

# Análisis estadístico de series de tiempo económicas *generadas con datos oficiales*

Víctor M. Guerrero

En este artículo se presentan algunos problemas que enfrentan los institutos nacionales de estadística en relación con la generación, difusión y análisis de los datos económicos oficiales de series de tiempo. Se mencionan las herramientas de análisis estadístico que pudieran aplicarse para darles solución y, aunque deberían ser aplicadas de preferencia por las propias instituciones, algunas podrían también ser empleadas por los usuarios interesados en la información. Además, se enuncian diversas circunstancias que dan origen a los problemas tratados y se ilustran de forma somera algunas aplicaciones con datos generados por el Instituto Nacional de Estadística y Geografía (INEGI) de México.

**Palabras clave:** ajuste estacional, desagregación, estimación rápida, imputación, índices de opinión, suavizamiento.

This paper presents some problems faced by national statistical institutes with regard to the generation, dissemination and analysis of official economic time series. The statistical analysis tools that may be applied to solve those problems are just mentioned. These tools should be preferably used by statistical institutes but, if needed, some may be employed by the information user. Several circumstances that give rise to such problems are also considered and the solutions are illustrated with data generated by Mexico's National Institute of Statistics and Geography (INEGI).

**Key words:** disaggregation, imputation, opinion indices, rapid estimates, seasonal adjustment, smoothing.

**Nota:** el autor desea agradecer la invitación hecha por Eduardo Sojo, presidente del INEGI, para escribir este artículo; asimismo, hace patente su agradecimiento a Blanca R. Sainz quien, junto con su equipo de trabajo, se encarga de realizar en buena medida los análisis aquí descritos para el INEGI; sus comentarios, al igual que los de Esperanza Sainz, ayudaron a aclarar las aplicaciones aquí descritas; la Asociación Mexicana de Cultura, AC apoyó esta investigación mediante la asignación de la Cátedra de Análisis y Pronóstico de Series de Tiempo en Econometría.

