Durante el 2013, y de acuerdo con la *TM_o*, en el país murieron 12.11 niños antes de cumplir el primer año de vida por cada mil nacidos vivos. En contraste, según la *TM_e*, esta cifra fue de 13.21. Como puede observarse en el cuadro 1, en todos los años existen brechas entre las tasas estimadas y observadas, lo que indica algún grado de subregistro y subestimación en la medida de la mortalidad infantil.

Las *TM_e* fueron más altas respecto a las *TM_o* en todo el periodo de estudio; ambas presentaron una tendencia decreciente, de manera que la tasa más alta estuvo al inicio de la serie (*TM_e* 34.25 vs. *TM_o* 24.72 en 1990) y la más baja, en el último año disponible (2013). Se mostró una disminución de 61.4 % según la estimada y de 51 % de acuerdo con la observada. Los intervalos de confianza de la *TM_e* se estrecharon de 25.0-32.7 a 13.0-16.5 para el 2013 (ver gráfica 1, panel izquierdo).

Cuadro 1

Continúa

Tasas de mortalidad infantil observadas y estimadas según sexo y área en México, 1990-2013

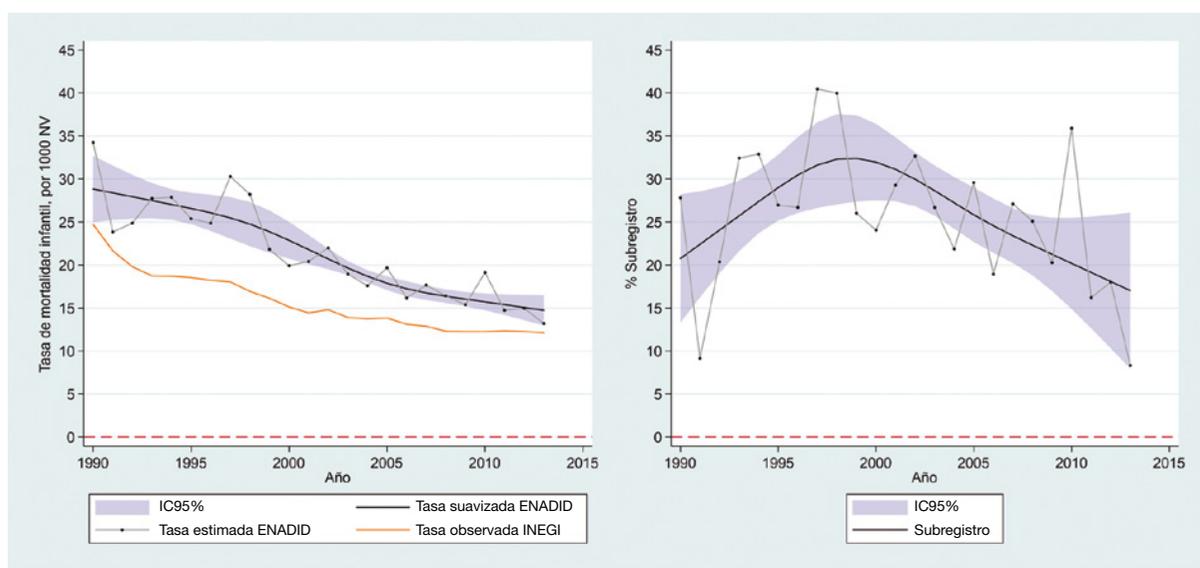
Año de ocurrencia	INEGI (tasa observada)					ENADID 2014 (tasa estimada)				
	Total	Sexo		Área		Total	Sexo		Área	
		Hombre	Mujer	Urbano	Rural		Hombre	Mujer	Urbano	Rural
1990	24.7	27.4	21.9	26.4	21.6	34.2	35.9	32.5	31.4	42.3
1991	21.7	24.1	19.1	22.8	19.4	23.8	26.5	21.0	20.7	32.7
1992	19.8	22.1	17.5	21.4	16.3	24.9	25.3	24.4	21.5	34.2
1993	18.7	20.8	16.6	19.9	16.1	27.7	33.1	22.2	27.8	27.4
1994	18.7	20.9	16.4	19.7	16.4	27.9	32.4	23.2	27.2	29.7
1995	18.5	20.7	16.3	19.6	16.1	25.4	23.3	27.5	22.3	33.7
1996	18.2	20.3	16.0	19.2	16.1	24.9	24.2	25.6	23.0	30.0

Tasas de mortalidad infantil observadas y estimadas según sexo y área en México, 1990-2013

Año de ocurrencia	INEGI (tasa observada)					ENADID 2014 (tasa estimada)				
	Total	Sexo		Área		Total	Sexo		Área	
		Hombre	Mujer	Urbano	Rural		Hombre	Mujer	Urbano	Rural
1997	18.0	20.1	15.9	19.0	15.8	30.3	34.5	25.8	27.6	38.2
1998	16.9	18.6	15.2	17.6	15.4	28.2	35.8	20.2	24.3	39.1
1999	16.1	17.9	14.3	17.0	14.0	21.8	27.1	16.4	20.6	25.0
2000	15.1	16.8	13.4	16.1	12.7	19.9	26.9	12.8	18.5	23.8
2001	14.4	16.1	12.7	15.3	12.3	20.4	25.0	15.4	18.6	25.4
2002	14.8	16.6	13.0	15.5	12.9	22.0	27.3	16.5	21.5	23.3
2003	13.9	15.7	12.1	14.7	11.8	18.9	17.6	20.3	17.7	22.3
2004	13.7	15.3	12.1	14.3	12.3	17.6	18.6	16.6	17.0	19.1
2005	13.8	15.3	12.4	14.2	12.9	19.7	19.1	20.2	17.5	25.6
2006	13.1	14.5	11.6	13.4	12.4	16.2	15.0	17.4	16.1	16.4
2007	12.9	14.3	11.4	13.0	12.6	17.7	19.6	15.7	16.8	20.1
2008	12.3	13.6	10.9	12.5	11.7	16.4	19.4	13.2	13.8	23.7
2009	12.3	13.5	11.0	12.8	10.8	15.4	17.4	13.2	15.4	15.4
2010	12.3	13.5	11.0	12.6	11.1	19.1	24.0	14.2	18.4	21.2
2011	12.3	13.6	11.1	12.8	11.1	14.7	16.9	12.5	14.6	15.1
2012	12.3	13.5	11.0	12.9	10.4	15.0	19.8	9.9	15.0	14.9
2013	12.1	13.4	10.8	12.5	11.0	13.2	15.4	10.8	13.2	13.3

Gráfica 1

Tasas de mortalidad infantil observadas vs. estimadas y estimación del subregistro general en México, 1990-2013



Subregistro general en la TMI

Para el 2013, se encontró un subregistro suavizado de 17 % (IC95 % 7.9, 26.1) con una disminución de 18.3 % con respecto a 1990 —20.8 % (IC95 % 13.3, 28.2)—; este presentó una tendencia ascendente al principio del periodo de estudio y su valor más alto fue en 1999 con 32.4 % (IC95 % 27.4, 37.4). A partir de ese año, la tendencia fue descendente y el más bajo se presentó al final del periodo. En el 2010, el subregistro observado tuvo un pico superior a 35 por ciento.

TMI según sexo en México

Las TMI estimadas y observadas fueron más altas en la población masculina durante todo el periodo, registrando sus valores más elevados en 1990 (35.9 estimada y 27.4 observada por cada mil nv), mientras en mujeres fue de 32.5 y 21.9, respectivamente (ver cuadro 1).

En el 2013, la *TMIe* para hombres fue de 15.4 por cada mil nv, mientras que la *TMIo*, 13.4. En mujeres, las dos fueron de 10.8 (ver cuadro 1). Se observaron disminuciones en la TMI en ambos sexos a lo largo del periodo de estudio; en mujeres, la reducción de la *TMIe* fue de 66.9 y 50.6 % en las observadas, mientras que en hombres fue de 57 y 51.3 %, respectivamente; el cuadro 1 presenta los datos absolutos de las tasas estimadas y observadas para todo el lapso de investigación.

Subregistro en la TMI según sexo

La tendencia de este tuvo un comportamiento diferente. Entre los hombres se observó un incremento entre 1990 y 1998, seguido de un descenso entre 1999 y el 2006. Desde el 2007 se evidencia una tendencia al aumento hasta el fin de la serie temporal estudiada. En general, el subregistro en población masculina presentó una elevación de 96.5 % entre 1990 y el 2013 —14.3 % (IC95 % 7.0, 21.5) a 28.1 % (IC95 % 15.1, 41.1), respectivamente—, mientras que en las mujeres fue a la baja pasando de 30.3 % (IC95 % 16.2, 44.4) en 1990 a -0.6 % (IC95 % -9.5, 8.3) en el 2013, con una disminución de 98 % durante el periodo (ver gráfica 3a).

El dato más elevado de subregistro suavizado para la población masculina fue en 1998 con 36.3 % (IC95 % 27.4, 45.3) y el más bajo, el de 1990 con 14.3 % (IC95 % 7.0, 21.5); por otro lado, en la población femenina, el más alto fue en el 2004 con 30.4 % (IC95 % 22.9, 37.9) y el menor, en el 2013 con -0.6 % (IC95 % -9.5, 8.3). Las diferencias entre hombres y mujeres fueron estadísticamente significativas a partir del 2010, llegando a ser casi 30 % mayor en ellos que en ellas durante el 2013 (ver cuadro 2 y gráfica 2).

TMI por ámbito en México

Las TMI fueron más altas en las zonas rurales en comparación con las urbanas durante la mayoría de los años. En 1990, la *TMIe* en las primeras fue de 42.3 y la *TMIo*, de 21.6 por mil nv, mientras en las segundas, fue de 31.4 y 26.4, respectivamente (ver cuadro 1). En el 2013, la *TMIe* fue de 13.3 y la *TMIo*, de 11.0 en áreas rurales y en las urbanas, de 13.2 y 12.5, en ese mismo orden (ver cuadro 1).

La tendencia en las TMI durante el periodo tuvo un descenso; este fue mayor en las *TMIe* en el ámbito rural (68.5 %) respecto al urbano (58 %), en tanto que en las *TMIo*, la disminución fue mayor en el urbano con 52.8 % en comparación con el rural, de 49.3 % (ver cuadro 1).

Subregistro por ámbito

Los porcentajes de subregistro de la TMI fueron mayores en el ámbito rural, siendo el punto más alto en 1998 con 51.5 (IC95 % 46.9, 56.0) y en el urbano en 1999 con 23.9 (IC95 % 18.1, 29.6), mientras que el menor para el área rural fue en 2013 con 25.1 (IC95 % 12.5, 37.7) y para la urbana en 1990 con 7.5 (IC95 % -4.1, 19.2). El ámbito rural tuvo una disminución de 44.8 % y el urbano un aumento de 117.3 % en el periodo. La diferencia entre el subregistro en el primero y el segundo fue durante todo el lapso de estudio con una tendencia al descenso, siendo en 1990 de casi 40 %, disminuyendo la brecha a menos de 10 % en el 2013 (ver cuadro 2 y gráfica 3).

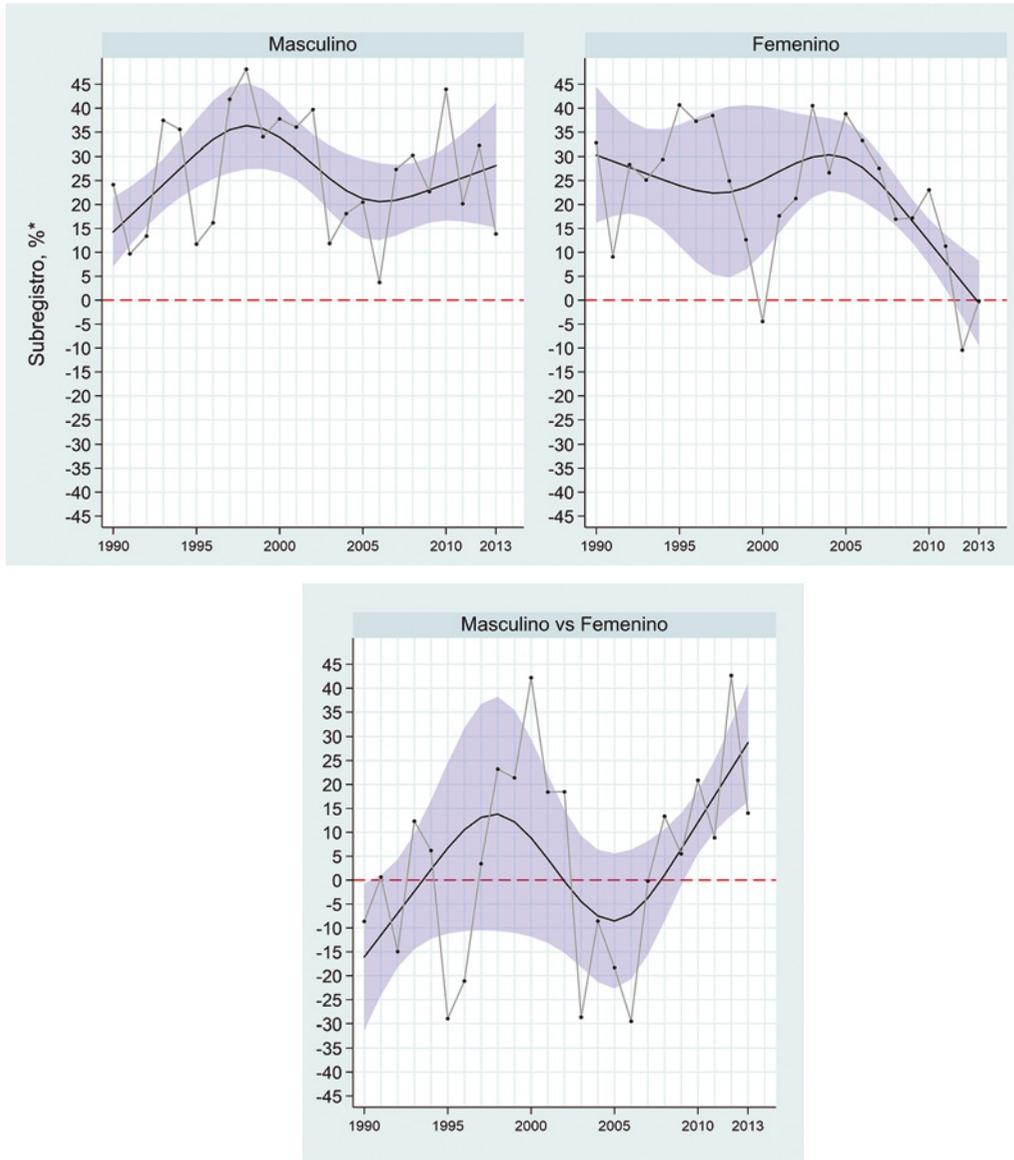
Cuadro 2

Subregistro de las tasas de mortalidad infantil observadas vs. estimadas según sexo y área en México, 1990-2013

Año de ocurrencia	INEGI (subregistro)						ENADID 2014 subregistro suavizado (IC95%)					
	Total	Sexo		Área		Total	Sexo		Área			
		Hombre	Mujer	Urbano	Rural		Hombre	Mujer	Urbano	Rural		
1990	27.8	23.6	32.6	15.7	48.8	20.8 (13.3, 28.2)	14.3 (7.0, 21.5)	30.3 (16.2, 44.4)	7.5 (-4.1, 19.2)	45.5 (41.5, 49.6)		
1991	9.1	9.2	8.8	-10.3	40.6	22.4 (16.2, 28.6)	17.6 (11.5, 23.7)	29.0 (17.6, 40.5)	9.7 (-0.1, 19.6)	46.4 (43.2, 49.7)		
1992	20.4	12.9	28.4	0.7	52.2	24.1 (19.1, 29.1)	20.8 (15.4, 26.3)	27.8 (18.1, 37.4)	12.0 (3.8, 20.1)	47.3 (44.7, 49.9)		
1993	32.4	37.2	25.1	28.4	41.4	25.7 (21.6, 29.8)	24.1 (18.8, 29.5)	26.5 (17.2, 35.7)	14.2 (7.5, 20.8)	48.2 (45.8, 50.6)		
1994	32.9	35.3	29.5	27.6	44.8	27.4 (23.7, 31.0)	27.4 (21.4, 33.4)	25.2 (14.8, 35.6)	16.4 (10.8, 21.9)	49.1 (46.4, 51.8)		
1995	27.0	11.4	40.7	11.9	52.3	29.0 (25.2, 32.8)	30.6 (23.6, 37.7)	23.9 (11.3, 36.6)	18.6 (13.5, 23.6)	50.0 (46.6, 53.4)		
1996	26.7	15.9	37.3	16.5	46.4	30.5 (26.1, 34.9)	33.5 (25.4, 41.6)	22.9 (7.9, 38.0)	20.6 (15.4, 25.8)	50.7 (46.7, 54.8)		
1997	40.5	41.7	38.4	31.1	58.7	31.6 (26.6, 36.6)	35.5 (26.6, 44.4)	22.4 (5.4, 39.4)	22.2 (16.6, 27.9)	51.3 (46.7, 55.8)		
1998	40.0	48.0	24.9	27.8	60.6	32.3 (27.1, 37.5)	36.3 (27.4, 45.3)	22.5 (4.7, 40.4)	23.4 (17.4, 29.3)	51.5 (46.9, 56.0)		
1999	26.0	33.8	12.7	17.6	43.8	32.4 (27.4, 37.4)	35.7 (27.5, 44.0)	23.5 (6.4, 40.7)	23.9 (18.1, 29.6)	51.2 (47.1, 55.4)		
2000	24.1	37.5	-4.3	13.1	46.8	32.0 (27.5, 36.4)	33.9 (26.8, 41.1)	25.1 (9.8, 40.4)	23.8 (18.7, 28.9)	50.6 (47.0, 54.2)		
2001	29.3	35.7	17.6	17.8	51.6	31.1 (27.4, 34.8)	31.4 (25.1, 37.6)	26.9 (14.1, 39.8)	23.3 (19.0, 27.5)	49.6 (46.3, 53.0)		
2002	32.7	39.3	21.2	27.7	44.8	30.0 (26.8, 33.1)	28.4 (22.3, 34.5)	28.6 (18.3, 39.0)	22.5 (19.1, 25.8)	48.3 (44.6, 52.0)		
2003	26.7	11.3	40.5	17.2	47.2	28.6 (25.7, 31.5)	25.4 (18.7, 32.2)	29.9 (21.5, 38.3)	21.5 (18.8, 24.2)	46.7 (42.2, 51.2)		
2004	21.9	17.7	26.7	16.1	35.8	27.2 (24.2, 30.2)	22.9 (15.2, 30.6)	30.4 (22.9, 37.9)	20.5 (18.0, 22.9)	44.9 (39.7, 50.2)		
2005	29.6	20.1	38.8	19.0	49.6	25.8 (22.7, 29.0)	21.2 (13.0, 29.4)	29.7 (22.4, 37.0)	19.6 (17.1, 22.1)	43.0 (37.4, 48.6)		
2006	18.9	3.2	33.2	16.9	24.6	24.6 (21.4, 27.7)	20.6 (12.5, 28.6)	27.7 (20.8, 34.7)	18.9 (16.4, 21.4)	40.9 (35.4, 46.4)		
2007	27.1	26.9	27.5	22.8	37.3	23.4 (20.2, 26.6)	20.9 (13.5, 28.3)	24.6 (18.5, 30.7)	18.4 (15.9, 20.9)	38.7 (33.4, 44.0)		
2008	25.1	29.8	16.9	9.2	50.4	22.3 (18.8, 25.8)	21.8 (15.0, 28.6)	20.8 (15.6, 25.9)	18.0 (15.2, 20.7)	36.5 (31.0, 41.9)		
2009	20.3	22.4	17.2	17.0	29.8	21.2 (17.0, 25.5)	23.0 (16.2, 29.9)	16.5 (12.1, 21.0)	17.6 (14.3, 20.9)	34.2 (28.0, 40.4)		
2010	35.9	43.7	23.0	31.1	47.6	20.2 (14.9, 25.5)	24.3 (16.6, 32.0)	12.2 (7.6, 16.9)	17.3 (13.2, 21.4)	31.9 (24.5, 39.3)		
2011	16.2	19.7	11.2	12.5	26.4	19.1 (12.6, 25.6)	25.6 (16.5, 34.7)	8.0 (2.3, 13.7)	17.0 (11.9, 22.0)	29.6 (20.7, 38.6)		
2012	18.0	31.8	-10.7	14.0	30.0	18.1 (10.3, 25.8)	26.8 (15.9, 37.8)	3.7 (-3.5, 10.8)	16.6 (10.6, 22.7)	27.4 (16.6, 38.1)		
2013	8.3	13.4	-0.3	5.2	17.7	17.0 (7.9, 26.1)	28.1 (15.1, 41.1)	-0.6 (-9.5, 8.3)	16.3 (9.2, 23.4)	25.1 (12.5, 37.7)		

Gráfica 2

Subregistro de las tasas de mortalidad infantil observadas vs. suavizadas según sexo en México, 1990-2013



* Unidades en puntos porcentuales para la comparación masculino vs. femenino.

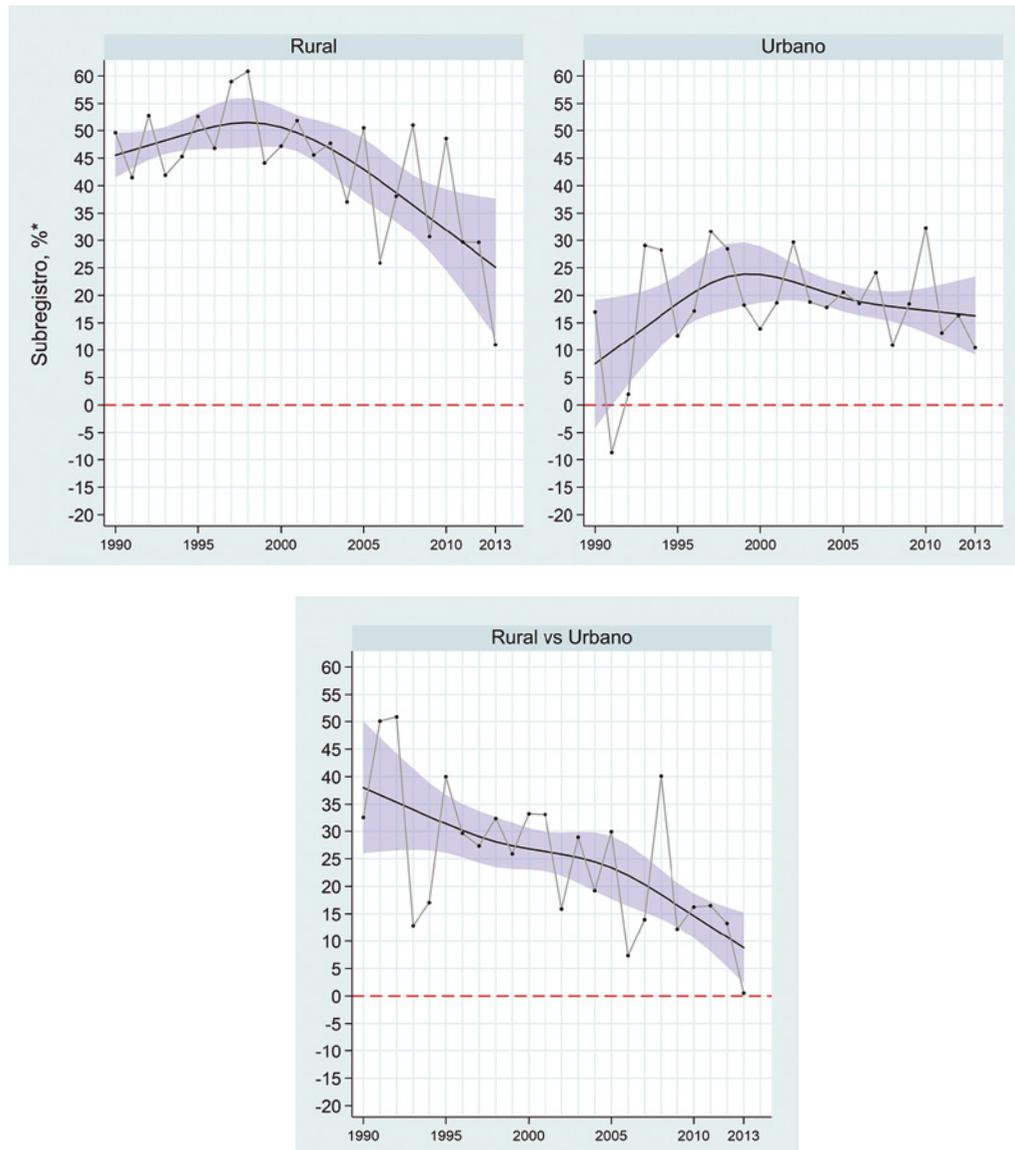
Discusión

Uno de los principales problemas para la medición de las tasas de mortalidad infantil es la calidad de los datos a los que se enfrentan los planificadores e investigadores de la salud, entre ellos el subregistro y la oportunidad de los datos, es decir, el atraso con el que son publicadas las estadísticas oficiales (Langer, A. *et al.*, 1990; Målqvist, M. *et al.*, 2008) y la

cobertura de estas. El análisis de la TMI a partir de datos observados en los que no se ha corregido el subregistro no mostraría la realidad demográfica de una población (Schmid, B., 2011), por lo cual es necesario estimar la TMI más cercana a la realidad. Para eso, en este trabajo se compararon los datos obtenidos de las estadísticas vitales con una estimación indirecta de la tasa de mortalidad infantil utilizando información de la ENADID 2014.

Gráfica 3

Subregistro de las tasas de mortalidad infantil observadas vs. suavizadas según ámbito en México, 1990-2013



*Unidades en puntos porcentuales para la comparación rural vs. urbano.

La tendencia de las TMI en el periodo fue a la disminución, tal vez como resultado de las políticas implementadas para cumplir con las metas del milenio de llegar a una tasa de 10.8 en el 2015 (INEGI, 2013). Según nuestros resultados, México estaba en el 2013 cerca de llegar a la meta establecida con una tasa de mortalidad infantil observada de 12.1 y una estimada de 13.2 defunciones por cada mil nv, lo cual representó un subregistro de

17 % (IC95 % 7.9, 26.1) en ese año. Diferencias de este tipo también fueron documentadas por Langer *et al.* en 1990, encontrando un subregistro de 52.9 % en el lapso 1982-1986, comparando los datos estimados con la Encuesta Mexicana de Fecundidad (*TMIe* 24.8 por mil nv) y los observados de las estadísticas vitales (*TMIo* 11.66). Según la OCDE, la tasa de mortalidad infantil en México en el 2013 fue de 12.2 por cada mil nv (OCDE, 2016) y los de

la Presidencia de México, de 13.0 (CONEVAL, 2018), indicador que era el mayor registrado en los países pertenecientes a dicha organización.

Nuestros resultados presentan una tendencia descendente del subregistro en los años analizados. Esto puede deberse a la implementación de estrategias en México como el Seguro Médico para una Nueva Generación a partir del 2006 (DOF, 2011 y 2012), el cual requiere el acta de nacimiento del niño para acceder a los beneficios del mismo. Estas medidas han contribuido a mejorar la cobertura y oportunidad del registro de nacimientos; también, han creado otras estrategias, como la del Embarazo Saludable, en el contexto del Sistema Nacional de Protección Social en Salud (Seguro Popular) (Secretaría de Gobernación, 2011), la cual ha permitido focalizar las acciones hacia la población infantil, además del Programa de Acción 2002-2010: un México Apropiado para la Infancia y la Adolescencia (SEDESOL, 2010), que se constituyó para disminuir los altos índices de morbilidad-mortalidad infantil de nuestro país (Aguirre Dugua, M. *et al.*, 2004). Aunado a esto, en el 2007 se implementó el *Certificado de nacimiento* como formato único a nivel nacional, gratuito, obligatorio, con carácter individual e intransferible, para constatar el nacimiento de un nacido vivo (DGIS-SSA, 2015). En el 2008 se desarrolló un estudio para la búsqueda intencionada de fallecimientos de niños en los 101 municipios con bajo Índice de Desarrollo Humano, realizada por la Dirección General de Información en Salud (DGIS) en conjunto con el Instituto Nacional de Salud Pública (INSP) (SSA, 2009). Asimismo, se planteó la meta de disminuir la mortalidad infantil 40 % como parte del Programa Nacional de Salud 2007-2012 (con la finalidad de avanzar en el cumplimiento de los ODM), donde se buscó facilitar el acceso a información confiable y oportuna que permitiera establecer un punto de partida de los indicadores de mortalidad infantil (SSA, 2007).

El subregistro de mortalidad infantil fue mayor en los niños que en las niñas; hallazgos similares han sido encontrados, y una posible explicación es que esto podría deberse a que, biológicamente, se presentan más nacimientos masculinos

(INEGI, 2017) y, por ende, más muertes (Fernández Cantón, S. B. *et al.*, 2012), dado que ellos tienden a ser más vulnerables que ellas (GBM, 2012). Sin embargo, estas diferencias solo fueron significativas del 2010 en adelante, ya que el subregistro en las niñas disminuyó de manera fuerte a partir del 2008. No se encuentra documentado en México qué políticas pudieron haberse implementado para disminuir el subregistro de mortalidad infantil en ellas, aunque pudiera relacionarse con algunas de inclusión y equidad de género establecidas por el gobierno federal. Sin embargo, es necesario investigar de forma más detallada estas diferencias para identificar factores que hicieron posible la reducción del subregistro en las niñas y aplicarlos a los niños, así como revisar, con la nueva Encuesta Nacional de Salud y Nutrición (ENSANUT) 2018,² que no fuera un problema de los datos.

Nuestros resultados indican que el subregistro fue más alto en las zonas rurales respecto a las urbanas; esto puede ser porque la calidad del registro sea mejor en estas, lo que da lugar a importantes sesgos en las estadísticas generales (Chen, L. M. *et al.*, 1998). Lo anterior puede deberse a factores culturales y de acceso tanto a los servicios de salud como a los del registro civil (Aguirre Dugua, M. *et al.*, 2004 y Hernández, B. *et al.*, 2012). Un trabajo realizado en Brasil sobre mortalidad neonatal menciona cómo los problemas relacionados con la distancia y accesibilidad geográfica impiden el acceso a los servicios de salud, que es donde se captan y registran eventos vitales como las defunciones y/o nacimientos, lo cual repercute en un subregistro y en errores en la calidad de la información registrada (Campos, D. *et al.*, 2007).

Cabe mencionar que, al estimar el subregistro con base en datos de una encuesta, las variaciones observadas están sujetas a errores de muestreo y sesgo de memoria; sin embargo, esto puede mejorarse incorporando información adicional, como datos censales y las mismas estadísticas vi-

² También levantada por el INEGI en coordinación con la Secretaría de Salud y el INSP.

tales (aunque incompletas), que permitan estimar con mayor precisión el subregistro y su desglose geográfico (Alarcón, D. y M. Robles, 2007).

No tener un buen registro de los nacimientos y las muertes en niños dimensiona las desigualdades en el acceso tanto de identidad legal como de salud (Chen, L. M. *et al.*, 1998 y Campos, D. *et al.*, 2007). La marginalidad y grandes distancias al registro civil y/o servicios de salud pueden ser factores que obstaculicen el registro de los nacimientos o defunciones en niños menores de un año y contribuyan al subregistro en las áreas rurales (Schmid, B., 2011; Campos, D. *et al.*, 2007 y Fernández, M. *et al.*, 2008). Lograr la universalización en el registro de las estadísticas vitales ocurridas en el país ayudaría a mejorar la calidad de la información sobre los mismos (OPS, 2020). Con esto se busca apoyar en el desarrollo de mejores políticas públicas y contribuir con una gobernanza mejor de los datos de los sistemas EEVV, que conlleva su producción, análisis, difusión y acceso (OPS, 2020).

Un error recurrente que se observa es el registro de la muerte de un nacido vivo como muerte fetal por el hecho de que murió de manera rápida, es decir, existe subregistro de nacimientos y de muertes infantiles e impacta directamente en la tasa de mortalidad infantil (OPS, 2017). Valdría la pena apostar por la capacitación del personal encargado del registro de estadísticas vitales, ya que esto puede ayudar a suavizar el subregistro de estos eventos.

La ENADID resultó ser una buena fuente de información para el cálculo de la TMI, sin embargo, está sujeta al sesgo de memoria de las madres seleccionadas. Realizar las estimaciones de tasas a través de la implementación de análisis estadístico puede ser una alternativa viable para la comparación de estas con el fin de trabajar, cada vez más, en la calidad y cobertura de las defunciones de los sistemas EEVV.

Algunas de las limitaciones del estudio, además del sesgo de memoria de las mujeres encues-

tadas, fueron el tipo de preguntas que hicieron en las encuestas para recabar información de la mortalidad; esto no permitió utilizar la serie de levantamientos de la dinámica demográfica para disminuirlo, pues las preguntas no eran comparables; es decir, existe el sesgo en la redacción de los cuestionamientos. También, debemos considerar como limitación el sesgo de carácter muestral, ya que la selección de la muestra puede incluir o excluir cierto tipo de resultados y, aunque dependa de la técnica del encuestador, podríamos considerar el error de respuesta, en caso de que lleven al encuestado a responder.

