

Retropolación hasta 1980 del PIB trimestral de México por entidad federativa y gran actividad económica

Retropolation up to 1980 of quarterly Mexican GDP by State and Grand Economic Activity

Víctor Manuel Guerrero Guzmán* y Francisco de Jesús Corona Villavicencio**

* Instituto Tecnológico Autónomo de México (ITAM), guerrero@itam.mx

**Instituto Nacional de Estadística y Geografía (INEGI), franciscoj.corona@inegi.org.mx

Nota: se agradecen los valiosos comentarios y sugerencias que brindaron Gerardo Leyva, Francisco Guillén, Lourdes Mosqueda y Jesús López, así como el apoyo brindado por Enrique Ordaz, todos ellos funcionarios del INEGI, durante el desarrollo de este trabajo. Víctor M. Guerrero agradece al Instituto Tecnológico Autónomo de México (ITAM) el otorgamiento de un año sabático que le permitió incorporarse como investigador invitado al INEGI para realizar el proyecto aquí descrito. Asimismo, agradece a la Asociación Mexicana de Cultura, AC por su apoyo mediante la Cátedra de Análisis de Series de Tiempo y Pronósticos en Econometría.

Diferentes tipos de gráficos y tablas en papel blanco/studioocasper/Getty Images



En este trabajo se *retropol*a el producto interno bruto estatal de México de 1980 a 1992 por gran actividad económica a partir de los datos oficiales disponibles para el público. El documento consta de seis etapas: 1) desagregación trimestral de la base de datos anual del 2003 al 2015, por estado; 2) conversión de la base de datos estatal y anual, de años base 1993 al 2008; 3) *retropolación* restringida de 1993 al 2002 con datos desagregados por estado que satisfacen restricciones temporales; 4) reconciliación de cifras estatales previamente *retropoladas* con los datos a nivel nacional; 5) *retropolación* restringida de 1980 a 1992 de la base ya reconciliada, por gran actividad económica, de manera que la suma de los estados produce el total nacional y, por último, 6) cambio de año base del 2008 al 2013, actualizando la información al 2016. Los resultados empíricos se validan tanto estadística como económicamente y se ilustran con datos de la Ciudad de México.

Palabras clave: conversión de series de tiempo; desagregación temporal; modelos AR y VAR; pronósticos restringidos; reconciliación de cifras; reconstrucción de valores pasados.

In this paper, we *retropolate* the Mexican Gross Domestic Product (GDP) by State and Economic Activity from 1980 to 1992 using all the official databases currently available. We apply 6 steps: i) temporal disaggregation of the state-annual database for years 2003-2015, ii) conversion of the state-annual database from base year 1993 to 2008, iii) restricted *retropolation* from 1993 to 2002 with the state-disaggregated database, satisfying temporal restrictions, iv) reconciliation of the *retropolated* quarterly data with the national database, v) restricted *retropolation* of the reconciliated database from 1980 to 1992 by Economic Activity, so that the sum of the state data yields the national figure; and finally, vi) change of base year from 2008 to 2013, with information updated to 2016. The empirical results are verified both statistically and econometrically, and illustrated with Mexico City's data.

Key words: time series conversion; temporal disaggregation; AR and VAR models; restricted forecasts; reconciliation; reconstruction of past values.

Recibido: 7 de diciembre de 2017.

Aceptado: 28 de junio de 2018.

Introducción

Este trabajo parte del reconocimiento de que, en México, el producto interno bruto (PIB) dentro de una cierta ventana temporal (digamos de 1993 al 2015), a precios constantes del año base 2008 —como aparece en el Banco de Información Económica (BIE) del INEGI— constituye una base de datos que se denota aquí como $PIB(Nac, Tri, Sub + Ram; 93-15)_{08}$ (el subíndice indica el año base), la cual indica la cobertura geográfica nacional (*Nac*), la periodicidad trimestral de los datos (*Tri*), la cobertura a nivel de subsectores —aunque en algunos casos alcanza el nivel de rama de actividad económica (*Sub + Ram*)— y el periodo de cobertura (93-15). A partir de las cifras trimestrales se pueden generar, mediante promedios, los datos anuales. De manera similar, con los datos de subsectores y ramas se pueden calcular, por suma, los datos de los sectores que, a su vez, se agregan en las tres grandes actividades (GA) económicas (primarias, secundarias y terciarias).

El INEGI presenta la base de datos del PIB estatal desagregada a nivel sector y por 12 grupos de subsectores manufactureros, donotada como $PIB(Est, Anu, Sec + Sub; 03-15)_{08}$. Asimismo, en la base $PIB(Est, Anu, GDiv + Div; 93-06)_{93}$ se presentan los datos del PIB de acuerdo con la Clasificación Internacional Industrial Uniforme (CIIU) de todas las actividades económicas (ver ONU, 1990), mientras que en México se usa desde 1997 el Sistema de Clasificación Industrial de América del Norte (SCIAN), que tiene categorías comparables con el nivel de dos dígitos de la CIIU, Rev. 3 (según se describe en INEGI, 2007). Además, se cuenta con el indicador trimestral de la actividad económica estatal (ITAE), que se denota como $ITAE(Est, Tri, GA; 03-15)_{08}$. Las bases de datos mencionadas se encuentran en formatos de hojas de cálculo, mientras que la siguiente se encuentra en uno de texto (ver INEGI, 1998), y a ésta se le denota como $PIB(Nac, Tri, GDiv + Div; 80-98)_{93}$ porque extiende la cobertura temporal hasta 1980, aunque con año base 1993. Además, a partir de octubre del 2017 se tienen disponibles también las bases $PIB(Est, Anu, Sec + Sub; 03-16)_{13}$, $ITAE(Est, Tri, GA; 03-16)_{13}$ y $PIB(Nac, Tri, Sub + Ram; 83-16)_{13}$, que corresponden al cambio de año base del 2008 al 2013.

Una vez conocidos los insumos de datos oficiales a los que se tiene acceso, se establecen los siguientes objetivos del estudio:

- Utilizar las tres bases de datos estatales — $PIB(Est, Anu, GDiv + Div; 93-06)_{93}$, $ITAE(Est, Tri, GA; 03-15)_{08}$ y $PIB(Est, Anu, Sec + Sub; 03-15)_{08}$ — para estimar la base de datos $PIB(Est, Tri, GA; 93-15)_{08}$, o sea, una base homogénea que contenga datos a nivel estatal y trimestral por GA , para el periodo 1993-2015 y con año base 2008.
- Complementar la base obtenida en el objetivo anterior con los de la base de datos nacional $PIB(Nac, Tri, Sub + Ram; 93-15)_{08}$, para tener en cuenta toda la información disponible del periodo 1993-2002 y mejorar, con ello, la estimación de la base $PIB(Est, Tri, GA; 93-15)_{08}$.
- Incorporar los datos de la base $PIB(Nac, Tri, GDiv + Div; 80-98)_{93}$ para extender los resultados de la estimación obtenida en el segundo objetivo, con el fin de estimar la base de datos $PIB(Est, Tri, GA; 80-15)_{08}$.
- Tener en cuenta el cambio de año base más reciente (del 2008 al 2013), así como los datos del 2016, para generar la base de datos actualizada $PIB(Est, Tri, GA; 80-16)_{13}$.

El tema de ampliar la longitud de las series con valores de periodos pasados y homogeneizar bases de datos se menciona en la literatura como *cálculo retrospectivo*, *reconstrucción histórica*, *estimación de valores pasados*, *retropolación*, *pronóstico hacia atrás*, *conversión o empalme de series de tiempo* (ver al respecto Correa *et al.*, 2003; Ponce, 2004; Caporin y Sartore, 2005; Roulin y Eidmann, 2007; Yuskavage, 2007; James, 2008; Buiten *et al.*, 2009; Hellberg, 2010; Arbués y López, 2011; DANE, 2013 y Gallegos-Rivas, 2015). Para estimar datos con mayor frecuencia de observación que la de los datos originales, se hace referencia a la *desagregación temporal de series*, tanto en versión univariada (ver Guerrero, 1990 y 2003 y Nieto, 1998) como multivariada (Guerrero y Nieto, 1999). También, se habla del ajuste de valores obtenidos con alta frecuencia de observación (comúnmente con poca exactitud, pero oportunos) a valores totales de baja frecuencia (más confiables, pero poco oportunos); a esta situación se le conoce como *ajuste a valores de referencia (benchmarking)*, como lo mencionan Dagum y Cholette (2006), quienes también consideran la *reconciliación de datos*, al igual que Di Fonzo y Marini (2005). Una metodología adicional que se utiliza es la de *pronósticos con restricciones*, según el enfoque multivariado de Guerrero y Nieto (1999). La idea general de los métodos empleados es combinar de manera eficiente las fuentes de información, como señalan Guerrero y Peña (2003).

Es importante señalar que existen trabajos previos para el caso mexicano, por ejemplo el de Puig y Hernández (1989), quienes aplican la metodología de Chow-Lin para desagregar el PIB anual nacional de 1970 a 1988 utilizando como variables auxiliares las series estatales de la captación en la banca comercial, o sea, los depósitos en cuentas de cheques, ahorro y a plazo. También, se cuenta con el trabajo de German-Soto (2005) quien, a partir tanto de datos oficiales como no oficiales, genera series retropoladas del PIB anual por entidad federativa de 1940 a 1992; en éste se obtienen estimaciones preliminares a través del uso de regresiones lineales ajustando el movimiento del PIB de cada entidad con ayuda del nacional. Posteriormente, las cifras así obtenidas se ajustan mediante un método de conciliación transversal.

La diferencia principal de esta investigación, con respecto a lo realizado en la literatura previa para el caso de México, es que la metodología de retropolación restringida que se propone aquí es óptima en términos estadísticos, como lo demuestran Guerrero y Peña (2003). También, es importante señalar que se realiza la estimación a nivel trimestral y que se tienen diversas ventajas que validan el uso de esta metodología como, por ejemplo, que se hace uso exclusivamente de información oficial, además de que se realiza la validación empírica de los modelos estadísticos utilizados.

Así, el objetivo de este trabajo radica en explicar a detalle el proceso de retropolación restringida óptima para el PIB trimestral de México, por entidad federativa, por gran actividad económica (de 1980 al 2016) y con el año base más reciente, que es el 2013.

Lo que resta de este documento está organizado como sigue: la segunda sección menciona los métodos a usar en las diferentes etapas del estudio; en la tres se introduce la notación y se describen formalmente los métodos utilizados en este trabajo; la siguiente presenta con detalle la aplicación de los métodos propuestos para el caso de la Ciudad de México, visto como un estado que se elige a manera de ilustración; por último, en la sección cinco se elaboran algunas conclusiones y recomendaciones que subrayan algunos aspectos de particular importancia acerca de la metodología utilizada y de los resultados alcanzados.

Propuesta metodológica

Los métodos que se utilizan en este trabajo siguen el enfoque de *macrodatos* e incluyen fases de conversión, desagregación temporal, retropolación restringida y reconciliación de cifras.

Método 1. Para combinar las bases de datos estatales de 1993 al 2002

A partir de $PIB(Est, Anu, Sec + Sub; 03-15)_{08}$ se obtiene, por agregación de los datos de sectores y subsectores a GA, la base de datos $PIB(Est, Anu, GA; 03-15)_{08}$; luego, se da la base $PIB(Est, Anu, GA; 93-06)_{93}$, expresada a nivel de GA, al aplicar una fase de conversión, ya que originalmente está clasificada en grandes divisiones y divisiones industriales. Esto se logra al encadenar las variaciones porcentuales obtenidas con los datos empalmados del 2003 provenientes de las dos bases de datos involucradas. Así, se llega a series de tiempo por estado que producen la base de datos convertida $\overline{PIB}(Est, Anu, GA; 93-02)_{08}$. Con ésta, unida a la de las observaciones generadas por cálculo directo de las fuentes básicas, se obtiene una homogénea que cubre el periodo completo 1993-2015, o sea, $\overline{PIB}(Est, Anu, GA; 93-15)_{08}$.