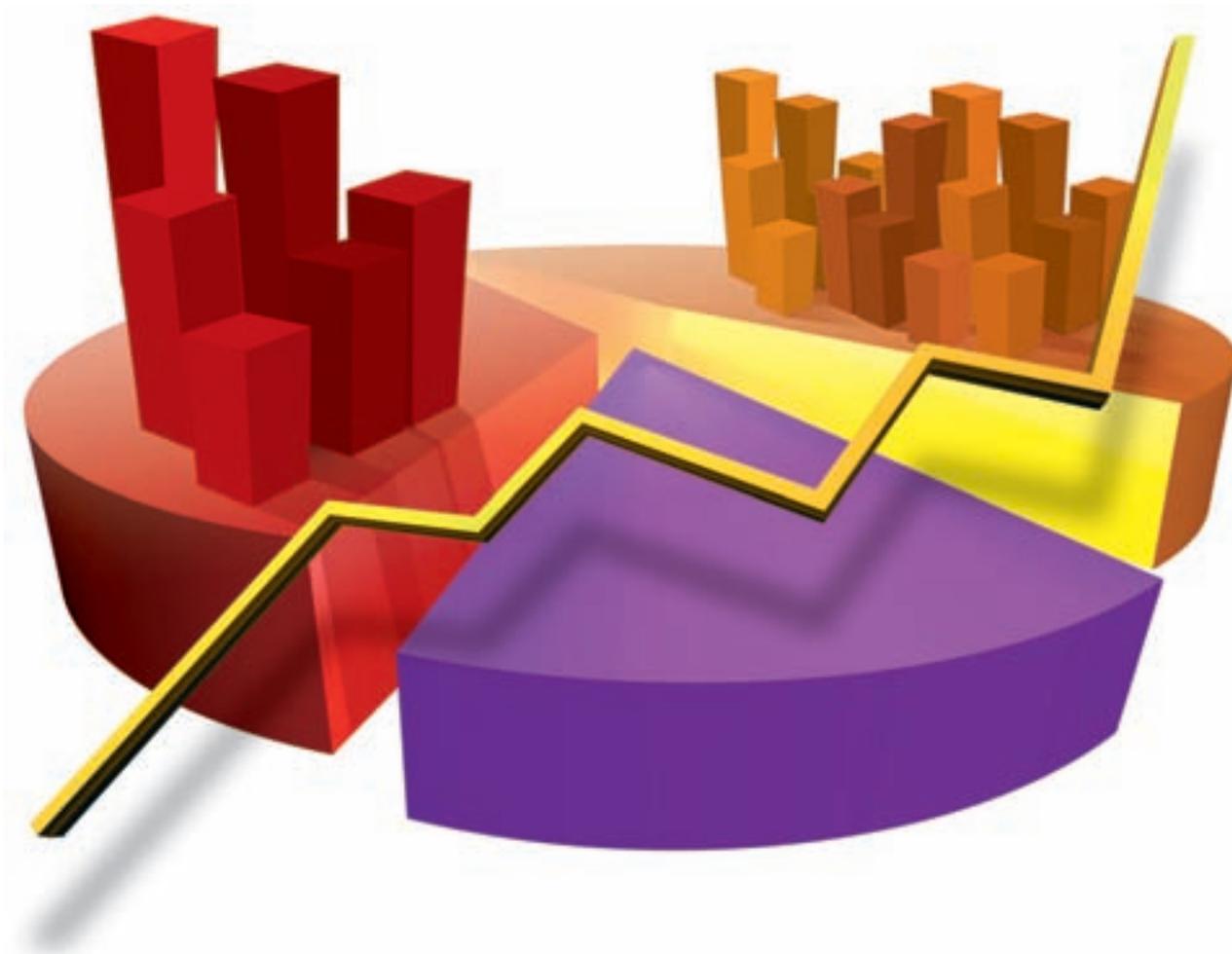
## 1. Introducción

Los institutos nacionales de estadística alrededor del mundo, al igual que el INEGI, producen datos oficiales que son de interés público acerca de fenómenos económicos, demográficos y sociales. Éstos son obtenidos directamente de los informantes involucrados en dichos fenómenos y, para ello, en México se cuenta con el respaldo de la *Ley del Sistema Nacional de Información Estadística y Geográfica* (INEGI, 2008), que establece la obligación legal de responder a los cuestionarios elaborados para tal fin por el Instituto. Los datos que se recaban tienen que ser de interés nacional, por lo cual, además de servir a todo público, deben ser usados de manera obligatoria por las instituciones gubernamentales para la formulación de políticas y la toma de decisiones en general. Cabe hacer notar que al INEGI se le ha dotado de la infraestructura física y de la capacidad técnica para generar los datos que le corresponden y garantizar su confidencialidad.

La Ley también establece que el INEGI debe producir información en forma regular y periódica (lo cual da origen a conjuntos de observaciones en forma de series de tiempo) con base en metodologías que tengan sustento científico (que se obtiene cuando se usan procedimientos con bases estadísticas sólidas).

En este trabajo se hará mención sólo de series que se refieren a la medición de fenómenos económicos<sup>1</sup> con cifras oficiales. Los institutos de estadística encargados de esta labor generan una gran cantidad de datos, lo que implica la necesidad de publicarlos en forma resumida y de manera que transmitan la mayor cantidad de información posible. Esto ha sido reconocido desde hace tiempo (e. g. Deville y Malinvaud, 1983) y para lograr este objetivo se aplican técnicas de ajuste estacional y

<sup>1</sup> La elección de los temas y las técnicas que aquí se presentan es arbitraria y refleja, en buena medida, las preferencias y la experiencia del autor de este trabajo.



de suavizamiento de datos que permiten apreciar con mayor claridad el comportamiento dinámico de las series, sin el oscurecimiento ocasionado por efectos estacionales y otras fluctuaciones que típicamente carecen de importancia para el usuario.

Aun cuando los datos oficiales sean generados mediante la aplicación de técnicas y procedimientos bien establecidos, probados y recomendados por organismos estadísticos internacionales, es común encontrar deficiencias en ellos, como puede ser la falta de coherencia, ya sea en relación con distintas variables para el mismo periodo de observación o para una sola variable en diferentes periodos; asimismo, es frecuente que haya cifras faltantes en las series. Estas dos carencias en la calidad de los datos pueden remediarse con la aplicación de procedimientos de edición e imputación, con los cuales se puede garantizar un cierto nivel de calidad de la información proporcionada.