Otra limitante podría ser el hecho de calcular la edad del fallecimiento, ya que no tenemos certeza de que sea la edad real pues no se recabó la fecha completa, debido a que solo preguntan mes y año; esto ayudaría para analizar subregistros más completos y en otros grupos de edad.

Los datos presentados en esta investigación brindan una oportunidad para corregir la TMI en México, así como la oportunidad de identificar aspectos que funcionaron para medir el subregistro de la mortalidad, y entre ellos encontramos que la ENADID 1997 fue la que mejor permitió medir este fenómeno, ya que cuenta con un módulo de mortalidad y con preguntas básicas de los fallecidos en los últimos cinco años (sexo, años cumplidos al fallecer y fecha de defunción), lo que permitió calcular la mortalidad general, por grupos de edad, sexo y lugar de residencia (INEGI, 2014). Por ello, sugerimos que estas variables se mantengan en los cuestionarios futuros de la ENADID, pues en las posteriores las preguntas cambiaron y en la edición del 2014 solo se pudo calcular la mortalidad infantil.

Otros operativos identificados con variables sobre mortalidad fueron la Encuesta Demográfica Retrospectiva (EDER) (INEGI, 2011) y la Encuesta Nacional sobre Salud y Envejecimiento en México (ENASEM) (INEGI, 2015), cuya principal limitante es que solo cuentan con una o dos ediciones, sin periodicidad definida, lo cual limita su uso; también está la Encuesta Nacional de Ocupación y Empleo

(ENOE), que tiene periodos definidos de levantamiento, pero con un diseño distinto debido a que su objetivo está enfocado en recopilar información sobre las actividades ocupacionales de la población económicamente activa en el país (15 años de edad o más) (INEGI, 2018).

Fuentes

- Aguirre Dugua, M., K. Gallo Campos y J. López Macedo. *A un año de distancia: el bienestar de las niñas, niños y adolescentes*. México, D. F., Secretaría de Desarrollo Social, 2004. p. 34.
- Alarcón, D. y M. Robles. *Los retos para medir la mortalidad infantil con registros civiles incompletos*. Washington, D. C., Instituto Interamericano para el Desarrollo Social, 2007.
- Blaxter, M. *The health of the children. A review of research on the place of health in cycles of disadvantage*. London, UK, Heinemann Educational Books, 1981, p. 272.
- Campos, D., R. H. Loschi y E. França. "Mortalidade neonatal precoce hospitalar em Minas Gerais: associação com variáveis assistenciais e a questão da subnotificação", en: *Revista Brasileira de Epidemiologia*. 10(2), 2007, pp. 223-238.
- Chen, L. M., C. A. Sun, D. M. Wu, M. H. Shen y W. C. Lee. "Underregistration of neonatal deaths: an empirical study of the accuracy of infantile vital statistics in Taiwan", en: *Journal of Epidemiology and Community Health*. 52(5), 1998, pp. 289-292.
- Chin, B., L. Montana y X. Basagaña. "Spatial modeling of geographic inequalities in infant and child mortality across Nepal", en: *Health & Place*. 17(4), 2011, pp. 929-936.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). *Estudio diagnóstico del derecho a la salud 2018*. Ciudad de México, 2018.
- Cordero, E. "La subestimación de la mortalidad infantil en México", en: *Estudios Demográficos y Urbanos*. 2(01), 1968, pp. 44-62.
- Diario Oficial de la Federación (DOF)*. "Acuerdo por el que se emiten las Reglas de Operación del Programa Seguro Médico para una Nueva Generación, para el ejercicio fiscal 2012. 2011".
- _____. "Acuerdo por el que se modifica el diverso por el que se emiten las Reglas de Operación del Programa Seguro Médico para una Nueva Generación, para el ejercicio fiscal 2012". 27 de diciembre de 2011.
- Dirección General de Información en Salud (DGIS)-Secretaría de Salud (SSA). *Manual de llenado del Certificado de nacimiento*. Subsistema de Información sobre Nacimientos (SINAC). México, D. F., Dirección General de Información en Salud. Secretaría de Salud, 2015, p. 91.
- Fernández Cantón, S. B., G. Gutiérrez Trujillo y R. Viguri Uribe. "Principales causas de mortalidad infantil en México: tendencias recientes", en: *Boletín Médico del Hospital Infantil de México*. 69, 2012, pp. 144-148.
- Fernández, M., C. Guevel, H. Krupitzki, E. Marconi y C. Massa. *Omisión de registro de nacimientos y muertes infantiles. Magnitud, desigualdades y causas*. 2008, p. 196.
- Fondo de las Naciones Unidas para la Infancia (UNICEF, por sus siglas en inglés). *Estado mundial de la infancia 2017. Niños en un mundo digital*. Nueva York, UNICEF, 2017, p. 215.
- UNICEF-Instituto Nacional de Estadística y Geografía (INEGI). *Derecho a la identidad. La cobertura del registro de nacimiento en México en 1999 y 2009*. Aguascalientes, Ags., UNICEF-INEGI, 2012, p. 73.
- González Cervera, A. S. y R. Cárdenas. "Una aproximación a la medición del subregistro de nacimientos en las estadísticas vitales de México", en: *Estudios Demográficos y Urbanos*. 20(3), 2005, pp. 619-625.
- Greene, W. H. *Econometric Analysis*. 7th ed. New Jersey, Pearson, 2002, July 10.
- Grupo Banco Mundial (GBM). *Mortalidad infantil: nuevos datos y avances más rápidos*. 2012 (DE) <https://blogs.worldbank.org/es/voices/mortalidad-infantil-nuevos-datos-y-avances-m-s-r-pidos>
- Harrell, F. E. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. EE. UU., Springer, 2001, p. 568.
- Hernández B., D. Ramírez-Villalobos, M. B. Duarte, A. Corcho, G. Villarreal, A. Jiménez *et al.* "Subregistro de defunciones de menores y certificación de nacimiento en una muestra representativa de los 101 municipios con más bajo Índice de Desarrollo Humano en México", en: *Salud Pública de México*. 54(4), 2012, pp. 393-400.
- Instituto Nacional de Estadística y Censos (INEC). *Ficha metodológica. Estimación del subregistro de mortalidad general*. Ecuador, INEC, 2016, p. 5.
- Instituto Nacional de Estadística y Geografía (INEGI). *Encuesta Demográfica Retrospectiva (EDER)*. México, INEGI, 2011.
- _____. *Encuesta Nacional de la Dinámica Demográfica (ENADID)*. México, INEGI, 2014.
- _____. *Encuesta Nacional de la Dinámica Demográfica (ENADID) 2014. Antecedentes*. México, INEGI, 2015.
- _____. *Encuesta Nacional de Ocupación y Empleo (ENOE)*. México, INEGI, 2018.
- _____. *Encuesta Nacional sobre Salud y Envejecimiento en México (ENASEM)*. México, INEGI, 2015.
- _____. *Natalidad y fecundidad*. México, INEGI, 2017.
- _____. *Resumen ejecutivo. Los Objetivos de Desarrollo del Milenio en México. Informe de avances 2013*. Aguascalientes, INEGI, 2013.
- _____. *Síntesis metodológica del XII Censo General de Población y Vivienda 2000*. México, INEGI, 2003, pp. 1-56.
- Langer, A., J. L. Bobadilla, L. Schlaepfer_Pedrazzini. "Limitaciones de la mortalidad infantil como indicador de salud", en: *Salud Pública de México*. 32(4), 1990, pp. 467-473.
- Mållqvist, M., L. Eriksson, N. T. Nga, L. I. Fagerland, D. P. Hoa, L. Wallin *et al.* "Unreported births and deaths, a severe obstacle for improved neonatal

- survival in low-income countries; a population based study”, en: *BMC International Health and Human Rights*. 8, 2008, pp. 1-7.
- Mathers, C. D., D. Ma Fat, M. Inoue, C. Rao, A. D. Lopez. “Counting the dead and what they died from: an assessment of the global status of cause of death data”, in: *Bulletin of the World Health Organization*. 8383(3), 2005, pp. 171-177.
- Organización Mundial de la Salud (OMS). *Objetivos de Desarrollo Sostenible: Metas. Objetivo 3: Garantizar una vida sana y promover el bienestar para todos en todas las edades*. 2020 (DE) <https://www.who.int/topics/sustainable-development-goals/targets/es/>
- _____. *Estadísticas sanitarias mundiales 2005*. Francia, 2005.
- Organización de las Naciones Unidas (ONU). *De los ODM a los ODS*. Sustainable Development Goals Fund, 2018.
- _____. *Manual X. Técnicas indirectas de estimación demográfica*. Nueva York, 1986, p. 343.
- _____. *Objetivos de Desarrollo del Milenio: una mirada desde América Latina y el Caribe*. Santiago de Chile, 2005.
- _____. *Objetivos del Desarrollo Sostenible. Objetivo 16: Promover sociedades, justas, pacíficas e inclusivas*. 2020 (DE) <https://www.un.org/sustainabledevelopment/es/peace-justice/>
- _____. *Principios y recomendaciones para un sistema de estadísticas vitales*. Nueva York, ONU, 2014, p. 240.
- Organización Panamericana de la Salud (OPS). *Lineamientos básicos para el análisis de la mortalidad*. Washington, D. C., 2017.
- _____. *Plan de acción para el fortalecimiento de las estadísticas vitales (PEV) 2017-2022*. 2020 (DE) <https://www.paho.org/relacsis/index.php/en/areas-de-trabajo/registro-adecuado-de-causas-de-muerte/item/305-plan-de-accion-para-el-fortalecimiento-de-las-estadisticas-vitales- pev-2017-2022>
- _____. *Situación de las estadísticas vitales, de morbilidad y de recursos y servicios en salud de los países de las Américas*. 2007.
- OPS-Organización Mundial de la Salud (OMS) (eds.). *160.ª Sesión del Comité Ejecutivo. Plan de Acción para el Fortalecimiento de las Estadísticas Vitales 2017-2022*. 2017. Washington, D. C., Estados Unidos de América (EE. UU.), OPS-OMS, 2017.
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). *Estudios de la OCDE sobre los sistemas de salud: México*. París, OECD iLibrary, 2016 (DE) <https://www.oecd-ilibrary.org/docserver/9789264265523-es.pdf?expires=1592330058&id=id&accname=guest&checksum=3C9DAD0C054E66E175B647171B7751FC>
- _____. *Perspectivas económicas de América Latina 2019: desarrollo en transición*. París, 2019.
- Schmid, B. *Estimation of live birth underreporting with a capture-recapture method*. 45(6), *Sergipe*, 2011, pp. 1-10.
- Secretaría de Desarrollo Social (SEDESOL). *Informe 2007-2009. Un México apropiado para la infancia y la adolescencia*. México, 2010.
- Secretaría de Gobernación (SEGOB)-SSA. *Sistema de Protección Social en Salud. Informe de resultados 2011*. 2011, pp. 1-125.
- SSA. *Búsqueda intencionada de muertes maternas y de menores de cinco años, y registro de nacimientos en los 101 municipios con bajo Índice de Desarrollo Humano*. SSA, 2009, p. 92.
- _____. *Programa Nacional de Salud 2007-2012. Por un México sano: construyendo alianzas para una mejor salud*. México, D. F., 2007, p. 188.
- Tomé, P., H. Reyes, C. Piña, L. Rodríguez y G. Gutiérrez. “Características asociadas al subregistro de muerte en niños del estado de Guerrero, México”, en: *Salud Pública de México*. 39(6), 1997, pp. 523-529.

Inclusión del calendario mexicano en el proceso de desestacionalización de series de tiempo: el caso del Indicador Global de la Actividad Económica

*Inclusion of
the Mexican Calendar in the Process
of Seasonal Adjustment of Time Series:
the Case of the Global Economic
Activity Indicator*

Francisco de Jesús Corona Villavicencio y Jesús López-Pérez*

* Instituto Nacional de Estadística y Geografía, francisjoj.corona@inegi.org.mx y jesus.lopezp@inegi.org.mx, respectivamente.

Nota: nuestro agradecimiento a Juan Martínez, Liliana Martínez y Oziel Martínez de la Dirección de Estudios Económicos del INEGI por su valiosa orientación en el desarrollo de este trabajo; también se agradecen los comentarios de Gerardo Leyva, director general adjunto de Investigación del INEGI, por varias conversaciones sobre el tema.



Hojas de las cuatro estaciones/LUHUANFENG/iStock

En este trabajo se presenta una alternativa para desestacionalizar series de tiempo económicas mexicanas utilizando, en la etapa de ajustes previos del paquete *X-13ARIMA-SEATS*, fechas móviles del calendario mexicano, como el Día del Trabajo, la Independencia de México y la interacción que existe entre Semana Santa con el natalicio de Benito Juárez. Se ejemplifican los resultados para el Indicador Global de la Actividad Económica (IGAE) y sus tres grandes actividades económicas. Se concluye que las variables del calendario mexicano son estadísticamente significativas para el IGAE y para las actividades terciarias, de tal forma que los resultados obtenidos indican que, en los meses donde ocurren estos eventos móviles de la economía mexicana, las variaciones mensuales difieren a las publicadas de manera oficial. De esta forma, se recomienda al INEGI evaluar la utilización del calendario mexicano en la etapa de ajustes previos en el proceso de desestacionalización para las series de tiempo económicas más relevantes.

Palabras clave: ajustes previos; calendario mexicano; eventos móviles; Indicador Global de la Actividad Económica; *X-13ARIMA-SEATS*.

Recibido: 10 de octubre de 2019.
Aceptado: 22 de septiembre de 2020.

This paper presents an alternative to deseasonalize Mexican economic time series using, in the prior-adjustment stage of the *X-13ARIMA-SEATS* package, movable dates of the Mexican calendar, such as Labor Day, Mexican Independence Day and the interaction between Easter Week and the birth of Benito Juárez. The results are exemplified for the Global Indicator of Economic Activity (IGAE) and its three main economic activities. It is concluded that the Mexican calendar variables are statistically significant for the IGAE and for the tertiary activities, in such a way that the results obtained indicate that, in the months where these mobile events of the Mexican economy occur, the monthly variations differ from those published officially. Thus, it is recommended that INEGI evaluate the use of the Mexican calendar in the stage of previous adjustments in the deseasonalization process for the most relevant economic time series.

Key words: prior-adjustments; Mexican calendar; moving holidays; Indicador Global de la Actividad Económica; *X-13ARIMA-SEATS*.

Introducción

En 1983, el Banco de México inició un proyecto formal y sistemático para realizar la desestacionalización de series de tiempo económicas para algunas variables relevantes del país (ver Guerrero, 1990 y 1992). En la actualidad, este proceso lo realiza de forma oficial el Instituto Nacional de Estadística y Geografía (INEGI) usando el procedimiento que otorga el programa *X-13ARIMA-SEATS*.¹ Asimismo, para las variables más relevantes de la coyuntura económica, como el Producto Interno Bruto (PIB) o el Indicador Global de la Actividad Económica (IGAE), la desestacionalización se realiza con el consenso del Banco de México y las secretarías de Hacienda y Crédito Público y de Economía. Como es sabido, el objetivo del ajuste estacional es proporcionar información libre de fenómenos temporales regulares, relacionados con la periodicidad o frecuencia de las series, de tal forma que, una vez modelados y ajustados dichos fenómenos, estas reflejen solo los efectos que pueden ser aludidos a políticas económicas.

En consecuencia, los analistas de información prefieren estudiar el comportamiento de series de tiempo desestacionalizadas por diferentes motivos; por ejemplo, en el contexto de pronósticos, suele ser conveniente que las variables estén desestacionalizadas para mejorar la precisión del modelo, evitando modelar datos atípicos o estructuras temporales que puedan afectar la estimación de modelos multivariados como componentes principales; ver, por ejemplo, los trabajos recientes de Elizondo (2019) y Gálvez-Soriano (2020). Por otro lado, los economistas dedicados al análisis de coyuntura prefieren trabajar con ellas para tener mayor certeza de aspectos estructurales de corto plazo (Banco de España, 2002). Además, para aquellos agentes que toman decisiones, quienes enfrentan un *shock* y observan una serie con una componente estacional y otra no estacional fijarán su variable de control en la segunda de una forma que esta refleje la primera (Cochrane, 2005). Por

tal motivo, mientras mejor sea la calidad del ajuste estacional, mejores elementos de decisión se proveen a los usuarios y analistas de información.

Aunque el procedimiento dentro del contexto del programa *X-13-ARIMA-SEATS* consta de una variedad de análisis y uso de técnicas estadísticas (resumidos más adelante), podemos centrarnos en dos pasos: 1) especificación y estimación del modelo ARIMA y 2) selección y utilización de filtros estacionales. En el primero, es relevante aislar los efectos del calendario, como la frecuencia de los días-calendario, la Semana Santa (SS), los años bisiestos, valores atípicos u *outliers*, entre otros; una vez modelados estos, se especifica el modelo ARIMA, donde es importante verificar los supuestos de parsimonia y la no presencia de autocorrelación serial y estacionalidad residual en los errores. En la segunda fase se realiza la selección y utilización de filtros estacionales que permiten extraer las series de tendencia-ciclo y los factores estacionales, los cuales coadyuvan a evaluar la calidad estadística del proceso de desestacionalización y, en consecuencia, obtener la serie ajustada por estacionalidad.

Enfatizando en la etapa de ajustes previos, es claro que, para el calendario gregoriano, los días-calendario, la SS y los años bisiestos son indistintos entre los países, motivo por el cual pueden ser aplicados de manera homogénea, sin importar el país al cual nos referimos. Más aún, es claro que, si en una nación las festividades son fijas, es decir, si siempre ocurren en el mismo día del mes, sus inclusiones como variables exógenas son irrelevantes al modelo ARIMA.

Para el caso de México, existe la particularidad de que, antes del 2006, los días inhábiles aludidos a eventos particulares en el país eran movibles, situación que cambió a partir de dicho año, donde se volvieron fijos, con excepción del Día del Trabajo (DT), que se presenta el 1 de mayo, y el de la Independencia de México (IM), que ocurre el 16 de septiembre. Asimismo, aunque el natalicio de Benito Juárez (BJ) es una festividad fija (tercer lunes de marzo), este puede caer en la misma semana

¹ ARIMA hace referencia a Modelos Autorregresivos, Integrados de Medias Móviles y SEATS, a *Seasonal Adjustment by Signal Extraction*.

que SS (BJ-SS), situación que, de forma tentativa, es posible que tenga un efecto en la economía, pues podría convertirse en un alargamiento vacacional o, incluso, resultar irrelevante, pues el descanso es absorbido por las vacaciones escolares de primavera.

En este trabajo se presenta un exhaustivo análisis estadístico que estudia el hecho de incluir estas particularidades del calendario mexicano en el proceso de desestacionalización, ejemplificando su aplicación para el IGAE y sus tres grandes actividades económicas (GA), el cual permite generar el precedente de análisis de esta problemática en el contexto de series económicas mexicanas. La Oficina Europea de Estadística (Eurostat, 2015, 13) recomienda que, cuando se detecten efectos derivados de festividades movibles y cuyos efectos no sean despreciables, se trate de corregirlos o calcular y remover las estimaciones de sus efectos.

Antecedente a esta investigación, para el caso de Estados Unidos de América (EE. UU.), es el de Findley *et al.* (2005), quienes analizaron las capacidades del *X-12-ARIMA* para detectar el efecto de Semana Santa en series de tiempo sintéticas; encontraron que cuatro años de información son suficientes para la identificación de impactos grandes y moderados entre 0.08 y 0.03, mientras que se requieren al menos ocho años para detectar los más pequeños, menores a 0.01. Por su parte, Roberts *et al.* (2010) compararon cómo se ajusta el efecto de días laborales y festivos en los paquetes *X-12-ARIMA* y el método de la Eurostat (inmerso en el paquete *Demetra*) utilizando herramientas de análisis espectral; los resultados favorecieron, en general, al de Eurostat. Otra aplicación es la que hicieron Sarhani y Afia (2014) para analizar el efecto de festividades islámicas en Marruecos, para lo cual utilizaron el criterio de información de Akaike para seleccionar entre diferentes modelos estimados con el paquete *X-13ARIMA-SEATS*; en particular, encontraron que, para las actividades de transporte, el efecto previo a las festividades es de solo un día, mientras que para el consumo, este se afecta desde una semana an-

tes. En fecha reciente, en otro trabajo relacionado, McElroy *et al.* (2018) analizaron los efectos de festividades movibles en series de tiempo de frecuencia diaria para EE. UU.; los días feriados que se probaron fueron Pascua, *Black Friday*, *Cyber Monday*, *Labor Day*, domingo de *Super Bowl* y año nuevo chino, utilizando modelos estructurales con componentes estocásticas; encontraron que la inclusión de estos nuevos regresores permitió una modelación superior en algunos casos.

No obstante, para el caso de México —con excepción del trabajo de Guerrero *et al.* (2018), que analiza el efecto de las interacciones entre días-calendario con SS, concluyendo que dicha interacción es significativa para los meses de abril en el IGAE—, no hay un estudio que analice, en particular, las fechas movibles en el calendario mexicano en el proceso oficial de desestacionalización.

Se presume que el lector de este trabajo conoce la notación de los modelos ARIMA, así como que está familiarizado con conceptos esenciales básicos de desestacionalización de series de tiempo; no obstante, en caso de requerir un acercamiento introductorio al tema, se invita a revisar, por ejemplo, a Heath (2012), Guerrero *et al.* (2018) y el manual de referencia del programa *X-13ARIMA-SEATS* que provee el *US Census Bureau* (2020) para conocer más detalles sobre la notación y los conceptos que aquí se discuten.

Lo que resta de este documento está organizado como sigue: la sección dos presenta las generalidades del procedimiento del ajuste estacional en el contexto del *X-13ARIMA-SEATS*; la tercera muestra con detalle la aplicación de los métodos descritos usando el IGAE y sus tres GA; por último, en la cuarta se incluyen las conclusiones y recomendaciones.

Desestacionalizando series de tiempo con *X-13ARIMA-SEATS*

El procedimiento de ajuste estacional dentro del programa *X-13ARIMA-SEATS* es la evolución del método inicialmente desarrollado por el *Census Bu-*

reau de EE. UU., conocido como *Census Method II*, el cual progresó a *X-11* y, con el tiempo, a *X-12-ARIMA*. El *X-13ARIMA-SEATS* es utilizado por una gran cantidad de países en el mundo, sobre todo en América. A continuación, se hace un breve recuento de esta metodología.

Series de tiempo: aditivas multiplicativas

Se define que la serie de tiempo de periodicidad mensual o trimestral (Y_t , para $t=1, \dots, N$) está compuesta por una componente de tendencia-ciclo (T_t , que incluye los movimientos a largo plazo de la serie y las fluctuaciones económicas no atribuibles a su periodicidad y sí a eventos económicos relevantes, como pueden ser políticas gubernamentales; la estacionalidad (E_t), que incluye los efectos calendario y la dinámica relacionada con su periodicidad; y la irregularidad (I_t), que representa la componente aleatoria, es decir, no atribuible a factores de índole estructural, motivo por el cual se le aíslan posibles *outliers*, cambios de nivel, etcétera.

En consecuencia, la serie de tiempo puede ser aditiva, si la suma de las componentes T_t , E_t , e I_t generan a Y_t , o multiplicativa si es atribuible al producto de estas. Representaciones híbridas pueden verse en Dagum (1980), Findley *et al.* (1998) y Ladiray y Quenneville (2000-2001).

En el contexto del programa *X-13ARIMA-SEATS* solo hacemos referencia a series aditivas o multiplicativas, de tal forma que cuando son del segundo tipo, se pueden aplicar logaritmos a Y_t , transformándolas en aditivas. Para diagnosticar lo anterior, se puede consultar un gráfico, de tal manera que, si el patrón estacional no varía a través del tiempo, se relaciona con una estacionalidad aditiva; por otra parte, si este va cambiando, se suele decir que la serie es multiplicativa. Existen pruebas formales para concluir sobre el carácter aditivo o multiplicativo de la serie, por ejemplo, la automática que porta el paquete *X-13ARIMA-SEATS*, que está basada en los criterios de Akaike y pruebas de transformación-potencia (ver Guerrero, 1993).

Especificación y estimación del modelo regARIMA

En la etapa de ajustes previos buscamos tener una serie ajustada (\hat{Y}_t) por efectos del calendario y interrupciones en esta. Para ello, se utiliza un modelo de regresión con errores ARIMA (regARIMA), es decir, donde la función se describe por una combinación lineal de regresores (x) y la estructura de la covarianza es la de un proceso ARIMA:

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D$$

$$(y_t - \sum_i \beta_i x_{it}) = \theta(B)\Theta(B^S)a_t. \quad (1)$$

Note que, si no se emplea ningún regresor, el regARIMA se reduce a uno ARIMA.

El programa *X-13ARIMA-SEATS* tiene ya implementados diversos regresores para estimar y aislar los efectos de calendario, por ejemplo: días hábiles (TD por *Trading Days*), año bisiesto (LY por *Leap Year*) y de festividades movibles (como SS para la Semana Santa). También, hay otros regresores para ciertos tipos de interrupciones en las series (*outliers*), los cuales se pueden clasificar como aditivos (AO por *Additive Outlier*); de cambio permanente de nivel (LS por *Level Shift*), cuyo efecto se mantiene una vez que aparece y se representan mediante una función de escalón; de cambio transitorio (TC por *Transitory Change*), que una vez que ocurren muestran un patrón de absorción en la serie y desaparecen a la larga; y de efecto rampa (RP por *Ramp*).

La forma como se construyen e introducen estas variables en la ecuación (1) se muestra en el cuadro 1.

Además de las variables aquí mostradas, el modelo regARIMA en el paquete *X-13ARIMA-SEATS* permite incorporar otras opciones predefinidas o bien otras definidas por el usuario. En este trabajo se utiliza la expresión (1) para incorporar los efectos de calendario que deseamos probar (en específico DT, IM, BJ-SS), lo cual se detalla en la tercera sección.

Variables predefinidas en el X-13ARIMA-SEATS

Regresor	Definición de la variable
Días hábiles	$T_{1,t} = (\#lunes) - (\#domingos), \dots, T_{6,t} = (\#sábados) - (\#domingos)$
Año bisiesto	$LY_t = \begin{cases} 0.75 & \text{para febrero bisiesto} \\ -0.25 & \text{para febrero no bisiesto} \\ 0.0 & \text{para cualquier otro mes} \end{cases}$
Semana Santa	$SS_{w,t} = \frac{1}{w} \quad (\# \text{ días antes del domingo de Pascua que caen en el mes } t)$
Outlier aditivo	$AO_t = \begin{cases} 1 & \text{para } t = t_0 \\ 1 & \text{para } t \neq t_0 \end{cases}$
Cambio de nivel	$LS_t = \begin{cases} -1 & \text{para } t < t_0 \\ 0 & \text{para } t \geq t_0 \end{cases}$
Cambio transitorio	$TC_t = \begin{cases} 0 & \text{para } t < t_0 \\ \alpha^{t-t_0} & \text{para } t \geq t_0 \end{cases}$ Donde a es la tasa en la que decae el efecto a su nivel previo.
Efecto rampa	$RP_t = \begin{cases} t_0 - t_1 & \text{para } t \leq t_0 \\ t - t_1 & \text{para } t_0 < t < t_1 \\ 0 & \text{para } t \geq t_1 \end{cases}$

Fuente: Census Bureau, 2020.

Una vez determinados los efectos de calendario, se sigue la metodología de Box y Jenkins (1970) para la construcción del modelo ARIMA, analizando las funciones de autocorrelación y de autocorrelación parcial, de tal forma que el ajuste de este incluya el número de diferencias para que la serie de tiempo sea estacionaria, posibles diferencias estacionales, el orden de parámetros autorregresivos y de medias móviles, tanto para el nivel de la serie como para la componente estacional no estable.

En esta fase se suele evaluar, además, que los coeficientes del modelo ARIMA sean significativos y, sobre todo, que el residuo obtenido no presente autocorrelación serial ni estacionalidad residual, lo cual implica estimaciones parsimoniosas y óptimas desde el punto de vista estadístico. Lo obtenido en la etapa de ajustes previos, aquí es

crucial para realizar pronósticos *hacia atrás* y *hacia adelante*, que permiten, en la etapa siguiente, aplicar los filtros estacionales.

Filtros estacionales

Para estimar las componentes de las series de tiempo (T_t , E_t e I_t), se aplican filtros estacionales de medias móviles a la serie \hat{Y}_t . Los filtros de tendencia son promedios ponderados de meses o trimestres consecutivos utilizados para estimar T_t , por otra parte, los estacionales consisten en medias móviles estacionales de valores consecutivos dentro de un mes o trimestre determinado. Una media móvil $n \times m$ es un promedio simple de m términos tomada en n tramos secuenciales consecutivos. De esta forma, se usan esos filtros para estimar la componente asociada a E_t .

El uso de los filtros es iterativo, suponiendo que la serie de tiempo sigue una estacionalidad aditiva, en la primera etapa se obtiene una estimación inicial de la tendencia usando filtros de tendencia para Y_t , obteniendo, así, $T_t^{(1)}$. En consecuencia, se estima la componente inicial de estacionalidad-irregularidad como $EI_t^{(1)} = Y_t - T_t^{(1)}$. Después, usando filtros estacionales para $EI_t^{(1)}$, se obtiene una primera estimación del factor estacional ($E_t^{(1)}$), de tal manera que el ajuste estacional inicial es $A_t^{(1)} = Y_t - E_t^{(1)}$. En el siguiente paso, el proceso se reinicia usando $A_t^{(1)}$ y el filtro de Henderson para obtener una nueva estimación de la tendencia, de tal forma que, iterando el proceso un paso más para estimar la tendencia, la descomposición final de la serie de tiempo es:

$$Y_t = \hat{T}_t + \hat{E}_t + \hat{I}_t, \tag{2}$$

donde $\hat{T}_t = T_t^{(3)}$, $\hat{E}_t = E_t^{(2)}$, $\hat{I}_t = A_t^{(2)} - \hat{T}_t$. El criterio para la selección de la media móvil estacional se basa en la \hat{I}/\hat{E} global, que mide el tamaño relativo de los movimientos irregulares y los estacionales promediados durante todos los meses o trimestres.

El criterio empírico para determinar si el filtro estacional es adecuado se muestra en el cuadro 2.

Resumen de desestacionalización de series de tiempo económicas

Una vez descritas las generalidades del procedimiento de desestacionalización usando el X-13ARIMA-SEATS, los siguientes son los pasos para hacerlo:

1. Revisar patrones gráficos tanto para la serie original como en logaritmos: tendencias, posible tipo de estacionalidad, presencia cambios estructurales, etcétera.
2. Estimar el modelo automático para diagnosticar si la serie es candidata para desestacionalizarse, las diferencias sugeridas (integración) y el tipo de estacionalidad indicada (aditiva o multiplicativa). De manera alternativa, se puede usar una prueba de raíces unitarias para diagnosticar, en su caso, el tipo de transformación estacionaria y el criterio de Guerrero (1993) para identificar el tipo de estacionalidad.
3. Aplicar algún criterio estadístico complementario para definir si la serie presenta estacionalidad, por ejemplo, el estadístico M7 (Lothian y Morry, 1978) o la prueba de estabilidad residual, D8-F.
4. Analizar si las variables de frecuencia de días-calendario (SS y LY) son significativas en el regARIMA. De forma alternativa, se puede explorar el uso de otras variables exógenas relevantes, por ejemplo, los eventos móviles específicos del país, como se hace en el presente trabajo.
5. Analizar la existencia de outliers.
6. Identificar el modelo ARIMA a través de la metodología *Box-Jenkins*. Estimar el modelo evaluando los siguientes criterios:
 - a. Significancia de parámetros: omitir la presencia de términos no significativos a través de pruebas de significancia conjunta.
 - b. No autocorrelación ni heterocedasticidad condicional en los residuos: pruebas *Ljung-Box* y efecto *ARCH*.

Cuadro 2

Criterio para la selección de los filtros estacionales

Criterio	Filtro estacional
$0 < \hat{I}/\hat{E} < 2.5$	3 x 3
$3.5 < \hat{I}/\hat{E} < 6.5$	3 x 5
$6.5 < \hat{I}/\hat{E}$	3 x 9

Fuente: Lothian, 1984.

- c. Normalidad de los residuales: pruebas *Shapiro-Wilk* y *Jarque-Bera*, entre otras. Este punto puede omitirse argumentando que la distribución asintótica de los estimadores por máxima verosimilitud es consistente en el límite.
 - d. Estacionariedad en el modelo ARIMA.
 - e. Baja correlación de los parámetros en el modelo ARIMA.
 - f. Estudiar la capacidad predictiva del modelo fuera de muestra, esto dado que en la etapa de aplicación de filtros estacionales se requiere extender la serie *hacia atrás y hacia adelante*.
7. Aplicar las pruebas explicadas en el punto 2 a los residuales del modelo ARIMA.
 8. Seleccionar los filtros estacionales de acuerdo con el cuadro 2 y aplicar estos para obtener las componentes de la serie de tiempo.
 9. Revisar la consistencia de los resultados en términos estructurales: variaciones mensuales, trimestrales y anuales concordantes a través del tiempo.
 10. Decidir si para los agregados debe aplicar el método directo o el indirecto de desestacionalización.