Enseguida, se aplica la desagregación univariada a la base $PIB(Est, Anu, GA; 03-15)_{08}$, con los datos de $ITAE(Est, Tri, GA; 03-15)_{08}$ usados como variables auxiliares. Con ello, se obtiene una estimación preliminar del PIB trimestral y estatal expresado en valores, a la cual se le debe hacer un ajuste para que satisfaga las restricciones lineales impuestas por el segmento temporal de la base de datos convertida, $\widehat{PIB}(Est, Anu, GA; 03-15)_{08}$. De esta forma, se logra la base estatal estimada $\widehat{PIB}(Est, Tri, GA; 03-15)_{08}$, con lo cual se tienen los dos insumos para la fase de retropolación restringida que se esquematiza en la figura 1, en la cual se puede apreciar cómo, con los datos desagregados, se pronostican irrestrictamente los PIB por GA hasta 1993, ejemplificados con las líneas azules que portan triángulos, cuadrados y círculos, los cuales, a su vez, están en color rojo, azul y negro, de forma respectiva. Estas series, a su vez, se ajustan a los valores de las series convertidas, representadas por la línea azul con círculos negros.

Las series de tiempo de la base anual $\widehat{PIB}(Est, Anu, GA; 93-02)_{08}$ se utilizan como restricciones temporales que deben satisfacer cada una de las GA, de manera que el promedio de los valores trimestrales de cada año sea igual al respectivo valor anual en cada uno de los estados durante la fase de retropolación restringida que sigue, la cual usa una estimación preliminar de las series que surge de la base $\widehat{PIB}(Est, Tri, GA; 03-15)_{08}$ al pronosticar hacia atrás. Así, las series pronosticadas del 2002 a 1993 determinan la dinámica simultánea de las series de tiempo de las GA. La retropolación restringida usa dichos pronósticos trimestrales de las tres GA en cada estado como series preliminares, las cuales deben ajustarse para cumplir con la restricción temporal impuesta por los datos anuales.

La estimación múltiple es tentativa porque satisface la restricción anual (que es un requisito contable), pero no necesariamente cumple con el criterio estadístico de optimalidad, o sea, no requiere alcanzar el error cuadrático medio (ECM) mínimo. Por tal motivo, a partir de la estimación tentativa se genera la definitiva, que es óptima en el sentido mencionado, con lo cual se produce la base de datos estimada $\widehat{PIB}_1(Est, Tri, GA; 93-02)_{08}$. Una vez logrado esto se completa dicha base con la de los datos desagregados $\widehat{PIB}(Est, Tri, GA; 03-15)_{08}$ para obtener la base del periodo completo que interesa, o sea, $\widehat{PIB}_1(Est, Tri, GA; 93-15)_{08}$ con lo cual se consigue el primer objetivo. El subíndice 1 se usa para indicar que este cálculo puede mejorarse al incorporar la información disponible de carácter nacional.

Método 2. Para combinar información nacional y estatal (1993-2002)

Después de aplicar el método previo a cada estado se utiliza la base de datos con cifras a nivel nacional $PIB(Nac, Tri, Sub + Ram; 93-15)_{08}$ para realizar una reconciliación de datos, que consiste en compatibilizar los resultados obtenidos de forma previa con los de esta nueva base; este segundo método complementa al anterior, pues utiliza información más agregada y considera a $\widehat{PIB}_1(Est, Tri, GA; 93-02)_{08}$ como una base estimada que se puede mejorar en términos estadísticos. Para ello, a esta última se le aplica el método de reconciliación para que cumpla con las restricciones contemporáneas impuestas por los datos de $PIB(Nac, Tri, GA; 93-15)_{08}$ con lo cual se alcanza el segundo objetivo. De manera esquemática, se tiene la situación mostrada en la figura 2, donde se combinan las 32 bases de datos trimestrales estatales estimadas con la nacional, que provee las restricciones de la reconciliación, ejemplificadas en dicha figura por la línea negra.

Nótese que la reconciliación dentro del proceso de la retropolación restringida es una etapa que consiste en *conciliar* los datos retropolados de forma estatal (que satisfacen restricciones temporales)

Figura 1

Bases de datos estatales para la retropolación restringida del 2002 a 1993

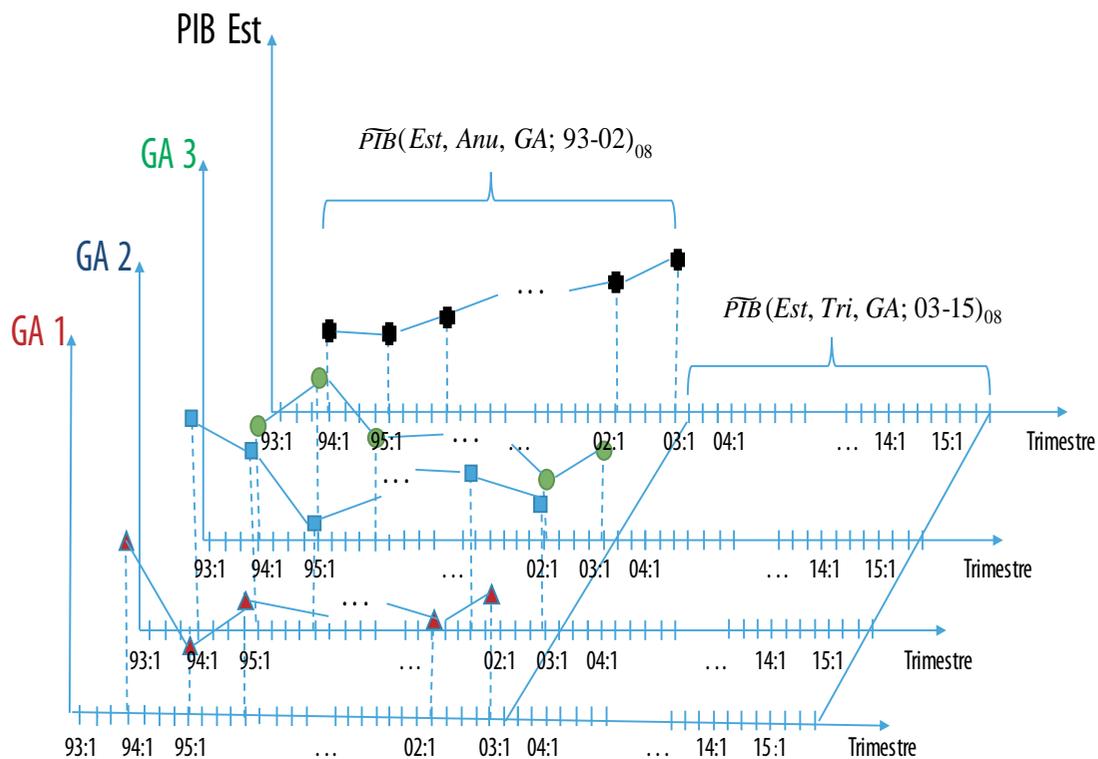
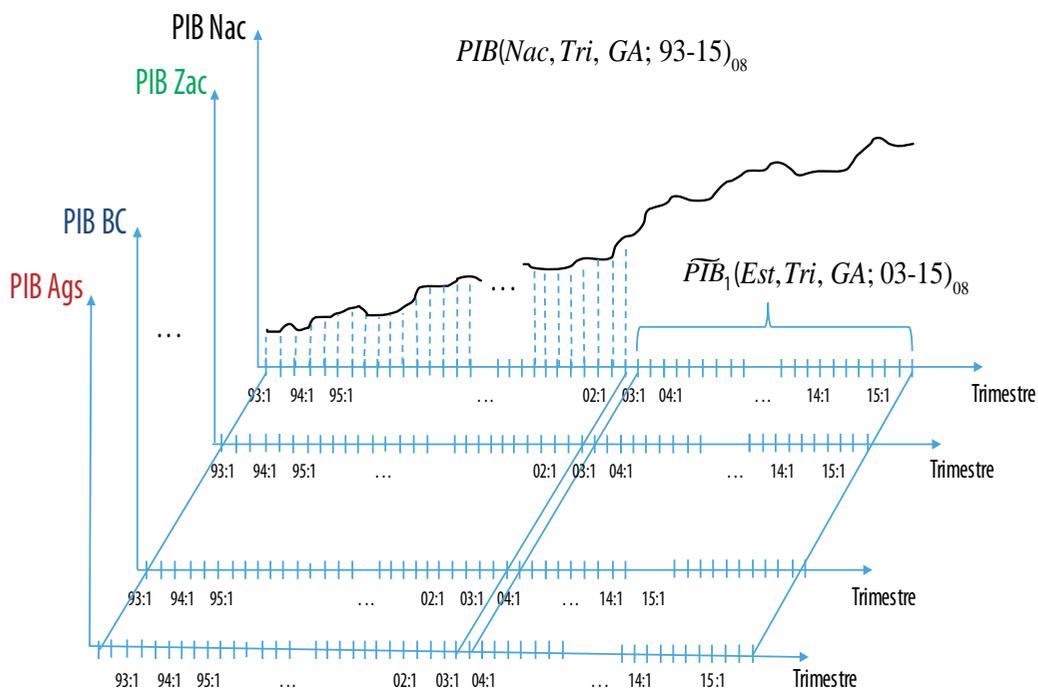


Figura 2

Bases de datos requeridas para la reconciliación de 1993 al 2002



con los valores contemporáneos de la base de datos nacional y que esta discrepancia es una regularidad empírica, pues los datos que satisfacen restricciones de carácter temporal son de un nivel de agregación mayor y la suma de cada periodo puede o no coincidir con el total nacional; es decir, ya que las restricciones temporales impuestas provienen de una fase de conversión, donde el proceso de empalme impone estructuras no necesariamente coincidentes con las observables de carácter estatal (y desconocidas), es necesario ajustar las cifras retropoladas de manera estatal, con los datos nacionales observados.

Método 3. Para retropolar de 1992 a 1980

Para alcanzar el tercer objetivo, se usan los datos disponibles en la base $PIB(Nac, Tri, GDiv + Div; 80-98)_{93}$ clasificados de acuerdo con la CIIU y con los que se genera la base de datos nacional del $PIB(Nac, Tri, GA; 80-92)_{93}$, de manera semejante a como se indicó antes para obtener $PIB(Est, Anu, GA; 93-02)_{93}$. Para trabajar con estas bases, primero se aplica una fase de conversión para generar datos con año base 2008, es decir, $\widehat{PIB}(Nac, Tri, GA; 80-92)_{08}$, los cuales sirven como el conjunto de restricciones lineales por satisfacer al aplicar la retropolación restringida. De hecho, con la base de datos $\widehat{PIB}(Est, Tri, GA; 93-02)_{08}$ se pueden construir tres modelos para series múltiples de dimensión 32, de manera que se incorporen los datos de todos los estados en el modelo de cada una de las tres GA.

Con esos modelos se pronostican hacia atrás los datos trimestrales de 1992 a 1980 y se usa la base $\widehat{PIB}(Nac, Tri, GA; 80-92)_{08}$ para restringir el comportamiento de los pronósticos; de esa forma, se produce la base de datos estimada $\widehat{PIB}(Est, Tri, GA; 80-92)_{08}$ que satisface la restricción de que la suma de los valores de todos los estados produce el PIB nacional para cada una de las GA. Al unir esta base de datos con la obtenida anteriormente se tiene $\widehat{PIB}(Est, Tri, GA; 80-15)_{08}$, con lo cual se alcanza el tercer objetivo.

Método 4. Para cambiar el año base al 2013 e incorporar los datos del 2016

El cuarto objetivo se consigue al aplicar primero una fase de conversión que utiliza las bases de datos $PIB(Est, Anu, GDiv + Div; 93-02)_{93}$ y $PIB(Est, Anu, Sec + Sub; 03-15)_{13}$ para generar la base $\widehat{PIB}(Est, Anu, GA; 93-15)_{13}$. Acto seguido, se realiza la fase de desagregación univariada en la base $PIB(Est, Anu, GA; 03-15)_{13}$ con apoyo en $ITAE(Est, Tri, GA; 03-15)_{13}$, cuyas series son usadas como estimaciones preliminares, lo que conduce a la base estimada $\widehat{PIB}(Est, Tri, GA; 03-15)_{13}$. Enseguida, se realiza la etapa de retropolación restringida, que usa 32 modelos de vectores autorregresivos (VAR, por sus siglas en inglés) trivariados para pronosticar hacia atrás la base de datos estimada en la fase anterior y que impone como restricciones temporales las obtenidas en la etapa de conversión, $\widehat{PIB}(Est, Anu, GA; 93-15)_{13}$, lo cual brinda como resultado la base $\widehat{PIB}(Est, Tri, GA; 93-15)_{13}$. Asimismo, se realiza una fase de conversión trimestral que emplea las bases de datos $\widehat{PIB}(Nac, Tri, Sub + Ram; 93-16)_{08}$ y $PIB(Nac, Tri, GA; 80-98)_{08}$ para generar la $\widehat{PIB}(Nac, Tri, GA; 80-16)_{13}$.