Las necesidades de los usuarios van en continuo aumento y requieren que se amplíe la cobertura de las bases de datos, ya sea en la dimensión temporal, sectorial, geográfica o en varias, simultáneamente. Estos requerimientos pueden cubrirse mediante un mayor esfuerzo para generar las cifras con la cobertura requerida y, como complemento de tal empeño, también a través de procedimientos que hagan uso de datos auxiliares que se tengan disponibles para usos internos en el instituto de estadística o de datos públicos en general. Cabe hacer notar que solicitar más datos al informante —al aumentar la frecuencia temporal o el detalle sectorial y geográfico— puede ser contraproducente, ya que las personas los proporcionan como una labor adicional a sus actividades propias del sitio de trabajo y al pedirle que brinde más o con mayor frecuencia, pueden perder calidad. En su lugar, los métodos de desagregación temporal de series resultan útiles para generar cifras con mayor frecuencia de observación a partir de datos observados con menor asiduidad. De igual manera, pueden aplicarse métodos de desagregación sectorial o geográfica para aumentar la cobertura respectiva. Además, si se consideran ambas necesidades en forma simultánea, surge un

problema que involucra series múltiples y que se conoce como desagregación temporal y contemporánea de series multivariadas; a esto también se le denomina *benchmarking*, ya que involucra la reconciliación de tablas de datos, así como el ajuste de cifras de alta frecuencia a otras con menor frecuencia de observación, pero que tienen mayor detalle y precisión.

Para complementar la oferta de información al público usuario, se debe considerar la obtención de datos cualitativos que permitan calcular índices de opinión oportunos sobre los fenómenos económicos. Este tipo de índices se construyen con la idea de que los informantes expresen opiniones que, al agregarse en un solo indicador, muestren la tendencia de la economía. Los informantes pueden ser productores o consumidores, pues cada uno de estos grupos señala y anticipa el comportamiento de variables relevantes en su ámbito de competencia. Asimismo, con el propósito de incrementar la oportunidad de aparición de los datos, se debe considerar la estimación rápida de variables que jueguen papeles fundamentales en el sistema económico, lo cual ocurre, por ejemplo, con el producto interno bruto (PIB) trimestral. Generar esta variable con mayor oportunidad que la actual y con la misma precisión en su cálculo requeriría de un gran esfuerzo y quizá no sería factible hacerlo debido a la gran cantidad de series que están involucradas. Por ello, es preferible sacrificar algo de precisión al generar una estimación rápida, que esperar el tiempo que tarda en publicarse el PIB trimestral, aunque su precisión sea mayor.

Por otro lado, se deben tener en cuenta las técnicas de reducción de dimensión, con las cuales se puede resumir el comportamiento de una gran cantidad de series en unas cuantas. Este tipo de técnicas son muy útiles para generar índices que pueden adelantar, coincidir o, incluso, ir retrasados respecto al comportamiento de variables que no son observables de forma directa, sino que se consideran latentes y que reflejan el *estado de la economía*, en términos generales. Dentro de los diversos métodos existentes en la actualidad, so-

bresalen los que se emplean a nivel internacional para construir índices cíclicos compuestos y que también se usan en el contexto nacional. Su objetivo principal es anticipar puntos de giro de la economía, más que la tendencia de la misma. Esta labor es en particular complicada porque se trata de predecir eventos discretos, más que valores de variables continuas.

El contenido de lo que resta de este artículo es como sigue: en la siguiente sección se mencionan los temas de ajuste estacional y suavizamiento de series de tiempo, los cuales se consideran semejantes en tanto que ambos tienen como objetivo apreciar con mayor claridad algún componente importante de la serie, que no sea observable directamente; el apartado tres se refiere a los problemas de edición e imputación de datos, con los cuales se procura mantener un estándar de calidad adecuado para la información; en el cuarto se trata el tema de la desagregación de datos, tanto en su dimensión temporal como en la sectorial, geográfica y bidimensional; la quinta sección se ocupa de los índices de opinión, de la estimación rápida de variables y de los indicadores cíclicos compuestos, ya que todas ellas son herramientas útiles para anticipar el comportamiento de variables macroeconómicas que se consideran de la más alta relevancia; se finaliza este documento con la alusión a otros temas no tratados de forma explícita y con algunas recomendaciones sobre las tareas de análisis estadístico que debe realizar una agencia oficial o, en algunos casos, el mismo usuario interesado en la información.

## **2. Suavizamiento de series y ajuste estacional**

La desestacionalización puede realizarse de diversas maneras, pero existen dos procedimientos que sobresalen por ser los más utilizados y recomendados a nivel internacional: el X-12-ARIMA de la Oficina del Censo de Estados Unidos de América ([www.census.gov/srd/www/x12a/](http://www.census.gov/srd/www/x12a/)) y el SEATS del Banco de España ([www.bde.es](http://www.bde.es)) descrito en Gómez y Maravall (1996). Una breve historia del desarrollo de los métodos de

ajuste estacional se presenta en Guerrero (1990) y, para mayores detalles, se recomienda consultar el texto de Ladiray y Quenneville (2001).