Note que los efectos de eventos-calendario movibles considerados en el punto 4 se mantienen fijos dentro de un trimestre, por lo que no afectan a las series trimestrales. Solo el efecto de Semana Santa puede ocurrir en el primer o segundo trimestres. En relación con el punto 10, es importante señalar que la desestacionalización que realiza el INEGI es de manera directa, es decir, el agregado es desestacionalizado sin importar sus desagregados. Lo anterior tiene su justificación en el hecho de que, al hacerlo de esta forma, se puede garantizar la optimalidad estadística en todo el proceso de desestacionalización (Eurostat, 2015), situación que, de manera indirecta, agregando las componentes a través de sumas, no se puede verificar de forma estadística. En ese sentido, la Eurostat también señala que, cuando se utiliza el método indirecto, es más recomendable incorporar los efectos de calendarios nacionales.

Desestacionalización del IGAE y sus componentes

Para desestacionalizar las series de tiempo mensuales del IGAE y sus tres GA (GA1: actividades primarias; GA2: secundarias y GA3: terciarias), para el periodo de enero de 1993 a julio de 2019, usamos los modelos utilizados de manera oficial por el INEGI.²

En este trabajo se considera que las variables incluidas en los *modelos INEGI* para la etapa de ajustes previos de días hábiles, año bisiesto, Semana Santa, *outliers* o los órdenes ARIMA han sido determinados con rigurosidad estadística, lo que valida su implementación y sus resultados, y sobre estos se adicionan las variables de interés en esta investigación (DT, IM y la interacción BJ-SS). Lo anterior brinda comparabilidad respecto a las series que fueron publicadas en su oportunidad.

La manera en que se incorporan los efectos de calendario del DT y la IM es descontándolos de la variable TD, como se definió en el cuadro 1. Por ejemplo, el DT en 1993 ocurrió en sábado, por lo que se descontaría ese día para contabilizar solo cuatro en lugar de cinco sábados de actividad económica para ese mayo; de la misma manera, el IM de ese año fue jueves, con lo que tendríamos tres y no cuatro jueves con actividad económica para septiembre. Con respecto a la interacción BJ-SS se incluye como variable exógena de la siguiente manera:

$$BJ - SS_t = \begin{cases} 1 & \text{cuando 21 de marzo coincide con SS} \\ 0 & \text{cuando no hay coincidencia} \end{cases}$$

El cuadro 3 muestra la ocurrencia de estas fechas para el periodo de análisis.

Para el caso del IGAE, los resultados de los modelos (el publicado de manera oficial y el obtenido al incluir los eventos movibles del calendario mexicano) se resumen en el cuadro 4.

² https://www.inegi.org.mx/est/bie/doc/Modelos_IGAE.pdf

Cuadro 3

Matriz de efectos de calendario para el modelo regARIMA

Año	DT	IM	BJ-SS	Año	DT	IM	BJ-SS
1993	Sábado	Jueves	0	2007	Martes	Domingo	0
1994	Domingo	Viernes	0	2008	Jueves	Lunes	1
1995	Lunes	Sábado	0	2009	Viernes	Miércoles	0
1996	Miércoles	Lunes	0	2010	Sábado	Jueves	0
1997	Jueves	Martes	0	2011	Domingo	Viernes	0
1998	Viernes	Miércoles	0	2012	Martes	Domingo	0
1999	Sábado	Jueves	0	2013	Miércoles	Lunes	0
2000	Lunes	Sábado	0	2014	Jueves	Martes	0
2001	Martes	Domingo	0	2015	Viernes	Miércoles	0
2002	Miércoles	Lunes	0	2016	Domingo	Viernes	1
2003	Jueves	Martes	0	2017	Lunes	Sábado	0
2004	Sábado	Jueves	0	2018	Martes	Domingo	0
2005	Domingo	Viernes	1	2019	Miércoles	Lunes	0
2006	Lunes	Sábado	0				

DT: Día del Trabajo; se observa siempre el primer día de mayo.

IM: Independencia de México; se observa siempre el 16 de septiembre.

BJ-SS: interacción del natalicio de Benito Juárez con Semana Santa.

Fuente: elaboración propia.

Cuadro 4

Continúa

Modelos ARIMA para desestacionalizar el IGAE

Variable	Modelo INEGI		Modelo INEGI con eventos móviles	
	Coefficiente	P valor	Coefficiente	P valor
Lun.	0.00069	0.3563	0.00045	0.5233
Mar.	0.00367	0.0000	0.00425	0.0000
Miér.	0.00197	0.0093	0.00164	0.0219
Jue.	0.00143	0.0606	0.00193	0.0068
Vie.	0.00258	0.0007	0.00222	0.0021
Sáb.	-0.00272	0.0004	-0.00274	0.0002
LY	0.01292	0.0000	0.01200	0.0000
SS[4]	-0.02740	0.0000	-0.02557	0.0000
LS1995.Feb.	-0.07269	0.0000	-0.07201	0.0000
A01995.Oct.	-0.03412	0.0000	-0.03232	0.0000
LS2009.Ene.	-0.03699	0.0000	-0.03813	0.0000
MA(1)	0.22081	0.0000	0.20201	0.0001
MA(4)	-0.19265	0.0002	-0.20120	0.0001
MA(7)	0.25924	0.0000	0.25512	0.0000
MAS(1)	0.48110	0.0000	0.47179	0.0000

Modelos ARIMA para desestacionalizar el IGAE

Variable	Modelo INEGI		Modelo INEGI con eventos movibles	
	Coefficiente	P valor	Coefficiente	P valor
BJ-SS			-0.00841	0.0282
BIC	808.6		803.3	
QS	0.9314		1.478	
Box-Ljung	13.9		18.8	
Shapiro-Wilk	0.9952		0.9949	

Fuente: elaboración propia con base en el programa X-13ARIMA-SEATS; el periodo de la serie es de enero de 1993 a julio del 2019.

Se puede apreciar que, a 5 % de significancia, el coeficiente BJ-SS es significativo, además de que los jueves se vuelven significativos respecto al modelo utilizado de forma oficial por el INEGI. Los demás coeficientes mantienen su interpretabilidad estadística. En los análisis de los modelos, en ambos casos, no existe ni estacionalidad residual ni autocorrelación serial según las pruebas QS y *Ljung-Box*, respectivamente. También, los residuales se comportan de manera normal de acuerdo con la prueba *Shapiro-Wilk*. Es de interés mencionar que el BIC o criterio bayesiano es menor para el modelo que incluye las variables del calendario mexicano, lo cual indicaría que es más informativo respecto al que no las incluye. Aunque los modelos no son totalmente anidados, dado que la frecuencia de los días-calendario no es la misma

entre modelos, a manera descriptiva (realizando una razón de verosimilitudes) concluimos que el *p* valor es 0.00, motivo por el cual rechazamos la hipótesis nula de que ambos modelos son igual de informativos; es decir, tenemos argumentos para concluir que el modelo que incluye los eventos movibles del calendario mexicano es más informativo.

Es importante analizar las variaciones mensuales discordantes entre ambos modelos, es decir, cuando las variaciones de las series desestacionalizadas oficialmente difieren del modelo alternativo. El cuadro 5 presenta los resultados: del lado izquierdo, aquellas con las que el modelo INEGI con eventos movibles arrojó discordancias cero o negativas y del lado derecho, las positivas.

Cuadro 5

Variaciones discordantes entre series desestacionalizadas: IGAE

Fechas	Modelo INEGI	Modelo INEGI con eventos movibles	Fechas	Modelo INEGI	Modelo INEGI con eventos movibles
1996/08	0.1	0.0	1994/02	0.0	0.1
1998/05	0.1	-0.2	2000/09	0.0	0.2
2000/10	0.1	-0.2	2008/07	-0.1	0.1
2004/06	0.0	-0.2	2010/04	-0.1	0.1
2005/03	0.2	-0.2	2015/06	-0.1	0.2
2007/08	0.1	0.0			
2013/11	0.1	0.0			
2019/05	0.1	0.0			

Fuente: elaboración propia con base en el programa X-13ARIMA-SEATS; el periodo de la serie es de enero de 1993 a julio del 2019.

Podemos apreciar que las series desestacionalizadas otorgan un total de 13 variaciones discordantes, de las cuales 10 son positivas y tres, negativas para el modelo INEGI, mientras que, cuando se incluyen eventos movibles del calendario mexicano resultan nueve positivas y cuatro negativas. Llama la atención analizar que la mayoría de los meses son circundantes entre marzo, mayo y septiembre, es decir, cuando se presentan los efectos BJ-SS, el DT y la IM. Estos resultados permiten identificar que, en general, la presencia de estas festividades tiende a disminuir las tasas de crecimiento, ya que, al haber menor número de días hábiles, se frena de forma ligera la actividad económica. Es conveniente comentar que estas discrepancias no pueden explicarse de manera directa a partir de los coeficientes del modelo debido a que, como hemos

visto, la etapa de los filtros estacionales rompe la interpretabilidad y su principal función es extender la serie en sus extremos, situación que la metodología *TRAMO-SEATS* trata de evitar.

Analizando las componentes del IGAE, es importante resaltar que el modelo ARIMA de la GA1 es un ARIMA ([4] 0 [5 10 11])(0 1 1) sin variables de TD, LY y SS, motivo por el cual no es necesario modelar los eventos movibles del calendario mexicano. Para la GA2 se utiliza un modelo ARIMA (0 1 [1 6])(0 1 1). Los resultados se resumen en el cuadro 6.

Para este modelo no resulta significativa la variable BJ-SS, además de que no se gana significancia en los días-calendario. Asimismo, el BIC es menor en el modelo que no incluye los eventos

Cuadro 6

Modelos ARIMA para desestacionalizar la GA2

Variable	Modelo INEGI		Modelo INEGI con eventos movibles	
	Coefficiente	P valor	Coefficiente	P valor
Lun.	0.00069	0.5328	0.00072	0.5195
Mar.	0.00079	0.4777	0.00122	0.2741
Miér.	0.00344	0.0019	0.00315	0.0049
Jue.	0.00223	0.0449	0.00228	0.0402
Vie.	0.00091	0.4125	0.00121	0.2787
Sáb.	-0.00094	0.4044	-0.00123	0.2868
LY	0.01835	0.0000	0.01797	0.0000
SS[4]	-0.03415	0.0000	-0.03315	0.0000
LS1995.Feb.	-0.10134	0.0000	-0.10088	0.0000
AO1995.Oct.	-0.07091	0.0000	-0.07004	0.0000
LS1995.Dic.	0.05822	0.0000	0.05868	0.0000
LS2009.Ene.	-0.04740	0.0000	-0.04732	0.0000
MA(1)	0.20050	0.0003	0.19870	0.0003
MA(6)	-0.15539	0.0047	-0.16299	0.0030
MAS(12)	0.55136	0.0000	0.55440	0.0000
BJ-SS			-0.00376	0.5294
BIC	1 055.0		1 059.0	
QS	0.0		0.0	
Ljung-Box	17.95		19.02	
Shapiro-Wilk	0.9974		0.9968	

Fuente: elaboración propia con base en el programa *X-13ARIMA-SEATS*; el periodo de la serie es de enero de 1993 a julio del 2019.

del calendario mexicano. Nótese que este parece tener ciertos problemas de estacionalidad residual; no obstante, de acuerdo con el contraste de *Ljung-Box*, los residuales están libres de autocorrelación serial, además de que estos se comportan de manera normal.

Para la GA3 se utiliza un modelo ARIMA (0 1 [1 4 7])(0 1 1), dado que es el sector con mayor participación en el total de la economía;³ son, *a priori*, esperables resultados similares a los del IGAE, los cuales se presentan en el cuadro 7.

De igual manera que para el IGAE, para la GA3, en el modelo que incluye los eventos móviles del

calendario mexicano, hay dos términos más significativos respecto al que no las considera, los cuales son los jueves y la interacción BJ-SS y, asimismo, el BIC es menor en dicho modelo.

Por lo anterior, aunado a que ambos modelos no presentan problemas de estacionalidad residual, autocorrelación serial y no normalidad en los errores, podemos concluir que el modelo que incluye las variables del calendario mexicano es más informativo que el que no las contempla, motivo por el cual las series de tiempo tenderían a estar desestacionalizadas de mejor manera.

El cuadro 8 representa las discrepancias en los signos de las variaciones porcentuales mensuales de cada modelo.

3 Para el 2019, la GA1 representaba 3 %; la GA2, 30 %; y la GA3, 67 % del total del PIB.

Cuadro 7

Modelos ARIMA para desestacionalizar la GA3

Variable	Modelo INEGI		Modelo INEGI con eventos móviles	
	Coefficiente	P valor	Coefficiente	P valor
Lun.	0.00093	0.2398	0.00046	0.5329
Mar.	0.00466	0.0000	0.00543	0.0000
Miér.	0.00210	0.0090	0.00155	0.0376
Jue.	0.00138	0.0883	0.00240	0.0012
Vie.	0.00264	0.0010	0.00191	0.0105
Sáb.	-0.00414	0.0000	-0.00431	0.0000
LY	0.01122	0.0000	0.00966	0.0001
SS[4]	-0.02463	0.0000	-0.02205	0.0000
LS1995.Feb.	-0.05156	0.0000	-0.04919	0.0000
LS2009.Ene.	-0.03807	0.0000	-0.03820	0.0000
MA(1)	0.24204	0.0000	0.19561	0.0003
MA(4)	-0.11951	0.0277	-0.14259	0.0088
MA(7)	0.17852	0.0009	0.18187	0.0007
MAS(1)	0.48994	0.0000	0.47425	0.0000
BJ-SS			-0.01318	0.0012
BIC	803.1		790.3	
QS	0.0		0.01522	
<i>Ljung-Box</i>	25.62		25.67	
<i>Shapiro-Wilk</i>	0.9773		0.9782	

Fuente: elaboración propia con base en el programa X-13ARIMA-SEATS; el periodo de la serie es de enero de 1993 a julio del 2019.

Cuadro 8

Variaciones discordantes entre series desestacionalizadas: GA3

Fechas	Modelo INEGI	Modelo INEGI con eventos movibles	Fechas	Modelo INEGI	Modelo INEGI con eventos movibles
1993/06	0.0	-0.2	1993/10	0.0	0.1
2002/04	0.4	-0.1	2001/06	0.0	0.1
2003/04	0.2	0.0	2006/08	-0.1	0.1
2008/02	0.1	-0.2	2006/09	-0.1	0.1
2008/03	0.2	-0.7	2007/02	0.0	0.1
2009/05	0.1	-0.4	2009/12	-0.1	0.1
2014/03	0.0	-0.2	2012/12	-0.1	0.0
2014/09	0.0	-0.4	2013/10	-0.1	0.0
2015/05	0.0	-0.4			
2016/03	0.4	-0.2			

Fuente: elaboración propia con base en el programa X-13ARIMA-SEATS; el periodo de la serie es de enero de 1993 a julio del 2019.

Al igual que el IGAE, la mayoría de los meses discordantes son cercanos a aquellos donde se presentan los tres eventos movibles que se han modelado en este trabajo, donde llama la atención cómo (ligeramente) el crecimiento económico está siendo sobrevalorado que si incluyéramos las variables del calendario mexicano.

Conclusiones y recomendaciones finales

La desestacionalización de series de tiempo proporciona, al público en general, información libre de patrones contaminados por la periodicidad de estas, lo cual es importante en materia de política económica, pues permite a la sociedad conocer el impacto de las políticas sobre el tema de mejor manera. En consecuencia, es necesario que esta acción se realice con las mejores prácticas disponibles. En este sentido, el INEGI, institución encargada de desestacionalizar de manera oficial las series de tiempo, utiliza el procedimiento del programa X-13ARIMA-SEATS, verificando cada uno de los supuestos que garantizan la correcta aplicación del método.

Sin embargo, es claro que dentro del proceso de desestacionalización se pueden incluir varia-

bles exógenas en la parte de ajustes previos, en particular, eventos movibles del calendario mexicano, lo cual puede ayudar a representar, de mejor forma, series de tiempo desestacionalizadas. Este trabajo analizó el efecto de incluir, de manera conjunta, el DT, la IM y el evento BJ-SS en la desestacionalización del IGAE.

Los resultados permiten concluir que, para el IGAE y su GA3, dichos eventos son relevantes estadísticamente, ya que resultan significativos y mejora la capacidad informativa de los modelos. Las series de tiempo que incluyen estos efectos presentan, de forma ligera, tasas de crecimiento mensuales menores respecto a los modelos que no los muestran, sobre todo, en los meses circundantes a la presencia de dichos eventos movibles.

Se recomienda al INEGI analizar la viabilidad de utilizar este enfoque, si no en todas, sí en las variables más relevantes de la coyuntura económica, como pueden ser el PIB, el IGAE, el consumo privado y la inversión fija bruta, entre otras, con el fin de otorgar información desestacionalizada que presente eventos particulares de México de manera significativa y, por ende, series de tiempo desestacionalizadas con un mayor sustento estadístico.

Fuentes

- Banco de España. *Ajuste estacional y extracción de señales en la contabilidad nacional trimestral*. Documentos de trabajo 0210. Banco de España, Servicio de Estudios, 2002.
- Box, G. y G. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day, 1970.
- Cochrane, John H. *Time series for macroeconomics and finance*. Manuscript. University of Chicago, 2005.
- Dagum, E. B. *The X-11-ARIMA Seasonal Adjustment Method*. Catalogue 12-564E. Ottawa, Canada, Statistics Canada, 1980.
- Elizondo, Rocío. "Estimaciones del PIB mensual en México basadas en el IGAE", en: *Estudios Económicos*. 34(2), 2019, pp. 197-241.
- Eurostat. *ESS guidelines on seasonal adjustment*. Luxembourg, Eurostat, 2015.
- Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto y B-Ch. Chen. "New capabilities and methods of the X-12-ARIMA Seasonal Adjustment Program", in: *Journal of Business and Economic Statistics*. 16(2), 1998, pp. 127-152.
- Findley, D., K. Wills and B. Monsell. *Issues in estimating Easter regressors using RegARIMA models with X-12-ARIMA*. Proceedings of the American Statistical Association, 2005.
- Gálvez-Soriano, Oscar de J. "Nowcasting Mexico's quarterly GDP using factor models and bridge equations", en: *Estudios Económicos*. 35(2), 2020, pp. 213-265.
- Guerrero, V. M. "Desestacionalización de series de tiempo: una introducción a la metodología", en: *Comercio Exterior*. 40(11), 1990, pp. 1035-1046.
- _____. "Desestacionalización de series de tiempo económicas: ajustes previos", en: *Comercio Exterior*. 42(11), 1992, pp. 1042-1053.
- _____. "Time series analysis supported by power transformations", en: *Journal of Forecasting*. 12(1), 1993, pp. 37-48.
- Guerrero, V. M., J. López-Pérez y F. Corona. "Ajuste estacional de series de tiempo económicas en México", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. 9(3), 2018, pp. 74-97.
- Heath, Jonathan. *Lo que indican los indicadores. Cómo utilizar la información estadística para entender la realidad económica de México*. México, INEGI, 2012.
- Ladiray, D. y B. Quenneville. "Desestacionalizar con el método X-11", en: *METHODOLOGICA*. Número especial, núm. 8-9. Bruxelles, Belgique, Universite Libre de Bruxelles, 2000-2001.
- Lothian, J. y M. Morry. *A Set of Quality Control Statistics for the X11-ARIMA88 Seasonal Adjustment Method*. Working Paper, 78-10-005, 1978.
- Lothian, Jack. "The identification and treatment of moving seasonality in X-11ARIMA", in: *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*. 1984, pp. 166-171.
- McElroy, Tucker. S., Brian C. Monsell y Rebecca. J. Hutchinson. *Modeling of holiday effects and seasonality in daily time series*. Reporte técnico. Center for Statistical Research & Methodology Research and Methodology Directorate U.S. Census Bureau, 2018.
- Roberts, Christopher G., Scott H. Holan y Brian C. Monsell. "Comparison of X-12-ARIMA Trading Day and Holiday Regressors with Country Specific Regressors", in: *Journal of Official Statistics*. 26(2), 2010, pp. 371-394.
- Sarhani, M. & A. El Afia. "An extension of X13-ARIMA-SEATS to forecast islamic holidays effect on logistic activities", in: 2014 International Conference on Logistics Operations Management. IEEE, 2014 June, pp. 79-84.
- Sax, C. y D. Eddelbuettel. "Seasonal Adjustment by X-13ARIMA-SEATS", in: *R. Journal of Statistical Software*. 87(11), 2018, pp. 1-17.
- US Census Bureau. *X-13ARIMA-SEATS. Reference Manual*. 2020.

Selección de transformación potencia en los indicadores trimestrales de actividad económica estatal: implicaciones en el análisis de factores

Power transformation selection in quarterly indicators of state economic activity: implications for factor analysis

Francisco de Jesús Corona Villavicencio y Jesús López-Pérez*

En este trabajo se presenta un análisis empírico de las implicaciones que tiene la selección de la transformación potencia en el proceso de ajuste estacional para el Indicador Trimestral de la Actividad Económica Estatal (ITAE). En primera instancia se comparan, en sentido estadístico, las series obtenidas a través de los modelos que arroja de manera automática el X-13ARIMA-SEATS, los de desestacionalización del Instituto Nacional de Estadística y Geografía y una alternativa de modelación que selecciona el tipo de transformación potencia a través del procedimiento de Guerrero (1993); las diferen-

This paper presents an empirical analysis of the implications of the selection of the power transformation in the seasonal adjustment process for the Quarterly Indicator of State Economic Activity (ITAE). First, a statistical comparison is made between the series obtained through the models automatically produced by X-13ARIMA-SEATS, the seasonal adjustment models of the National Institute of Statistics and Geography and an alternative model that selects the type of power transformation through Guerrero's (1993) procedure; the different statistical tests indicate that there are in-

* Instituto Nacional de Estadística y Geografía, franciscoj.corona@inegi.org.mx y jesus.lopezp@inegi.org.mx, respectivamente.

Nota: nuestro agradecimiento a los valiosos comentarios de Juan Martínez, Liliana Martínez y Oziel Martínez de la Dirección de Estudios Económicos del INEGI, los cuales nos ayudaron a hacer de este un mejor trabajo; también, nuestra gratitud por sus aportaciones durante todo el desarrollo de esta investigación a Víctor M. Guerrero del Departamento de Estadística del ITAM; y, por último, a Gerardo Leyva Parra, director general adjunto de Investigación del INEGI.

tes pruebas estadísticas indican que hay indicios en favor del enfoque alternativo. Posteriormente, se realiza la estimación de factores comunes con los tres grupos de series desestacionalizadas, la cual muestra que existen ligeras diferencias entre los tres enfoques, la más importante, en relación con el número de factores estimado, donde existe uno más cuando se utilizan las series desestacionalizadas con el nuevo enfoque sugerido.

Palabras clave: errores idiosincráticos; factores comunes; series de tiempo aditivas; series de tiempo multiplicativas; X-13ARIMA-SEATS.

Recibido: 12 de diciembre de 2019.

Aceptado: 7 de octubre de 2020.

dications in favor of the alternative approach. Subsequently, the estimation of common factors is carried out with the three groups of seasonally adjusted series, which shows that there are slight differences between the three approaches, the most important one, in relation to the number of estimated factors, where there is one more when using the seasonally adjusted series with the new suggested approach.

Key words: idiosyncratic noises; common factors; additive time series; multiplicative time series; X-13ARIMA-SEATS.



Mexican market of traditional products/Maxhastello/iStock

Introducción

El objetivo estadístico del procedimiento de desestacionalización es proporcionar series de tiempo libres de patrones sistemáticos aludidos a su frecuencia temporal, es decir, sin movimientos repetitivos año con año de carácter semestral, trimestral, mensual, quincenal, etc., según la serie de tiempo. El objetivo, desde un punto de vista económico, es que estas permitan realizar un análisis más certero sobre el comportamiento conceptual de la variable, importante para la formulación de políticas públicas; por ejemplo, puede ser significativo estudiar y entender la dinámica del Producto Interno Bruto (PIB) para el diseño de políticas contracíclicas en momentos en los cuales la economía decrece o bien crece, pero de manera mínima. Para estos fines, es de interés enfocarse en la serie del PIB trimestral desestacionalizado, la cual permite centrarse en la componente de tendencia-ciclo más los efectos irregulares, libre de movimientos de naturaleza no económica que influyen sobre el comportamiento de dicha serie. De esta forma, las fluctuaciones del PIB reflejan de mejor manera los efectos de carácter económico, como las políticas macroeconómicas. Lo mismo puede ser para las series de desempleo, ventas y producción industrial, entre otras.

En México, el Instituto Nacional de Estadística y Geografía (INEGI) es el organismo encargado de desestacionalizar las series de tiempo de manera oficial, para lo cual utiliza como base el *X-13ARIMA-SEATS* del *Census Bureau* de los Estados Unidos de América. El detalle de dicha metodología aplicada para el caso de nuestro país puede ser consultado en Guerrero *et al.* (2018).

Este procedimiento se basa en varios pasos que podemos sintetizar en: 1) visualizar patrones gráficos de la serie, 2) realizar pruebas formales para determinar si la serie de tiempo tiene patrones estacionales, 3) establecer si esta es aditiva o multiplicativa, 4) determinar los ajustes previos (efectos calendario y *outliers*), 5) construir y estimar el modelo ARIMA con la metodología de *Box-Jenkins*, 6) seleccionar y aplicar los filtros estacionales, 7) ve-

rificar que los resultados cumplan los supuestos y, en su caso, hacer ajustes y 8) determinar si es necesario aplicar el método indirecto de desestacionalización para el caso de conjunto de series agregadas.

Enfocándonos en el punto 3, este se encuentra directamente relacionado con métodos de selección de transformación potencia al estilo Box y Cox (1964) en series de tiempo, donde no solo nos limitamos a la disyuntiva entre mantener lineal la serie de tiempo o aplicarle logaritmos, sino que puede extenderse también, por ejemplo, para transformaciones cuadráticas o cúbicas. No obstante, dentro del contexto del *X-13ARIMA-SEATS* sí requerimos enfocarnos solo en determinar si es aditiva o bien, si es necesario estabilizar la varianza cuya dependencia es proporcional a su media, es decir, realizar transformación logarítmica con lo cual se determina que es multiplicativa. La manera en cómo se realiza esto está relacionada con seleccionar aquel modelo que tiene un menor valor en el criterio de información de Akaike (AIC); intrínsecamente, buscamos minimizar la varianza residual o maximizar el logaritmo de la función de verosimilitud invariante a la magnitud de la serie en el modelo ARIMA, lo cual es válido si suponemos que los modelos subyacentes tienen la misma cantidad de parámetros. Este procedimiento suele dar resultados razonables en la práctica (Shulman y McKenzie, 1984).

Por otra parte, Guerrero (1993), bajo pocos supuestos en los primeros dos momentos de la serie de tiempo, propone un criterio para encontrar la constante que minimiza el coeficiente de variación en esta, el cual es visto como una extensión a la transformación Box y Cox (1964) y representa una alternativa natural para identificar de qué tipo es; en otras palabras, permite decidir si se realiza o no la transformación logarítmica bajo una búsqueda restringida de la constante que minimiza el coeficiente de variación. Aunque existen representaciones híbridas (entre aditivas y multiplicativas) que pueden verse en Dagum (1980), Findley *et al.* (1998) y Ladiray y Quenneville (2000-2001), no son materia de análisis de este trabajo, ya que no se utilizan con frecuencia en la práctica.

Nótese que el uso del AIC para decidir si una serie es aditiva y multiplicativa (en el contexto del *X-13ARIMA-SEATS*) puede resultar erróneo, dado que la maximización del logaritmo de la función de verosimilitud depende de la distribución de los errores del modelo paramétrico subyacente que, en términos prácticos, suele verificarse con pruebas de normalidad, dados los supuestos del modelo ARIMA. Lo anterior implica que si la serie es multiplicativa (es decir, procede de un proceso generado con distribución *Log-normal*) estaría siendo comparada como si fuese generada por un proceso con distribución normal (Guerrero *et al.*, 2018, 79). Al momento no se encuentra en la literatura una investigación que compare métodos de selección de transformación potencia en el proceso de desestacionalización.

En consecuencia, dada la importancia del uso de las series desestacionalizadas con el método *X-13ARIMA-SEATS* y la relevancia del Indicador Trimestral de Actividad Económica Estatal (ITAEE) para analizar la coyuntura económica como variables *proxy* al PIB trimestral por entidad federativa, en este trabajo pretendemos cumplir con los siguientes objetivos: 1) comparar estadísticamente los modelos que arroja de manera automática el *X-13ARIMA-SEATS*, los modelos elaborados por el INEGI y los que seleccionan la transformación potencia a través del procedimiento de Guerrero (1993) para este conjunto de series de tiempo y, una vez desestacionalizadas las series con los tres enfoques, 2) analizar las implicaciones con base en un análisis de factores, esto con el fin de contribuir a la explicación de las variaciones a corto plazo de cada una de las series de tiempo.

Lo que resta de este documento está organizado como sigue: la sección dos resume las metodologías abordadas en este trabajo para realizar las transformaciones potencia de las series de tiempo; la siguiente presenta con detalle la aplicación de los métodos descritos previamente y los resultados de la estimación de factores comunes; en el apartado final se muestran las conclusiones y recomendaciones.

Metodología de transformación potencia

Podemos definir a la serie de tiempo por desestacionalizar como Y_t , la cual, en el contexto del *X-13ARIMA-SEATS*, puede ser de frecuencia mensual o trimestral y presenta los siguientes componentes: la tendencia-ciclo (T_t) que incluye los movimientos de largo plazo y las fluctuaciones alrededor de estos; la estacionalidad (E_t) que contiene efectos calendario y patrones sistemáticos dados por la frecuencia temporal de la serie; y la irregularidad (I_t) que, para nuestros fines, se supone ruido blanco.

La manera de ligar Y_t con sus respectivas componentes viene dada, primero, por la siguiente representación aditiva:

$$Y_t = T_t + E_t + I_t. \quad (1)$$

Por otra parte, la descomposición multiplicativa puede ser expresada como:

$$Y_t = T_t \times E_t \times I_t. \quad (2)$$

Nótese que, en este caso, se puede representar a la ecuación (2) en su forma *log*-aditiva:

$$\log(Y_t) = \log(T_t) \times \log(E_t) \times \log(I_t). \quad (3)$$

Asimismo, como comentamos en la introducción, en la literatura se han considerado otras opciones híbridas, por ejemplo, del tipo señalado por Ladiray y Quenneville (2000-2001):

$$Y_t = T_t \times (1 + E_t) \times (1 + I_t). \quad (4)$$

o del considerado por Findley *et al.* (1998):

$$Y_t = T_t \times (E_t + I_t - 1). \quad (5)$$

Podemos apreciar que, para el caso del modelo aditivo, suponen que la estacionalidad es estable e independiente de la componente de tendencia, mientras que en las opciones (2), (3) y (4), que la estacionalidad depende del nivel de la tendencia.

Es esperable que el último patrón esté presente en una mayor cantidad de series de tiempo económicas.

En el entorno del *X-13ARIMA-SEATS* la expresión de la serie de tiempo puede ser generalizada como sigue:

$$Y_t = \mu_t + Z_t, \quad (6)$$

donde $\mu_t = E(Y_t) = \sum_{i=1}^K \beta_i x_{it}$, es decir, son la suma de los efectos asociados a las $K \geq 1$ variables explicativas, y $Z_t = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t = \theta(B)\Theta(B^s)a_t$ sigue un modelo ARIMA multiplicativo estacional, donde B es el operador de retraso tal que $BZ_t = Z_{t-1}$; $\nabla = 1 - B$ es el operador de diferencia regular; $\nabla_s = 1 - B^s$ es el operador de diferencia estacional, con periodo estacional s ; $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ es un polinomio autorregresivo regular de orden p ; $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$ es un polinomio autorregresivo estacional de orden P ; $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ es un polinomio de promedios móviles regular de orden q ; $\Theta(B) = 1 - \Theta_1 B - \dots - \Theta_Q B^Q$ es un polinomio de promedios móviles estacional de orden Q ; a_t son variables aleatorias independientes y distribuidas de forma idéntica como una normal con media 0 y varianza σ^2 ; y d y D son los órdenes de diferencia regular y estacional, respectivamente. Los parámetros β_i , del modelo de regresión, así como los ϕ , Φ , θ , Θ y σ^2 del modelo ARIMA son estimados de manera simultánea por el método de máxima verosimilitud.

De esta manera, si $L_{f(Z)} = L_{\bar{Z}} + \sum_{t=1}^T \log\left(\frac{df(Z_t)}{dZ_t}\right)$ es el logaritmo de la función de verosimilitud, donde el primer término es maximizado para Z_1, \dots, Z_T condicionado al pasado, es decir, a la parametrización del modelo ARIMA, y el segundo es la corrección por el tipo de transformación, es decir, si $f(Z_t) = \log Z_t$, entonces este es $-\sum_{t=1}^T \log(Z_t)$; el procedimiento del *X-13ARIMA-SEATS* para decidir si una serie de tiempo es aditiva o multiplicativa viene dado por $L_Z > L_{\log(Z)}$ en el caso del modelo aditivo o bien, $L_Z < L_{\log(Z)}$ en el del multiplicativo (ver Mood *et al.*, 1974). De esta manera, $L_{f(Z)}$ es invariante a la escala de la serie de tiempo.