Después, se lleva a cabo otra etapa de retropolación restringida, ahora mediante tres modelos VAR de dimensión 32 (uno para cada GA), en los cuales se verifica la validez de la regionalización mediante la idea de cointegración entre series de tiempo (ver Johansen, 1991). De esta forma, se usa $\widehat{PIB}(Est, Tri, GA; 93-15)_{13}$ y, como restricciones contemporáneas, los datos de la base \widehat{PIB}

$(Nac, Tri, GA; 80-16)_{13}$, con lo que se genera $\widehat{PIB}(Est, Tri, GA; 80-15)_{13}$. Es claro que, aunque existan relaciones de cointegración entre las entidades que forman cada una de las regiones no implica que esto valide algún concepto teórico de regionalización o, en su caso, que ésta sea la única forma de regionalizar (ver, por ejemplo, la propuesta por Bassols-Batalla, 1983). No obstante, en conceptos de series de tiempo, la presencia de cointegración sí garantiza que existen relaciones empíricas predecibles, es decir, que no es espurio pronosticar el PIB de cada estado en función de las otras entidades pertenecientes a la región; además, que cada una esté cointegrada valida que se puedan expresar modelos VAR en niveles para cada región y, así, realizar los pronósticos irrestrictos (ver Lütkepohl, 2005).

Por último, para obtener la base estimada $\widehat{PIB}(Est, Tri, GA; 80-16)_{13}$, se utilizan las restricciones contemporáneas que surgen de $PIB(Nac, Tri, Sub + Ram; 93-16)_{08}$ junto con los resultados de un ejercicio de regresión lineal simple para relacionar las series anuales de $PIB(Est, Anu, GA; 03-15)_{13}$ con las de $ITAE(Est, Anu, GA; 03-15)_{13}$; con ello, se produce la base de datos trimestral $\widehat{PIB}(Est, Tri, GA; 16)_{13}$ a partir de los datos de $ITAE(Est, Tri, GA; 16)_{13}$. Por lo tanto, la unión de $\widehat{PIB}(Est, Tri, GA; 80-15)_{13}$ y $\widehat{PIB}(Est, Tri, GA; 16)_{13}$ genera finalmente la base estimada $\widehat{PIB}(Est, Tri, GA; 80-16)_{13}$.

Descripción técnica de los métodos estadísticos

Los métodos se presentan de manera compacta y uniforme con la siguiente notación: para cada uno de los estados, $Z_{i,4(T-1)+q}$ denota la observación de la GA i , con $i = 1, 2, 3$, donde $T = 1, \dots, n$ indica años y $q = 1, \dots, 4$ señala trimestres, o sea, el año $T = 1$ incluye los trimestres $t = 1, \dots, 4$, $T = 2$ considera a $t = 5, \dots, 8$ y así, sucesivamente, hasta $T = n$, que incluye los trimestres $t = 4(n-1) + 1, \dots, 4n$. De igual forma, las observaciones anuales correspondientes se escriben como $Y_{i,T}$, con $i = 1, 2, 3$ y $T = 1, \dots, n$, que están relacionadas con las observaciones trimestrales mediante promedios anuales, o sea, como:

$$Y_{i,T} = \sum_{q=1}^4 Z_{i, 4(T-1)+q} / 4 \text{ para } i=1, 2, 3 \text{ y } T = 1, \dots, n. \quad (1)$$

Conversión

Aquí se usa el método proporcional (conocido también como de *empalme* de series de tiempo) con dos vertientes: anual y trimestral. En ambos casos, cuando el clasificador cambia de CIU a SCIAN, se aplica la técnica al nivel de desglose de actividades económicas más detallada posible (ver Roulin y Eidmann, 2007), que en el presente caso corresponde al de sectores, o sea, a dos dígitos de la clasificación del SCIAN. Si el clasificador no cambia a nivel de GA (lo cual ocurre del SCIAN 2007 al SCIAN 2013), no hay necesidad de utilizar mayor desglose que el de gran actividad, al que se requieren los resultados (ver Castillo, 2013).

La técnica consiste en asignarle a la serie de tiempo con año base más nuevo el crecimiento temporal (anual, trimestral, mensual, etc., según sea el caso) de la serie de tiempo con año base más viejo. Para el anual, $Y_{i,T,j}$ denota la observación del PIB de la GA i en el año T , a precios constantes del año base j , con $j = 0, 1$, donde $j = 0$ denota la base antigua y $j = 1$, la más reciente. El método que se usa se conoce como *primer periodo común* —o, en equivalencia, *primer año común*, porque la serie es anual—; para

más detalles, ver Parrot y McKenzie (2003). Se calcula un coeficiente de conversión, c_i , que relaciona las dos series involucradas, como cociente de los valores registrados el primer año de empalme de las series, o sea:

$$c_i = Y_{i,e,1} / Y_{i,e,0} \quad (2)$$

donde e denota el año inicial del empalme de las dos bases de datos del PIB. Dicho coeficiente es una constante de proporcionalidad que se aplica a toda la serie de datos con año base 0, para obtener la nueva serie expresada con año base 1, o sea:

$$Y_{i,T,1} = c_i \times Y_{i,T,0} \quad \text{para } T = 1, \dots, e \quad (3)$$

Para apreciar el efecto del cambio de año base propuesto sobre las tasas de variación, se considera la tasa de variación anual del PIB en el año T , con año base j , que está dada por:

$$R_{i,T,j} = \frac{Y_{i,T,j} - Y_{i,T-1,j}}{Y_{i,T-1,j}} = \frac{Y_{i,T,j}}{Y_{i,T-1,j}} - 1, \quad \text{para } j = 0, 1. \quad (4)$$

Al aplicar la conversión se tiene que:

$$R_{i,T,1} = \frac{c_i \times Y_{i,T,0}}{c_i \times Y_{i,T-1,0}} - 1 = \frac{Y_{i,T,0}}{Y_{i,T-1,0}} - 1 = R_{i,T,0}, \quad (5)$$

para $T = 1, \dots, e,$

de manera que las tasas de variación de la serie con año base 0 son las mismas al convertir la serie al año base 1. En cambio, la tasa de variación para el año $e + 1$, con año base 1, o sea:

$$R_{i,e+1,1} = \frac{Y_{i,e+1,1}}{Y_{i,e,1}} - 1 = \frac{Y_{i,e+1,1}}{c_i \times Y_{i,e,0}} - 1, \quad (6)$$

es diferente de $R_{i,e+1,0}$ excepto si se cumple que $Y_{i,e,1} = c_i \times Y_{i,e,0}$, lo cual ocurriría si el año de empalme no se afecta por el cambio de año base (lo cual no es razonable que ocurra). En general, las tasas de variación $R_{i,2,0}, R_{i,3,0}, \dots, R_{i,e-1,0}$ y $R_{i,e,1}, R_{i,e+1,1}, \dots, R_{i,N,1}$, con N siendo la última observación registrada con año base 1 se mantienen inalteradas y la única que registra el cambio es la del año de empalme, $R_{i,e,0}$; por ello, se afirma que se preserva el movimiento.

Para el trimestral, se supone que hay datos del PIB en algún trimestre de empalme para los dos años base, $i = 0, 1$. En este caso, la constante de conversión se calcula como cociente de los valores del primer trimestre en común del año de empalme, o sea:

$$c_i = Z_{i,4(e-1)+1,1} / Z_{i,4(e-1)+1,0} \quad (7)$$

Con este coeficiente se realiza el cálculo de la serie antigua, con año base 0, convertida al año base 1, como sigue:

$$Z_{i,4(T-1)+q,1} = c_i \times Z_{i,4(T-1)+q,0} \text{ para } T=1, \dots, e$$

$$\text{y } q = 1, \dots, 4, \quad (8)$$

lo cual, como se hizo notar en el caso anual, modifica el nivel del PIB trimestral con año base 0, de acuerdo con el nivel del PIB trimestral con año base 1. No obstante, el PIB con año base 0 mantiene su evolución histórica (su crecimiento) tanto anual como trimestral durante los años que abarca el periodo de conversión, $T = 1, \dots, e$, mientras que el PIB trimestral con año base 1 no modifica su comportamiento.

Desagregación univariada

La desagregación puede efectuarse con el método de Denton-Cholette (ver Dagum y Cholette, 2006), que es recomendado por organismos multinacionales (e. g. IMF, 2014) por ser automático. Por lo mismo, tiene opciones que se deciden de manera subjetiva, como la inclusión de un parámetro autorregresivo con valor 0.99³. Aquí se usa el método de Nieto (1998), el cual surge de un modelo que debe ser validado con los datos y, por ello, es menos subjetivo. La validación de los supuestos ha sido reconocida por otros autores, entre los que destacan Bikker *et al.* (2013). La desagregación, al igual que la retropolación restringida, pueden ser de tipo multivariado o univariado (ver Guerrero y Peña, 2003). El método univariado usa el vector $Y=(Y_1, \dots, Y_n)'$ que contiene los datos anuales y $Z=(Z_1, \dots, Z_{4n})'$, los trimestrales, donde el apóstrofo indica transposición. Con estas definiciones, las restricciones temporales se escriben como:

$$Y = (I_n \otimes c') \begin{pmatrix} Z_1 \\ \vdots \\ Z_{4n} \end{pmatrix} = CZ \text{ con } c' = \frac{1}{4} (1, 1, 1, 1), \quad (9)$$

donde I_n denota la matriz identidad de dimensión n , mientras que \otimes , el producto Kronecker. También, se usa el vector de estimaciones preliminares $W=(W_1, \dots, W_{4n})'$.

La representación que se postula es que la serie por estimar es una preliminar más un ruido aleatorio $Z = W + S$, con S siendo el ruido que es un proceso estacionario con media 0 y que sigue un modelo autorregresivo y de promedios móviles (ARMA, por sus siglas en inglés). Estos son los supuestos del método de Guerrero (2003), y en Nieto (1998) se supone que las series trimestrales $\{Z_t\}$ y $\{W_t\}$ se representan con el mismo modelo AR, aunque con distintas varianzas de los errores aleatorios. La no *estacionariedad*, ya sea en la tendencia o en la estacionalidad, se captura por los elementos determinísticos del modelo. De hecho, la expresión que liga al vector por estimar con la estimación preliminar es $\pi(Z-W)=e$, con e siendo un vector aleatorio de tamaño $4n$ asociado a un proceso de ruido blanco con media 0, de forma que $E(e|W)=0$ y $E(ee'|W)=\sigma_e^2 I_{4n}$. En esta expresión aparece la matriz π formada por los parámetros AR del modelo para la serie $\{W_t\}$, que se considera conocida.