La necesidad de suavizar los datos de una serie de tiempo surge porque la información muchas veces presenta movimientos erráticos y fluctuaciones indeseables que, básicamente, oscurecen la dinámica de la variable a la que se refiere la serie; por ejemplo, en el panel izquierdo de la figura 1 se muestra la serie trimestral del PIB real (a precios constantes del 2003) considerada como la original, junto con su correspondiente serie ajustada por efectos estacionales, mientras que el panel derecho presenta la desestacionalizada, junto con la de tendencia-ciclo respectiva. Estas series fueron obtenidas del Banco de Información Económica (BIE) del INEGI ([www.inegi.org.mx](http://www.inegi.org.mx)), donde la serie original está previamente corregida por efectos de Semana Santa, ya que las fluctuaciones debidas a este fenómeno se consideran de tipo determinista, por lo cual pueden predecirse con relativa facilidad y típicamente carecen de importancia para un analista de datos económicos. La serie ajustada por estacionalidad muestra con mayor claridad los movimientos a corto y mediano plazo del PIB, ya que se le han eliminado los efectos periódicos que se repiten cada año y que tienden a exagerar los movimientos de la serie; como ejemplo de esto, considérese la caída del PIB iniciada en el último trimestre del 2008 y que se aprecia de menor magnitud al considerar los datos ajustados por estacionalidad, lo cual implica que una parte de la caída en los datos originales se debe a efectos estacionales.

Por su lado, la serie de tendencia-ciclo se obtiene al cancelar los movimientos irregulares de la serie ajustada por estacionalidad, de manera que la nueva permite observar mejor los movimientos a largo plazo del PIB. Las variaciones porcentuales anuales que se presentan en la figura 2 muestran los efectos de la desestacionalización (panel izquierdo), donde es evidente que los datos originales exageran los movimientos de la serie, mientras que la ajustada por estacionalidad presenta un comportamiento más conservador. En cambio,

Figura 1

**Serie original del PIB real trimestral y desestacionalizada (izquierda).  
Serie desestacionalizada y tendencia-ciclo del PIB (derecha).  
Periodo 2000:01-2011:04**

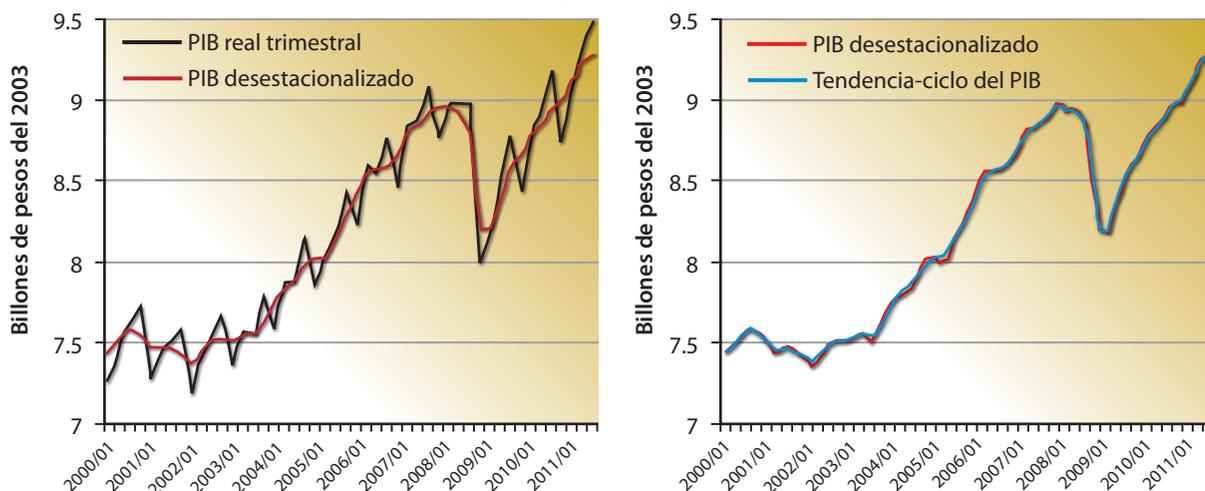
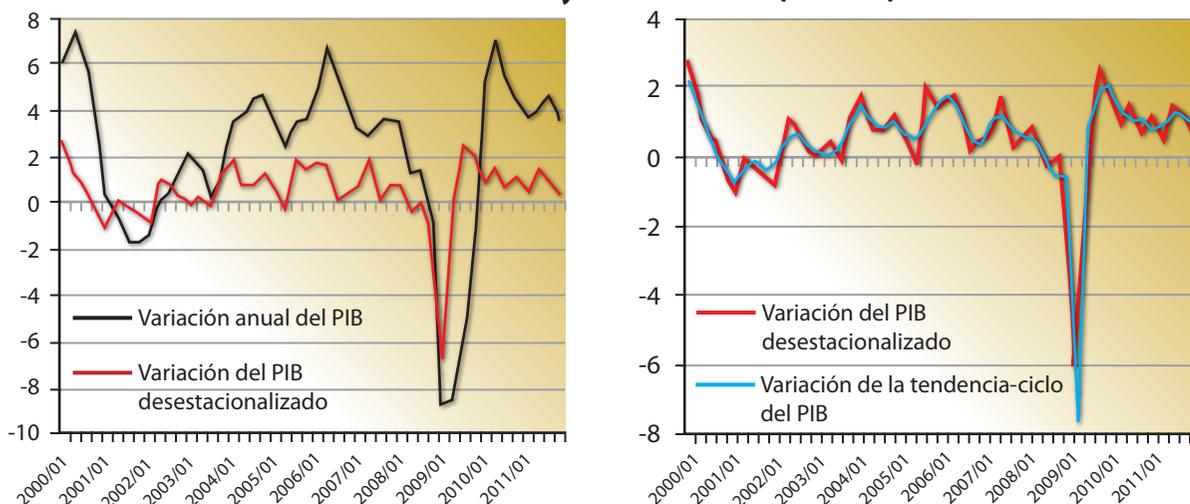