De Guerrero *et al.* (2018) se concluye que, dados los supuestos del modelo, el cálculo de $L_{f(Z)}$

está sujeto a suponer que tanto a_t como $\log(a_t)$ están distribuidas como una normal, lo cual resulta contradictorio dados los supuestos implícitos del modelo. En beneficio y en contra de este criterio, Findley *et al.* (1998, 37) comentan lo siguiente: "En el caso de elegir una transformación potencia en la serie de tiempo sin restricciones, cuando se usa el mismo tipo de modelo ARIMA y el mismo número de parámetros estimados, es conveniente centrarnos en el logaritmo de la función de verosimilitud que sea más grande. En esta situación, uno puede examinar un intervalo de valores para identificar la maximización del logaritmo de la función de verosimilitud, ver Ansley *et al.* (1977). Aunque este procedimiento parece arrojar resultados razonables (Shulman y McKenzie, 1984), la teoría de la distribución asintótica sobre la cual descansa no se ha verificado y, si es válido, puede requerir argumentos sutiles para su prueba cuando $d > 0$ o $D > 0$...".

Por otra parte, la transformación Box y Cox (1964) es una técnica de transformación potencia para reducir anomalías en las series de tiempo, como no aditividad (no linealidad), no normalidad y heterocedasticidad. Definimos $y^\lambda = (y^\lambda - 1)/\lambda$ si $\lambda \neq 0$ y $y^\lambda = \log(y)$ si $\lambda = 0$, lo cual puede ser usado como una aproximación para estabilizar la varianza.

Guerrero (1993) deduce que para una variable aleatoria $X > 0$, donde $E(X) = \mu_x < \infty$ y $var(X)$ depende de la media, la transformación estabilizadora de la varianza satisface que para una λ :

$$[var(x)]^2 / \mu_x^{1-\lambda} = a, \quad (7)$$

para una constante $a > 0$, es decir, λ es tal que el coeficiente de variación es una constante positiva. En la práctica se utilizan los respectivos estimadores de la varianza y la media, en consecuencia, podemos estimar (para submuestras $h=1, \dots, H$) el coeficiente de variación muestral como $S_h / \bar{Z}_h^{1-\lambda}$, donde:

$$\bar{Z}_h = \frac{\sum_{r=1}^R Z_{hr}}{R} \quad \text{y} \quad S_h = \left[\frac{\sum_{r=1}^R (Z_{hr} - \bar{Z}_h)^2}{R-1} \right]^{1/2} \quad (8)$$

con Z_{hr} siendo la r -ésima observación de la sub-serie h cuya dimensión es R . Es claro observar que nos interesa encontrar el parámetro λ , donde el coeficiente de variación muestral es estadísticamente constante para una serie de submuestras H . Existen otras estrategias para seleccionar λ , las cuales son detalladas en Guerrero (1993), pero la esencia es similar al enfoque comentado con anterioridad.

Análisis empírico

Análisis de desestacionalización

Para desestacionalizar de forma individual las series de tiempo del ITAEE, se comparan tres modelos: 1) los automáticos que arroja el *X-13ARIMA-SEATS*, 2) los publicados por el INEGI de manera oficial y 3) una variante, que tiene como motor la transformación potencia de Guerrero (1993) que, posteriormente, busca el modelo que mejor satisfaga los supuestos que validan el correcto ajuste estacional dentro de un amplio conjunto de modelos ARIMA definidos de forma previa. Lo anterior se realiza para el periodo de 2003-T1 a 2019-T2, dado que es para esta longitud en la que se pueden utilizar los modelos del Instituto I para poder replicar los resultados oficiales. No se considera el periodo de 1980-T1 a 2002-T4, que es donde el INEGI re-topoló las series de tiempo a través de modelos econométricos y estadísticos y en el cual se utilizan otros modelos que no son publicados oficialmente en la actualidad (INEGI 2019, 4). No omitimos señalar que los modelos del Instituto no utilizan de forma directa los resultados generados de manera automática por el *X-13ARIMA-SEATS*, sino que se especifican y estiman de acuerdo con los pasos que fueron comentados en la introducción, situación que los modelos automáticos no contemplan en su totalidad debido a la búsqueda sistemática y heurística de los modelos ARIMA.

La transformación potencia de las series de tiempo a través de la metodología de Guerrero (1993) se realiza restringiendo la búsqueda de parámetros λ entre 0 y 1 y redondeando el caso decimal

para fijar entre la selección de modelos aditivos ($\lambda = 0$) y multiplicativos ($\lambda = 1$).

Para comparar estadísticamente los tres enfoques, se analizan las siguientes pruebas:

- Estadístico QS de estacionalidad: p valores de la prueba que indica bajo hipótesis nula que no hay evidencia de estacionalidad residual.
- Estadístico de Friedman de estacionalidad: p valores de la prueba que indica bajo hipótesis nula que no hay evidencia de estacionalidad residual.
- Estadístico Ljung-Box: p valores de la prueba que indica bajo hipótesis nula que no hay evidencia de autocorrelación serial en los errores del modelo ARIMA.
- Estadístico Shapiro-Wilk: p valores de la prueba que indica bajo hipótesis nula que los residuales se distribuyen normal.
- Criterio de AIC: estadístico que indica el tamaño de la varianza residual penalizado por el número de parámetros del modelo subyacente.
- Tipo de transformación potencia sugerida: ninguna (serie aditiva) o logarítmica (serie multiplicativa).

Los resultados para los estadísticos QS, Friedman, Ljung-Box, Shapiro-Wilk y AIC se resumen en el cuadro 1, mientras que en los anexos se muestran con detalle los de cada entidad federativa y enfoque.

Se puede apreciar que, en términos generales, los tres enfoques tienen buenas propiedades estadísticas, considerando como buenos modelos aquellos libres de estacionalidad (QS y Friedman) y autocorrelación residual (Ljung-Box) y donde los residuales se distribuyen normalmente (Shapiro-Wilk). No obstante, se puede apreciar que, en términos absolutos, el enfoque alternativo, que considera la transformación potencia con el enfoque de Guerrero (1993), otorga p valores más cercanos a 1 en las cuatro pruebas, lo que nos indica evidencia más fuerte de ausencia de estacionalidad y autocorrelación serial, además de que los

Cuadro 1

Promedios de p valores de la evaluación estadística para los tres enfoques de desestacionalización del ITAEE

Enfoque	QS	Friedman	Ljung-Box	Shapiro-Wilk	AIC
Automático	0.9377	0.6665	0.6248	0.3817	278.5165
INEGI ^a	0.9747	0.6392	0.5237	0.5548	279.6063
Alternativo	0.9948	0.8488	0.7177	0.8120	285.3254

^a No considera al estado de Campeche, ya que no se identifica un patrón estacional.

residuales siguen de mejor manera una campana gaussiana. También, es claro que el enfoque automático es el que otorga mejores AIC dado que es su objetivo. Por último, es interesante comentar, además, que el enfoque automático parece ser ligeramente mejor que el modelo INEGI, con excepción del estadístico QS. Una razón puede deberse a que los del Instituto se actualizan con cierta periodicidad, lo cual, como corolario, daría indicios de que requieren ponerse al día con mayor frecuencia.

Al evaluar caso por caso (ver anexos), los resultados indican que el modelo automático detecta que en 20 entidades federativas la transformación potencia óptima es la logarítmica, el enfoque alternativo determina 15 casos, mientras que el del INEGI afirma que, en las 31 ocasiones, el mejor modelo es el multiplicativo (transformación logarítmica).¹ Por otra parte, a un nivel de significancia de 5 %, el enfoque automático rechaza en una ocasión la hipótesis nula de autocorrelación y cuatro veces la prueba de normalidad, mientras que la modelación del Instituto lo hace en una y tres, respectivamente, aunque también en una se rechaza la hipótesis nula según el estadístico de Friedman. Por otra parte, en el enfoque alternativo, en ninguna ocasión se rechaza la hipótesis nula para cualquier prueba. Los resultados se mantienen para los subperiodos 2003-T1 a 2010-T2, 2003-T1 a 2013-T2 y 2003-T1 a 2016-T2, los cuales están disponibles bajo petición.²

En conclusión, de acuerdo con los criterios tradicionales para medir la calidad en el proceso

1 Para Campeche, el INEGI no ajusta un modelo estacional.

2 Agradecemos el comentario de un árbitro anónimo que motivó realizar este ejercicio.

de desestacionalización, el enfoque alternativo proporciona mejores resultados para las series del ITAEE.³

Modelo de factores

El análisis de factores tiene como objetivo resumir el comportamiento de N series de tiempo en un $r < N$ factores comunes, más componentes individuales, llamadas términos idiosincráticos. En este contexto, partimos de la siguiente representación:

$$X_t = \Psi F_t + \varepsilon_t, \quad (9)$$

$$F_t = AF_{t-1} + \eta_t, \quad (10)$$

$$\varepsilon_t = \Gamma \varepsilon_{t-1} + u_t, \quad (11)$$

donde las observaciones (X_t), el error idiosincrático (ε_t) y su error asociado (u_t) son de dimensión $N \times 1$. Por otra parte, los factores comunes (F_t) y su término de error (η_t) son de tamaño $r \times 1$, mientras que la contribución de los factores sobre las observaciones (Ψ) es una matriz $N \times r$. Para finalizar, A y Γ son las matrices de coeficientes de modelos de vectores autorregresivos de dimensión $r \times r$ y $N \times N$, respectivamente. Se supone que los factores y el error idiosincrático pueden ser $I(1)$.

Nótese que el número de factores r es desconocido, por lo tanto, recurrimos al método de Onatski

3 En un ejercicio adicional, se estimó esta parte para las series del Indicador Global de la Actividad Económica (IGAE) del INEGI y sus componentes, las cuales arrojaron evidencia en favor del método propuesto. Dichos resultados están disponibles bajo petición a los autores.

(2010) para estimarlo. Se utiliza este procedimiento debido a que Corona *et al.* (2017) prueban que, en muestras finitas, dicha metodología funciona aun cuando el término idiosincrático es I(1) para distintos procesos generadores de datos.

Por otra parte, la estimación de los factores se realiza siguiendo el procedimiento de Barigozzi *et al.* (2017), de tal forma que el estimador de Ψ viene dado por los r valores propios de la matriz de $\Delta X \Delta X'$, donde X es de dimensión $N \times T$. Por lo tanto, puede mostrarse que el estimador de F_t viene dado por la siguiente expresión:

$$\hat{F} = \frac{\hat{\Psi}X}{N}. \quad (12)$$

Usando el procedimiento de Onatski (2010) para los datos oficialmente desestacionalizados y los que otorgan los modelos automáticos, encontramos que $\hat{r} = 5$, mientras que para el enfoque alternativo se determina que $\hat{r} = 6$. La implicación econométrica más importante es que este último enfoque indica que se requiere de un factor más respecto a los otros con el fin de poder representar la variabilidad total de los ITAEE para las 32 entidades.

Es importante señalar que, econométricamente, si al menos un factor es I(1) y, además, podemos concluir que el error idiosincrático es I(0), implicaría que existe al menos un par de economías estatales cointegradas (Bai, 2004). Nótese que en este caso, cuando los factores son I(1) y el término idiosincrático es I(0), los factores comunes son ahora las tendencias comunes.

Para verificar la naturaleza estocástica de los factores y el error idiosincrático, se estima el ejercicio PANIC (Bai y Ng, 2004) para determinar cuántos factores son I(1) y si el error idiosincrático es I(0). Los resultados indican que, para los tres conjuntos de datos, los factores comunes son I(1), mientras que los errores idiosincráticos son I(0).

El conjunto de gráficas 1 muestra el comportamiento de los términos idiosincráticos. Econométricamente, estos representan la desviación respecto a la tendencia común; de esta forma, pueden in-

terpretarse como la componente a corto plazo de cada una de las economías.

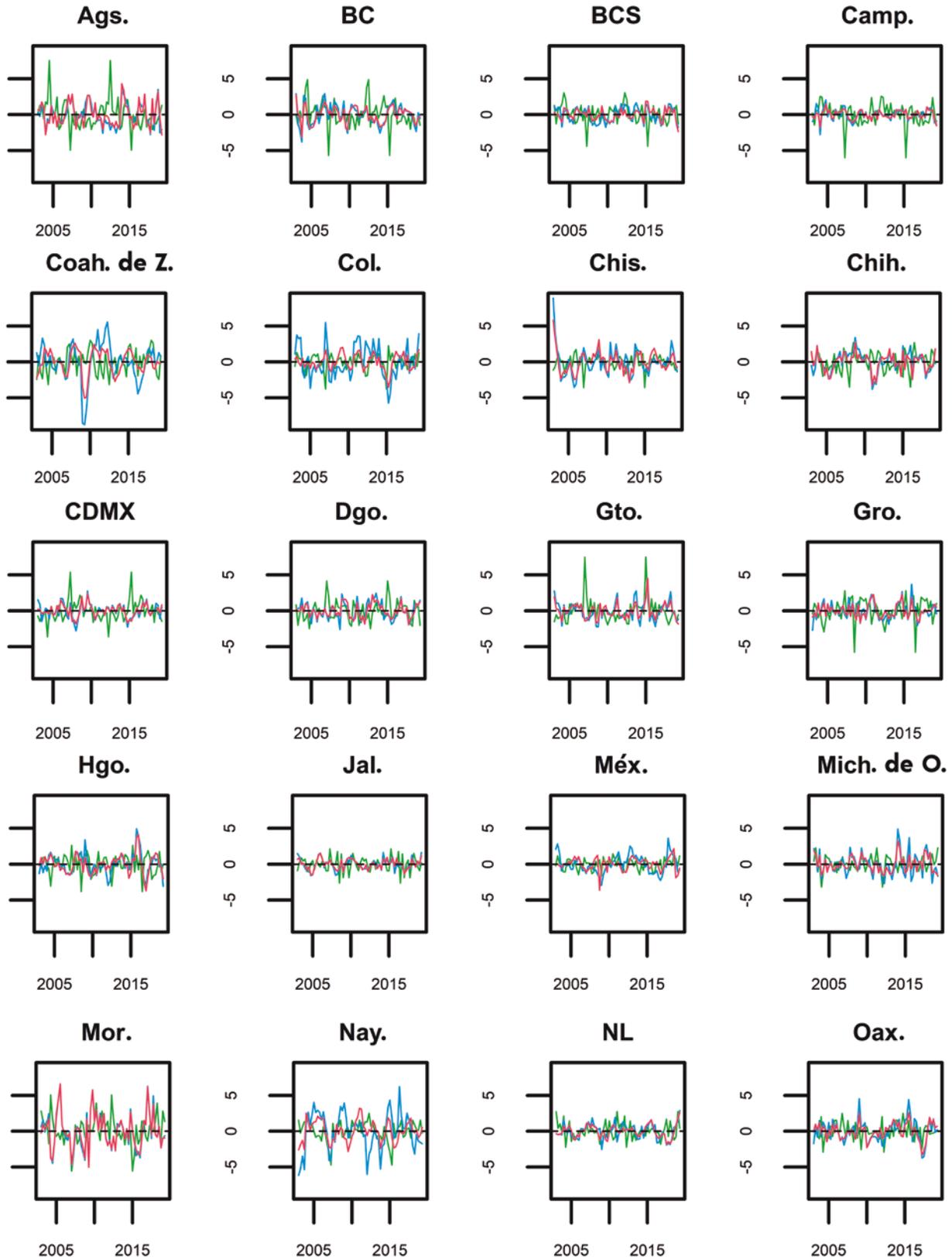
Se puede apreciar que los comportamientos a corto plazo de las economías cuando utilizamos modelos automáticos son muy volátiles, lo que da indicios de ciertos problemas en poder captar la variabilidad residual a través de los factores estimados. Por otra parte, para el conjunto de datos desestacionalizados con los modelos del INEGI y para la metodología alternativa sugerida en este trabajo, las series son similares, aunque destaca que, para Coahuila de Zaragoza, Colima y Nayarit, los movimientos de los comportamientos idiosincráticos presentan una mayor variabilidad, hecho atribuible, posiblemente, a la existencia de un factor común más para explicar la variabilidad en la economía. En consecuencia, el nivel mínimo de los términos idiosincráticos cuando usamos cifras desestacionalizadas con los modelos del INEGI es -8.8, mientras que el enfoque alternativo nos ofrece que es -5.5. Por otra parte, el nivel máximo se alcanza en 8.9 cuando se usan los modelos del Instituto y 7.8 para el otro procedimiento.

Como los factores comunes son sensibles a la estimación, es decir, a las condiciones de identificabilidad para poder ser estimados, nos centramos en el comportamiento del primer factor, el cual es invariante a la posible rotación utilizada, a la par de que, por construcción, es el que aporta mayor porcentaje de explicación respecto a la variabilidad total observada.

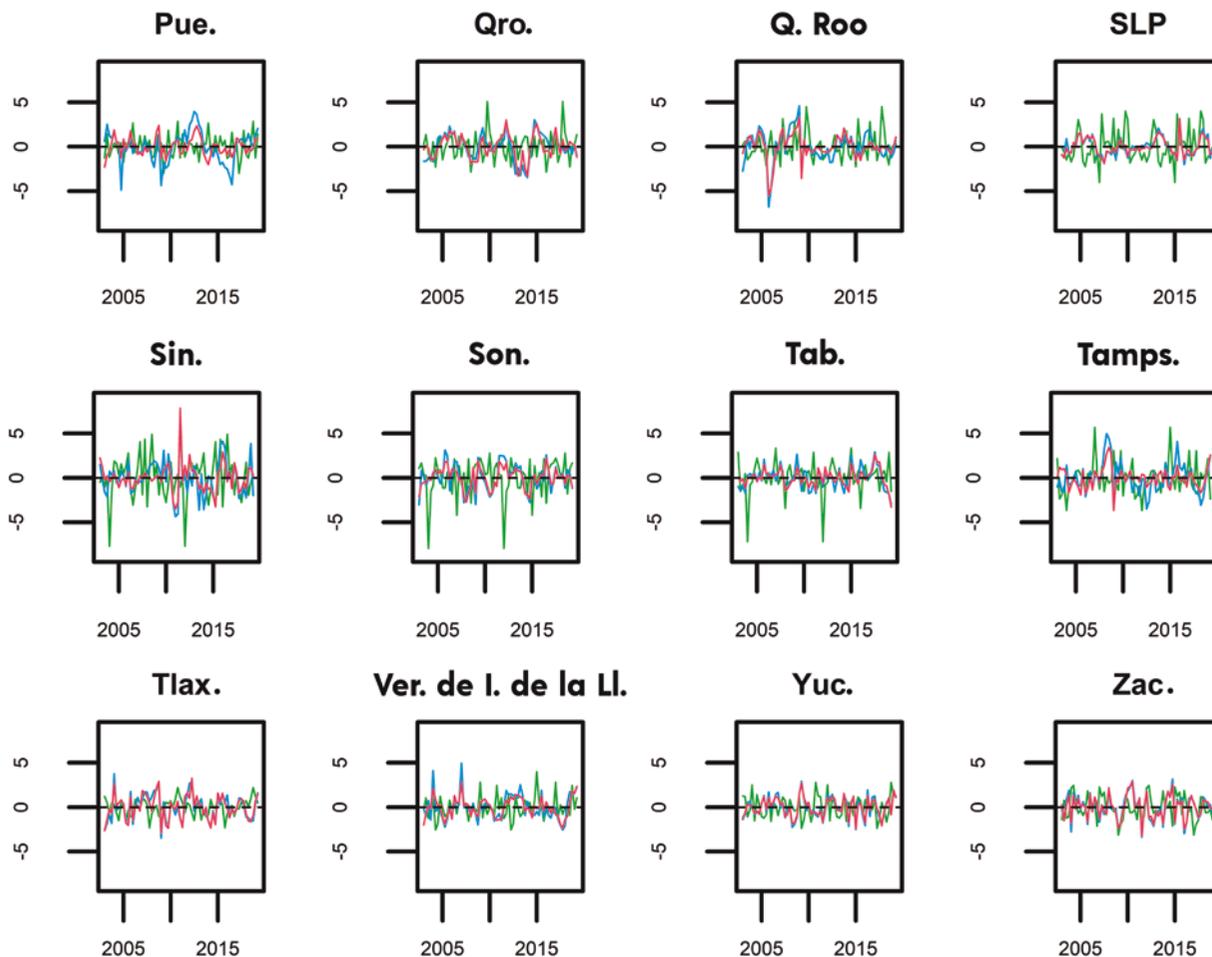
La gráfica 2 muestra el comportamiento de la variación trimestral del primer factor común estimado para los tres conjuntos de datos.

Se puede apreciar que los factores comunes presentan variaciones muy similares, donde llama la atención la caída asociada al último periodo, la cual es menor cuando se estiman aquellos con las series desestacionalizadas usando los modelos del INEGI. Utilizando estos y al centrar la atención en los últimos cuatro trimestres, las variaciones trimestrales indican crecimientos de 0.3, 0.1, 0.1 y -0.6 %, mientras que si empleamos el enfoque alternativo,

Errores idiosincráticos estimados para ambos conjuntos de datos desestacionalizados



Errores idiosincráticos estimados para ambos conjuntos de datos desestacionalizados



Nota: la línea verde se refiere a los modelos automáticos; la azul, a los del INEGI; y la roja, al método automático X-13ARIMA-SEATS seleccionando la transformación potencia con método de Guerrero (1993).

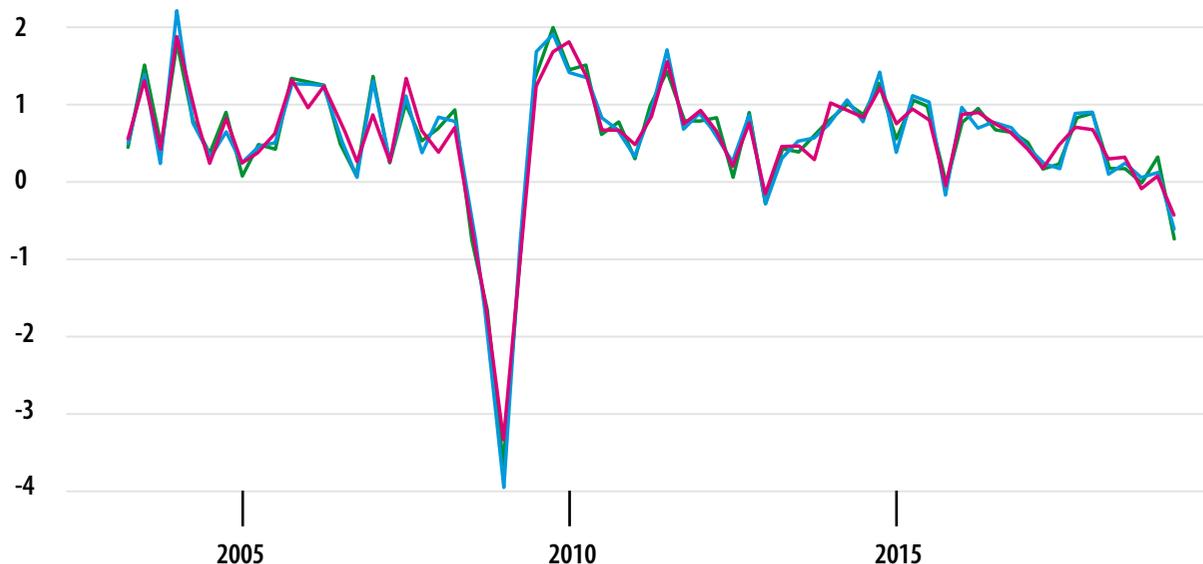
obtenemos variaciones de 0.3, -0.1, 0.1 y -0.4 %, lo cual, aunque de manera estadística puede resultar igual, no necesariamente lo es en términos económicos, sobre todo cuando las tasas de crecimiento son cercanas a 0 por ciento. Los modelos automáticos otorgan fluctuaciones más cercanas a las cifras oficiales, es decir, 0.2, 0.0, 0.3 y -0.7 por ciento.

En resumen, el análisis de factores utilizado en este trabajo nos muestra que, cuando usamos el enfoque alternativo, encontramos que existe un

factor más que explica la variabilidad de las series que si empleamos el automático o los modelos del INEGI. Esto, en sí, no es una razón en favor o en contra de utilizar alguno de los procedimientos, sino que indica que las estimaciones pueden ser sensibles al conjunto de datos con los cuales se realiza el análisis de factores. Lo que sí señala es que hay mayor variabilidad en el conjunto de datos desestacionalizados con el enfoque alternativo, por lo cual requerimos un factor más que, en términos econométricos, representa estimar términos idio-

Gráfica 2

Variación trimestral del primer factor común estimado para los tres conjuntos de datos desestacionalizados



Nota: la línea verde se refiere a los modelos automáticos; la azul, a los del INEGI; y la roja, al método automático *X-13ARIMA-SEATS* seleccionando la transformación potencia con método de Guerrero (1993).

sincráticos con menor volatilidad. Por otra parte, la primera tendencia común de la economía muestra ligeramente tasas de crecimiento menores en el último periodo de la serie cuando se utilizan los modelos del INEGI o los automáticos para desestacionalizar las series de tiempo, que cuando se utiliza el enfoque alternativo.

Conclusiones

La contribución de este trabajo es mostrar un análisis que mida los efectos de la transformación potencia en el proceso de desestacionalización bajo el contexto del *X-13ARIMA-SEATS*.

Se evaluó el funcionamiento de su método automático, los modelos que proporciona el INEGI y una alternativa que se basa en elegir la transformación potencia con el método de Guerrero (1993); lo anterior, para las series del ITAEE. Los

resultados obtenidos permiten concluir que el enfoque alternativo otorga mejores resultados, en el sentido de que hay evidencia más fuerte de ausencia de estacionalidad y autocorrelación serial en las series desestacionalizadas. Además, dada la importancia que tiene el proceso de desestacionalización para hacer análisis de coyuntura económica, el estudio de factores indicó que desestacionalizar de mejor forma las series de tiempo (es decir, mediante nuestra propuesta), nos otorga errores idiosincráticos menos volátiles, lo que genera ciertas discrepancias en las variaciones trimestrales de los factores comunes; en otras palabras, resultados econométricos más robustos.

Esta investigación permitió corroborar que el método de Guerrero (1993) es una herramienta valiosa para discernir entre realizar el ajuste estacional con una descomposición aditiva o bien multiplicativa. No obstante, en este caso solo se presentó para una

variable, por lo que limita poder extender estas conclusiones a otras series económicas relevantes, como el PIB, el IGAE y el consumo privado, entre otras. Por ello, se recomienda, como líneas futuras de investigación, realizar un experimento más robusto para series de tiempo estacionales simuladas, que permita evaluar el funcionamiento de la técnica propuesta abarcando una amplia gama de escenarios y no solo en una aplicación empírica.

Fuentes

Ansley, C. F., W. A. Spivey y W. J. Wroblewski. "A class of transformations for Box-Jenkins seasonal models", en: *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 26(2). 1977, pp. 173-178.

Bai, Jushan y Serena Ng. "A PANIC attack on unit roots and cointegration", en: *Econometrica*. 72, núm. 4. 2004, pp. 1127-1177.

Bai, Jushan. "Estimating cross-section common stochastic trends in nonstationary panel data", en: *Journal of Econometrics*. 122, núm. 1. 2004, pp. 137-183.

Barigozzi, Matteo, Marco Lippi y Mateo Luciani. *Dynamic factor models, cointegration, and error correction mechanisms*. Working paper. 2017. arXiv:1510.02399v3.

Box, George. E. y David R. Cox. "An analysis of transformations", en: *Journal of the Royal Statistical Society: Series B (Methodological)*. 6(2). 1964, pp. 211-243.

Corona, Francisco, Pilar Poncela y Esther Ruiz. "Estimating non-stationary common factors: implications for risk sharing", en: *Computational Economics*. 2017, pp. 1-24.

Dagum, Estela Bee. *The X-11 ARIMA Seasonal Adjustment Method*. Catalogue 12-564E. Ottawa, Canada, Statistics Canada, 1980.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto y Bor-Chung Chen. "New capabilities and methods of the X-12-ARIMA seasonal-adjustment program", en: *Journal of Business & Economic Statistics*. 16(2). 1998, pp. 127-152.

Guerrero, Victor M. "Time series analysis supported by power transformations", en: *Journal of Forecasting*. 12(1). 1993, pp. 37-48.

Guerrero, Víctor M., Jesús López-Pérez y Francisco Corona. "Ajuste estacional de series de tiempo económicas en México", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. 9(3). 2018, pp. 74-97.

INEGI. *Indicador Trimestral de la Actividad Económica Estatal. Producto Interno Bruto trimestral, anual y anual por entidad federativa. Retropolación hasta 1980. Síntesis metodológica*. Aguascalientes, INEGI, 2019.

Ladiray, Dominique y Benoit Quenneville. "Desestacionalizar con el método X-11", en: *Methodologica*. 8-9. Bruxelles, Belgique, Universite Libre de Bruxelles, 2000-2001.

Mood, Alexander M., Franklin A. Graybill y Duane C. Boes. *Introduction to the theory of statistics*. McGraw-Hill, 1974.

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

Shulman, Holly B. y Sandra K. McKenzie. *A Study of Pre-Adjustment Transformations*. Bureau of the Census. 1984.

Anexos

Anexo 1

Continúa

Evaluación estadística: método automático X-13ARIMA-SEATS

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Ags.	1.0000	0.8073	0.9073	0.3476	306.9818	log
BC	1.0000	0.7055	0.9779	0.2406	287.0846	none
BCS	1.0000	0.9454	0.9161	0.0128	362.2961	log
Camp.	0.2022	0.4798	0.4823	0.3293	299.3089	log

Evaluación estadística: método automático X-13ARIMA-SEATS

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Coah. de Z.	1.0000	0.9600	0.2357	0.0735	266.0060	<i>none</i>
Col.	0.7436	0.0832	0.4846	0.5431	322.3133	<i>none</i>
Chis.	1.0000	0.6504	0.3838	0.6164	293.2155	<i>log</i>
Chih.	0.9984	0.3373	0.8058	0.7163	253.9994	<i>log</i>
CDMX	1.0000	0.8779	0.8992	0.4475	242.5887	<i>none</i>
Dgo.	1.0000	0.2658	0.5131	0.2714	250.1194	<i>log</i>
Gto.	1.0000	0.9735	0.9574	0.6702	257.7001	<i>log</i>
Gro.	1.0000	0.8013	0.0327	0.4270	270.1018	<i>log</i>
Hgo.	1.0000	0.4798	0.8372	0.3616	277.0527	<i>log</i>
Jal.	1.0000	0.0887	0.8055	0.7292	235.9591	<i>none</i>
Méx.	1.0000	0.2191	0.5363	0.0421	236.1930	<i>log</i>
Mich. de O.	1.0000	0.8126	0.9524	0.0578	295.4476	<i>log</i>
Mor.	1.0000	0.6997	0.7459	0.4705	321.0442	<i>log</i>
Nay.	0.8641	0.9735	0.7392	0.2727	288.1156	<i>none</i>
NL	1.0000	0.6997	0.6477	0.4061	236.5760	<i>none</i>
Oax.	1.0000	0.6651	0.9066	0.1852	270.2739	<i>log</i>
Pue.	0.9879	0.5828	0.0925	0.7935	271.4867	<i>none</i>
Qro.	1.0000	0.6997	0.9313	0.2255	247.4691	<i>log</i>
Q. Roo	1.0000	0.7819	0.0654	0.4752	277.0635	<i>none</i>
SLP	1.0000	0.9735	0.6639	0.3141	239.0184	<i>log</i>
Sin.	0.9994	0.8791	0.9958	0.0229	310.4772	<i>log</i>
Son.	1.0000	0.8399	0.8530	0.6079	264.7505	<i>none</i>
Tab.	1.0000	0.5988	0.5612	0.4833	271.8826	<i>log</i>
Tamps.	0.6542	0.6651	0.1751	0.3750	271.9959	<i>log</i>
Tlax.	1.0000	0.6823	0.3584	0.9423	381.7335	<i>none</i>
Ver. de I. de la LI.	1.0000	0.4663	0.7374	0.0371	256.2821	<i>none</i>
Yuc.	0.5555	0.8791	0.1782	0.3452	249.6721	<i>log</i>
Zac.	1.0000	0.7530	0.6161	0.3719	298.3189	<i>log</i>

Evaluación estadística: modelos INEGI

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Ags.	1.0000	0.7626	0.6075	0.7909	301.6523	log
BC	1.0000	0.5011	0.8885	0.6319	298.6417	log
BCS	1.0000	0.0487	0.6780	0.7124	385.5282	log
Camp. ^{al}	NA	NA	NA	NA	NA	NA
Coah. de Z.	0.8292	0.8399	0.5007	0.7514	265.9317	log
Col.	0.7731	0.5479	0.2085	0.7034	290.6847	log
Chis.	1.0000	0.8590	0.1416	0.8821	293.0361	log
Chih.	1.0000	0.0979	0.8243	0.1382	255.1412	log
CDMX	1.0000	0.5641	0.5129	0.0437	273.6339	log
Dgo.	1.0000	0.8073	0.0502	0.7430	245.6831	log
Gto.	1.0000	0.9454	0.9487	0.6578	257.4206	log
Gro.	1.0000	0.5479	0.1045	0.5467	271.7846	log
Hgo.	1.0000	0.8590	0.8801	0.8719	275.7459	log
Jal.	1.0000	0.1594	0.3459	0.9388	240.3622	log
Méx.	1.0000	0.2191	0.5364	0.0421	236.1930	log
Mich. de O.	1.0000	0.5808	0.8124	0.3986	299.8557	log
Mor.	1.0000	0.8073	0.2957	0.7715	317.7741	log
Nay.	1.0000	0.6651	0.9147	0.1649	316.4814	log
NL	0.7156	0.9735	0.6085	0.2791	241.6011	log
Oax.	1.0000	0.9600	0.5956	0.5864	269.6494	log
Pue.	0.9334	0.7626	0.3653	0.5520	245.9769	log
Qro.	1.0000	0.2988	0.9067	0.4297	236.8483	log
Q. Roo	0.9767	0.5077	0.0394	0.7985	269.3532	log
SLP	1.0000	0.6504	0.4167	0.4007	229.7664	log
Sin.	1.0000	0.5988	0.8766	0.1394	313.2004	log
Son.	0.9987	0.8779	0.3942	0.6403	270.9363	log
Tab.	1.0000	0.8254	0.1598	0.5874	270.1237	log
Tamps.	0.9876	0.9484	0.3695	0.0029	299.8285	log

Anexo 2

Concluye

Evaluación estadística: modelos INEGI

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Tlax.	1.0000	0.3373	0.6141	0.3283	392.5801	<i>log</i>
Ver. de l. de la Ll.	1.0000	0.7710	0.3431	0.9980	240.5965	<i>log</i>
Yuc.	1.0000	0.8399	0.4978	0.9500	260.1765	<i>log</i>
Zac.	1.0000	0.6504	0.7974	0.7175	301.6085	<i>log</i>

*/ El INEGI indica que no encuentra patrón estacional en esta serie.