Así, dados los vectores W y Y , el estimador lineal con ECM mínimo está dado por:

$$\hat{Z} = W + A(Y - CW),$$

donde $A = \pi^{-1}\pi^{-1}'C'(C\pi^{-1}\pi^{-1}'C')^{-1}$. (10)

La matriz de ECM de este estimador está dada por:

$$MSE(\hat{Z}|W, Y) = \sigma_e^2(I_{4n} - AC)\pi^{-1}\pi^{-1}'. \quad (11)$$

En estas condiciones, la aplicación de (10) se realiza después de estimar el modelo AR para la serie preliminar, de tal forma que se cuenta con los coeficientes del modelo y se pueden calcular las matrices $\hat{\pi}$ y \hat{A} , con las cuales se obtiene \hat{Z} . Además, para estimar la matriz (11), se requiere de un estimador de la varianza de la serie de discrepancias entre la estimación recién obtenida y la preliminar $\{e_t\}$. El estimador factible que obtuvo Nieto (1998) es:

$$\hat{\sigma}_e^2 = \hat{e}'\hat{e}/tr[(\hat{\pi}\hat{A}C\hat{\pi}^{-1})'], \quad (12)$$

donde $tr(\cdot)$ denota la traza de una matriz y el vector estimado de discrepancias es:

$$\hat{e} = \hat{\pi}\hat{A}(Y - CW). \quad (13)$$

La verificación de que el procedimiento empleado produce resultados óptimos ocurre si el modelo AR para la serie preliminar es adecuado. Para ello, se debe cumplir que: a) los errores del modelo sean estacionarios y b) se comporten como si fueran generados por un proceso de ruido blanco, con media 0. La estacionariedad se cumple si las raíces de la ecuación característica del polinomio AR están fuera del círculo unitario (en el plano complejo). La no-autocorrelación de los errores se verifica con el estadístico de Ljung-Box y se rechaza el supuesto cuando el estadístico produce un p -valor menor a 5 por ciento. Si eso ocurre, habrá que modificar el modelo hasta que dicho valor sea mayor o igual que ese umbral.

Nótese que se propone desagregar las series utilizando un enfoque univariado (Nieto, 1998) debido a que la calidad de las series preliminares (en este caso el ITAEE) puede considerarse una excelente aproximación oficial del PIB estatal. Por citar un ejemplo, el ITAEE para la Ciudad de México tiene una cobertura de 71.5, 100 y 79.5%, respectivamente, para cada una de las GA. De forma alternativa, pueden proponerse técnicas multivariadas, como la de Guerrero y Nieto (1999); sin embargo, otro de los objetivos de este trabajo es comparar los resultados de metodología de Nieto (1998) con respecto a otros utilizados en el campo de la estadística oficial, como es el de Denton-Cholette.

Retropolación restringida

Se presenta el método multivariado, ya que el univariado se obtiene al reducir la dimensión de $k \geq 1$ variables en general, a solo $k = 1$. Se denota a los datos trimestrales (no-observados y por estimar) de cada estado, como el siguiente vector que incluye las tres GA:

$$\mathbf{Z}_{4(T-1)+q} = (Z_{1,4(T-1)+q}, Z_{2,4(T-1)+q}, Z_{3,4(T-1)+q})'$$

para $T = 1, \dots, n$ y $q = 1, \dots, 4$. (14)

También, se puede escribir ese vector como:

$$\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})' \text{ para } t = 1, \dots, 4n \quad (15)$$

y, al apilar estos vectores, se obtiene todo el conjunto de observaciones, expresado como:

$$\mathbf{Z} = ((Z_{1,1}, Z_{2,1}, Z_{3,1}), \dots, (Z_{1,4n}, Z_{2,4n}, Z_{3,4n}))' = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_{4n})'. \quad (16)$$

Los datos anuales de las tres GA de cada estado, son:

$$\mathbf{Y}_T = (Y_{1,T}, Y_{2,T}, Y_{3,T})' \text{ para } T = 1, \dots, n \quad (17)$$

y al apilar estos datos anuales, se obtiene el vector:

$$\mathbf{Y}^T = (Y'_1, \dots, Y'_n)'. \quad (18)$$

Se define ahora la matriz de restricciones temporales:

$$C_0 = \frac{1}{4}(I_3, I_3, I_3, I_3), \quad (19)$$

donde I_3 es la matriz identidad de dimensión 3 x 3. Entonces, las restricciones temporales pueden expresarse como:

$$\mathbf{Y}^T = \begin{pmatrix} C_0 & 0 & \dots & 0 \\ 0 & C_0 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & C_0 \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_{4n} \end{pmatrix} = (I_n \otimes C_0) \mathbf{Z}. \quad (20)$$

Por otro lado, las restricciones contemporáneas para cada uno de los trimestres son tales que la suma de las cifras de las GA es igual que el total del PIB estatal; es decir, el total del estado sirve para imponer la restricción contemporánea en forma de suma, o sea:

$$Y_i^C = \sum_{i=1}^3 Z_{i,t} = \mathbf{b}' \mathbf{Z}_t \text{ con } \mathbf{b}' = (1, 1, 1). \quad (21)$$

Tales sumas forman entonces el vector de restricciones contemporáneas:

$$\mathbf{Y}^C = (Y_1^C, \dots, Y_{4n}^C)' = \begin{pmatrix} \mathbf{b}' & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{b}' & \dots & \mathbf{0}' \\ & & \ddots & \\ \mathbf{0}' & \mathbf{0}' & \dots & \mathbf{b}' \end{pmatrix} \begin{pmatrix} Z_{1,1} \\ Z_{2,1} \\ Z_{3,1} \\ \dots \\ Z_{1,4n} \\ Z_{2,4n} \\ Z_{3,4n} \end{pmatrix} = (I_{4n} \otimes \mathbf{b}') \mathbf{Z}. \quad (22)$$

Como consecuencia, los dos conjuntos de restricciones se expresan de la siguiente forma:

$$Y = \begin{pmatrix} Y^T \\ Y^C \end{pmatrix} = \begin{pmatrix} I_n \otimes C_0 \\ I_{4n} \otimes b' \end{pmatrix} Z = CZ. \quad (23)$$

Éste es, en esencia, el planteamiento que hacen Guerrero y Nieto (1999) en el contexto de desagregación temporal y contemporánea de series de tiempo múltiples. En cambio, ahora se hace mención al problema de retropolación restringida, que difiere de la desagregación tan solo en la manera como se genera la estimación preliminar. El vector de estimación preliminar de Z se define como:

$$W = (W'_1, \dots, W'_{4n})' \text{ con } W_t = (W_{1,t}, W_{2,t}, W_{3,t})' \\ \text{para } t = 1, \dots, 4n. \quad (24)$$

Para la desagregación se obtiene mediante el uso de variables auxiliares, mientras que en la retropolación restringida se utilizan pronósticos hacia atrás en el tiempo; o sea, el vector de estimación preliminar es el pronóstico irrestricto de un VAR de dimensión tres, estimado con los datos de la base estatal y trimestral de las tres GA, con el tiempo invertido.

Los vectores W y Y se consideran conocidos al aplicar los procedimientos y en lo que sigue se supondrá que $Z_t = W_t + S_t$, donde $\{S_t\}$ es un proceso estacionario vectorial que admite la misma representación VAR de $\{W_t\}$, como se justifica en el trabajo de Guerrero y Nieto (1999). Ahora, la matriz Π corresponde a la representación VAR para la serie múltiple $\{W_t\}$ que permite escribir $\Pi(Z - W) = a$ donde a es un vector de dimensión $12n$ tal que $E(a|W) = 0$ y $E(aa'|W) = P \otimes \Sigma_a$ con $\Sigma_a = E(a_t a_t' | W)$. Entonces, como se cumple el conjunto de restricciones, $Y = CZ$ y W es el estimador preliminar (el pronóstico irrestricto en su caso) de Z ; el mejor estimador lineal e insesgado de Z , basado en W y Y , está dado por:

$$\hat{Z} = W + A(Y - CW) \\ \text{con } A = \Pi^{-1}(P \otimes \Sigma_a) \Pi^{-1} C' [C \Pi^{-1}(P \otimes \Sigma_a) \Pi^{-1} C']^+ \quad (25)$$

y su correspondiente matriz de varianza-covarianza es:

$$Cov(\hat{Z} - Z | W, Y) = (I_{12n} - AC) \Pi^{-1}(P \otimes \Sigma_a) \Pi^{-1}, \quad (26)$$

donde el supraíndice $+$ denota la inversa generalizada de Moore-Penrose.

Los resultados (25) y (26) son válidos si el supuesto de que la serie de tiempo múltiple de las diferencias $\{\hat{Z}_t - W_t\}$ se comporta como ruido blanco, es decir, no presenta estructura de autocorrelación, lo que se traduce en que $P = I_{4n}$. De otra manera, no se obtiene el estimador óptimo, ya que existe otro con $P \neq I_{4n}$ que sí es óptimo porque minimiza el ECM. Para obtenerlo, Guerrero y Nieto (1999) aplican un método bietápico del tipo de mínimos cuadrados generalizados (MCG) factibles: en la primera etapa se usa el resultado previo, en el supuesto tentativo $P = I_{4n}$, y se verifica si las diferencias entre las series estimada y preliminar se comportan como ruido blanco; si esto se cumple, el procedimiento

termina y la estimación resultante es óptima. De no ser así, en la segunda fase se incorpora la estructura de autocorrelación y se aplica el procedimiento anterior, pero con la matriz P deducida de los datos de las diferencias de la primera etapa; el estimador de ésta se denomina tentativo y se mejora al construir el modelo VAR para las discrepancias.

Finalmente, la verificación de los supuestos de los modelos VAR involucrados en la retropolación restringida se realiza como en la desagregación univariada. En principio, debe verificarse la estacionariedad mediante el cálculo de las raíces de las matrices de los polinomios asociados con las ecuaciones determinantes de tales modelos. Enseguida, se debe probar que no existe estructura de autocorrelación con la prueba de Ljung-Box apropiada para el caso multivariado.

Reconciliación de cifras estatales y nacionales

Para realizar esta acción, se escriben los datos de la variable en cada trimestre τ , con orden inverso en el tiempo (del 2002:4 a 1993:1), como sigue:

$$Z^\tau = W^\tau + D^\tau, \quad (27)$$

donde, $Z^\tau = (Z_1^\tau, \dots, Z_{32}^\tau)'$, es el vector de datos estatales por estimar en el trimestre τ , con W^τ y D^τ de forma semejante. Estos vectores corresponden a la estimación preliminar (que surge de la retropolación restringida con los datos estatales) y a la discrepancia entre información estatal y nacional, respectivamente. La restricción contemporánea que se debe cumplir cada trimestre proviene de la base de datos nacional y está dada por:

$$Y^\tau = c_\tau Z^\tau, \quad (28)$$

donde se usa el vector $c_\tau = (1, 1, \dots, 1)$ de dimensión 1×32 , de manera que se satisfaga la restricción de que la suma de valores de todos los estados para la GA correspondiente sea igual que el valor nacional de esa gran actividad.

En lugar de usar la misma ponderación para todos los estados para distribuir la discrepancia entre el dato nacional Y^τ (observado) y la suma de los valores estatales $c_\tau W^\tau$ (estimados) que es lo que implica usar $\frac{1}{32} c'_\tau$, se utiliza un promedio ponderado por la proporción de GA de cada estado en el total nacional. Esto se logra al definir las ponderaciones:

$$p_{j,\tau} = W_{j,\tau} (c_\tau W^\tau)^{-1} \text{ para } j=1, \dots, 32, \quad (29)$$

donde $W_{j,\tau}$ es el valor de la GA en el estado j y el trimestre τ , proveniente de la base de datos estatal estimada por retropolación restringida, mientras que $c_\tau W^\tau = \sum_{j=1}^{32} W_{j,\tau}$ es el total nacional estimado de la gran actividad. Por ello, en la expresión anterior se sustituye al vector $\frac{1}{32} c'_\tau = (\frac{1}{32}, \dots, \frac{1}{32})'$ por $p_\tau = (p_{1,\tau}, \dots, p_{32,\tau})' = W^\tau (c_\tau W^\tau)^{-1}$, con lo que se obtiene:

$$\hat{Z}^\tau = W^\tau + p_\tau (Y^\tau - c_\tau W^\tau), \quad (30)$$

que cumple con la restricción contable, o sea, $c_\tau \hat{Z}^\tau = c_\tau W^\tau + c_\tau p_\tau (Y^\tau - c_\tau W^\tau) = Y^\tau$, ya que $c_\tau p_\tau = \sum_{j=1}^{32} W_{j,\tau} (c_\tau W^\tau)^{-1} = 1$.

Aplicación empírica de los métodos

Para ilustrar los resultados que surgen de los métodos descritos, en este apartado se muestran los que se obtuvieron para la Ciudad de México, cuya proporción de actividad terciaria es la más alta e importante del país (86% en el 2013), expresados en millardos de pesos, a precios constantes del 2008. Una versión más amplia y detallada de este artículo aparece en el documento de investigación del INEGI, que considera los 32 estados del país (ver Guerrero y Corona, 2017).