Figura 2

**Variaciones porcentuales anuales del PIB: original y desestacionalizado (izquierda);  
desestacionalizado y tendencia-ciclo (derecha)**



nótese que las variaciones porcentuales del panel derecho son muy semejantes, lo cual es atribuible al hecho de que, en este caso, las fluctuaciones irregulares afectan poco al comportamiento de la serie desestacionalizada.

La obtención de tendencias está ligada muchas veces al ajuste estacional de series, pero eso no tiene por qué ser siempre así, ya que suavizar series de tiempo puede requerirse también para

otras que no contengan efectos estacionales —de las que se desea eliminar las fluctuaciones indeseables, que oscurecen su patrón dinámico de largo plazo— o bien el suavizamiento podría ser aplicado a una serie que no haya sido desestacionalizada de forma previa. Existen varios procedimientos para suavizar los datos, la mayoría de los cuales parten del criterio de sencillez de aplicación para su uso repetitivo y masivo. En Guerrero (2012) se describen los más usados por diversos

institutos de estadística a nivel internacional: 1) el de Tendencia con Promedio de Fases (PAT, por sus siglas en inglés) usado por el National Bureau of Economic Research (NBER) y también por el INEGI en relación con el cálculo de índices cíclicos;<sup>2</sup> 2) el filtro de Hodrick-Prescott (HP) doble que se utiliza en la Organización para la Cooperación y Desarrollo Económicos (OCDE) (véase Hodrick y Prescott, 1997 y OECD, 2008);<sup>3</sup> 3) el filtro HP con suavidad controlada por el usuario (véase Guerrero, 2008 y 2011); y 4) el filtro de Christiano y Fitzgerald (2003) empleado por Eurostat, que permite elegir el paso de banda para la frecuencia de los ciclos.

Los citados procedimientos se denominan filtros, ya que se aplican a una serie de datos para conseguir un fin en particular, sin hacer referencia necesariamente a un modelo estadístico ni a supuestos que pudieran ser verificables con los datos disponibles. De esta manera, un filtro se considera útil si sirve para lo que fue construido; por ejemplo, el procedimiento propuesto por Guerrero (2008) tiene como base el filtro HP y la idea de que la principal característica que debe mostrar

una serie de tendencia es que su comportamiento sea más suave que el de la original. Esto se traduce en que a la tendencia se le pueda asignar un porcentaje de suavidad deseado (uno cercano a cero hace que la serie de tendencia sea prácticamente igual que la original, mientras que uno alto hace que sea casi una línea recta) y que al fijar dicho porcentaje sea posible deducir la constante de suavizamiento involucrada, que es lo único necesario para poder utilizar el filtro HP; un ejemplo de su aplicación aparece en la figura 3, donde el panel izquierdo muestra la tendencia del PIB real con suavidad de 70% y el del panel derecho, de 90 por ciento.