Anexo 3

Continúa

Evaluación estadística: enfoque alternativo

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Ags.	0.9970	0.8964	0.9117	0.9533	304.4682	<i>log</i>
BC	1.0000	0.9735	0.8758	0.8851	274.3108	<i>none</i>
BCS	1.0000	0.9776	0.7364	0.7971	338.2677	<i>log</i>
Camp.	1.0000	0.9852	0.6149	0.6969	295.1092	<i>log</i>
Coah. de Z.	1.0000	0.9911	0.1521	0.9834	258.8206	<i>none</i>
Col.	1.0000	0.9319	0.5666	0.9287	313.1694	<i>log</i>
Chis.	1.0000	0.6997	0.8026	0.8228	298.6462	<i>log</i>
Chih.	0.9065	0.7173	0.8962	0.8718	257.6747	<i>log</i>
CDMX	1.0000	0.5672	0.9787	0.7123	248.8935	<i>none</i>
Dgo.	1.0000	0.9911	0.1640	0.8660	322.8667	<i>log</i>
Gto.	1.0000	0.9776	0.8903	0.5334	258.7408	<i>none</i>
Gro.	1.0000	0.8791	0.5027	0.8135	273.7549	<i>none</i>
Hgo.	1.0000	0.9893	0.9786	0.8659	279.3237	<i>none</i>
Jal.	1.0000	0.0971	0.9385	0.9140	247.7539	<i>none</i>
Méx.	1.0000	0.9297	0.6891	0.9859	263.6581	<i>log</i>
Mich. de O.	1.0000	0.8254	0.5346	0.8950	329.4463	<i>log</i>
Mor.	1.0000	0.9134	0.9512	0.5695	319.9894	<i>log</i>

Evaluación estadística: enfoque alternativo

Entidad	QS	Friedman	Ljung-Box	Shapiro	AIC	Trans
Nay.	1.0000	0.9776	0.8609	0.7781	303.1116	<i>none</i>
NL	1.0000	0.7244	0.7684	0.9823	249.2211	<i>none</i>
Oax.	1.0000	0.9735	0.7757	0.7920	271.2719	<i>log</i>
Pue.	1.0000	0.7819	0.4179	0.8675	278.5381	<i>none</i>
Qro.	1.0000	0.9454	0.8429	0.7755	251.3195	<i>none</i>
Q. Roo	1.0000	0.8435	0.3358	0.5534	293.2648	<i>none</i>
SLP	1.0000	0.9319	0.9112	0.8110	224.5664	<i>none</i>
Sin.	1.0000	0.9776	0.9419	0.1899	308.4488	<i>log</i>
Son.	1.0000	0.7710	0.9686	0.6513	273.1348	<i>log</i>
Tab.	1.0000	0.7351	0.2754	0.9683	287.7929	<i>none</i>
Tamps.	1.0000	0.6481	0.6553	0.9413	278.4726	<i>none</i>
Tlax.	1.0000	0.9454	0.8423	0.8280	386.8301	<i>none</i>
Ver. de I. de la Ll.	1.0000	0.8779	0.5190	0.9333	288.7431	<i>none</i>
Yuc.	0.9300	0.8791	0.7911	0.9487	245.5384	<i>log</i>
Zac.	1.0000	0.8073	0.8762	0.8684	305.2635	<i>log</i>

Impacto de la crisis económica del 2008 en el empleo y salarios de las industrias manufacturera y automotriz de la región sureste de Coahuila de Zaragoza¹

Impact of the 2008 Economic Crisis on Employment and Wages in the Manufacturing and Automotive Industries in the Southeast Region of Coahuila de Zaragoza

Reyna Elizabeth Rodríguez Pérez,* Liliana Meza González** y David Castro Lugo*

¹ Los autores agradecen el apoyo financiero recibido por la Universidad Autónoma de Coahuila a través del proyecto semilla número C01-2019-4.

* Universidad Autónoma de Coahuila, reynarodriguez@uadec.edu.mx y david.castro@uadec.edu.mx, respectivamente.

** Instituto Nacional de Estadística y Geografía, liliana.meza@inegi.org.mx



Body of car on conveyor top view. Modern Assembly of cars at the plant. The automated built process of the car body/Trainmak_Ivan/Stock

El objetivo de este trabajo es analizar las consecuencias de la perturbación económica del 2008 en industrias altamente integradas con la economía mundial, como es el caso de la automotriz por la fuerte integración de esta actividad de la Zona Metropolitana de Saltillo con el mercado de Estados Unidos de América, donde las remuneraciones (sin importar su nivel de calificación) de los trabajadores de este subsector resultaron más afectadas que las de los del sector manufacturero en general de la misma región. La investigación se llevó a cabo con datos de la Encuesta Nacional de Ocupación y Empleo del 2005, 2009 y 2018 y a través de la metodología de DiNardo, Fortin y Lemieux.

Palabras clave: salarios; empleo; diferenciales salariales; integración comercial; México.
JEL Classification: J01, J31, R33.

Recibido: 25 de junio de 2020.
Aceptado: 24 de septiembre de 2020.

Introducción

La implementación del modelo económico con orientación hacia el exterior mediante la incorporación al Acuerdo General sobre Aranceles Aduaneros y Comercio (GATT, por sus siglas en inglés) en 1985 y, después, al Tratado de Libre Comercio de América del Norte (TLCAN) en 1994 (además de otros tratados comerciales bilaterales y multilaterales) generaron en la economía mexicana un mayor grado de integración a nivel global y, en particular, con la de Estados Unidos de América (EE. UU.), misma que se expresa en la magnitud de flujos comerciales, servicios e inversión extranjera directa (IED).

A nivel espacial y sectorial, la integración comercial generó en México una relocalización de las actividades económicas provocando que ya no fuera solo el centro del país el motor económico más importante; de hecho, impulsó el desarrollo económico de diversas regiones y sectores (Castro y Félix, 2010, Figueroa *et al.*, 2018). Sin embargo, esta

The objective of this paper is to analyze the consequences of the 2008 economic turmoil in industries highly integrated with the world economy, as is the case of the automotive industry due to the strong integration of this activity in the metropolitan area of Saltillo with the U.S. market, where the salaries (regardless of their level of qualification) of workers in this subsector were more affected than those of the manufacturing sector in general in the same region. The research was carried out with data from the National Occupation and Employment Survey of 2005, 2009 and 2018 and through the methodology of DiNardo, Fortin and Lemieux.

Key words: wages; employment; wage differentials; trade openness; Mexico.
JEL Classification: J01, J31, R33.

misma acción ocasionó cambios económicos que afectaron a trabajadores de ciertas industrias de manera diferenciada. Esto implica que la globalización requiere también de un Estado fuerte, que implemente políticas industriales y laborales acordes con las necesidades y circunstancias. La historia muestra que las crisis económicas mundiales son cíclicas, y México debe tomar esto en cuenta para afrontarlas de una mejor manera. La emergencia sanitaria del 2020 (COVID-19) dejará muchas lecciones en este sentido.

La crisis financiera internacional del 2008, considerada (hasta antes del 2020) la de mayor magnitud desde la gran depresión de 1929, tuvo importantes repercusiones sobre la actividad económica global, y su efecto ha tenido implicaciones relevantes, en el mediano plazo, sobre el comportamiento de la economía mundial. De acuerdo con Juárez *et al.* (2015), entre el 2008 y 2012 esta creció 6 puntos porcentuales menos que el ritmo observado en los 28 años previos. La actividad

económica de México no estuvo exenta de tal perturbación; así, durante el segundo trimestre del 2009, las actividades económicas se contrajeron 8.9 % a tasa anual, y el ritmo de crecimiento se redujo posteriormente respecto al observado en los años previos. Por otra parte, dada la heterogeneidad espacial y sectorial en el grado de integración a la economía internacional, el impacto de la crisis fue diferenciado, lo cual hace suponer que, en caso de perturbaciones, las prioridades de atención a través de la política pública deberían centrarse en aquellas regiones e industrias más expuestas.

La zona de la frontera norte de México ha sufrido fuertes modificaciones económicas relacionadas con esta crisis, como estancamiento y declive, lo que generó cambios en la estructura ocupacional y salarial (desempleo, aumento en la informalidad y precariedad salarial). La economía de Coahuila de Zaragoza, en específico, se ha visto marcada por estos problemas, a pesar de las grandes inversiones en manufactura, que comenzaron con el establecimiento del complejo industrial en la región sureste del estado² en la segunda década del siglo XX.

El Grupo Industrial Saltillo (GIS) se considera el primer antecedente de una nueva cultura industrial y laboral en la entidad y, con ello, de la especialización de la mano de obra. Desde que se estableció en 1928, comenzaron a gestarse las condiciones necesarias de infraestructura, con circuitos carreteros y ferroviarios, para el posterior desarrollo estatal. Cabe recordar que en la década de los 80, con la llegada de *General Motors* y *Chrysler*, se empezó a formar en la Zona Metropolitana de Saltillo (ZMS) el clúster de empresas automotrices (Aboites y Castro, 2010).

Durante los últimos años, la actividad de la industria automotriz (IA) se ha consolidado en la región; de acuerdo con información de los Censos Económicos (CE) 2014,³ la IA (subsector 336) en la ZMS concentra 43 % del empleo manufacturero,

57 % de las remuneraciones y más de 65 % del valor agregado censal bruto generado en el sector de la manufactura. Lo anterior muestra su relevancia pero, también, hace prever que la estabilidad económica de la región puede ser vulnerable ante modificaciones en el entorno económico. Por ello, resulta pertinente preguntarse: ¿qué efectos tuvo la crisis del 2008 en diferentes industrias de México, definidas estas a través de su grado de integración con la economía mundial? Para responder, este trabajo analiza el caso de la IA localizada en la ZMS y la compara con los resultados laborales en la industria manufacturera en general de la misma región que, si bien no implica que esté orientada de manera exclusiva al mercado interno, sí presenta un menor grado de integración con el exterior respecto a la automotriz y, así, está menos expuesta a los vaivenes de la economía global. De ahí que otras preguntas relevantes pueden ser las siguientes: ¿cuáles son las características del empleo en esta industria?, ¿qué impactos tuvo la crisis económica del 2008 sobre el empleo y salarios de este subsector?, ¿los cambios difieren por tipo de calificación laboral?, ¿cómo se comparan estos cambios con los que se observan en toda la industria manufacturera de la ZMS?

La hipótesis a probar es que, debido a la fuerte integración de la actividad automotriz de la ZMS al mercado de EE. UU., el empleo y los salarios de los trabajadores en este subsector resultaron más afectados que en el resto de la manufactura por la recesión que se presentó en la economía estadounidense durante el 2008-2009.

Los datos que se utilizaron son de la Encuesta Nacional de Ocupación y Empleo (ENOE)⁴ del 2005, 2009 y 2018 correspondientes a la ZMS, mientras que la metodología correspondió a la de DiNardo, Fortin y Lemieux (DFL) (1996) que, a diferencia de otros métodos, nos permite observar las distribuciones de los salarios de los trabajadores ubicados en la IA frente al sector manufacturero (no automotriz), con contrafactuales, para conocer si habría diferencia salarial si las características

2 Para efectos de este documento, se utilizará de manera indistinta región sureste de Coahuila de Zaragoza y Zona Metropolitana de Saltillo, la cual está integrada por los municipios de Saltillo, Ramos Arizpe y Arteaga.

3 Programa del Instituto Nacional de Estadística y Geografía (INEGI).

4 Ídem.

(en términos de edad, años de experiencia y calificación) fuesen las mismas.

Los resultados indican que las perturbaciones del 2008 y la reducción de las exportaciones automotrices tuvieron efectos adversos sobre la actividad económica de este subsector, impactando en particular sobre el mercado laboral. En el análisis descriptivo, se observa una reducción del empleo no calificado de la IA durante la crisis y una continuación en años posteriores. Mediante la metodología DFL se confirma la existencia de cambios salariales, no solo en el nivel medio del salario, sino a lo largo de toda la distribución. Por otra parte, el desempeño de la brecha salarial por tipo de calificación parece mostrar una tendencia a la reducción, caracterizada por una caída relativa del salario real por hora, sobre todo de los trabajadores calificados.

El artículo se presenta en el siguiente orden: la primera parte describe el proceso de integración comercial y los efectos de la crisis del 2008 en la región sureste de Coahuila de Zaragoza, y presenta un resumen de los estudios que han abordado el tema de manera empírica; la segunda sección describe la metodología y los datos empleados; en la tercera, resultados, se hace un análisis descriptivo del empleo y los salarios de la ZMS, se exploran las diferencias salariales y sus determinantes, además de que se analizan las funciones de densidad salarial de los trabajadores por tipo de calificación en las industrias manufacturera y automotriz, y en la parte empírica se hace un análisis contrafactual con el método DFL; por último, se detallan las conclusiones.

Efecto de la crisis en el mercado laboral

Integración comercial

En general, la globalización se ve como una oportunidad para el progreso de las economías en vías de desarrollo y, de acuerdo con la teoría del comercio internacional (sostenida en el teorema de Stol-

per-Samuelson), estos países tendrían una ventaja relativa en las exportaciones de bienes intensivos en mano de obra no calificada. Conforme a este esquema teórico, los efectos de la integración comercial sobre el mercado laboral serían: 1) un incremento en la generación de empleos, en especial de los trabajadores no calificados; 2) un aumento relativo en los salarios de estos; y 3) una reducción en la desigualdad salarial por tipo de calificación. No obstante, también es un proceso que puede tener efectos negativos. La dinámica relacionada con el resto del mundo y el amplio intercambio comercial a nivel internacional provocan que aquellas ocupaciones que se encuentran en los sectores económicos centrados en la exportación se vean vulnerables ante las crisis globales, poniendo en riesgo el empleo, la calidad de este y su remuneración (Rodríguez y Castro, 2014).

Acemoglu (2003) desarrolló un modelo para analizar la diferencia en los retornos a las calificaciones a lo largo del tiempo y entre países; estudió el impacto del comercio internacional sobre la desigualdad salarial. En la revisión teórica, argumentó que los retornos estuvieron determinados por la tecnología, la oferta relativa de mano de obra calificada y el comercio, asumiendo que la tecnología era endógena y respondía a incentivos de ganancias y al aumento de trabajadores calificados. El resultado reveló que el incremento del comercio internacional indujo un cambio tecnológico sesgado por calificación y aumentó la desigualdad tanto en EE. UU. como en naciones menos desarrolladas, entre las que se incluye a México, esto a partir de una elevación en los precios relativos de los productos que requieren mano de obra más calificada.

México entró en un periodo de liberalización comercial desde mediados de la década de los 80, que se profundizó a través de la entrada en vigor del TLCAN, lo cual provocó que 80 % de sus exportaciones fueran hacia su vecino del norte y alrededor de 60 % de la IED proviniera de este mismo país (López *et al.*, 2010 y Rodríguez *et al.*, 2017).

El enfoque del modelo económico orientado hacia la exportación de la economía mexicana pro-

vocó que, durante la crisis mundial iniciada en el 2008, esta resultara más afectada que el resto de los países de América Latina (Salgado, 2011) debido, sobre todo, a que las exportaciones se redujeron (y, por lo tanto, la producción) afectando de manera directa al mercado laboral interno.

Robbins (2003) menciona que, en México, los salarios y la oferta relativa de los trabajadores calificados aumentaron alrededor de 30 y 50 %, respectivamente, entre 1987 y 1993. Estos resultados parecen contrarios a las predicciones habituales de Stolper-Samuelson para un país relativamente dotado de mano de obra no calificada. Sin embargo, hubo reformas importantes que regularon la inversión extranjera, lo que complicó la identificación del aumento de los salarios relativos con la reforma comercial. Por su parte, López *et al.* (2010), Rodríguez y Castro (2014) y Rodríguez *et al.* (2017) coinciden en mencionar que los salarios reales de todos los trabajadores, sin importar su nivel de calificación, cayeron en el 2008 y 2009; además, se presentó un efecto diferenciado por regiones en México de acuerdo con el grado de integración económica internacional, pues "... cuanto más globalizada e integrada es la región a la economía de Estados Unidos, mayor es el impacto de las recesiones..." (López *et al.*, 2010:13). En este sentido, en la zona norte del país y en las localidades urbanas en las que se ubica la IA (donde gran parte de lo que se fabrica se destina a la exportación), el descenso de la demanda externa dio lugar a una caída de la producción y a una gran pérdida de empleos, contrario a lo sucedido en la región sur del país.

Lo anterior permite explicar que, si bien en el periodo estudiado (en específico el 2009) la variación en los salarios relativos por tipo de calificación pudiera atribuirse básicamente a los efectos de la crisis, se podría argumentar, también, que estos han sido diferentes en distintas zonas del país, sobre todo en función de la integración con el exterior que las regiones presentan. De manera clara, Coahuila de Zaragoza estaría entre los estados más afectados por la perturbación del 2008 y, en especial, la región sureste y la industria automotriz que se ubica en ella.

Efectos de la crisis del 2008 en el mercado laboral de las industrias manufacturera y automotriz

México se vio afectado por la recesión, detonada en EE. UU. El empleo de ciertas manufacturas resultó ser más vulnerable, pues a consecuencia de ella se produjeron paros técnicos, despidos de empleados y reducción salarial, así como una disminución de los inventarios y volúmenes de producción.

Guevara (2012) argumenta que esta crisis cambió los patrones de crecimiento de la economía a nivel mundial y puso de manifiesto que la expansión que se había registrado en el periodo anterior a ella (en términos de producción y empleo) estaba sustentada en la flexibilidad laboral, misma que, luego de la recesión, tuvo que ser reactivada de manera paulatina a través de suspensión temporal de personal y reducción de empleo, además de una marcada estratificación salarial, elementos a través de los cuales se dinamizó de nuevo la producción.

Ríos y Carrillo (2014) encontraron que la manufactura tuvo una participación relevante en la economía mexicana, ya que aportó alrededor de 19 % del Producto Interno Bruto (PIB) y del empleo durante el periodo 2005-2009. Respecto a este, se tiene que 82.7 % lo generaban cuatro subsectores de manufactura, cuya aportación en el total del sector en el 2009 fue la siguiente: de productos metálicos, maquinaria y equipo, 31.4 %; de alimentos, bebidas y tabaco, 23 %; de textiles y prendas de vestir, 17.2 %; y de sustancias químicas y productos de plástico y caucho, 11.1 por ciento. Sin embargo, durante la crisis económica del 2008, solo los subsectores de productos metálicos, maquinaria y equipo y de textiles y prendas de vestir presentaron decrecimiento en el nivel de empleo.

La expansión de la IA en México ha sido consecuencia del establecimiento de grandes empresas orientadas al mercado mundial, las cuales ven un atractivo en la mano de obra barata y calificada, así como en la cercanía geográfica con el de EE. UU. (Mendoza, 2011; Félix, *et al.*, 2018). Por ello, la cri-

sis de este subsector no solo afecta al país en su conjunto (en términos de las exportaciones y divisas) sino, de manera particular, a 15 zonas urbanas que concentran poco más de 50 % del empleo de la industria automotriz de la nación. Así, esta actividad se encuentra muy concentrada en términos espaciales (Félix *et al.*, 2018) y se ha convertido en el corazón del modelo de producción basado en la integración al comercio internacional, pues fue la industria más dinámica y con mayor participación de las exportaciones de México.

Como ejemplo, se puede señalar que, de acuerdo con información presentada por el periódico *El Economista* en su sección *Empresa*,⁵ la IA contribuyó con más de un tercio de las exportaciones manufactureras y durante la última década avanzó más de 10 puntos porcentuales en su participación, hecho que pone de manifiesto su relevancia dentro del sector y de la economía en general, pero también da cuenta de la vulnerabilidad que este presenta, y en particular la IA, respecto a las perturbaciones económicas del exterior.

De Bell (2012) menciona que el peso económico del clúster de empresas automotrices en la región sureste de Coahuila de Zaragoza en el total de las exportaciones del estado se acrecentó a partir del TLCAN, ya que más de 60 % del total de la nueva inversión en la entidad benefició a la fabricación de vehículos. Sin embargo, durante el 2008, este subsector de la ZMS fue muy golpeado. De acuerdo con Félix *et al.* (2018), durante el lapso de junio del 2008 a junio del 2009, en esta zona se perdieron 7 688 puestos de trabajo en la IA, una de las mayores cifras dentro de los 15 centros urbanos con mayor presencia de esta industria, equivalentes a 56.7 % de los empleos perdidos en la región sureste.

Gachúz (2011) hace una síntesis sobre los determinantes más recientes que derivaron en el colapso del subsector de los automóviles durante la crisis del 2008: en primera instancia fue el incremento de los precios de los combustibles (el del petróleo au-

mentó de manera repentina); después, vino la crisis financiera, cuando la escasez de crédito comenzó a afectar primero a EE. UU. y luego se extendió por Europa, Asia y alcanzó niveles mundiales; y al final, la recesión económica global trajo como consecuencia una disminución en la demanda de vehículos sin precedentes. La IA enfrentó uno de sus peores momentos en la historia.

Quiroz (2009), abonando a lo anterior, menciona que la crisis comenzó a vislumbrarse desde febrero del 2008 cuando la tasa de crecimiento anual de la producción de vehículos en México empezó a decaer de 25.9 a 15.5 %; para julio del 2008, fue de -2.8; y en enero del 2009, de -50.9 por ciento. A raíz de esta situación, entre el 19 de diciembre de 2008 y el 8 de enero de 2009 se efectuaron paros técnicos en diferentes armadoras del país, con pago de 50 % del salario, con el fin de aminorar los despidos masivos.

Félix y Castro (2015) encontraron que la IA en México funcionaba de acuerdo con la teoría de la fragmentación (Jones y Kierzkowski, 1990, 2000 y 2005), la cual se sustenta en dos conceptos clave: bloques de producción y servicios de vinculación (*production blocks* y *service links*). El primero hace referencia a que el bien puede fabricarse combinando varios segmentos, mientras que el segundo facilita el funcionamiento de cada bloque, así como la interacción entre ellos, donde los bloques y servicios que intervienen en el proceso de producción pueden ubicarse en más de una región o país, lo que permite aumentar la competitividad y aprovechar las ventajas en dotaciones de recursos y costos de factores.

La globalización permite la fragmentación de los procesos de producción y el desarrollo de las tecnologías de información y comunicación, es decir, la mayor apertura de los países hace posible la integración de estos,⁶ fragmentados a nivel internacional para la reducción de costos.

⁵ *El Economista*. 16 de abril de 2017.

⁶ También se denominan *redes internacionales de producción-distribución o red integradora de producción* (Félix, *et al.*, 2018).

Algunos autores, como Félix y Castro (2015), señalan que ese tipo de relaciones se presentan en la IA mexicana, donde se observa la prevalencia de clústeres localizados en ciudades con accesibilidad a las principales salidas de exportación. De acuerdo con los autores, las empresas automotrices combinan varias ventajas: a) la importación temporal de maquinaria, equipo, insumos y componentes provenientes de bloques localizados en otros países —facilitada por medio del Programa de la Industria Manufacturera, Maquiladora y de Servicios de Exportación (IMMEX)—; b) la localización de clústeres en la ciudad, o en regiones cercanas, de empresas integradas como proveedoras; c) la accesibilidad para exportar bienes finales o intermedios a bloques de distribución en otros países; y d) la disposición de fuerza laboral calificada de forma aceptable, pero menos costosa de manera sustancial que en naciones desarrolladas.

Considerando que las plantas armadoras automotrices en México fungen como la industria alrededor de la cual se organiza un conjunto de empresas proveedoras (algunas localizadas local o regionalmente y otras fuera del país) y que la mayoría de las compañías interrelacionadas en la zona utilizan procesos intensivos en mano de obra, se puede establecer qué perturbaciones económicas externas o modificaciones en las relaciones comerciales con el resto del mundo, pero en especial con EE. UU., tendrían efectos adversos sobre las economías regionales con fuerte presencia de la IA y sobre su mercado laboral. La magnitud y permanencia del efecto dependerá de la relevancia de la IA en la región y del tipo de las características de la propia perturbación (Félix *et al.*, 2018).

En síntesis, el comportamiento del empleo en la manufactura frente a las crisis económicas, en específico en el subsector de producción automotriz, depende de las condiciones productivas y de las características de las empresas. Siendo la IA uno de los pilares en los que se ha apuntalado el avance de la economía nacional, en particular del sureste de Coahuila de Zaragoza, resulta esencial analizar qué ha pasado en materia laboral en esta zona.

El mercado laboral en la ZMS

Dado el peso que tiene la IA en esta área, es importante conocer la importancia de este subsector como receptor de IED y destacar los efectos que los problemas en estos frentes han tenido sobre el mercado laboral de la zona.

Aguirre *et al.* (2018) mencionan que la región sureste de la entidad registra un alto crecimiento económico debido al dinamismo de las actividades de la industria manufacturera, ya que desde la década de los 80, con el establecimiento del complejo de *General Motors* en Ramos Arizpe, se dio un repunte en la actividad económica regional. La producción automotriz se ha convertido en la principal fuente de divisas del estado, así como en importante generadora de empleos y de derrama económica. Además, son significativas en esta zona las industrias metal-mecánica, químico-farmacéutica, del cemento, de productos minerales no metálicos y, en menor medida, la textil y del vestido, así como la papelería.

Salazar (2015) comenta que en el clúster de empresas automotrices que se ubica en la región sureste de Coahuila de Zaragoza se localizan 10 de los más importantes parques industriales del estado, donde se incluye a más de 300 fabricantes de autopartes instalados en el área conurbada de Ramos Arizpe y Saltillo; 37.8 % del PIB estatal está representado por este subsector, donde se producen 25 de cada 100 automóviles del total nacional. En la entidad se cuenta con dos plantas armadoras de vehículos y una dedicada a la fabricación de tractocamiones, por lo que más de 70 mil empleos dependen de la IA en la ZMS.

Dávila (2010) encontró que, tanto en la economía mexicana como en la de Coahuila de Zaragoza, la integración comercial modificó las formas de producir y propició una acelerada organización territorial y sectorial de la actividad económica. Con la mayor integración de México al mercado mundial, en especial al de EE. UU., se expandieron de forma acelerada las exportaciones e importaciones de bienes y servicios.

Aboites y Castro (2010) mencionan que la integración comercial —que comenzó durante la década de los 80— impactó a la ZMS por el arribo de la IA (1980-1994) con tres plantas instaladas en Ramos Arizpe. La estructura económica de la región se transformó, ampliando el abanico de empleo, de las ramas tradicionales (como la minería, textil y agroindustrial) hacia la manufactura, dentro de las cuales se encuentra la metal-mecánica y la automotriz.

Castro *et al.* (2013), tomando como base la información presentada en los CE 2009, mencionan que, en el sureste del estado, la actividad del sector manufacturero concentraba a 72 mil personas ocupadas, lo que representó 36.3 % de la población económicamente activa. Para el 2014 (con datos de los CE 2014, se tiene que estas pasaron a 37.3 % del total, lo cual significó un aumento relativamente bajo en ese periodo.

Dávila (2010) encontró que, durante la era del TLCAN y hasta el 2008, el PIB del estado creció 4 % anual, cifra superior a la nacional (3 %); sin embargo, la situación cambió en el 2009, ya que la economía norteamericana experimentó una contracción de 2.4 %, provocando una caída en las exportaciones de -14.8 por ciento. Al final del 2009, el PIB nacional cayó 6.5 por ciento. En el caso de Coahuila de Zaragoza, su mayor dependencia en exportaciones, sobre todo en el subsector de producción automotriz, causó un desplome del PIB de 13.5 por ciento.

Medina (2009) menciona que Coahuila de Zaragoza, al ser un estado con vocación industrial, ha sufrido los efectos de la recesión económica mundial, lo que ha afectado el empleo pues, de acuerdo con el Instituto Mexicano del Seguro Social (IMSS), en la entidad se perdieron 36 318 empleos formales de diciembre del 2008 a febrero del 2009. Esto implica que la crisis causó que la actividad industrial (en especial la IA) trabajara por debajo de su capacidad habitual, dando como resultado el despido de trabajadores.

Con lo anterior, se puede deducir la importancia que para Coahuila de Zaragoza tiene el subsector

automotor en términos de crecimiento económico y empleo, y de forma más específica para la región sureste, ya que los efectos de la crisis del 2008 causaron una pérdida de empleo en la manufactura y una reducción del PIB estatal.

En el siguiente apartado, se analizará, mediante una revisión de literatura empírica, el impacto de la crisis económica en el empleo a niveles nacional y regional para observar cómo se ha medido y qué resultados se han obtenido.

Revisión de literatura

A continuación, se describen algunos estudios que se han realizado sobre los efectos de la integración económica de México con el exterior y el impacto sobre el mercado laboral. En especial, nos interesa entender el efecto de la crisis del 2008 sobre la desigualdad salarial por tipo de trabajador, definido este por su nivel de calificación laboral, en distintas regiones de México, según su exposición a la integración comercial.

En un estudio a nivel nacional para el periodo 1988-2000, Esquivel y Rodríguez (2003) encontraron un incremento en la brecha salarial por tipo de calificación laboral en el periodo anterior a la firma del TLCAN, la cual es consistente con la evidencia que presentan investigaciones como las de Meza (1999, 2005), Cragg y Epelbaum (1994), Airola y Juhn (2005), entre otras. Además, los autores averiguaron que, posterior a los tratados de liberalización en la economía, la brecha se redujo, coincidiendo en cierta forma con varias épocas de recesión en la economía mexicana y con lo que otros afirman respecto a la caída en la desigualdad en México a partir de la firma del TLCAN (Castro, 2007; Castro y Huesca, 2007; Bosch y Manacorda, 2010 y Campos-Vázquez, 2013).

Campos-Vázquez (2013) investigó las fuerzas que estaban detrás de la caída de la desigualdad en el periodo posterior al TLCAN. Para medirla, utilizó la metodología cuantílica de Machado y Mata (2005) y aplicó la descomposición Bound y Johnson (1992)

para el periodo 1996-2006. Los resultados mostraron que la disminución de la desigualdad salarial, causada por una caída en los retornos a la educación, se explicaba por una mayor oferta relativa en los trabajadores con nivel universitario. Sus hallazgos sugieren que los aumentos en la oferta de empleados más calificados fueron la fuente principal de la reducción de la desigualdad salarial.

Otro estudio que aborda este tema, pero por tipo de calificación laboral, es el de Campos-Vázquez, Esquivel y Lustig (2014), quienes analizaron el periodo de 1989 al 2010. Con la metodología de Oaxaca-Blinder y una descomposición del coeficiente de Gini, encontraron un incremento significativo de las brechas salariales entre trabajadores calificados y no calificados en el primer lapso hasta el 2000. Después, revelaron una reducción porcentual en esta brecha, que explicaron a través de la relación que tiene el mercado laboral con la integración comercial. Además, mencionan la importancia de las capacidades productivas que la nueva estructuración laboral requería, lo que causa, en un primer tiempo, incrementar brechas para después comenzar a reducirse.

Cuando se hace el análisis a nivel regional, Andalón y López Calva (2002) estudiaron la evolución en la desigualdad de ingreso entre el sur y el resto de México. Utilizaron de manera agregada los datos de las encuestas nacionales de Ingresos y Gastos de los Hogares (ENIGH) y la de Empleo Urbano (ENEU) de 1989-2000⁷ y estimaron un coeficiente de Gini, así como el índice de Theil. Los resultados indicaron que la desigualdad en el ingreso disminuyó en México durante la década de los 90, siendo aún menor disminución en la región sur.