Conversión de año base 1993 a 2008

Los datos de esta fase se muestran en la figura 3, donde aparecen las series generadas de manera indirecta por agregación de los sectores a GA y las que se obtienen por conversión directa de las grandes actividades. La apropiada es la indirecta, cuyos resultados fueron considerados razonables por funcionarios del INEGI a cargo del Sistema de Cuentas Nacionales de México.

Desagregación temporal univariada de cada GA y del total

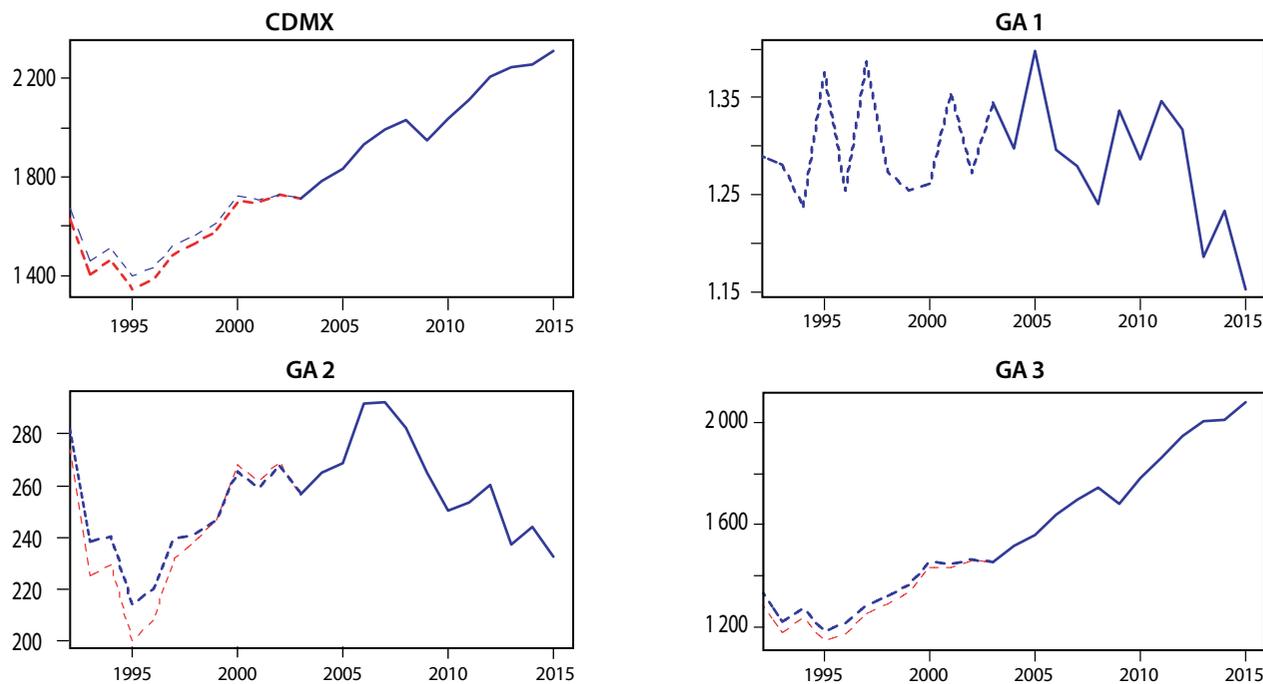
Ésta se aplicó a cada una de las GA en forma individual y al sumar sus resultados se obtuvo la desagregación del total estatal. El método aplicado se basa en modelos estadísticos y, por ende, se requiere validar los supuestos que los sustentan. De manera específica, los imprescindibles son: estacionariedad del modelo (que se valida si las raíces del polinomio AR están fuera del círculo unitario y no-autocorrelación del error, que se verifica con la prueba de Ljung-Box) y el de no-correlación cruzada entre residuos del modelo para la serie preliminar con las discrepancias entre las series preliminar y desagregada. De no cumplirse este supuesto, lo que se debe hacer es buscar un mejor modelo para la preliminar o sustituirla. Sin embargo, si ya se verificó que el modelo para esta serie se justifica de forma empírica, se debería buscar una preliminar alternativa, lo cual en este caso no es posible, porque solo existe la de la base de datos oficial.

La figura 4 muestra las series desagregadas para la Ciudad de México, donde se observan resultados muy parecidos con los métodos de desagregación de Nieto y de Denton-Cholette, excepto en el caso de GA 1. Para ésta, se ven discrepancias de mayor magnitud entre las series preliminar y desagregada con el segundo método, por lo que se prefiere el método de Nieto para la desagregación univariada, pues preserva mejor el movimiento de la serie preliminar. Al sumar los resultados por trimestre, de cada una de las GA se obtiene la serie desagregada por trimestre del total de la actividad económica del estado.

Para los modelos AR de las series preliminares de las GA de la Ciudad de México, se obtuvieron los resultados del cuadro 1; en él se aprecia que se cumple el supuesto de estacionariedad, pues todas las raíces de los modelos AR están fuera del círculo unitario. También, se cumple el supuesto de no-autocorrelación del error y todos los coeficientes son significativamente distintos de 0 al compararlos con sus respectivos errores estándar. De esta forma, se concluye que estos modelos son válidos de forma razonable (desde el punto de vista de la teoría estadística) y sus resultados tienen soporte empírico. Por su lado, la validación del método de Nieto indica verificar ausencia de correlación entre los residuos del modelo para las series preliminar y de discrepancias entre la desagregada y la preliminar: GA 1, $Q = 7.43$, p -valor = 0.96; GA 2, $Q = 5.49$, p -valor = 0.99; GA 3, $Q = 7.48$, p -valor = 0.96, que no muestran significancia y brindan soporte empírico al método.

Figura 3

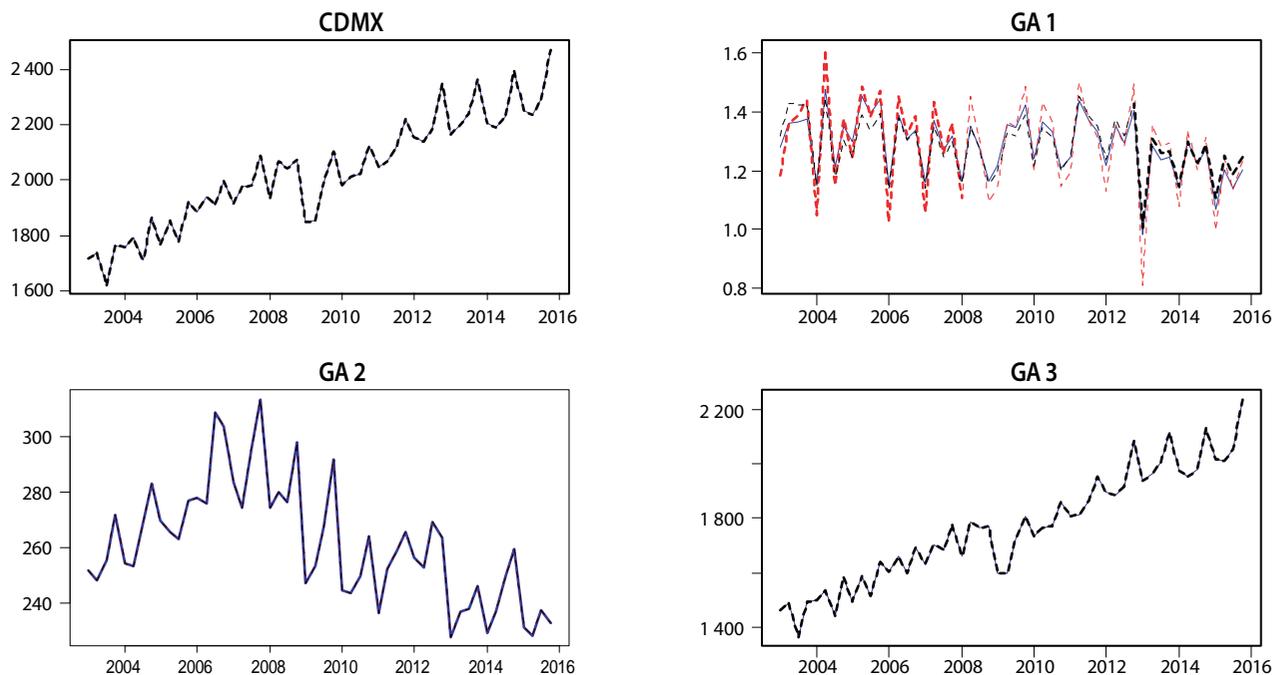
Conversión por grandes actividades para la Ciudad de México



Nota: datos originales con base 2008 (línea azul sólida), serie convertida en forma directa (línea roja) y serie convertida en forma indirecta (línea azul punteada)

Figura 4

Desagregación por gran actividad para la Ciudad de México



Nota: preliminar (línea negra), Denton-Cholette (línea roja) y Nieto (línea azul).

Retropolación restringida hasta 1993

Los datos desagregados se usan para construir modelos VAR trivariados que generen pronósticos hacia atrás, del cuarto trimestre del 2002 al primero de 1993, con origen en el trimestre I del 2003. Para ello, se consideran las tres GA de manera simultánea; lo primero que se hace es transformar los datos mediante la aplicación del logaritmo natural a cada una de las series que entran al modelo, lo cual se hace con el fin de evitar que los pronósticos tomen valores negativos y porque, al transformar de esta manera, también se mejora la estabilidad de la varianza del error involucrado. Después de obtener los pronósticos en la escala logarítmica, se retransforman a la escala original de las GA con el antilogaritmo.

Luego, se elige el orden del modelo VAR, para lo cual se usa el paquete de Pfaff (2008), que se basa en aplicar los criterios de información de Akaike, Schwarz y Hannan-Quinn. Después, se especifica la parte determinística del modelo, es decir, si lleva constante, tendencia, ambas o ninguna y si se deben incluir variables artificiales para capturar la estacionalidad de los datos. El modelo VAR estimado para cada estado se somete a la verificación de estacionariedad y no-autocorrelación residual.

El cuadro 2 muestra los resultados de la estimación del modelo VAR de orden 1 que produce los pronósticos irrestrictos. Éste incluye constante y variables artificiales para la estacionalidad, que son significativas para explicar a cada una de las tres GA. El efecto estacional es básicamente lo que hace que el coeficiente R^2 ajustado por grados de libertad sea tan alto en las tres ecuaciones. A GA 1 la explica de forma adicional GA 3, y GA 2 es explicada por sí misma, mientras que GA 3 no lo es por ninguna de las otras, lo cual indica que se comporta de manera prácticamente exógena. Por otro lado, los supuestos de estacionariedad y no-autocorrelación del error tienen soporte empírico y el modelo puede considerarse estadísticamente válido.

La prueba para verificar que la serie múltiple de discrepancias entre valores estimados y preliminares se comporta como ruido blanco brindó como resultado: valor de la J_i -cuadrada = 362.61, con 135 grados de libertad, lo que produjo el p -valor = 0.00, por lo que se decidió aplicar la segunda etapa. En ésta se produjeron los resultados de la estimación del modelo VAR(1) para las discrepancias, que se muestran en el cuadro 3. El modelo resumido en él cumple con los supuestos de estacionariedad y de no-autocorrelación del error, por lo que se le considera estadísticamente válido. Además, los pronósticos irrestrictos y restringidos se muestran en la figura 5, donde se aprecia el beneficio de incorporar las restricciones, ya que los pronósticos irrestrictos solo marcan la tendencia y la estacionalidad de las respectivas series de GA y carecen de credibilidad.

Reconciliación con las cifras trimestrales nacionales 1993-2002

La reconciliación de las bases de datos estimadas mediante retropolación restringida se aplica a cada una de las GA para incorporar la información de la base de datos nacional. Como resultado, se obtiene un ajuste de los datos retropolados tal que la suma de valores de cada trimestre brinda el total de las tres grandes actividades de la entidad respectiva. De igual forma, se obtiene el promedio de los valores trimestrales para cada una de las GA. Lo importante a reconocer aquí es que el patrón de las series obtenidas con la retropolación restringida sufre algunas modificaciones, como puede apreciarse en las gráficas de la figura 6.