En los dos casos de la figura 3 se aplicó el filtro a la serie desestacionalizada, aunque bien pudo haberse hecho a la serie sin ajustar por efectos estacionales, como lo muestra la figura 4, pues la tendencia que se logra es casi la misma cuando el porcentaje de suavidad es relativamente alto (al respecto, véase Guerrero, 2008). En caso de que se desee cancelar la tendencia para realizar análisis de ciclos económicos, la citada constante de suavizamiento puede elegirse de manera que la serie cíclica resultante tenga las características esperadas de una serie de ciclos. Para conocer más sobre el tema, se recomienda consultar el texto de Kaiser y Maravall (2001).

2 El INEGI lo usa para el Sistema de Indicadores Cíclicos Compuestos Coincidente y Adelantado que se difunde a través del BIE.  
3 El INEGI también calcula los indicadores cíclicos (que se difunden en el comunicado de prensa y mediante el *Reloj de los ciclos económicos* que aparece en el sitio de Internet) con estos filtros.

Figura 3

**Tendencia del PIB desestacionalizado con 70% de suavidad (izquierda) y con 90% de suavidad (derecha). Periodo del 2000:01 al 2011:04**

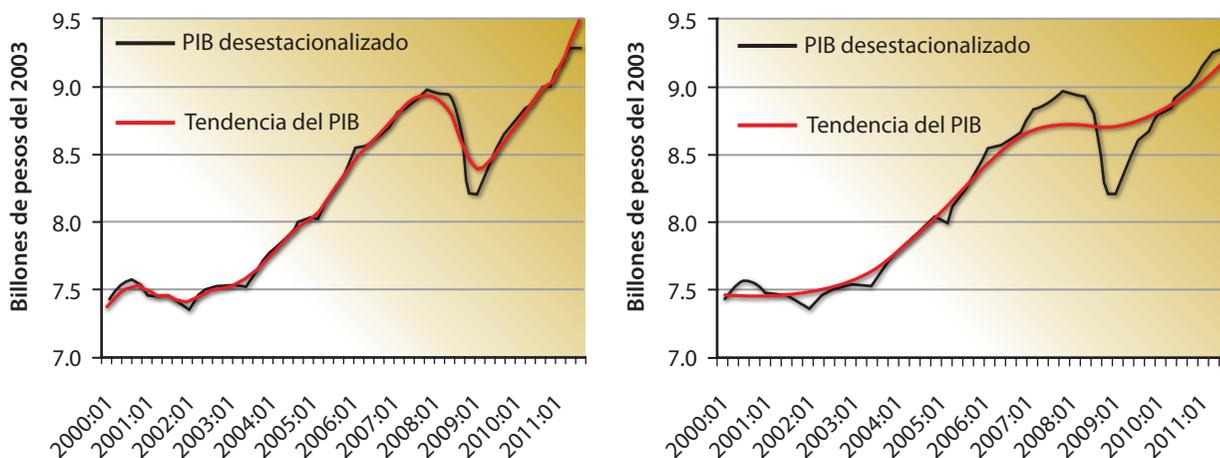
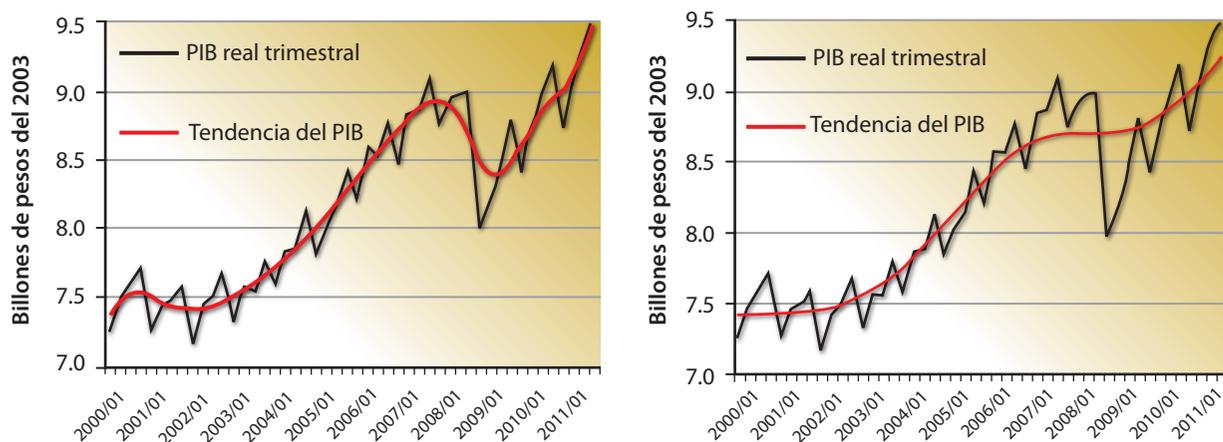