Chiquiar (2004), a través de los censos industriales de los CE del periodo 1990-2000, analizó las diferencias salariales regionales en la década de los 90. Sus resultados indican que estas parecen ser un reflejo del impacto heterogéneo que el TLCAN tuvo sobre el potencial de las empresas a través de sus regiones. México, tras experimentar un au-

mento en su comercio, IED y un mayor impacto de la migración en los salarios locales, no parece que haya experimentado una reducción en los retornos a la educación de los trabajadores más calificados y, sin duda, la globalización pudo haber llevado a los incrementos de la desigualdad salarial dentro de los distintos grupos de empleados por nivel de calificación.

Por su parte, Garza y Quintana (2014), a través de una regionalización del país de acuerdo con el nivel salarial de cada uno de los estados y, después, apoyada en una regresión cuantílica para el 2005 y 2010, encontraron que los factores político-ideológicos (abandono de políticas de protección al empleo y promoción de la desregulación de mercados laborales) son elementos importantes que ayudan a explicar la desigualdad salarial.

Rodríguez y González (2015) analizaron la evolución de la desigualdad entre trabajadores calificados y no calificados a partir de la apertura comercial en regiones de México y, de este modo, observaron si la teoría de Stolper-Samuelson se cumplía. Ellos determinaron que en el periodo en el que se sentirían más los efectos de la crisis (2009-2015) disminuyó la desigualdad salarial entre empleados calificados y no calificados en línea con el teorema de Stolper-Samuelson. Sin embargo, consideraron que esta reducción se debió más a la precarización de los salarios de los calificados.

Levy y López Calva (2016) desarrollaron un trabajo empírico con el objetivo de hacer un análisis del impacto de la mala asignación de recursos en los ingresos y los retornos a la educación en México, ya que se observó un aumento de la oferta de trabajadores con mayor nivel educativo, así como una mala asignación del recurso humano en empresas menos productivas. Estas dos tendencias generaron un desajuste entre la oferta y la demanda de empleados calificados. En su trabajo, estos autores utilizaron datos de la Encuesta Nacional de Empleo (1996-2004)⁸ y la ENOE (2005-2015). Al realizar el análisis de descomposición de las ganancias

⁷ Ambas del INEGI.

⁸ Del INEGI.

de los trabajadores en características observables y no observables de la empresa, se simuló una distribución de ganancias contrafactual en ausencia de una mala asignación de capital humano. En el contexto hipotético, los diferenciales de ingresos entre los niveles de escolaridad aumentarían, al igual que los retornos a la educación, si los empleados se ubicaran en empresas más productivas y acordes con su nivel de calificación.

En un estudio a nivel sectorial realizado por Meza (2003) se encontró que durante la década de los 90 la integración comercial indujo poco o nulo impacto sobre los cambios laborales en el sector manufacturero, para el trabajo calificado en especial, y concluyó que el cambio tecnológico sesgado y factores estructurales internos (crisis económicas) fueron los elementos que condicionaron este mercado. En sus estimaciones, la autora utilizó la ENEU y la Encuesta Industrial Anual (EIA).⁹

Siguiendo con estudios en el sector manufacturero de México, Ríos y Carrillo (2014) analizaron el impacto de los cambios en el producto de sus ramas o subsectores, sucedidos a raíz de la crisis del 2008, sobre el empleo calificado y no calificado. Con información de datos de panel, estos autores aplicaron un modelo de efectos fijos. Sus resultados indicaron que, en periodos de crisis, las empresas tienden a conservar el trabajo calificado y a despedir a trabajadores no calificados, siendo más afectadas las maquiladoras y las industrias tradicionales intensivas en empleo no calificado.

Por su parte, Rodríguez *et al.* (2017) analizaron la evolución de la desigualdad salarial a nivel regional de acuerdo con su grado de exposición a la apertura comercial. Al aplicar el índice de descomposición de Theil, con los datos de la ENOE ediciones 2005-2015, encontraron una disminución en la desigualdad salarial en el país y en todas las regiones, sobre todo en las menos expuestas a la integración comercial, obteniendo que su fuente era la calificación laboral y el sexo de los empleados.

⁹ Ídem.

En síntesis, se revisó el proceso de incorporación del país a la globalización que se dio a finales de la década de los 80, y teniendo como mayor punto de referencia la firma del TLCAN en 1994. En este sentido, México ha buscado ser una de las naciones más abiertas en materia comercial y, de este modo, generar desarrollo y aumentos en la productividad; sin embargo, los resultados no han sido del todo satisfactorios. A partir de la época de cambio estructural, se comenzaron a desarrollar diversos trabajos de investigación para observar el impacto de este cambio de modelo económico en la desigualdad en el país, en especial la que se observa entre trabajadores calificados y no calificados. El resultado más relevante fue que esta ha caído, en particular, por el deterioro de los salarios de los más calificados. Esta caída se explica por el aumento en la oferta de trabajo altamente calificado y por la mala asignación de recursos, donde al parecer se están privilegiando con política fiscal a las empresas menos productivas. Esta investigación busca contribuir a esta literatura analizando la IA de la ZMS y comparándola con la manufacturera no automotriz de la misma región. Los resultados pueden revelar cómo perturbaciones económicas externas afectan de manera diferenciada a una industria altamente integrada con la economía internacional frente a otra que, si bien no puede decirse que esté orientada en exclusiva al mercado interno, sí presenta menor grado de integración.

Datos y metodología

Datos

Los usados en esta investigación provienen de la ENOE ediciones 2005, 2009 y 2018, tercer trimestre de cada año. Los datos se limitan a la ZMS, que es una de las áreas autorrepresentadas en las muestras. Para este trabajo, el uso de los años se justifica, dado que los primeros dos corresponden a los periodos pre y postcrisis económica; la elección del 2018 obedece a que es el año más reciente disponible a cuando se realizó el estudio.

Las variables de interés son: población ocupada que se desempeña en la manufactura no automoto-

triz, así como en la fabricación de maquinaria y equipo (industria automotriz 336) que corresponden al sector automotor. También, se consideraron los años de experiencia laboral y de educación, además de salario promedio por hora.

Los salarios de los trabajadores se deflataron al 2018 y se capturaron en forma mensual para, después, convertirlos a salarios por hora. Se consideró trabajo de tiempo completo (es decir, si la jornada laboral es igual o mayor a 40 horas a la semana); la población ocupada se clasificó en tres grupos con base en los años de escolaridad: no calificados (0-9), semicalificados (10-15) y calificados (15 y más). También, se contemplaron la experiencia potencial, los rangos de edad y el sexo.

La ENOE sufre el problema de no respuesta salarial, por lo que se realizó un proceso de imputación donde, para poder trabajar con la base de datos que se utilizó en esta investigación, primero se creó un *pool* de observaciones que estuvieran empleadas y que, además, tuvieran como posición en el empleo ser trabajadores subordinados y remunerados, empleadores y trabajadores por cuenta propia, esto para los cuatro trimestres del año en cuestión; después, a cada uno se le asignó un número aleatorio entre 0 y 1.

A partir del *pool* se creó un conjunto de datos de donadores (que declararon un monto de ingresos por trabajo) y otro de receptores (que no declararon ingresos o bien solo un rango de estos). A los receptores que manifestaron menos de 1 salario mínimo (SM) se les imputó el SM multiplicado por el número aleatorio ya asignado; a los que declararon 1 SM se les imputó este. Para cada individuo contenido en el conjunto de datos de receptores que aún no se le había imputado un ingreso, se buscó un donador en el conjunto de datos correspondiente. Esta búsqueda pudo haber arrojado más de un donador, por lo que se seleccionó entre ellos aquel que tuviera el número aleatorio mayor (que le fue asignado al formar el *pool*). Las variables de empate para la búsqueda de posibles donadores fueron las siguientes: sexo, condición de ocupación, sector de actividad económica (agropecuario,

manufacturas, comercio, servicios y otros), estrato, tipo de unidad económica (sociedades privadas y públicas), así como posición en la ocupación. Dado que era casi imposible que todos los receptores tuvieran un donador que incluyera todas las variables de empate (más de 95 % sí encontró) fue necesario repetir el proceso tres veces más, pero eliminando variables en cada ocasión: primero rama, después rama y estrato y, por último, rama, estrato y que el tipo de unidad económica fuera diferente a *Agricultura de autosubsistencia*.

Antes de la aplicación del modelo se realizó un análisis estadístico con el objetivo de conocer los cambios en los salarios y el empleo de los trabajadores por tipo de calificación en la manufactura y el subsector de producción automotriz en la región sureste de Coahuila de Zaragoza.

Estimación de densidades de Kernel

Con el objetivo de saber qué pasó con los salarios de los trabajadores, más allá de los promedios, se realizó un análisis para conocer la densidad salarial a lo largo de la distribución de los empleados ubicados en la industria manufacturera y el subsector de producción automotriz para el 2005, 2009 y 2018. Para lograr este objetivo, se utilizaron las curvas de densidad Kernel, metodología que basa su estimación de densidades en histogramas suavizados que permiten calcular si los trabajadores localizados en la manufactura o en el subsector específico (IA) muestran un cambio en la distribución empírica del ingreso; es esencialmente no paramétrica, ya que se pueden obtener distribuciones del ingreso sin suponer ninguna forma funcional para las mismas.

Los datos se utilizaron para estimar la distribución de la variable de interés sobre la población, sin embargo, la aplicación del método de Kernel permitió que una observación fuera incluida en un número infinito de grupos que dio como resultado una suavización en la forma de distribución.

La principal restricción de una función de Kernel ($K(t)$) es que está integrada a 1. Cualquier función

de densidad de probabilidad satisface esta condición. El estimador de Kernel tiene las mismas funciones de diferenciación que la de Kernel elegida. Aun cuando K es, por lo regular, simétrica, la densidad estimada resultante no adquiere esta característica. La suavidad en la distribución de la densidad resultante depende de la longitud de la ventana del intervalo (llamado ancho de banda), denotado por h_n . Para consistencia del estimador, el ancho de banda debería decrecer conforme la muestra se incrementa. Sin embargo, es difícil escoger el óptimo porque la escasez de datos varía a lo largo de la distribución y la de los datos subyacentes no depende del tamaño de la muestra dibujada.

Las colas de las distribuciones típicamente tienen menores observaciones vecinas con respecto a las que se ven en la mitad de la distribución, por lo tanto, utilizar un ancho de banda con base en las áreas con datos escasos provocará una sobreesuavización en las áreas más densas. Se usó un procedimiento en dos etapas para calcular h_n de adaptación con estimaciones de densidad piloto obtenidas por medio de un ancho de banda fijo para el cálculo de factores con los que dichos h_n son ajustados.

Permitiendo a $x_i, i = 1, \dots, n$ denotar los valores observados de la muestra (w_i las ponderaciones muestradas y λ_i , el ancho de banda ajustado de los factores), el estimador de densidad de Kernel en valores del ingreso z está dado por

$$\hat{f}_n(z) = \frac{1}{n} \sum_{i=0}^n \frac{1}{h_n \lambda_i} w_i k \left(\frac{z-x_i}{h_n \lambda_i} \right) \quad (1)$$

El ancho de banda ajustado de los factores λ_i está definido por:

$$\lambda_i = \left[\frac{\exp\left(\frac{i}{n} \sum_{j=1}^n w_j \log \hat{f}_n(x_j)\right)}{\hat{f}_n(x_i)} \right]^{\frac{1}{2}} \quad (2)$$

donde $\hat{f}_n(x)$ es la estimación piloto de la densidad calculada usando el ancho de banda fijo, h_n . El inicial de este se calcula sobre la base del mínimo de la escala de rango intercuartil y la desviación estándar.

No obstante la pertinencia de utilizar las gráficas Kernel para observar la densidad salarial a lo largo de la distribución, es importante ir más allá en el análisis y aplicar una metodología como la DFL. En el siguiente apartado se explica en qué consiste.

Método DFL

La hipótesis a probar es que, debido a la fuerte integración de la actividad automotriz de la ZMS al mercado de EE, UU., el empleo y los salarios de los trabajadores en el subsector de producción automotriz, más que en el resto de la manufactura, resultaron en particular afectados por la crisis que se presentó en la economía estadounidense durante 2008-2009. Se analizan las diferencias en las remuneraciones más allá de los valores medios, pues se comparan las distribuciones considerando el comportamiento del logaritmo de salario por hora original y contrafactual del personal ocupado por tipo de calificación e industria; el primero se obtiene de la remuneración observada y el segundo se calcula a partir de considerar las características de los empleados por tipo de calificación ubicados en la manufactura, pero remunerados como si se encontraran en la IA, antes y después de la crisis del 2008.

La metodología propuesta (DFL) es semiparamétrica; una de sus mayores aportaciones es que muestra una clara representación de las diferencias salariales en la distribución, ya que se trabaja con la totalidad de la densidad salarial. Sus autores utilizan las distribuciones Kernel no paramétricas; la gran bondad de este método es que la distribución original es recalculada para ajustar por diferencias en características observables presuponiendo retornos constantes. No obstante, su desventaja es que solo puede controlar por diferencias en características observables.

Esta metodología parte de la generación de un escenario contrafactual, el cual se define como la distribución de salarios de los trabajadores en manufactura no automotriz vs. la automotriz, y lo que se busca contestar es: ¿hay diferencia salarial si las

características en términos de edad, años de experiencia y calificación fuesen las mismas?; también, ¿la relación salarial cambia antes y después de la crisis del 2008?

El razonamiento del método DFL propone utilizar distribuciones Kernel no paramétricas, por lo que se construye a partir de considerar la distribución salarial de j , donde j está compuesto por dos grupos: (C_M) y (C_A) ,¹⁰ los cuales contienen características determinadas de trabajadores por tipo de calificación en la manufactura y en la industria automotriz; su distribución salarial se describe como:

$$\begin{aligned}
 FC_M(W) &= \int F^{C_M}(W|X) j(X|i=C_M) dx \\
 FC_A(W) &= \int F^{C_A}(W|X) j(X|i=C_A) dx
 \end{aligned}
 \tag{3}$$

donde:

$F^j(W|X)$ = respuesta del salario ante cambios en las características observables de los trabajadores.

W = logaritmo del salario por hora.

Estas distribuciones son estimables con funciones de densidad Kernel que cumplen con:

$$\begin{aligned}
 \int_{-\infty}^{\infty} K(z) dz &= 1. \\
 FC_M(W) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{h} K \left(\frac{W-w_i}{h} \right) \\
 FC_A(W) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{h} K \left(\frac{W-w_i}{h} \right)
 \end{aligned}
 \tag{4}$$

donde:

h = ancho de banda de Kernel.

N = número de observaciones.

$FC_M(W)$

y $FC_A(W)$ = corresponden a la contrafactual, la cual se definiría como la distribución salarial de los trabajadores por tipo de calificación en la manufactura, considerando sus características, pero re-

muneradas de acuerdo con la estructura salarial de los trabajadores, en su respectiva calificación laboral, pero en la industria automotriz:

$$\begin{aligned}
 FC_M(W) \text{ y } FC_A(W) &= \int F^{C_M}(W|X) j(X|i=C_A) dx. \\
 FC_M(W) \text{ y } FC_A(W) &= \int F^{C_M}(W|X) \psi(X) j(X|i=C_M) dx
 \end{aligned}
 \tag{5}$$

donde la función de reponderación $\psi(X)$ (densidad contrafactual) puede ser definida como:

$$\begin{aligned}
 \psi(X) &= \frac{j(X|i=C_A)}{j(X|i=C_M)} \\
 \psi(X) &= \frac{j(i=C_M|X)}{j(i=C_A|X)} \cdot \frac{j(i=C_A)}{j(i=C_M)}
 \end{aligned}
 \tag{6}$$

La estimación de $\psi(X)$ se lleva a cabo mediante un modelo logarítmico, donde la variable dependiente es el logaritmo del salario de los trabajadores por tipo de calificación laboral y las explicativas, las características observables (X).

La función reponderada estimada es utilizada para estimar la densidad contrafactual por medio de la función Kernel reponderada:

$$(W) = \frac{1}{N} \sum_{i=1}^N \frac{\psi(X)}{h} K \left(\frac{W-w_i}{h} \right)
 \tag{7}$$

En el siguiente apartado se analizan algunos valores promedio de las variables que caracterizan a los trabajadores en el mercado laboral de la manufactura y la industria automotriz para, después, llevar a cabo el ejercicio empírico derivado al aplicar las curvas Kernel y el análisis contrafactual con la metodología DFL.

Resultados

Estadística descriptiva

Con la finalidad de conocer las características de empleo y salarios de la ZMS, se presenta una des-

¹⁰ C_M son los trabajadores de la industria manufacturera y C_A los de la automotriz.

cripción general de la población que labora en el subsector de producción automotriz y en el resto de la manufactura antes y después de la crisis del 2008 y en el 2018. En especial, es importante saber cuáles son las características del empleo en la IA de la ZMS y cómo se diferencian los trabajadores de esta de los que se encuentran en la manufactura en general. Para ello, el cuadro 1 muestra, en primer término, que la dinámica ocupacional entre la IA y los otros subsectores fue diferente durante el periodo de crisis y su posterior recuperación. La ocupación en la manufactura aumentó, después de la perturbación económica, 9.2 % y del 2009 al 2018 se observa otro aumento de 4.9 %, mientras que, en el subsector competente a esta investigación se vio una caída de 7.4 % del 2005 al 2009; sin embargo, en el 2018, se muestra que los empleos se han recuperado 59.9 por ciento.

Cuando se realizó el análisis dentro de las ramas por sexo, se encontró que había una mayor proporción de hombres en la IA en comparación con la manufactura en general, y estos aumentaron su participación en el 2009, para caer en el 2018. Respecto a los rangos de edad, en la manufactura existían más empleados en el grupo de 36-45 años y en el subsector de producción automotriz la mayor parte se encontraba en el de 26-35; es decir eran, en promedio, 10 años más jóvenes respecto al resto de los subsectores. También, es importante destacar que en los dos ámbitos se

revela una reducción en la participación de ambos grupos de trabajadores en el 2018.

De igual manera, se observa que el salario promedio por hora era mayor en la IA respecto a la manufactura, aunque en los dos se muestra un ligero aumento en el 2009 y que cayó en el 2018. Es importante destacar que en el cuadro 1 se presenta el salario promedio de todos los trabajadores, sin distinguir por nivel de calificación, distinción que sí se hace en el 2, que resulta de interés para cumplir con el objetivo de esta investigación.

Cuando se analiza la ocupación por calificación laboral, se observa que los empleados en la IA contaban con más años de educación que en el resto de la manufactura, aunque había un mayor número de trabajadores no calificados en los dos ámbitos. No obstante, la participación de estos decreció después del 2009, aunque mostraron un ligero aumento en la IA en el 2018. Por su parte, los semicalificados aumentaron su participación, aun con la declaratoria de crisis en el subsector y el resto de las manufacturas, contrario a lo que sucede con los calificados, que disminuyeron en estos dos en el 2018.

A pesar de los cambios en la participación laboral que se muestran en el cuadro 1, es importante conocer el comportamiento de los salarios de los trabajadores por tipo de calificación laboral.

Cuadro 1

Continúa

Características promedio de los trabajadores de la IA y el resto de manufactura en la ZMS, Coahuila de Zaragoza, 2005, 2009 y 2018

Variables	Resto de la manufactura			Industria automotriz		
	2005	2009	2018	2005	2009	2018
Hombre	0.703	0.714	0.681	0.800	0.844	0.779
Años educativos	9.598	10.34	10.46	10.80	11.11	11.08
Edad 14-25	0.273	0.207	0.233	0.285	0.172	0.266
Edad 26-35	0.270	0.293	0.249	0.376	0.437	0.321
Edad 36-45	0.283	0.294	0.275	0.245	0.273	0.231
Edad 46-55	0.124	0.151	0.176	0.067	0.096	0.154
Edad 56-65	0.048	0.053	0.065	0.025	0.019	0.025
Salario por hora	44.27	48.00	40.99	49.61	50.51	48.06
Jornada laboral	45.86	44.43	47.58	47.98	47.34	48.97

Características promedio de los trabajadores de la IA y el resto de manufactura en la ZMS, Coahuila de Zaragoza, 2005, 2009 y 2018

Variables	Resto de la manufactura			Industria automotriz		
	2005	2009	2018	2005	2009	2018
Experiencia	18.95	19.74	19.82	15.48	16.69	17.05
No calificados	0.631	0.566	0.549	0.497	0.433	0.454
Semicalificados	0.263	0.271	0.322	0.308	0.374	0.381
Calificados	0.105	0.162	0.128	0.194	0.191	0.164
Total ponderado	41 451	45 664	48 064	26 852	25 001	62 343

Fuente: elaboración propia con base en datos de la ENOE, años respectivos.

En el cuadro 1 se muestra que el empleo, sin distinguir por tipo de calificación, aumentó del 2005 al 2009 en la manufactura y cayó ligeramente en la IA, sin embargo, del 2009 al 2018 incrementó el empleo para ambos, con mayor impacto en la industria automotriz. Nótese que existía un mayor número de trabajadores calificados en esta antes de la crisis, lo cual se revirtió en el 2018. Sin embargo, vale la pena preguntarnos, ¿qué impactos tuvo la perturbación económica del 2008 en los salarios de los trabajadores del subsector de producción automotriz en la ZMS por tipo de calificación laboral?

El cuadro 2 revela que, del 2005 al 2009, las retribuciones de los empleados calificados disminuyeron en la manufactura y aumentaron en la IA; sin embargo, entre el 2009 y 2018 tendieron a la baja en ambos. Para el personal semicalificado, se observa que los salarios se mantuvieron sin cambios en la manufactura y aumentaron en la industria

automotriz del 2005 al 2009, aunque disminuyeron en el 2018 en los dos. Para los trabajadores no calificados se ve un ligero aumento en la IA y el resto de los subsectores del 2005 al 2009, aunque en el 2018 solo se recuperaron en la industria automotriz.

Resumiendo, la IA mostró una ligera pérdida de empleo del 2005 al 2009 y una recuperación entre el 2009 y 2018. También, se destaca un aumento en la participación de mujeres. Asimismo, aunque en los dos ámbitos existía una mayor proporción de empleo no calificado, hubo una disminución de este tipo de mano de obra en el 2009 que no se recuperó en el 2018, por lo que son ellos los que han sido los más perjudicados por los efectos de la crisis. Para los trabajadores calificados y semicalificados, se observa un aumento en su demanda relativa entre el 2005 y 2009 en la IA y el resto de la manufactura, aunque en el 2018 se encontró una caída de los calificados. A pesar de los cambios en

Cuadro 2

Salario promedio de los trabajadores de la IA y el resto de manufactura en la ZMS, Coahuila de Zaragoza, 2005, 2009 y 2018

Variables	Resto de la manufactura			Industria automotriz		
	2005	2009	2018	2005	2009	2018
No calificados	33.66	36.48	35.45	36.64	34.04	36.70
Semicalificados	49.44	49.75	39.77	45.59	46.77	43.35
Calificados	94.87	85.21	67.71	89.08	95.07	90.33
Total ponderado	41 451	45 664	48 064	26 852	25 001	62 343

Fuente: elaboración propia con base en datos de la ENOE, años respectivos.

el empleo, estos no se vieron aparejados con un aumento en su salario real por hora. Nótese cómo las remuneraciones por hora cayeron en los dos bloques para todos los empleados en el periodo posterior a la crisis.

Con lo anterior, se deduce que, aunque los empleos de personal no calificado fueron los que más se perdieron, el salario de todos los trabajadores se redujo después de la crisis económica, por lo que resulta interesante indagar qué pasó con los pagos de los empleados más allá de los valores medios; en la siguiente sección se analizan los salarios de los trabajadores por tipo de calificación en las dos industrias, a lo largo de la distribución de salarios, antes y después de la perturbación del 2008.

Análisis de la densidad Kernel

Con el objetivo de conocer qué está pasando con los salarios de los trabajadores más allá de los promedios, se realizó un análisis de la densidad salarial a lo largo de la distribución. La gráfica 1 muestra los resultados obtenidos al efectuar la prueba de densidad Kernel sobre la ocupación y salarios por hora de los empleados no calificados en la indus-

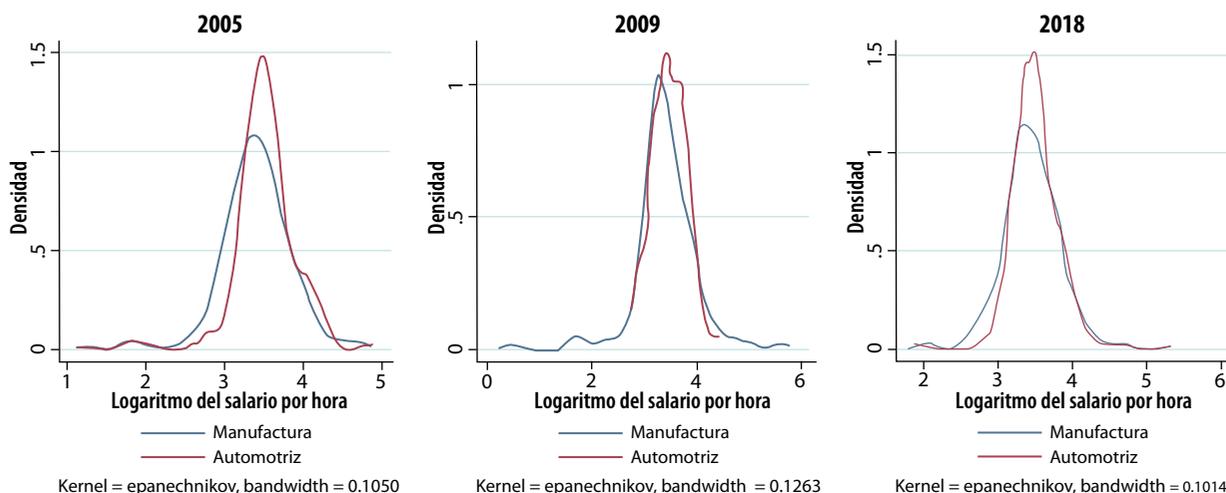
tria automotriz y el resto de la manufactura antes y después de la crisis del 2008. Revelan que, en el 2005, el personal no calificado en la manufactura mostraba una menor remuneración respecto al ubicado en la IA, sobre todo en el logaritmo del salario por hora 4 y 5, lo que muestra la existencia de una concentración de trabajadores no calificados con pago mayor en la IA respecto al resto de la manufactura.

En el 2009 se observa una disminución de la densidad ocupacional en ambas industrias, lo que significa una pérdida de empleo no calificado, en especial en el sector manufacturero, aunque con un desplazamiento de la curva a la derecha en el logaritmo del salario por hora 3.5 y 4, así como una reducción en el 5, lo que podría deberse a una caída en la ocupación de los trabajadores por los efectos de la crisis económica.

Por último, en el 2018 se ve un aumento en la densidad ocupacional en el resto de la manufactura, pero no sucede lo mismo en la IA, donde se observa de forma clara que los empleos que se perdieron con los efectos de la crisis se recuperaron en el 2018 y que había una mayor concentración de trabajadores con salarios más altos.

Gráfica 1

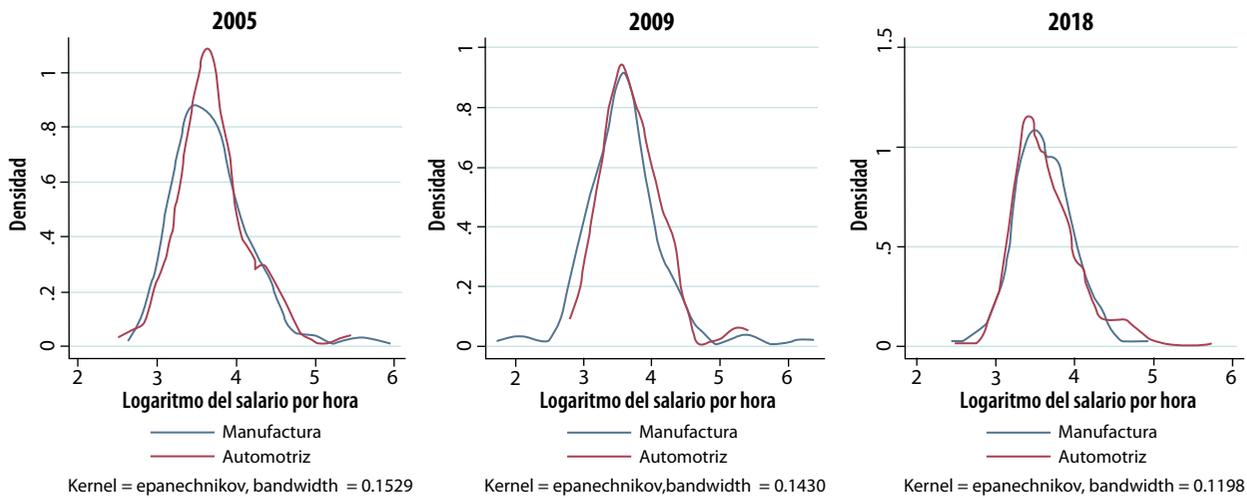
Funciones de densidad salarial por hora de los trabajadores no calificados en la manufactura y en la IA en la ZMS, 2005, 2009 y 2018



Fuente: elaboración propia con base en la ENOE, años respectivos.

Gráfica 2

Funciones de densidad salarial por hora de los trabajadores semicalificados en la manufactura y en la IA en la ZMS, 2005, 2009 y 2018



Fuente: elaboración propia con base en la ENOE, años respectivos.

La gráfica 2 muestra a quienes se desempeñan bajo el esquema de semicalificados en el resto de la manufactura y en la industria automotriz. Los resultados revelan que en el 2005 existía una mayor densidad ocupacional con salarios más altos para los trabajadores en el subsector de producción automotriz, dado que se observa un pico más alto en la curva, con un leve desplazamiento a la derecha respecto a los ubicados en el resto de la manufactura. Por lo tanto, había una mayor ocupación y salarios más altos en la fabricación de equipo de transporte para este tipo de trabajadores.

Con los efectos de la crisis que se vivieron en el 2009, se revela igualmente una reducción en el empleo y desplazamiento de la densidad ocupacional a la derecha de los trabajadores ubicados en la IA respecto a los ocupados en el resto de la manufactura, donde los más afectados por la perturbación económica fueron los que tenían menor salario y que se localizaban en la industria automotriz.

Sin embargo, en el 2018 se observan menores diferencias en las funciones de densidad para ambas industrias, pero con una menor densidad ocupacional en la manufacturera, incluso, con leves des-

plazamientos de la curva a la derecha e izquierda, es decir, en salarios más altos y bajos, por lo cual se puede deducir que hubo una disminución de la desigualdad de los trabajadores semicalificados ubicados en una y otra después de la crisis.

La gráfica 3 muestra las funciones de densidad de los trabajadores calificados en la industria automotriz y en el resto de la manufactura. Los resultados revelan que, en el 2005, la de mayor magnitud la presentó el personal calificado de la IA dentro del grupo de los que más ganaban, excepto en el logaritmo del salario 6 y 7, donde se ubicaba una proporción menor respecto a los ocupados en el resto de la manufactura.

En el 2009, la densidad de la distribución correspondiente a los trabajadores calificados ubicados en el resto de la manufactura se vuelve más plana respecto a los ocupados en la industria automotriz. Asimismo, se observa un desplazamiento a la derecha de ambas curvas, lo cual revela que los salarios de los que más ganaban se volvieron más homogéneos, por lo que se puede deducir que se redujo la desigualdad dentro del grupo de salarios altos.

Para los empleados ubicados en la IA, en el 2018 se observa una disminución de la densidad ocupacional, con una curva más plana y desplazamiento de la misma hacia la derecha. Por lo tanto, los datos revelan que hubo una pérdida en el empleo de los trabajadores calificados, aunque se ve un aumento de la densidad en el tramo del logaritmo del salario 4 y 5 y no en salarios más altos.

Hasta esta fase del análisis hemos encontrado que los efectos de la crisis son distintos por tipo de empleo según su nivel de calificación. En el 2005 se observa una concentración de los salarios en la parte media de la distribución en el caso de la IA respecto a la manufactura, lo cual significa que, en ese año, la ocupación por tipo de calificación en el subsector automotor estaba concentrada en salarios medios y no en altos. No obstante, en el 2009, las densidades se vuelven más planas, pero con un desplazamiento de la curva ligeramente hacia la izquierda, lo que denota una reducción del número de empleos, pero manteniendo salarios relativamente altos, sobre todo en la manufactura y, en especial, para los trabajadores no calificados y semicalificados. Para el 2018, se observa que las

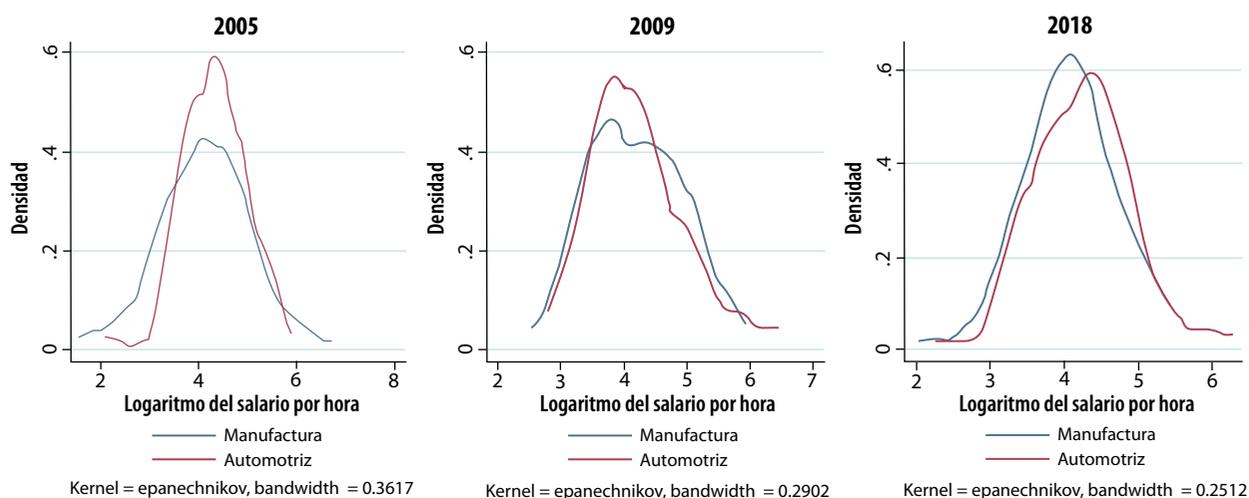
curvas salariales se desplazan hacia la izquierda para estos, lo cual denota que la concentración de empleados se encontraba en salarios más bajos; asimismo, en la cola derecha de la distribución parece haber un ligero aumento de empleados en la manufactura con salarios altos. Por su parte, los calificados reflejan una pérdida en el empleo en el subsector de producción automotriz y una concentración solo en la parte media de la distribución.