Cuadro 1

Validación de modelos AR para series preliminares

GA 1	$\hat{W}_{1,t} = 1.32 + 0.34W_{1,t-2} - 0.14d_{1,t} + 0.05d_{2,t} - 0.02d_{3,t}$
	(0.02) (0.14) (0.03) (0.02) (0.03)
	con $\hat{\sigma}_{W,1}^2 = 0.005$, raíces: 1.72, -1.72
	$Q(16) = 16.20$, p -valor = 0.44
GA 2	$\hat{W}_{2,t} = 271.48 + 0.87W_{2,t-1} - 24.21d_{1,t} - 22.16d_{2,t} - 10.67d_{3,t}$
	(9.88) (0.07) (2.71) (3.06) (2.64)
	con $\hat{\sigma}_{W,2}^2 = 104.8$, raíz: 1.15
	$Q(16) = 24.24$, p -valor = 0.08
GA 3	$\hat{W}_{3,t} = 1\ 870.66 + 0.84W_{3,t-1} + 0.56W_{3,t-4} - 0.41W_{3,t-5} - 98.30d_{1,t} - 84.16d_{2,t} - 108.20d_{3,t}$
	(280.49) (0.09) (0.13) (0.14) (24.03) (27.23) (23.70)
	con $\hat{\sigma}_{W,3}^2 = 2\ 062$, raíces: 1.01, -1.17, 0.02-1.17i, 1.49, 0.02+1.17i
	$Q(16) = 11.14$, p -valor = 0.80

Cuadro 2

Resultados de la estimación del modelo VAR(1) para generar pronósticos irrestrictos (estimaciones preliminares)

Variable explicada	Significancia de pruebas F para explicar variabilidad				R^2 ajustada
	GA 1	GA 2	GA 3	Estacionalidad	
GA 1	0.46	0.24	0.01	0.00	0.95
GA 2	0.39	0.00	0.63	0.00	1.00
GA 3	0.48	0.69	0.31	0.00	1.00

Raíces de la ecuación determinante para estacionariedad: 0.75, 0.34, 0.08.

Prueba de no-autocorrelación: Ji -cuadrada (135) = 142.12, p -valor = 0.32.

Cuadro 3

Resultados de la estimación del modelo VAR(1) para la serie múltiple de discrepancias (segunda fase de MCG)

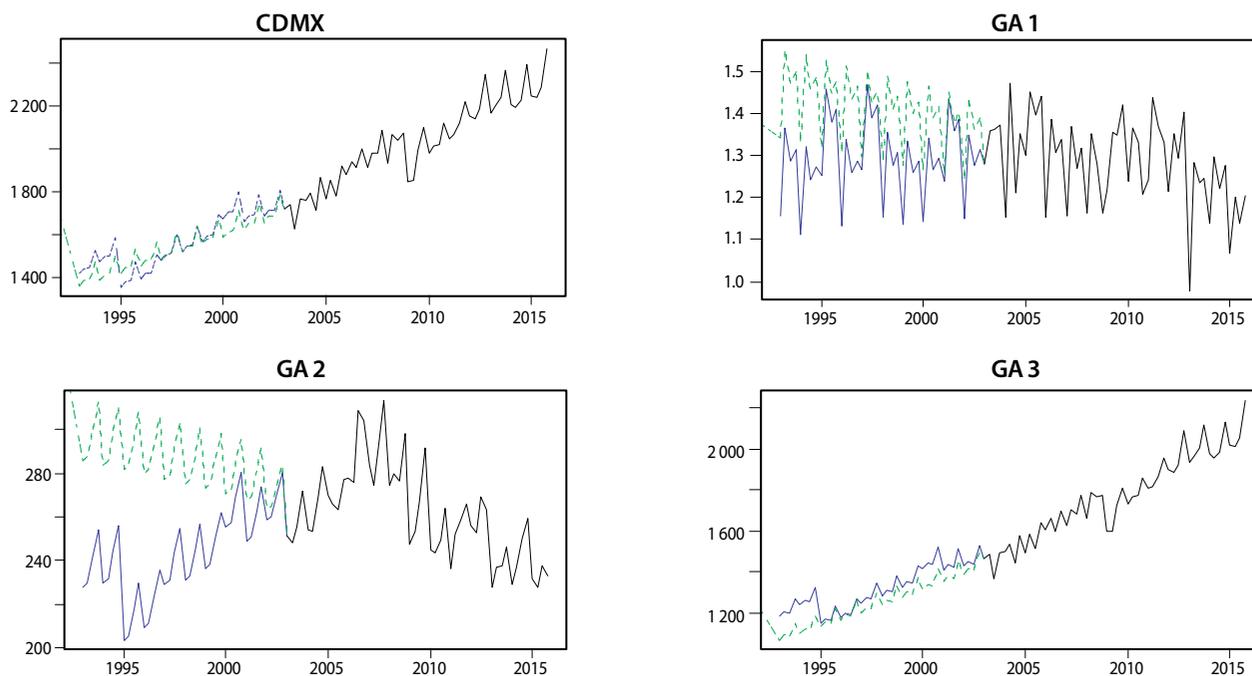
Discrep. explicada	Pruebas F de significancia para explicar variabilidad				R^2 ajustada
	Disc(GA 1)	Disc(GA 2)	Disc(GA 3)	Estac.	
Disc(GA 1)	0.01	0.38	0.33	0.44	0.48
Disc(GA 2)	0.14	0.01	0.52	0.20	0.92
Disc(GA 3)	0.25	0.37	0.41	0.60	0.64

Raíces de la ecuación determinante para estacionariedad: 0.81, 0.81, 0.65.

Prueba de no-autocorrelación: Ji -cuadrada (135) = 102.55, p -valor = 0.98.

Figura 5

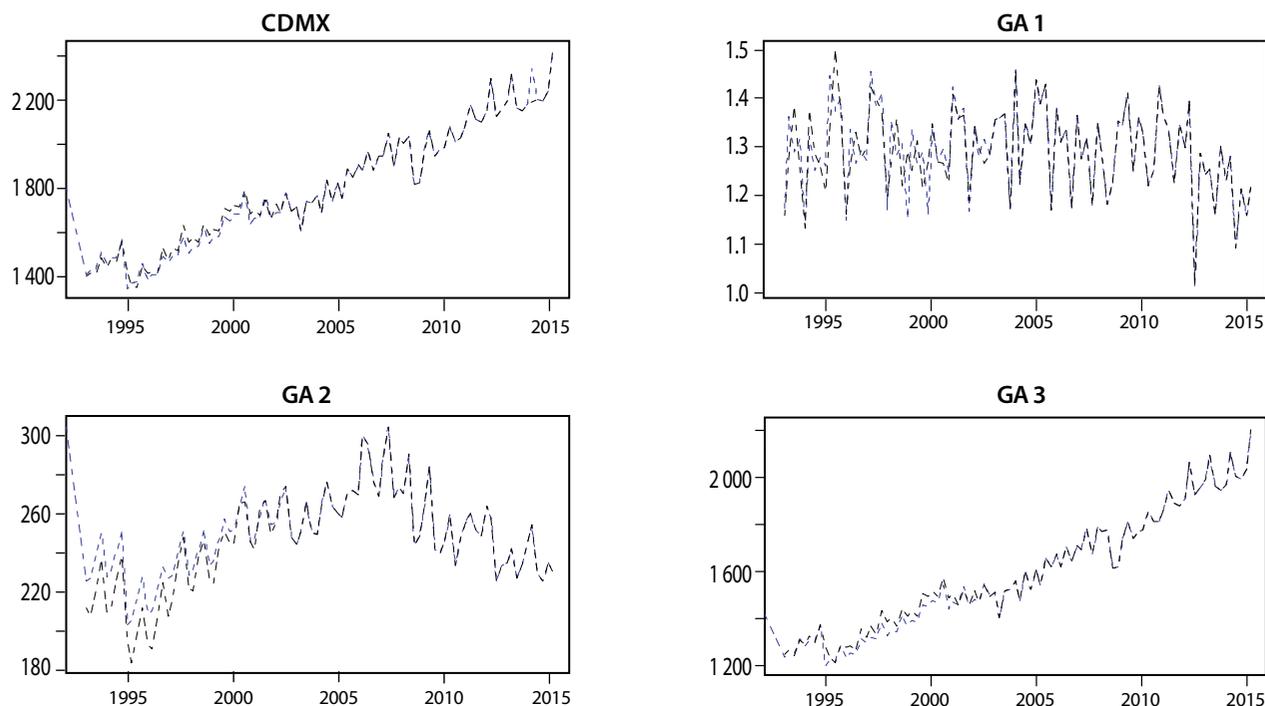
Retropolación restringida multivariada: GA y total de la Ciudad de México



Nota: desagregada (línea negra), pronósticos irrestringidos (línea verde) y restringidos (línea azul).

Figura 6

Reconciliación de cifras estatales y nacionales: GA y total de la Ciudad de México



Nota: retropolada (línea azul punteada) y reconciliada (línea negra).

La misma figura 6 muestra series más creíbles, ya que su patrón dinámico no se observa extraño en algún sentido, además de que satisfacen la restricción de que la suma de todos los estados es el total del país. Por otro lado, cabe destacar el hecho de que la GA 1 se ve alterada muy poco (en términos relativos) al aplicar la reconciliación. En cambio, la GA 2 es la que (de nuevo, en términos relativos) se ve más afectada al reconciliar los datos de cada una de las tres entidades con los datos de la base nacional. Para enfatizar el hecho de que las cifras reconciliadas cumplen con la restricción contable de que la suma de valores de las tres GA brinda el PIB total del estado de cada trimestre (excepto por redondeos en la presentación de las cifras), se presenta el cuadro 4. Algo que debe resaltarse es que el promedio anual sigue la dinámica del PIB estatal anual convertido de año base 1993 a 2008. Además, la suma de valores de las GA por entidad es igual al total de la base de datos nacional.

El total del país se obtiene como suma de los resultados estatales de las respectivas GA excepto para la 1. En consulta directa con los funcionarios encargados de calcular el PIB de manera oficial en México, se mencionó que: "En los datos base 2008 coinciden 733 de las 734 actividades económicas con las que se integran los cálculos del PIB, a excepción de la agricultura, que en el ITAEE se mide por *año calendario*; en el PIB por entidad federativa se mide por *año agrícola* y en el PIB trimestral por *año calendario*, por ello no debe coincidir el ITAEE con el PIB del estado..." (Lourdes Mosqueda, 2017, comentario hecho de manera personal a los autores). El *año agrícola* que se usa en el subsector de agricultura se refiere al hecho de que la producción (desde la preparación de la tierra hasta el levantamiento de la cosecha) abarca más de un año calendario y el valor agregado de los cultivos se considera proporcional al costo de los insumos empleados en la producción, como la fuerza de trabajo y los insumos intermedios, lo cual conduce a distribuir el valor total de la producción en proporción a los costos incurridos cada trimestre (ver INEGI, 2013). Por ello, las series reconciliadas de la GA 1 para cada estado se pueden ajustar para satisfacer el criterio usado en el INEGI durante el periodo 1993-2002 y obtener valores referidos al *año agrícola*.

Retropolación restringida hasta 1980

Para la retropolación trimestral hasta 1980 se usa la base de datos que surge de la reconciliación para 1993-2002, como se indica en el apartado anterior. Ahora, se utiliza la base de datos nacional $PIB(Nac, Tri, GDiv + Div; 80-98)_{93}$ que emplea la clasificación del CIIU de 1993 para obtener por conversión la base $\overline{PIB}(Nac, Tri, GA; 80-92)_{08}$. Esta última brinda las restricciones lineales por satisfacer (la suma de los valores de los 32 estados en cada una de las GA debe ser el total nacional de la gran actividad respectiva) cuando se aplica la retropolación restringida, de manera semejante a lo que se muestra en el esquema de la figura 2. Las series a nivel país obtenidas por conversión directa se muestran en las gráficas de la figura 7, para cada una de las tres GA y el total.