Figura 4

**Tendencia del PIB sin ajuste estacional, con 70% de suavidad (izquierda) y con 90% de suavidad (derecha). Periodo del 2000:01 al 2011:04**



### 3. Edición e imputación de datos

El problema de edición de datos se presenta cuando hay inconsistencia en su recolección, debido a que las respuestas al cuestionario son inválidas, de acuerdo con algún criterio usado para validarlas; el de imputación ocurre cuando faltan datos de una o más de las variables en el cuestionario. Ambas tareas conviene realizarlas a nivel de cuestionarios individuales para evitar la agregación de errores de dos o más de ellos, lo cual implica que esta labor deba hacerse de manera masiva y repetitiva; por ello, los métodos preferibles son los de más fácil aplicación y que consideren la imputación de dos o más variables de forma simultánea, es decir, que incluyan series de tiempo múltiples. Zhang (2003) presenta una revisión de la literatura sobre métodos de imputación múltiple en uso en la actualidad, mientras que Pfeifferman y Nathan (2002) muestran un método basado en modelos que permite llevar a cabo la imputación múltiple, pero cuya aplicación masiva y repetitiva se dificulta debido a lo complicado de los modelos de espacio de estados que utiliza y por la estimación de parámetros asociada. En cambio, Caporello y Maravall (2002) se enfocan en la edición e imputación univariada de series de tiempo y presentan un método muy sencillo de emplear.

Conviene mencionar que el problema de datos faltantes está relacionado con la imputación, pero ha sido tratado básicamente en relación con algún tipo de análisis estadístico específico y con los efectos que produce su presencia al utilizar herramientas de inferencia diseñadas para el caso de datos completos; la referencia básica sobre el tema es Little y Rubin (1987). Por el contrario, la imputación de datos que efectúa una agencia de estadística oficial no parte de la necesidad de realizar un tipo de análisis en particular, sino que se hace con la intención de completar los datos faltantes en los cuestionarios de forma lo más razonable que sea posible.

El método propuesto por Guerrero y Gaspar (2010) considera de manera integral los problemas de edición e imputación y resulta ser hasta cierto punto simple, ya que está basado en modelos de tipo Vector Auto-Regresivo (VAR) y en la técnica de pronósticos con restricciones; esto es, la edición de datos se basa en pruebas de compatibilidad que permiten validar los datos, pues detectan los que difieren significativamente de lo esperado, de acuerdo con la historia de las variables y el modelo estimado —con el cual se resume la información histórica del sistema de variables en consideración—; los valores que resulten, digamos con

significancia de 1%, podrían reemplazarse por los estimados por el modelo, mientras que los datos que no difieran con una significancia de 10% deberían permanecer inalterados y los que se encuentren entre 1 y 10% podrían ser editados (modificados) con una combinación lineal entre el dato y el valor estimado por el modelo. Además, para estimar los datos faltantes, se incorpora toda la información disponible, tanto la histórica de las variables con datos faltantes —a través del modelo VAR— como la de las otras variables que sí tienen cifras reportadas para el periodo en el que faltan datos (mediante pronósticos restringidos). Por tal motivo, y dado que el modelo es de tipo lineal, el procedimiento de imputación es de fácil implementación y tiene características deseables desde el punto de vista de teoría estadística, ya que brinda el mejor estimador lineal e insesgado del dato faltante.

#### **4. Desagregación de series de tiempo**

Surgió originalmente en la dimensión temporal con el fin de estimar datos que no son observables con alta frecuencia (digamos mensuales) a partir de datos que sí lo son, pero cuya frecuencia de observación es baja (digamos trimestrales). El requisito fundamental es que los desagregados satisfagan la restricción contable impuesta por el dato agregado, por ejemplo, si la desagregación se refiere a distribuir la cifra trimestral entre los meses, entonces la suma de los datos desagregados cada mes debe ser el dato observado del trimestre si la serie es de flujos —lo que ocurre con una variable de producción—; si la serie es de un índice, la restricción impuesta es que el promedio de los datos mensuales coincida con la cifra del trimestre, en cambio, si la serie es de saldos —por ejemplo, la deuda del sector público—, al problema se le conoce como interpolación de datos y, en tal caso, algún valor mensual del trimestre, típicamente el último, se restringe a ser igual que el valor trimestral.