Los resultados anteriores revelan que los trabajadores ubicados en la IA tenían mejores salarios respecto a los que se empleaban en el resto de la manufactura en el 2005, aunque esta situación cambió en el 2009 y 2018. Se observan impactos heterogéneos por tipo de calificación laboral a lo largo de la distribución salarial.

En la siguiente sección se estiman las diferencias salariales a través del cálculo contrafactual con la metodología DFL con el objetivo de tener una medida más exacta de los efectos de la crisis económica en el empleo por tipo de calificación laboral.

Gráfica 3

Funciones de densidad salarial por hora de los trabajadores calificados en la manufactura y en la IA en la ZMS, 2005, 2009 y 2018



Fuente: elaboración propia con base en la ENOE, años respectivos.

Estimación de las diferencias salariales con el modelo DFL

La evidencia empírica revela que la economía de México (en especial en las regiones que se encuentran más integradas con el mundo, comercialmente hablando) ha sufrido de manera drástica los impactos de las crisis económicas mundiales. Esto se atribuye, por un lado, a su fuerte integración al mercado estadounidense y, por el otro, a un aumento de la oferta relativa del trabajo calificado, lo que influye en una disminución de la desigualdad salarial entre trabajadores calificados y no calificados. A nivel sectorial, la ocupación y los salarios parecen ser muy heterogéneos, por lo que interesa conocer, en especial para el sureste de Coahuila de Zaragoza (donde existe una fuerte integración con el sector externo y los salarios en el subsector de producción automotriz son más altos respecto a la industria manufacturera), qué pasaría con los pagos de los empleados ubicados en el resto de la manufactura, dadas sus características, si fueran remunerados como si se emplearan en la IA, esto en el contexto de la perturbación económica del 2008. Este ejercicio nos ayudará a entender los efectos de la crisis del 2008 en un subsector muy integrado con la economía internacional, como el automotriz de la ZMS, y compararlo con los efectos en un sector que tiene más lazos con el mercado interno. Para realizar este análisis, se emplearon las estimaciones de diferencias salariales con el modelo DFL.

A continuación, se muestran las estimaciones de densidades empíricas de las remuneraciones fac-

tuales (observadas) y contrafactuales (estimadas) de los trabajadores no calificados antes y después de la crisis del 2008.

Primero, en el cuadro 3 se presentan los coeficientes y la significancia estadística de las funciones de densidad salarial; en la gráfica 4 se observa el comparativo entre las densidades salariales factual y contrafactual de los trabajadores no calificados ubicados en la manufactura y el subsector automotor. Los resultados indican que, en el 2005, la densidad contrafactual se encuentra desplazada a la derecha, después del logaritmo del salario por hora 4. En el 2009 y 2018, la curva se mueve hacia la izquierda, ya que el cambio favorece a los empleados no calificados en el resto de la manufactura a partir del logaritmo 2; por lo tanto, estos resultados estarían indicando que las condiciones salariales han sido mejores en la manufactura, sobre todo, en el 2009 y 2018. Al revisar los cambios de ambas densidades, se revela que en el 2005 y 2009, si a los trabajadores de la manufactura que se ubican en la parte media de la distribución se les pagara como a los automotrices, estos ganarían más. Esta relación cambió en el 2018, ya que los empleados ubicados en los extremos de la distribución tenían un mayor beneficio.

En el cuadro 3 se presentan los coeficientes de las variables independientes utilizadas en modelo *logit* para el *propensity score*. Los modelos permiten conocer la calidad de la técnica de emparejamiento, la cual muestra que estos son significativos a 99 % de confianza.

Cuadro 3

Coefficientes y significancia estadística de las funciones de densidad salarial por hora de los trabajadores no calificados

VARIABLES	2005	2009	2018
Años educativos	0.0654***	-0.0078532***	0.0301517***
Experiencia	0.0724***	0.1331248***	0.021241***
Experiencia2	-0.00194***	-0.0030977***	-0.000802***
Constante	-1.57***	-1.788693***	-0.0977321***

*** p < 0.01.

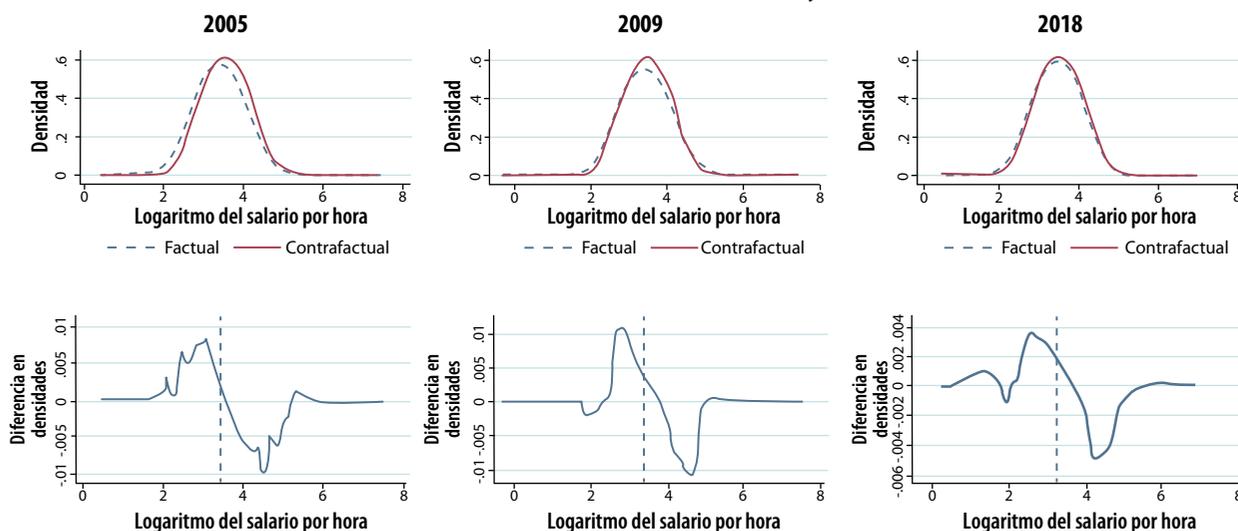
** p < 0.05.

* p < 0.1.

Fuente: elaboración propia con base en la ENOE, años respectivos.

Gráfica 4

Funciones y cambio factual y contrafactual de los salarios por hora de trabajadores no calificados en la manufactura en la ZMS, 2005, 2009 y 2018



Nota: la línea vertical que cruza la gráfica de cambios corresponde a la mediana de la distribución factual.

Fuente: elaboración propia con base en la ENOE, años respectivos.

El cuadro 4 muestra los coeficientes y la significancia estadística de la densidad salarial; en la gráfica 5 se presenta qué hubiera pasado con los salarios de los trabajadores semicalificados, antes y después de la crisis del 2008, ubicados en la manufactura, pero pagados como si se encontrasen en la IA, de acuerdo con los resultados obtenidos por el método DFL.

A partir del comparativo entre las dos densidades salariales se puede establecer que existe un efecto significativo después de la crisis, ya que en el 2005 y 2009 se observa un desplazamiento de la curva contrafactual a la derecha de la distribu-

ción, lo cual implica que los trabajadores semicalificados se encontrarían mejor remunerados si estuvieran en la IA. Este escenario no se mantiene en el 2018, ya que no parece haber mayor diferencia salarial entre el personal ubicado en la industria manufacturera respecto al subsector de producción automotriz, aunque después del logaritmo 4, se observa un desplazamiento hacia arriba de la curva contrafactual en el grupo de los salarios más altos, lo que implica una mejoría para los que más ganan.

Cuando se analizan los cambios de las densidades factual y contrafactual, se revela que en el 2005

Cuadro 4

Coefficientes y significancia estadística de las funciones de densidad salarial por hora de los trabajadores semicalificados

Variables	2005	2009	2018
Años educativos	0.1053658***	-0.0435834***	0.2305331***
Experiencia	0.0810911***	0.0613197***	0.0410392***
Experiencia ²	-0.0036247***	-0.0020858***	-0.0013016***
Constante	-1.628294***	0.0241896***	-2.392885***

*** p < 0.01.

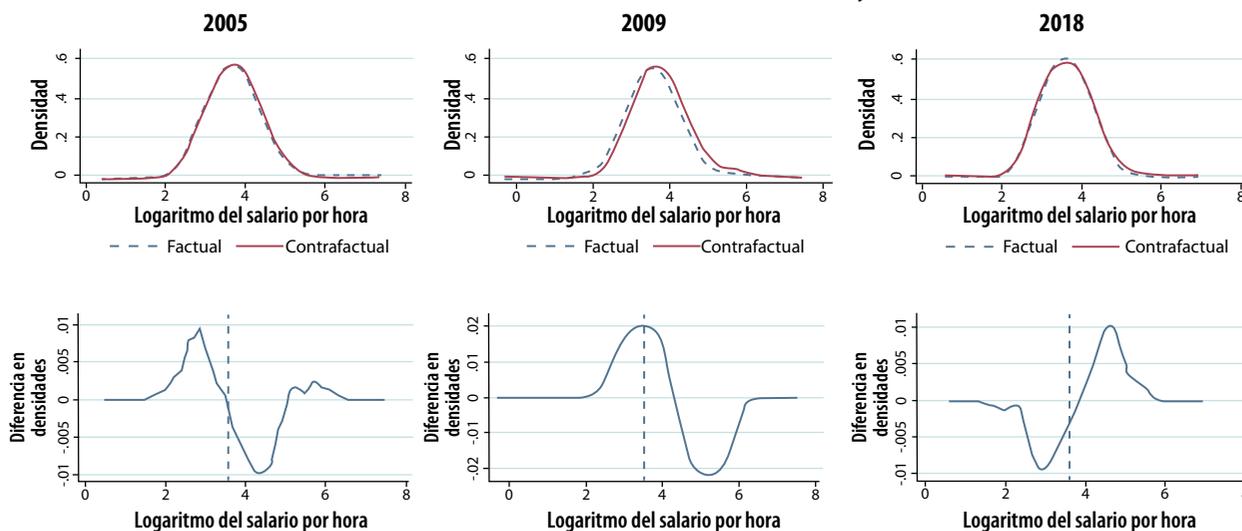
** p < 0.05.

* p < 0.1.

Fuente: elaboración propia con base en la ENOE, años respectivos.

Gráfica 5

Funciones y cambio factual y contrafactual de los salarios por hora de trabajadores semicalificados en la manufactura en la ZMS, 2005, 2009 y 2018



Nota: la línea vertical que cruza la gráfica de cambios corresponde a la mediana de la distribución factual.

Fuente: elaboración propia con base en la ENOE, años respectivos.

y 2009, si a los trabajadores en la manufactura que se ubican en la parte derecha de la distribución se les pagara como a los automotrices, estos ganarían más. Esta relación cambia en el 2018. En este grupo de empleados se observan desplazamientos de la curva contrafactual a la derecha en el 2005 y 2009, y a la izquierda en el 2018. Lo anterior implica que los semicalificados se encontrarían mejor remunerados si se ubicaran en la IA en el 2009 dentro del grupo de los salarios altos, aunque después de la crisis las remuneraciones tienden a mejorar en la parte media de la distribución, lo cual podría significar que, para este tipo

de trabajadores, la crisis no trajo consecuencias tan drásticas como en el caso de los no calificados.

En el cuadro 5 se revelan los coeficientes y la significancia estadística de las funciones de densidad y la gráfica 6 presenta el resultado del escenario hipotético sobre la variable salarios por hora, manteniendo constantes las características de capital humano; en este caso, se muestra cómo hubiese sido el salario de los trabajadores calificados en la manufactura, antes y después de la crisis del 2008, pero remunerados como si se encontraran en la industria automotriz.

Cuadro 5

Coeficientes y significancia estadística de las funciones de densidad salarial por hora de los trabajadores calificados

Variables	2005	2009	2018
Años educativos	0.0053388***	0.2460713***	0.0649439***
Experiencia	0.0088698***	-0.0642593***	0.0977589***
Experiencia2	-0.0010577***	0.0007659***	-0.0035858***
Constante	0.1499945***	-3.789712***	-0.9377789***

*** p < 0.01.

** p < 0.05.

* p < 0.1.

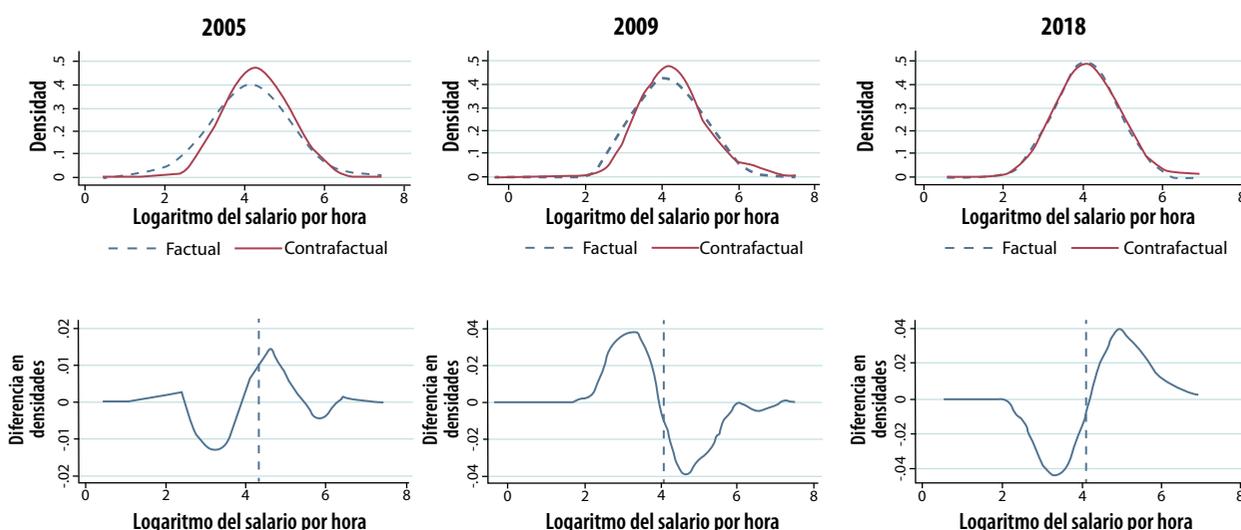
Fuente: elaboración propia con base en la ENOE, años respectivos.

Se observa que en el 2005 hay un desplazamiento de la curva contrafactual a la derecha de la densidad salarial, lo cual implica que los trabajadores ubicados en el resto de la manufactura mejorarían de forma notable sus salarios si se encontraran en la IA. No obstante, en el 2009, las curvas se vuelven más homogéneas, aunque en este caso el desplazamiento se observa en la parte media de la densidad. Asimismo, en el 2018 se ve un movimiento hacia la izquierda y en la parte media. Al revisar los cambios de las densidades factual y contrafactual, se revela que en el 2005 y 2009, si a los trabajadores en la manufactura que se ubican en la parte media de la distribución se les pagara como a los automotrices, estos ganarían más. Esta relación cambia en el 2018, ya que los ubicados en la parte media e izquierda de la distribución tienen un mayor beneficio.

Con los resultados de las distribuciones factuales y contrafactuales para los trabajadores por tipo de calificación en la manufactura, pero remunerados como los ocupados en la IA antes y después de la crisis del 2008, mostrados en las gráficas 4-6, se pueden apreciar los siguientes aspectos relevantes:

1) para todos los empleados, sin importar su nivel de calificación, el encontrarse en la IA en el 2005 y 2009 les permitía tener una mejor remuneración, en particular los ubicados en la parte baja de salarios, lo cual indica que el rendimiento por las características de capital humano no son homogéneas a lo largo de la distribución; 2) cuando se realiza el análisis de las funciones de densidades factual y contrafactual, así como sus cambios, se observa que el personal en la manufactura en la parte de salarios bajos y medios estaría mejor remunerado si se encontrara en la industria automotriz aunque, en el 2018, el beneficio cambia por tipo de calificación laboral; 3) en los tres tipos de calificación laboral, la curva contrafactual se encuentra desplazada hacia la derecha en el 2005 y 2009, lo cual significa que si los trabajadores estuvieran en el subsector automotor recibirían mejores salarios. Sin embargo, el desplazamiento del contrafactual se da en sentido inverso en el 2018; de hecho, en empleos semicalificados y calificados, la curva se desplaza hacia la izquierda y arriba, lo que implica que, en ese año, los salarios en la parte media de la distribución en el resto de la manufactura eran mejores que los que se encontraban en la industria automotriz.

Gráfica 6
Funciones y cambio factual y contrafactual de los salarios por hora de trabajadores calificados en la manufactura en la ZMS, 2005, 2009 y 2018



Nota: la línea vertical que cruza la gráfica de cambios corresponde a la mediana de la distribución factual.

Fuente: elaboración propia con base en la ENOE, años respectivos.

La literatura empírica revisada indica que los efectos de la crisis en el mercado laboral mexicano se han reflejado en una mayor pérdida de empleo y una caída en los salarios reales de los trabajadores (ver Esquivel y Rodríguez, 2003 y Campos-Vázquez, Esquivel y Lustig, 2014). Cuando se analizan los salarios y la ocupación por regiones, la evidencia ha mostrado que los impactos de la crisis son más evidentes en la frontera norte del país, dado que en esta zona existe una mayor exposición a la economía mundial, tal como lo mencionan Rodríguez *et al.* (2017). Los resultados de estos autores coinciden en mencionar que la desigualdad salarial ha caído a lo largo del tiempo y se han incrementado los empleos informales.

Conclusiones

El objetivo del documento es analizar el efecto que la perturbación económica del 2008 tuvo sobre el empleo y los salarios en actividades integradas a las cadenas globales de valor, como lo es la industria automotriz; para ello, se estudió el caso de la ZMS, donde esta actividad tiene una importante presencia.

Los resultados indican que las características de los trabajadores de la IA presentan diferencias respecto al resto de la manufactura, las cuales se expresan en una mayor preponderancia de trabajadores de sexo masculino, concentración en grupos de edad más jóvenes, así como mayor presencia de trabajadores semicalificados y calificados.

Durante la crisis del 2008, la industria automotriz presentó mayor contracción del empleo en relación con el resto de las manufacturas, aunque también esta actividad mostró mayor dinamismo en el periodo posterior a la perturbación económica, lo cual aporta evidencias en favor de la hipótesis de trabajo y confirma la mayor volatilidad de los sectores vinculados a las redes globales de valor.

En lo relativo a los salarios, las remuneraciones promedio fueron mayores en la IA en relación con

el resto de las manufacturas, brecha que se redujo durante la crisis, hecho que se explica especialmente asociado a la reducción en los trabajadores no calificados; sin embargo, posterior a la crisis, las diferencias se ampliaron, en este caso, vinculadas con una mayor contracción de los salarios en la manufactura no automotriz.

El análisis de densidades permitió identificar que, para los años previos y durante la perturbación económica, los trabajadores del resto de las manufacturas se encontrarían en mejores condiciones salariales si fueran retribuidos como en la IA; sin embargo, esto no se cumplió para el 2018.

A partir de lo anterior, se puede establecer que la crisis económica tuvo implicaciones relevantes en el mercado laboral de la ZMS, impactando coyunturalmente sobre los niveles de empleo, en particular en la IA que presentaba mayor integración a las cadenas globales de valor que el resto de la manufactura y, de manera permanente, sobre los salarios que no obstante que el último año de estudio está a 10 años de la perturbación económica, estos aún no recuperan los niveles salariales registrados en ese momento.

El análisis llevado a cabo para la ZMS permitió identificar que las perturbaciones económicas externas ejercen efectos diferenciados sobre el empleo y los salarios de los sectores económicos regionales que presentan distintos grados de integración con el exterior, hecho que tiene implicaciones sobre las condiciones de estabilidad y crecimiento regional.

Ejercicios como este pueden ser relevantes en otras regiones del país que permitan identificar la presencia de elementos similares, así como capturar particularidades propias de ellas.

Finalmente, es pertinente señalar que, no obstante que este documento aporta elementos que permiten una mejor comprensión de la dinámica regional y su comportamiento frente a perturbaciones económicas, es necesario realizar más investigaciones sobre el tema.

Fuentes

- Aboites, Gilberto y David Castro. "Nuevos obreros y viejas tradiciones en la Zona Metropolitana de Saltillo. Coahuila: la revolución y su gente. Coahuila (1910-2010)", en: Cerutti, Mario y Javier Villareal Lozano (coords.). *Economía, historia económica y empresa*. 2010, pp. 132-168.
- Acemoglu, D. "Patterns of Skill Premia", in: *The Review of Economic Studies*. 70 (2), April 2003, pp. 199-230.
- Aguirre Villaseñor, Luis; Gloria Tobón de Garza y René Mendoza Alfaro. "Dinámica de las regiones de Coahuila: entre las fuerzas de la cuarta revolución industrial, el asalto a sus recursos naturales y las luchas por el espacio", en: *Ciudad, Género, Cultura y Educación en las Regiones*. México, Universidad Nacional Autónoma de México-Asociación Mexicana de Ciencias para el Desarrollo Regional A. C., 2018. ISBN UNAM: 978-607-30-0002-4, AMECIDER: 978-607-96649-9-2.
- Airola, J. and C. Juhn. "Wage Inequality in Post-Reform Mexico", in: *IZA Discussion Papers 1525*. Institute for the Study of Labor (IZA), 2005.
- Andalon, Mabel y Luis López-Calva. *The Many Mexicos: Income inequality and polarization in urban Mexico during the 90. Paper prepared for the Cornell-LSE-Wider Conference on Spatial Inequality*. London, June 28-30 2002.
- Bosch, M. and M. Manacorda. "Minimum Wages and Earnings Inequality in Urban Mexico", in: *American Economic Journal: Applied Microeconomics*. 2(4), 2010, pp. 128-149.
- Bound, John and G. Johnson. "Changes in the structure of wages during the 80s: An evaluation of alternative explanations", in: *American Economic Review*. 82, 1992, pp. 371-392.
- Campos-Vázquez, R., G. Esquivel y N. Lustig. "The rise and fall of income inequality in México, 1989-2010", in: *Falling Inequality in Latin America: Policy Changes and Lessons*. Coria, Giovanni Andrea (ed.). Oxford University Press, UNU-Wider Studies in Development Economics, 2014.
- Campos-Vázquez. "Why Did Wage Inequality Decrease in Mexico after NAFTA?", en: *Economía Mexicana. Nueva Época*. Vol. XXII, núm. 2. Centro de Investigación y Docencia Económicas, A. C., 2013, pp. 245-278.
- Castro Lugo, David. "Disparidad salarial urbana en México, 1992-2002", en: *Estudios Sociales (Hermosillo, Son.)*. 15(29), 2007, pp. 117-153.
- Castro Lugo, David, & Luis Huesca Reynoso. "Desigualdad salarial en México: una revisión", en: *Papeles de Población*. 13(54), 2007, pp. 225-264.
- Castro Lugo, David & Gustavo Félix Verduzco. "Apertura comercial, relocalización espacial y salario regional en México", en: *Estudios Fronterizos*, 11(21), 2010, pp. 43-79.
- Castro Lugo, David; Gilberto Aboites Manrique y Reyna Rodríguez Pérez. "Cambio tecnológico en la industria automotriz de Saltillo: una exploración", en: *Innovación y desarrollo regional en México: resultados y avances*. Plaza y Valdez, 2013, pp. 119-144. ISBN:978-607-506-167-2.
- Chiquiar, D. *Globalization, regional wage differentials and the Stolper-Samuelson Theorem: Evidence from Mexico*. Documentos de trabajo. Banco de México, 2004, pp. 2004-2006.
- Cragg, M. and M. Eppelbaum. *The premium for skills in LDCs: Evidence from Mexico*. Mimeo. Columbia University, School of International and Public Affairs, November 1994.
- Dávila Flores, Alejandro. "La economía de Coahuila en el gozne de dos siglos: de la atalaya al ágora. Coahuila: la revolución y su gente. Coahuila (1910-2010)", en: Cerutti, Mario y Javier Villareal Lozano (coords.). *Economía, historia económica y empresa*. 2010, pp. 13-63.
- De Bell, Leendert. "Inversión extranjera directa y desarrollo local. El estado de Coahuila bajo el TLCAN. IIED-América Latina", en: *Medio Ambiente y Urbanización*. Volumen 76, núm. 1, 2012, pp. 109-132.
- DiNardo, J., N. Fortin and T. Lemieux. "Labor market institutions and the distributions of wages, 1973-1992: A semiparametric Approach", in: *Econometrical*. Vol. 65, no. 5, 1996, pp. 1001-1044. DOI: 10.2307/2171954.
- Esquivel, Gerardo y José Rodríguez. "Technology, trade, and wage inequality in Mexico before and after NAFTA", in: *Journal of Development Economics*. 72(2), December 2003, pp. 543-565.
- Félix, Gustavo y David Castro. "Demanda laboral en las industrias integradas en cadenas de valor internacional: un análisis con panel dinámico no estacionario para el caso de México", en: *International Conference of Regional Science: Innovation and geographical spillovers: new approaches and evidence*. Reus, España, Universitat Rovira i Virgili, 2015.
- Félix Verduzco, Gustavo, David Castro Lugo & Gilberto Aboites Manrique. "Nuevas barreras comerciales de EU a las importaciones automotrices y posibles efectos en el empleo regional", en: *Economía: Teoría y Práctica*. (spe4), 2018, pp. 125-175.
- Figuroa, Jorge; Tomás Arroyo & Aneliss Aragón. "Relocalización de la industria manufacturera en México en la apertura comercial, 1980-2014", en: *Estudios Regionales en Economía, Población y Desarrollo*. Núm. 44. Cuadernos de trabajo de la Universidad Autónoma de Ciudad Juárez, 2018.
- Gachúz, Juan Carlos. "La crisis mundial en el sector automotriz, China: ¿jaliado estratégico de México?", en: *México: Análisis Económico*. Vol. XXVI, núm. 63, 2011, pp. 105-128.
- Garza, A. B. E y L. Quintana. "Determinantes de la desigualdad salarial en las regiones de México: 2005-2010. Una visión alternativa a la teoría del capital humano", en: *Paradigma Económico*. Año 6, núm. 1, mayo-junio del 2014, pp. 33-48.
- Guevara, S. *Reactivación de la movilización obrera en la industria terminal automotriz (2004-2011). Recuperación parcial del salario con persistencia en la flexibilización laboral*. Argentina, Núcleo Básico, 2012.
- Instituto Nacional de Estadística y Geografía (INEGI). *Encuesta Nacional de Ocupación y Empleo (ENOE)*. México, INEGI, ediciones 2005, 2009 y 2017.
- Jones, Ronald y Henryk Kierzkowski. "The role of services in production and international trade: A theoretical framework", in: Jones, R. W. y A. Krueger (eds.). *The political economy of international trade: Essays in honor of Robert E. Baldwin*. Oxford, Basil Blackwell, 1990.

- _____. "A framework for fragmentation", in: Arndt, S. y H. Kierzkowski (eds.). *Fragmentation: New Production Patterns in the World Economy*. Oxford University Press, 2000.
- _____. "International fragmentation and the new economic Geography", in: *North American Journal of Economics and Finance*. 16, 2005, pp. 1-10.
- Juárez, Gloria de la Luz, Alfredo Sánchez Daza & Jesús Zurita González. "La crisis financiera internacional de 2008 y algunos de sus efectos económicos sobre México", en: *Contaduría y Administración*. 60(Supl. 2), 2015, pp. 128-146.
- Levy, Santiago y Luis Felipe López-Calva. "Labor Earnings, Misallocation, and the Returns to Education in Mexico", in: *IDB Working Paper Series*. No. IDB-WP-671. Washington, D. C., Inter-American Development Bank (IDB), 2016.
- López Peláez, J. y E. Arrazola. *Crisis global y su impacto en el mercado de trabajo de la frontera norte y sur de México*. 2010.
- Machado, J. A. F. and J. Mata. "Counterfactual decomposition of changes in wage distributions using quantile regression", in: *Journal of Applied Econometrics*. 20, 2005, pp. 445-465.
- Medina Luna, C. *El impacto de la crisis financiera global en el sector automotriz de la región sureste de Coahuila como oportunidad de desarrollo en el futuro*. Tesis de maestría. Instituto Tecnológico y de Estudios Superiores de Monterrey, 2009.
- Mendoza, J. *La crisis de la industria automotriz en México en el marco de la integración económica con Estados Unidos*. México, COLEF, 2011.
- Meza, L. "Apertura comercial y cambio tecnológico. Efectos del mercado laboral mexicano", en: *El Trimestre Económico*. Vol. LXX, núm. 279, 2003, pp. 457-505.
- _____. "Cambios en la estructura salarial de México en el periodo 1988-1993 y el aumento en el rendimiento de la educación superior", en: *El Trimestre Económico*. Vol. 66, núm. 262, 1999, pp. 189-226.
- _____. "Mercados laborales locales y desigualdad salarial en México", en: *El Trimestre Económico*. 72(1), 2005, pp. 133-178.
- Quiroz Trejo, José Othon. "La crisis de la industria automotriz en México: ¿paradigma o caso aislado?", en: *El Cotidiano*. Núm. 158, 2009, pp. 115-123.
- Ríos-Almodóvar, Jesús Gerardo y Salvador Carrillo-Regalado. "El empleo calificado y no calificado en la manufactura de México ante la crisis de 2009", en: *Economía, Sociedad y Territorio*. Vol. XIV, núm. 46, 2014, pp. 687-714.
- Robbins, D. *The impact of trade liberalization upon inequality in developing countries: a review of theory and evidence*. ILO working paper No. 13. Geneva, 2003.
- Rodríguez, Reyna y David Castro. "Discriminación salarial de la mujer en el mercado laboral de México y sus regiones", en: *Economía Sociedad y Territorio*. Vol. XIV, núm. 46, 2014, pp. 655-686.
- Rodríguez Pérez, Reyna, Vicente Germán Soto y Joel Cuatianquis. "Desigualdad salarial en el sector manufacturero de México y sus regiones, según su exposición a la apertura comercial, 2005. 2015", en: *Equilibrio Económico*. Vol. 13, núm. 43, 2017, pp. 49-76.
- Rodríguez Pérez, Reyna Elizabeth y Christian Joel González Cuatianquis. "Skilled-Unskilled Workers Wage Gap in Mexico: a Regional Approach of the Exposure to trade Openness", in: *Journal of Economics and Development Studies*. Vol. 3, no. 4, December 2015, pp. 41-50.
- Rodríguez, E. y B. López. "Imputación de ingresos laborales: una aplicación con encuestas de empleo en México", en: *El Trimestre Económico*. Vol. LXXXII (1), núm. 325, 2015, pp. 117-146.
- Salazar Luna, Karina Selene. "Clúster automotriz en Coahuila", en: *Revista de Tecnologías de la Información*. Vol. 2, núm. 5, 2015, pp. 229-236.
- Salgado García, Lorenzo. "El mercado de trabajo y la crisis en México. Una mirada desde el desarrollo e impactos locales", en: *Economía y Sociedad*. Vol. 14, núm. 28, 2011, pp. 35-50.

Colaboran en este número

Raúl Octavio Martínez Rincón

De nacionalidad mexicana. Es biólogo general por la Universidad de Ciencias y Artes de Chiapas, cuenta con la Maestría en Manejo de Recursos Marinos y el Doctorado en Ciencias Marinas por el Centro Interdisciplinario de Ciencias Marinas. Tiene experiencia en modelos de distribución de especies, percepción remota y fotogrametría. En la actualidad, labora en el Centro de Investigaciones Biológicas del Noroeste (CIBNOR), SC como catedrático CONACYT.

Contacto: rrincon@cibnor.mx

César Augusto Salinas-Zavala

De nacionalidad mexicana. Es biólogo marino por la Universidad Autónoma de Baja California Sur y maestro en Ciencias en Biología de Sistemas y Recursos Acuáticos, así como doctor en Ecología Marina por el Centro de Investigación Científica y de Educación Superior de Ensenada. Cuenta con experiencia en análisis de sistemas socioambientales marinos y terrestres del noroeste de México. Actualmente, trabaja en el CIBNOR, SC, donde se desempeña como investigador titular C.