Los pronósticos irrestrictos se obtienen mediante tres modelos VAR de dimensión 32, uno para cada GA, en los que se introduce la estructura regional utilizada en el INEGI para dividir a México en cinco regiones. De esta forma, lo que se obtiene es una estructura de modelo VAR por bloques, constituidos por las regiones del país, la cual sirve para que solo las variables de los estados dentro de una región puedan tener efecto sobre las de la misma zona. El empleo de la regionalización equivale a trabajar con cinco modelos VAR para cada una de las GA, sin embargo, para introducir la restricción de que la suma de los valores estatales de cada gran actividad produzca el total nacional, se requiere usar un solo VAR de dimensión 32 para cada GA; es por ello que se mantiene la referencia a un modelo VAR por gran actividad. La estimación de

Reconciliación de cifras trimestrales para años seleccionados de 1993 al 2002

Actividad	Trimestre de 1993				Prom. anual	...	Trimestre del 2002				Prom. anual
	1	2	3	4			1	2	3	4	
GA 1	1.2	1.3	1.4	1.3	1.3		1.2	1.4	1.3	1.3	1.3
GA 2	212.8	208.4	222.3	239.4	220.7		253.9	261.0	276.1	280.6	267.9
GA 3	1 195	1 215	1 192	1 260	1 216		1 425	1 490	1 436	1 516	1 467
Total	1 409	1 425	1 416	1 501	1 438		1 680	1 752	1 714	1 798	1 736

los tres VAR que contienen a las GA de la región centro (constituida por la Ciudad de México y el estado de México) produce los resultados del cuadro 5, que se deben interpretar como ecuaciones individuales por fila, de manera que la primera indica que la ecuación de la GA 1 de la Ciudad de México emplea dos retrasos de las variables y se explica de forma significativa por las GA 1 de ella misma y del estado de México, así como por efectos estacionales, con el valor de R^2 ajustada de 100 por ciento. Lo mismo ocurre con las grandes actividades 2 y 3 de la Ciudad de México.

Dicho cuadro muestra, también, que el supuesto de no-autocorrelación del error se satisface razonablemente bien en los tres casos. Por su lado, la verificación de estacionariedad de los modelos se realiza con las raíces de la ecuación determinante para cada uno de los cinco VAR de cada GA. Esta verificación se resume en los valores del cuadro 6 para las tres GA. En todos los casos, las raíces son menores que la unidad, lo cual valida la estacionariedad del modelo respectivo.

Después de validar los supuestos fundamentales de cada ecuación, se les considera instrumentos útiles para capturar las regularidades empíricas de los datos; la inspección visual de los pronósticos, irrestrictos y restringidos les brinda credibilidad empírica. El resultado final de este ejercicio se resume con las gráficas de las series retropoladas en forma restringida, junto con sus intervalos de predicción de dos errores estándar, que aparecen en la figura 8.

Cambio de año base a 2013 y actualización de datos al 2016

En todos los estados del país se aplicaron las mismas herramientas de análisis y se obtuvieron resultados semejantes a los que aquí se muestran para la Ciudad de México. En particular, se impuso la restricción de que la agregación de dichos resultados proporcionara los valores que se obtuvieron por conversión de las bases de datos con año base 1993 al nuevo 2013, para el total nacional. Las figuras 9 y 10 presentan los resultados de la retropolación restringida para la Ciudad de México, considerando el cambio de año base de 2008 a 2013 y la actualización de la información al 2016, los cuales se dan en niveles (millardos de pesos), así como en variaciones porcentuales anuales, para observar tendencias y crecimientos anuales, de manera respectiva. La figura 9 permite observar que la GA 3 se comporta de manera más irregular e independiente de las otras, mientras que la 1 muestra un patrón estacional muy marcado, pero consistente para todos los años considerados. Por su lado, la figura 10 presenta patrones de fluctuaciones parecidos para todo el periodo y cada una de las grandes actividades.

Cuadro 5

Resultados de la estimación de modelos VAR(2) para GA 1, VAR(3) para GA 2 y VAR(3) para GA 3 usados en la retopología restringida

Variable explicada	Significancia de las pruebas F para explicar variabilidad con las GA			R^2 ajustada
	CDMX	Méx.	Estac.	
GA 1	0.00	0.00	0.00	1.00
Ljung-Box $Q(16) = 24.60, p\text{-valor} = 0.08$				
GA 2	0.00	0.00	0.00	1.00
Ljung-Box $Q(16) = 16.76, p\text{-valor} = 0.40$				
GA 3	0.00	0.00	0.00	1.00
Ljung-Box $Q(16) = 20.27, p\text{-valor} = 0.21$				

Cuadro 6

Raíces de la ecuación determinante de los modelos VAR para la región centro: dos raíces por cada ecuación y estado para GA 1, tres para GA 2 y tres para GA 3

Actividad	Raíces de la ecuación determinante					
	GA 1	0.95	0.62	0.57	0.42	
GA 2	0.96	0.93	0.47	0.37	0.27	0.27
GA 3	0.99	0.76	0.69	0.69	0.31	0.31

Figura 7

Conversión de año base 1993 a 2008, de los datos trimestrales a nivel nacional para las tres GA y el total de la actividad económica durante 1980-1992

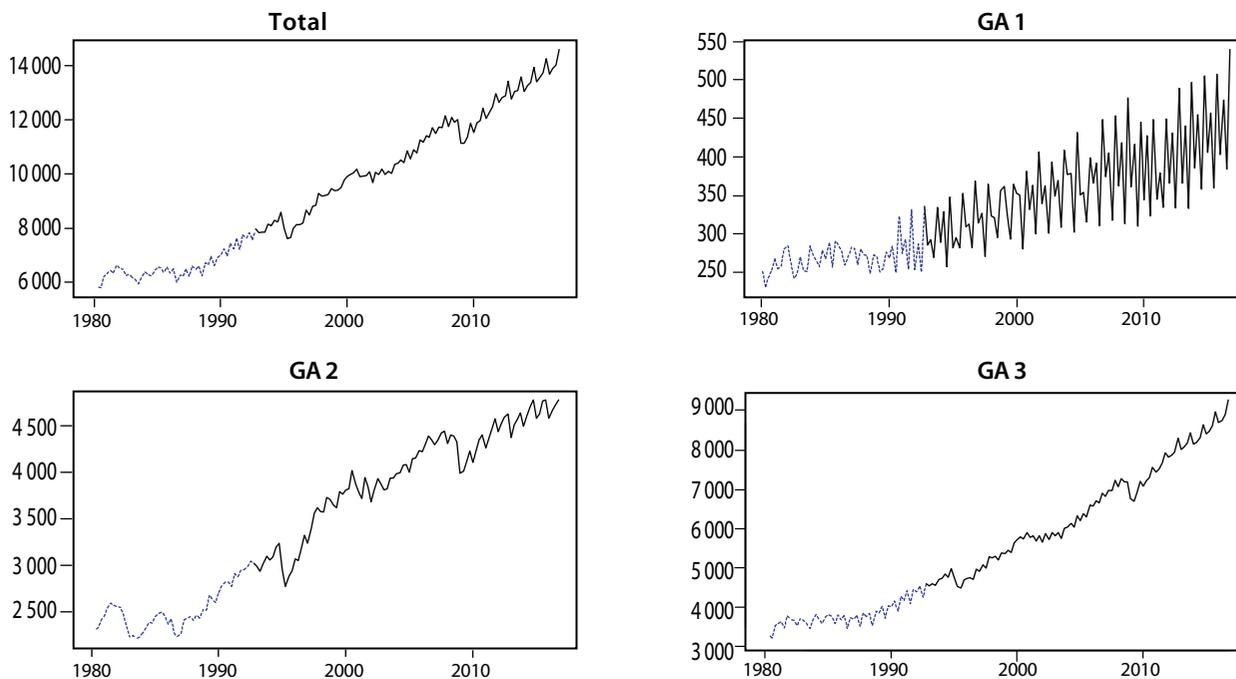
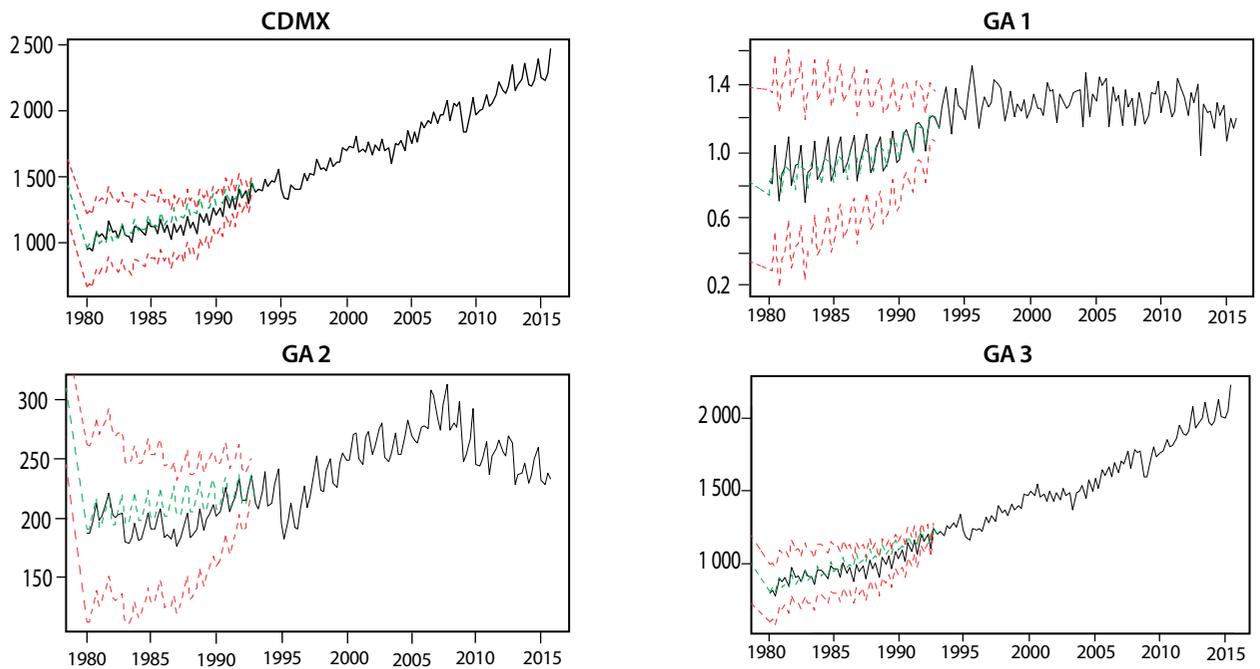


Figura 8

Retropolación hasta 1980 por GA y total



Nota: restringida (línea negra sólida), con intervalos e irrestricta (líneas punteadas).

Figura 9

Retropolación hasta 1980 por GA y total para la Ciudad de México (millardos de pesos del 2013)

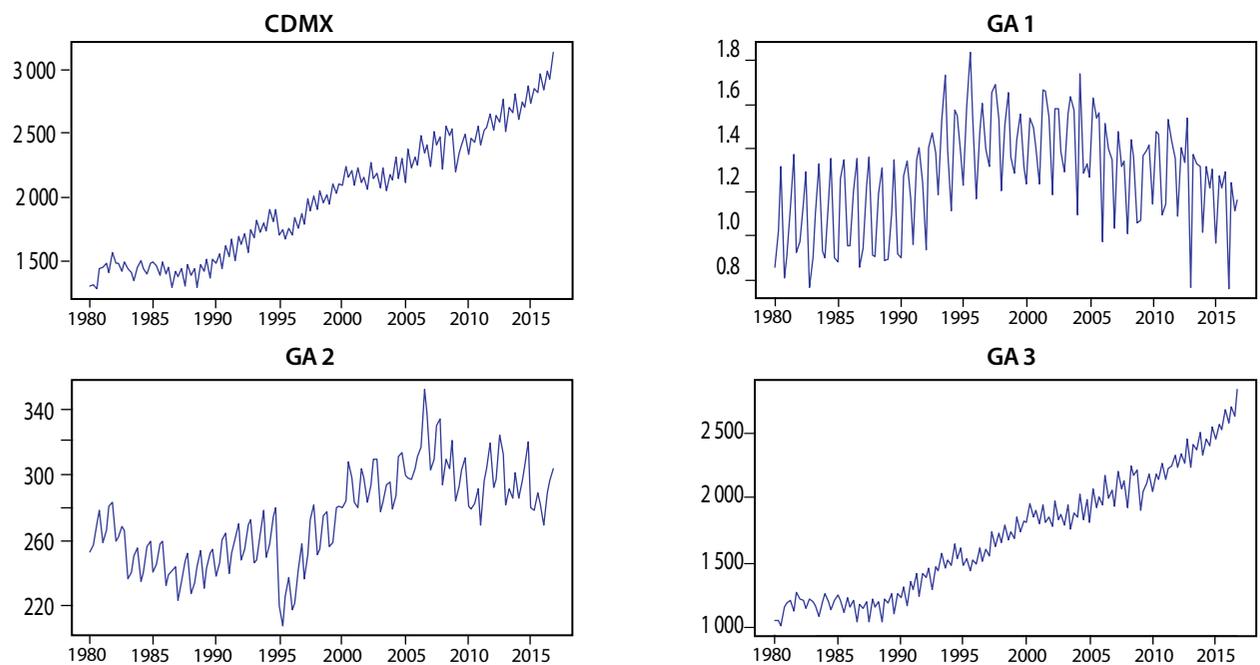
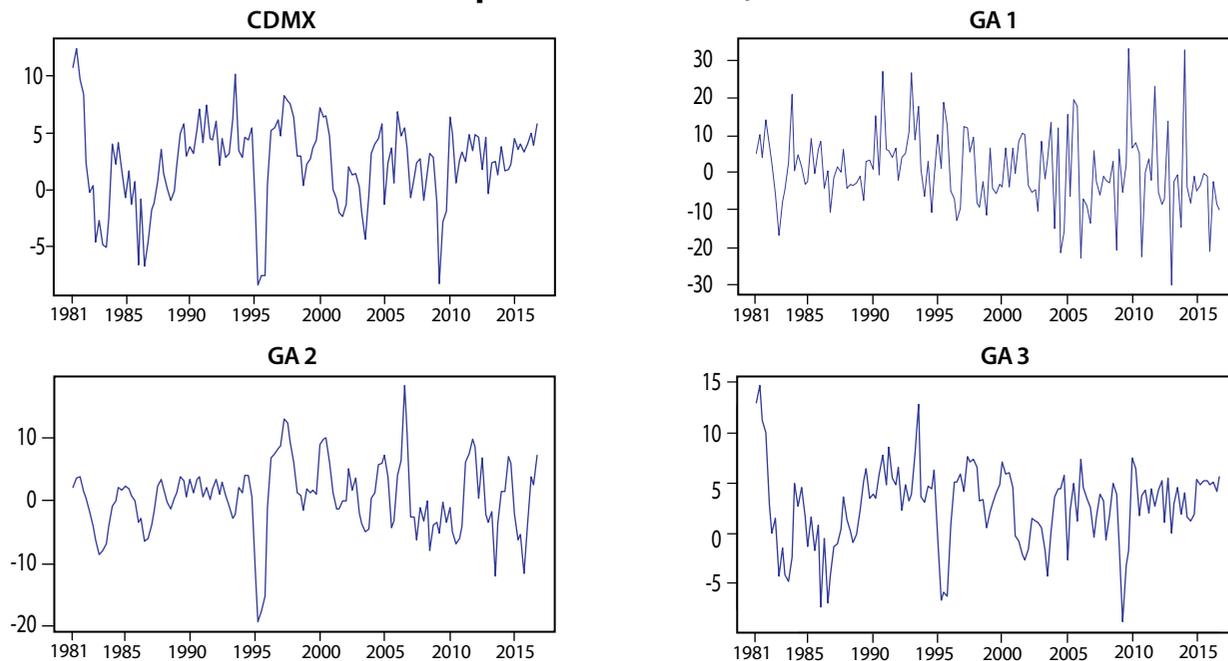


Figura 10

Retropolación hasta 1980 por GA y total para la Ciudad de México. Variaciones porcentuales anuales, año base 2013



En términos de la interpretación económica, los resultados fueron validados de forma exitosa realizando un análisis que relaciona el comportamiento de las series retropoladas con acontecimientos relevantes para México y la entidad, como pueden ser crisis petroleras o financieras, devaluaciones de la moneda, implementación de políticas económicas y desastres naturales, entre otros. En específico, se evaluó que el año donde sucedieron dichos acontecimientos tuviesen una relación económicamente directa con las fechas de las crestas o valles que presentan los cambios porcentuales de las series retropoladas.

Conclusiones

La más importante es, desde luego, que se logró combinar de manera óptima en términos estadísticos todas las bases de datos oficiales disponibles y se generó, así, una sola base *completa y homogénea* con cifras compatibles entre sí por estados, trimestres, actividades económicas y año base, por lo que se cumplió también con las restricciones impuestas por la contabilidad nacional. Con esto se considera que el trabajo fue exitoso y sus resultados se podrían dar a conocer al público que requiere de este tipo de información para elaborar análisis macroeconómicos de diversos tipos, ya sea históricos, regionales, etcétera.

Los métodos utilizados en esta investigación producen resultados que reflejan, en esencia, la información que contiene los datos usados como insumos, mientras que los modelos estadísticos que respaldan a las metodologías se conciben como herramientas que permiten a aquéllos expresarse con libertad. Los datos empleados provienen de las fuentes oficiales disponibles, de manera que, si se modifican éstas por revisiones o actualizaciones del tipo que sea, deberán alterarse también las cifras estimadas que aquí se reportan y las metodologías mostradas podrían usarse para ese fin.

Como línea futura de estudio se propone el empleo de metodologías que usen modelos no estacionarios, para lo cual se puede utilizar, por ejemplo, la de Guerrero (1990) en la desagregación temporal y los modelos de vectores con corrección de error (VEC) para la retropolación multivariada. Asimismo, ya que el objetivo de este trabajo fue retropolar de manera restringida con ayuda solo de información oficial, se pueden incorporar los resultados de German-Soto (2005) como restricciones de carácter temporal para obtener series estatales y retropoladas de forma restringida hasta 1940. Esto coadyuvaría para tener información regional en periodos anteriores a 1992. Por último, se podrían utilizar también variables exógenas en la fase de realización de pronósticos irrestrictos.

Fuentes

- Arbués I. y N. López. *On the error of backcast estimates using conversion matrices under a change of classification (working papers 02/2011)*. España, Instituto Nacional de Estadística, 2011.
- Bassols-Batalla, Á. *México, formación de regiones económicas. Influencias, factores y sistemas*. México, Universidad Nacional Autónoma de México, Instituto de Investigaciones Económicas, 1983.
- Bikker, R., J. Daalman and N. Mushkudiani. "Benchmarking large accounting frameworks: A generalized multivariate model", en: *Economic Systems Research*. Vol. 25, 2013, pp. 390-408.
- Buiten, G., J. Kampen, A. Neef y S. Vergouw. *Producing historical time series for STS-statistics in NACE Rev. 2: Theory with an application in industrial turnover in the Netherlands (1995-2008) (discussion paper 09001)*. Statistics Netherlands, 2009.
- Caporin, M. y D. Sartore. *Methodological aspects of time series back-calculation (Eurostat Working Papers and Studies)*. Luxembourg, Eurostat, 2005.
- Castillo, N. E. "El Sistema de Clasificación Industrial de América del Norte (SCIAN), ¿un traje hecho a la medida?", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 4, núm. 3, 2013, pp. 74-89.
- Correa, S. V., A. A. Escandón, P. R. Luengo y M. J. Venegas. "Empalme de series anuales y trimestrales del PIB", en: *Economía Chilena*. Vol. 6, núm. 1, 2003, pp. 77-86.
- Dagum, E. B. y P. A. Cholette. "Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series", en: *Lecture Notes in Statistics 186*. New York, Springer-Verlag, 2006.
- Departamento Administrativo Nacional de Estadística (DANE). *Documento metodológico y resultados de la retropolación 1975-2005. Base 2005*. Bogotá, Colombia, DANE, 2013.
- Di Fonzo, T. y M. Marini. *Benchmarking a system of time series: Denton's movement preservation vs. a data based procedure (Eurostat Working Papers and Studies)*. Luxembourg, Office for Official Publications of the European Communities, 2005.
- Gallegos-Rivas, A. E., I. Santacruz-Villaseñor y G. Tapia-Tovar. "Aspectos metodológicos de estudios del PIB en series de tiempo largas", en: *Mundo Siglo XXI*. México, CIECAS-IPN. Vol. XI (37), 2015, pp. 89-96.
- German-Soto, V. "Generación del producto interno bruto mexicano por entidad federativa, 1940-1992", en: *El Trimestre Económico*. Vol. 72 (287), 2005, pp. 617-653.
- Guerrero, V. M. "Temporal Disaggregation of Time Series: An ARIMA-based Approach", en: *International Statistical Review*. Vol. 58, no. 1, 1990, pp. 29-46.
- _____. "Monthly Disaggregation of a Quarterly Time Series and Forecasts of Its Unobservable Monthly Values", en: *Journal of Official Statistics*. Vol. 19, no. 3, 2003, pp. 215-235.
- Guerrero, V. M. y F. Corona. *Retropolación óptima de series de tiempo de las tres grandes actividades económicas de México, por estado y trimestre, a precios constantes de 2013, para 1980-2016 (documento de investigación)*. DGAI-DGIAI, 17-02. Ciudad de México, INEGI, 2017.
- Guerrero, V. M. y F. H. Nieto. "Temporal and contemporaneous disaggregation of multiple economic time series", en: *Test*. Vol. 8, 1999, pp. 459-489.
- Guerrero V. M. y D. Peña. "Combining multiple time series predictors: A useful inferential procedure", en: *Journal of Statistical Planning and Inference*. Vol. 116, 2003, pp. 249-276.
- Hellberg, O. *Backcasting Swedish Industrial Production (paper presented during the Workshop on Survey Sampling Theory and Methodology)*. Lithuania, Vilnius, 2010.
- IMF. "Chapter 6. Benchmarking and reconciliation", en: *Quarterly National Accounts Manual: Concepts, Data Sources and Compilation (update)*. Washington, International Monetary Fund, 2014.
- INEGI. *Sistema de Clasificación Industrial de América del Norte, México (SCIAN) 2007*. Aguascalientes, México, INEGI, 2007.

- _____. *Sistema de Cuentas Nacionales de México. Series históricas del producto interno bruto trimestral, 1980.I a 1998.I. Base 1993*. Aguascalientes, México, Instituto Nacional de Estadística, Geografía e Informática, 1998.
- _____. *Sistema de Cuentas Nacionales de México. Indicador trimestral de la actividad económica estatal. Fuentes y metodologías*. Aguascalientes, México, Instituto Nacional de Estadística y Geografía, 2013.
- James, G. *Backcasting, for use in Short Term Statistics*. UK Office for National Statistics, 2008.
- Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. Germany, Springer-Verlag, 2005.
- Nieto, F. H. "Ex-post and Ex-ante Prediction of Unobserved Economic Time Series: A Case Study", en: *Journal of Forecasting*. Vol. 17, 1998, pp. 35-58.
- ONU. "Clasificación Internacional Industrial Uniforme de todas las actividades económicas (CIIU)", en: *Informes estadísticos. Serie M, número 4, revisión 3*. New York, ONU, 1990.
- Ponce, J. "Una nota sobre empalme y conciliación de series de cuentas nacionales", en: *Revista de Economía*. Segunda época. Vol. XI, núm. 2, 2004, pp. 178-210.
- Parrot, F. y R. McKenzie. "Linking factors for gross and seasonally adjusted series", en: *Note, Short Term Economic Statistics Division*. OECD, 2003.
- Pfaff, B. "VAR, SVAR and SVEC Models: Implementation Within R Package vars", en: *Journal of Statistical Software*. 27, no. 4, 2008.
- Roulin, E. y U. Eidmann. *Back Casting Handbook*. Luxembourg, Eurostat, 2007.
- Yuskavage, R. E. *Converting historical industry time series data from SIC to NAICS (paper prepared for the Federal Committee on Statistical Methodology 2007 Research Conference)*. Arlington, VA, EE.UU., 2007.