Contacto: csalinas@cibnor.mx

María Verónica Morales-Zárate

De nacionalidad mexicana. Tiene el Doctorado en Ciencias en el Uso, Manejo y Preservación de los Recursos Naturales por el CIBNOR, SC. Cuenta con experiencia en el campo de ordenación territorial, modelación dinámica y análisis de sistemas socioambientales. En la actualidad, labora en el CIBNOR, SC con el nivel de técnico titular D.

Contacto: mzarate04@cibnor.mx

Federico Salinas-Zavala

De nacionalidad mexicana. Es ingeniero en Agronomía por la Universidad Autónoma de Sinaloa y maestro en Ciencias en el Uso, Manejo y Preservación de los Recursos Naturales por el CIBNOR, SC. Cuenta con experiencia en hidrología, caracterización de flora, paisaje, sistemas de información geográfica e impacto ambiental. Actualmente, ocupa un puesto de técnico titular C en el CIBNOR, SC.

Contacto: fsalinas04@cibnor.mx

Miguel Muñoz

De nacionalidad mexicana. Es ingeniero en Recursos Naturales por la Universidad Autónoma Chapingo con la Maestría en Ciencias Forestales por el Colegio de Postgraduados. Su experiencia principal se basa en el uso de herramientas de sensores remotos para el monitoreo de la cobertura terrestre. En la actualidad, a través del Servicio Forestal de los Estados Unidos (USFS, por sus siglas en inglés), colabora con la Unidad Técnica de Monitoreo Reporte y Verificación (UTEMRV) de la Comisión Nacional Forestal (CONAFOR) en el análisis de los procesos y determinantes de la deforestación a nivel nacional.

Contacto: miguelmrv.munoz@gmail.com

Miguel Castillo

De nacionalidad mexicana. Es ingeniero Forestal por la Universidad de Chapingo; estudió la Maestría en Estadística en el Colegio de Postgraduados y obtuvo el Doctorado en Ciencias en el Instituto de Biología de la Universidad Nacional Autónoma de México (UNAM). Sus principales líneas de investigación incluyen el análisis de la estructura de los bosques usando datos satelitales, la modelación del cambio de uso del suelo y el análisis de las causas de la deforestación. Desde hace 20 años trabaja en el Departamento de Observación y Estudio de la Tierra, la Atmósfera y el Océano, antes Laboratorio de Análisis Información Geográfica y Estadística, del Colegio de la Frontera Sur.

Contacto: mcastill@ecosur.mx

Felicia Line

De nacionalidad británica. Es licenciada en Geografía por la Universidad de Bristol en el Reino Unido y maestra en Tecnología Ambiental por el *Imperial College* con especialidad en Cambio Global Ambiental y Políticas Públicas. Tiene más de 10 años de experiencia de trabajo en México y fue coordinadora del proyecto *Forests 2020* en México y Colombia. En la actualidad, es consultora de Ecosistemas Consultoría y coordinadora del proyecto de UK PACT en torno al fortalecimiento de uso de plataformas de información geográfica en tres estados de México.

Contactos: felicia@ecosistemasconsultora.org y flicline@gmail.com

Edith Mondragón

De nacionalidad mexicana. Es egresada de la Licenciatura en Ciencias de la Tierra en la Universidad de Ciencias y Artes de Chiapas. Desde hace cinco años trabaja en el Departamento de Observación y Estudio de la Tierra, la Atmósfera y el Océano, antes Laboratorio de Análisis Información Geográfica y Estadística del Colegio de la Frontera Sur, como especialista en análisis espacial y percepción remota con enfoque en la clasificación de los tipos de cobertura y usos del suelo, así como su dinámica de cambio.

Contacto: edith.mondragonvazquez@gmail.com

Oswaldo Carrillo

De nacionalidad mexicana. Es maestro en Estadística Aplicada por el Colegio de Postgraduados. Tiene más de siete años de experiencia en materia de monitoreo, reporte y verificación (MRV) de emisiones de gases de efecto invernadero (GEI) del sector forestal y es miembro del *roster* de expertos de la Convención Marco de las Naciones Unidas sobre Cambio Climático. En los últimos tres años ha coordinado el diseño e implementación del Sistema Satelital de Monitoreo Forestal y actualmente funge como titular de la Unidad Técnica Especializada en MRV de la CONAFOR.

Contactos: ocarrillo.ute@conafor.gob.mx y oswaldisma@gmail.com

Richard Tipper

De nacionalidad británica. Su formación incluye estudios de Maestría en Ciencias y Gestión y de Doctorado en Economía de Cambio Tecnológico en Agricultura en la Universidad de Stirling en Escocia. Tiene más de 20 años de experiencia en asesorar y evaluar los esfuerzos de los sectores público y privado para mitigar el cambio climático a través de mejoras en el manejo forestal, uso de suelo y bioenergía. Durante 12 años realizó investigación en los sectores forestal y agrícola en el sur de México. En la actualidad, es director de *Ecometrica* y líder del proyecto *Forests 2020*.

Contactos: Richard.tipper@ecometrica.com y richard.tipper@resilienceconstellation.com

Catherine Navarro

De nacionalidad mexicana. Es egresada de la Licenciatura en Biología por la Universidad Autónoma Metropolitana (UAM)-Xochimilco. Realizó la especialidad y Maestría en Geomática en el Centro de Investigación en Ciencias de Información Geoespacial (CENTROGEO), A. C. Durante sus estudios de posgrado participó como asistente de investigación en el Centro de Estudios Demográficos, Urbanos y Ambientales de El Colegio de México, A. C. Contribuyó como responsable de geomática en la CONAFOR y en la actualidad es responsable de la Gerencia de Información Forestal.

Contacto: catherine.navarro@conafor.gob.mx

Mario Camberos Castro

De nacionalidad mexicana. Es licenciado en Economía por la UdeG, así como maestro y doctor en Ciencias Económicas por la UNAM y UABC-Tijuana, respectivamente. En la actualidad, se desempeña como profesor-investigador emérito en el Departamento de Economía del CIAD, A. C. Es miembro del SNI con nivel II; sus áreas de interés son bienestar económico, desigualdad, mercados laborales, informalidad, salarios y evaluación de políticas de combate a la pobreza y la desigualdad. Con el doctor Bracamontes, obtuvo el primer lugar en el *Premio Nacional a la Investigación Social y Opinión Pública 2011* (1.ª edición) y fue acreedor a mención honorífica en el *Premio Nacional Solidaridad 1994*.

Contacto: mcamberos@ciad.mx

Joaquín Bracamontes Nevárez

Nació en México. Es licenciado, maestro y doctor en Economía por la Universidad de Sonora, de Guadalajara (UdeG) y Autónoma de Baja California (UABC), respectivamente. Es profesor-investigador titular C en el Departamento de Economía del Centro de Investigación en Alimentación y Desarrollo (CIAD). Pertenece al Sistema Nacional de Investigadores (SNI) con nivel I; sus áreas de interés son desarrollo económico regional, política pública, desigualdad socioeconómica, empleo y pobreza.

Contacto: joaco@ciad.mx

Juan Carlos Gomez

De nacionalidad mexicana. Es maestro en Astrofísica y doctor en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica. Fue investigador posdoctoral y asistente de investigación en la *KU Leuven* en Bélgica y en el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). En la actualidad, es profesor de tiempo completo en la Universidad de Guanajuato (UG) y miembro adherente de la Academia Mexicana de Computación. Pertenece al SNI con nivel I; sus áreas de interés incluyen el aprendizaje automático, la minería de datos, la computación evolutiva y la recuperación de información.

Contacto: jc.gomez@ugto.mx

Luis Miguel López Santamaría

De nacionalidad mexicana. Es ingeniero en Sistemas Computacionales por la UG. En la actualidad, es estudiante de la Maestría en Ingeniería Eléctrica en la misma Universidad. Sus áreas de interés incluyen la minería de datos y el aprendizaje automático.

Contacto: lm.lopezsantamaria@ugto.mx

Mario Alberto Ibarra Manzano

De nacionalidad mexicana. Es ingeniero en Comunicaciones y Electrónica y maestro en Ingeniería Eléctrica por la UG y doctor en Microelectrónica y Microsistema por el *Institut National des Sciences Appliquées*, en Toulouse, Francia. Sus intereses de investigación son el diseño digital en FPGA para procesamiento de imágenes, aplicado en robots autónomos y sistemas en tiempo real. En la actualidad, es profesor de tiempo completo en el Departamento de Ingeniería Electrónica de la UG.

Contacto: ibarram@ugto.mx

Dora Luz Almanza Ojeda

De nacionalidad mexicana. Es ingeniera en Comunicaciones y Electrónica y maestra en Ingeniería Eléctrica por la UG; además, es doctora en Sistemas Informáticos y Sistemas Integrados por la *University of Toulouse III: Paul Sabatier*, en Toulouse, Francia. Sus intereses de investigación son la visión por computadora embebida para el control de vehículos autónomos y sistemas en tiempo real. En la actualidad, es profesora de tiempo completo en el Departamento de Ingeniería Electrónica en la UG.

Contacto: dora.almanza@ugto.mx

Tonatiuh Suárez -Meaney

De nacionalidad mexicana. Tiene las licenciaturas en Diseño de los Asentamientos Humanos por la Universidad Autónoma Metropolitana, así como la de Criminología y Criminalística por la Academia Internacional de Formación en Ciencias Forenses, es pasante de Técnico Superior Universitario en Matemáticas por la Universidad Abierta y a Distancia de México y es candidato a maestro en Urbanismo por la UNAM. En la actualidad, es coordinador de Impacto Social en Plurmac y responsable de Análisis Especial en la unidad GITS del Instituto de Geografía de la UNAM; además, es profesor de Geoestadística en la Universidad Autónoma del Estado de México.

Contacto: tonatiuhsmeaney@gmail.com

Israel Albarrán Martínez

De nacionalidad mexicana. Es planificador territorial egresado de la Universidad Autónoma Metropolitana y estudia la ingeniería en Desarrollo de *Software* en la Universidad Abierta y a Distancia de México. En el ámbito laboral, se desempeña como director general de la empresa consultora Geoestrategias Electrónicas de Mercado SA. de CV. Tiene más de 18 años de experiencia en los campos de análisis territorial y geoestadísticas, así como en manejo de cartografía digital, sistemas de información geográfica para el análisis en disciplinas de desarrollo urbano, *geomarketing*, geografía y disciplinas afines.

Contacto: israel.8am@gmail.com

Verónica Pérez Barrón

De nacionalidad mexicana. Tiene la Licenciatura en Psicología por la Universidad Autónoma Metropolitana. Se desempeña como investigadora en la División de Investigación del Hospital Psiquiátrico Infantil Dr. Juan N. Navarro. Participa en protocolos de investigación que versan sobre el desarrollo normal y psicopatología del desarrollo en la infancia y la adolescencia. Las principales líneas de investigación son el funcionamiento, la adaptación, el suicidio adolescente y la discapacidad.

Contacto: veronicaperezbarron@gmail.com

Ioannis Chatziioannou

De nacionalidad griega. Tiene la Licenciatura en Ingeniería Mecatrónica y Telemática por la Universidad de Oeste de Ática, la Maestría en Tecnología de la Información por la Universidad de Oeste de Escocia y el Doctorado en Ingeniería de Sistemas de Transporte por la UNAM. En la actualidad, es investigador postdoctoral en el Instituto de Ingeniería de la UNAM. Sus áreas de investigación incluyen planificación urbana a través de geotecnología, políticas de transporte sostenible, teoría de sistemas, planificación de movilidad y accesibilidad, estudios prospectivos, exclusión social y grupos vulnerables.

Contacto: ioannis.gits@gmail.com

Alejandro Martínez López

De nacionalidad mexicana. Tiene la Licenciatura en Actuaría y una especialización en Estadística Aplicada, ambas por la UNAM. En la actualidad, es gerente de Inteligencia de Negocio en el Registro Único de Vivienda y profesor de Estadística en la Universidad La Salle.

Contacto: alejandrolpz@gmail.com

Luis Chías Becerri

De nacionalidad mexicana. Tiene la Licenciatura y Maestría en Geografía por la UNAM, y el grado de Doctor en Geografía le fue otorgado por la Universidad de Toulouse, Le Mirail, Francia. Desde 1974, labora como investigador en el Instituto de Geografía de la UNAM. Sus principales líneas de investigación son el transporte, la organización territorial, la seguridad vial, así como el diseño y desarrollo de sistemas de información geográfica utilizando *software* libre y de código abierto.

Contacto: luis.chias@gmail.com

Lina Sofía Palacio-Mejía

De nacionalidades mexicana y colombiana. Es doctora en Estudios de Población por El Colegio de México y maestra en Demografía por El Colegio de la Frontera Norte, así como profesional en Gerencia en Sistemas de Información en Salud por la Universidad de Antioquia, Medellín, Colombia. En la actualidad, es investigadora catedrática en el Consejo Nacional de Ciencia y Tecnología (CONACYT), comisionada en el Instituto Nacional de Salud Pública (INSP), sede Cuernavaca, y coordinadora académica titular del Programa de Maestría en Salud Pública con Área de Concentración en Bioestadística y Sistemas de Información en Salud, modalidad presencial, del INSP. Pertenece al SNI de México con nivel II; su área de interés es en el diseño y evaluación de los sistemas de información en salud para el análisis demográfico y estadístico de los principales problemas en salud, como diabetes, cáncer cérvico-uterino, cáncer de mama y VIH/SIDA, entre otros temas.

Contacto: lpalacio@insp.mx

Diana Molina-Vélez

De nacionalidad mexicana. Es maestra en Salud Pública con área de concentración en Bioestadística y Sistemas de Información en Salud por la Escuela de Salud Pública de México/INSP. En la actualidad, es investigadora/coordinadora adjunta de la Maestría en Salud Pública en Bioestadística y Sistemas de Información en Salud, modalidad virtual, en el INSP. Sus áreas de interés son los sistemas de información en salud, mortalidad, VIH, gerencia en salud, subregistro de la información, auditoría de bases de datos y análisis de información.

Contacto: ciee06@insp.mx

Concepción García-Morales

De nacionalidad mexicana. Es maestra en Salud Pública con área de concentración en Bioestadística y Sistemas de Información en Salud por la Escuela de Salud Pública de México/INSP. Actualmente, es analista de riesgos sanitarios en la COPRISEM/Servicios de Salud del estado de Morelos (SSM) y participa como investigadora, docente y consultora independiente. Líneas de investigación: subregistro de mortalidad, egresos hospitalarios, auditoría de bases de datos, análisis estadístico y enfermedades degenerativas y crónicas no transmisibles (ECNT).

Contacto: ciee03@insp.mx

Amado David Quezada-Sánchez

De nacionalidad mexicana. Es licenciado en Economía por la Universidad de las Américas, Puebla, y tiene la Maestría en Estadística Aplicada por el ITESM. En la actualidad, es investigador en ciencias médicas nivel C en el Centro de Investigación en Evaluación y Encuestas (CIEE) del INSP. Sus áreas de interés son los métodos estadísticos, enfermedades crónicas y el análisis de mediación.

Contacto: amado.quezada@insp.mx

Leonel González-González

De nacionalidad mexicana. Es licenciado en Economía por la UdeG, maestro en Desarrollo Regional por El Colegio de la Frontera Norte y doctor en Ciencias Sociales por El Colegio de México; asimismo, cuenta con el Diplomado en Formación Docente por el INSP. En la actualidad, es director tanto de Planeación como del Centro Colaborador Cochrane del INSP. Forma parte del SNI y del Sistema de Investigadores en Ciencias Médicas de la Secretaría de Salud de México. Es miembro de la línea de investigación Sistemas de Información en Salud y docente de doctorado y maestría en la Escuela de Salud Pública de México.

Contacto: Leonel.gonzalez@insp.mx

Maylen Rojas-Botero

De nacionalidad colombiana. Es profesional en Gerencia en Sistemas de Información en Salud, maestra en Epidemiología por la Universidad CES de Colombia y candidata a doctora en Epidemiología, todas en la Universidad de Antioquia, Colombia. En la actualidad, es miembro del Grupo de Demografía y Salud de la Facultad Nacional de Salud Pública de la misma Universidad y epidemióloga de la Dirección de Epidemiología y Demografía del Ministerio de Salud y Protección Social de Colombia. Se interesa en los determinantes sociales de la salud, las desigualdades sociales y territoriales en salud, así como la relación entre la migración y la salud.

Contacto: maylen.rojas@udea.edu.co

Juan Eugenio Hernández-Ávila

De nacionalidad mexicana. Es maestro en Ciencias en Bioestadística por la Escuela de Salud Pública de *Bloomberg* de la Universidad de Johns Hopkins y doctor en Epidemiología por el INSP. En la actualidad, es investigador en Ciencias Médicas con nivel E del INSP y coordinador académico titular del Programa de Maestría en Salud Pública con área de concentración en Bioestadística y Sistemas de Información en Salud, modalidad virtual, en el INSP. Pertenece al SNI con nivel II, y su área de interés se orienta en el uso de sistemas de información geográfica y de inteligencia de negocios aplicados a la salud pública, para la producción de información relevante y conocimiento en esta materia a nivel microrregional para la toma de decisiones y la planeación estratégica de los servicios de salud en el país. Ha liderado

diversos proyectos con el objetivo de fortalecer los sistemas de vigilancia de la salud pública y la producción de información integrada de diversos programas, como el de control vectorial, calidad del agua y salud materno-infantil. Ha colaborado en la evaluación del desempeño de los sistemas de información en diversas naciones de la Región Mesoamericana.

Contacto: juan_eugenio@insp.mx

**Francisco de Jesús Corona
Villavicencio**

De nacionalidad mexicana. Es licenciado en Economía por la UABC, maestro en Estadística Aplicada por el ITESM y doctor en Economía y Métodos Cuantitativos por la Universidad Carlos III de Madrid (UC3M). En la actualidad, es investigador en el Instituto Nacional de Estadística y Geografía (INEGI) y sus líneas de estudio están relacionadas con el análisis econométrico y pronóstico de series de tiempo, así como en *Sport Analytics*; de ambas, ha publicado trabajos en revistas arbitradas de circulación internacional. Pertenece al SNI, con nivel de candidato.

Contacto: franciscoj.corona@inegi.org.mx

Jesús López-Pérez

De nacionalidad mexicana. Es licenciado en Economía y maestro en Estadística Aplicada por el ITESM. Ocupó diversos cargos en áreas de administración y análisis de riesgo crediticio en instituciones financieras de los sectores público y privado. En la actualidad, es investigador invitado en el INEGI en temas relacionados con el análisis econométrico de series de tiempo.

Contacto: jesus.lopezp@inegi.org.mx

**Reyna Elizabeth Rodríguez
Pérez**

De nacionalidad mexicana. Es doctora en Ciencias y maestra en Desarrollo Regional (con mención honorífica) por el Centro de Investigación en Alimentación y Desarrollo, así como licenciada en Economía por la Universidad de Sonora. En la actualidad, es profesora-investigadora en la Facultad de Economía de la Universidad Autónoma de Coahuila (UadeC). Es perfil deseable integrante del Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP) y líder de un cuerpo académico consolidado de la Facultad de Economía de la UadeC. Pertenece al SNI con nivel 1.

Contacto: reynarodriguez@uadec.edu.mx

Liliana Meza González

De nacionalidad mexicana. Es doctora y maestra en Economía por la Universidad de Houston, así como licenciada y maestra en Economía por el Instituto Tecnológico Autónomo de México (ITAM). En la actualidad, es investigadora en el INEGI y profesora de asignatura en los departamentos de Economía y de Estudios Internacionales de la Universidad Iberoamericana en la Ciudad de México. Es miembro del SNI con nivel 1.

Contacto: liliana.meza@inegi.org.mx

David Castro Lugo

De nacionalidad mexicana. Es doctor en Economía Aplicada por la Universidad Autónoma de Barcelona, maestro en Economía Aplicada por el Colegio de la Frontera Norte y licenciado en Economía por la Universidad Autónoma de Sinaloa. En la actualidad, es profesor-investigador del Centro de Investigaciones Socioeconómicas de la UadeC y subdirector de investigación en la misma institución. Es perfil deseable integrante del PRODEP. Pertenece al SNI con nivel 1.

Contacto: david.castro@uadec.edu.mx

Política y lineamientos editoriales

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA es una publicación cuatrimestral que sirve de enlace entre la generación de la información estadística y geográfica oficial y la investigación académica para compartir el conocimiento entre especialistas e instituciones con propósitos similares.

Se publicarán sólo artículos inéditos y originales relacionados con la situación actual del uso y aplicación de la información estadística y geográfica a nivel nacional e internacional.

Es una revista técnico-científica, bilingüe, cuyos trabajos son arbitrados por pares (especialistas), bajo la metodología doble ciego, con los siguientes criterios de evaluación: trabajos inéditos, originalidad, actualidad y oportunidad de la información, claridad en la definición de propósitos e ideas planteadas, cobertura de los objetivos definidos, estructura metodológica adecuada y congruencia entre la información contenida en el trabajo y las conclusiones.

El resultado del proceso de dictaminación se comunica por correo electrónico y contempla tres variantes: recomendado ampliamente (con modificaciones menores), recomendado (pero condicionado a modificaciones sugeridas) y no recomendado (rechazado). Dos dictámenes aprobados, se notifica al autor que se publica y se envía a corrección de estilo; un aprobado y uno rechazado, se le solicita realizar cambios; y dos rechazados, se notifica la no publicación.

Indizaciones y registros

- LATINDEX Catálogo (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal).
- CLASE (Citas Latinoamericanas en Ciencias Sociales y Humanidades).
- REDIB (Red Iberoamericana de Innovación y Conocimiento Científico).

Lineamientos para publicar

Se publicarán trabajos en español e inglés: artículos de investigación, revisión y divulgación; ensayos; metodologías; informes técnicos; comunicaciones cortas; reseñas de libros; revisiones bibliográficas y estadísticas, entre otros.

1. El artículo —o cualquier otro tipo de escrito de los mencionados— deberá entregarse con una carta dirigida al editor responsable de REALIDAD, DATOS Y ESPACIO. REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA en la que se proponga el texto para su publicación, que se declare que es inédito y que no ha sido postulado de manera paralela en otro medio. Asimismo, deben incluirse los datos completos del(os) autor(es), nacionalidad(es), institución(es) de adscripción y cargo(s) que ocupa(n), domicilio(s) completo(s), correo(s) electrónico(s) y teléfono(s). Esto debe dirigirse a la atención de la M. en C. Virginia Abrin Batule, virginia.abrin@inegi.org.mx (tel. 5278 10 00, ext. 1161).
2. El trabajo se debe presentar en versión electrónica (formato *Word* o compatible) con: a) extensión no mayor de 20 cuartillas; b) letra Helvética, Arial o Times de 12 puntos y c) interlineado de 1.5 líneas. El material adicional al texto se requiere por separado: a) las imágenes, con resolución de 300 ppp y un tamaño no menor a 17 centímetros de base (ancho) en formato JPG o TIF —no remuestrear (ampliar) imágenes de menor resolución—; si son líneas o mapas, deben entregarse en formato vectorial (EPS o Ai), en caso de incluirse imágenes en mapa de bits, incrustarlas o enviarlas con el nombre con el cual se creó el vínculo (conservando los requerimientos de resolución y tamaño estipulados); y para fotografías, éstas no deben ser menores a 5 megapíxeles; b) las fórmulas o expresiones matemáticas tienen que elaborarse con el editor de ecuaciones propio de *Microsoft*[™], pero en caso de usar *software* de terceros, incluir en la entrega PDF testigo en el cual figuren exactamente cómo deben representarse; c) las gráficas, que incluyan el archivo en *Excel* con el cual se desarrollaron o, en su defecto, la imagen JPG legible, de origen, en alta resolución; y d) los cuadros, que sean editables, no se deben insertar como imagen.
3. La colaboración debe incluir: título del trabajo (en español e inglés o viceversa); resúmenes del trabajo en español e inglés (que no excedan de un párrafo de 10 renglones); palabras clave en español e inglés (mínimo tres, máximo cinco); bibliografía u otras fuentes; así como breve(s) semblanza(s) del(os) autor(es) que no exceda(n) de un párrafo de cinco renglones y que incluya(n) nacionalidad(es), grado(s) académico(s), principal(es) experiencia(s) profesional(es), adscripción(es) laboral(es) actual(es) y dirección(es) electrónica(s) de contacto.
4. Las referencias bibliográficas u otras fuentes deberán presentarse al final del artículo de la siguiente manera: nombre(s) del(os) autor(es) comenzando por el(los) apellido(s); título de la publicación con cursivas (si se trata de un artículo, debe estar entrecomillado, seguido de coma y la preposición en con dos puntos y, enseguida, el título de la revista o libro donde apareció publicado, con cursivas); país de origen; editorial; lugar y año de edición; página(s) consultada(s). En el caso de las fuentes electrónicas (páginas web) se debe seguir el mismo orden que en las bibliográficas, pero al final se pondrá entre paréntesis DE (dirección electrónica), la fecha de consulta y la liga completa. Se tienen que omitir aquellas que se mencionen como notas a pie de página. Si se aplica la opción de incluir en cuerpo de texto la referencia de nombre de autor y año de la fuente consultada entre paréntesis, sí deben aparecer todas las referencias mencionadas.

Página electrónica: <http://rde.inegi.org.mx>

Editorial Guidelines and Policy

REALITY, DATA AND SPACE INTERNATIONAL JOURNAL OF STATISTICS AND GEOGRAPHY is a four-monthly publication that connects statistics and geographic official information with academic research in order to share knowledge among specialists and institutions with similar aims.

We will publish only original and unpublished articles related to the current use and appliance of statistical and geographical information at both national and international levels.

It is a technical-scientific and bilingual magazine, with articles previously peer-reviewed by specialists under a double-blind methodology with the following evaluation criteria: unpublished works, originality, information related to opportunity and current affairs, we expect clarity in the definition of aims and ideas stated, defined objectives coverage, accurate methodological structure and coherence between the information of the paper as well as its conclusions.

The result of the paper-assessment process is delivered by email, and it involves three possibilities: fully recommended (with slight modifications), recommended (on condition of suggested modifications) and not recommended (i.e. rejected). When there are two reports of approval, the author gets notified that his/her paper will be published and it is sent to a style editing process. When one report approves the paper for publication and another one rejects it, the author is requested to make some changes for the text to be published. If the text submitted receives two non-favourable reports, the author is notified that the text will not be published.

Index and Registers

- LATINDEX Catalogue (Online Regional Information System for Scientific Journals from Latin America, the Caribbean, Spain and Portugal).
- CLASE (Latin American Quotations in Humanities and Social Sciences)
- REDIB (Latin American Net of Innovation and Scientific Knowledge)

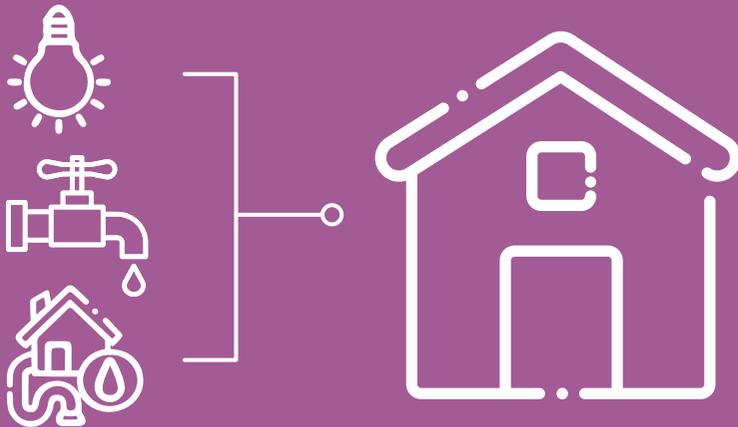
Publishing Guidelines

Articles will be published in Spanish or English: research, revision and scientific-spreading articles; methodologies; technical reports; short texts; book reviews; and bibliographical and statistical revisions, among others.

1. The article —or any other kind of text from those aforementioned— must be delivered with an attached letter addressed to the chief editor of Reality, Data and Space. International Statistics and Geography Magazine in which the text intended for publication will be submitted. There it must be stated that the text has not been published, and that it has not been submitted for publication in any other media. The names in full of the authors must be included, as well as their nationalities, adscription institutions, position in those institutions, postal address, e-mail address, and telephone numbers. This must be addressed to MSc Virginia Abrin Batule, virginia.abrin@inegi.org.mx (tel (+52) (55) 52.78.10.00, extension 1161).
2. The article must be submitted in an electronic version (a Microsoft Word file or a compatible one) with the following format: a) the text should not exceed the 20 pages of length; b) typography must be Helvetic, Arial or Times (12 points); and c) there should be a 1.5 line spacing in each paragraph. Additional material to the text will be delivered separately: a) images with a resolution of 300 ppp and no smaller than 17 cm width will be delivered in format JPG or TIF —please do not amplify images with lower resolution—. If the added materials are lines or maps, these must be delivered in vectorial format (EPS or Ai). If there are images in bits map, these must be embedded or attached with the name of the original file with which the link was created (keeping the resolution and size requirements above stated). As regards to photographs, these should not be inferior as 5 megapixels; b) mathematical expressions or formulae have to be created with the equations editor by *Microsoft*[™], but in case of using third-parties software, please attach a witness PDF in which the exact representation of mathematical formulae or expressions is contained; c) graphics must include the Excel file in which they were created or a legible image in the original JPG format in high resolution; and d) charts must be editable, and must not be inserted as images.
3. The text must include the following: the article's title (both in English and Spanish); the abstract of the article—both in English and Spanish (not longer than a 10-line paragraph); key words—both in English and Spanish (three as minimum and five as maximum); bibliography and other sources; as well as brief biographical sketches of the authors not exceeding a five-line paragraph each including nationalities, academic titles, main professional experiences, current work-related affiliations, and electronic addresses for the authors to be contacted.
4. Bibliographical references and other sources must be included at the end of the article in the following way: author's name (Surname first), and publication's title (in italics). If it is an article, the title must be in quotation marks followed by a comma and the preposition "en" with semicolon (in Spanish), then it should appear the title of the book or magazine in which the article was published (in italics); country of origin; publishing house, edition year, and consulted pages. As regards to electronic sources (web pages) the same order of the bibliographical references must be followed, but at the end the word "EA" (as for Electronic Address) ("DE" in Spanish) must be added within parenthesis followed by consultation date and the complete reference link. Those web links referred previously as footnotes, must be omitted in this section. However, if the name of the author and the year of the consulted source were included in the main body of the text within parenthesis, all these must be included as part of the bibliographical references.

Webpage: <http://rde.inegi.org.mx>

Servicios con que cuentan:

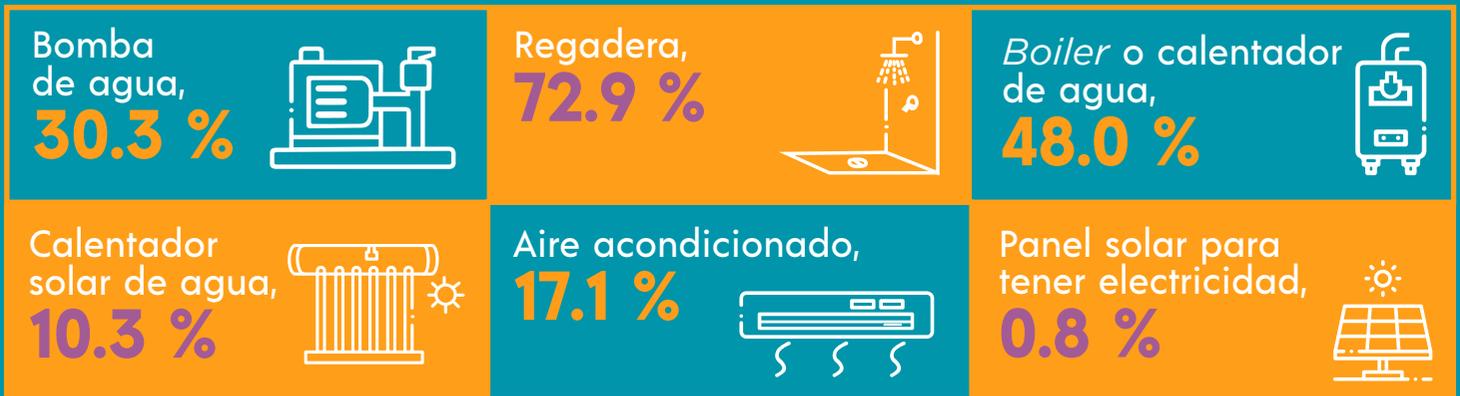


99 de cada 100 viviendas particulares habitadas tienen **energía eléctrica**; en 2010 eran **98**.

96 de cada 100 tienen **agua entubada** dentro de la vivienda, o en el patio o terreno; en 2010 eran **88**.

78 de cada 100 cuentan con drenaje conectado a **red pública**; en 2010 eran **72**.

Equipamiento:



Los tres principales bienes:



Conociendo
México

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

 **INEGI** Informa

Fuentes: INEGI. Censo de Población y Vivienda 2010. Cuestionario básico. // INEGI. Censo de Población y Vivienda 2020. Cuestionario básico y Cuestionario ampliado.