La desagregación temporal puede realizarse con métodos que requieran o no de variables auxiliares, las cuales deben contar con observaciones

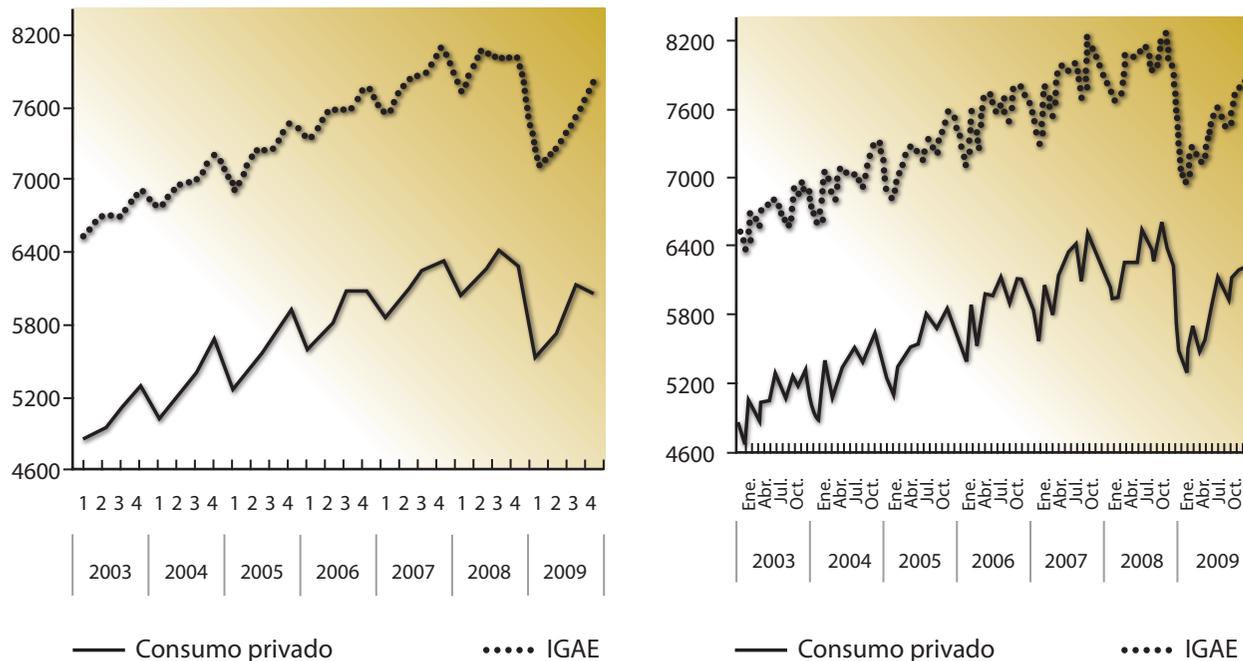
mensuales cuya publicación sea oportuna, que tengan una interpretación adecuada en relación con la variable a desagregar y que, agregadas al trimestre, estén correlacionadas de manera lineal con la serie trimestral. Desde luego, es preferible hacer uso de variables auxiliares para darle mayor contenido económico y validez empírica a la desagregación, aunque los métodos correspondientes son relativamente más complicados que los que no las usan. Para referencias sobre el tema se recomienda consultar el trabajo de Guerrero (2003), donde se presenta una metodología basada en modelos para series de tiempo univariadas que permite desagregar una serie de datos históricos en bloque, así como los datos del momento actual en forma recursiva, según se van registrando, y también realizar pronósticos de la variable desagregada. Por otro lado, Quilis (2009) desarrolló una librería de programas elaborados en el lenguaje de Matlab, que puede ser útil para decidir el procedimiento más adecuado a utilizar, dentro de los métodos de desagregación que ahí se consideran.

Una posible aplicación inmediata de esta metodología podría darse en algún sector económico de México donde no se cuente con un indicador que tenga la frecuencia de observación deseada. Para ilustrar el tipo de resultados que produce la desagregación temporal de series, la figura 5 muestra la serie del consumo privado (CONPRI) trimestral y la misma serie desagregada por mes, que surge al usar como serie auxiliar el índice global de actividad económica (IGAE) expresado como valor agregado bruto, en millones de pesos del 2003. Es de notar que la relación proporcional de los datos trimestrales se mantiene en la serie mensual, pero los desagregados siguen la dinámica del IGAE dentro de los trimestres.

De manera semejante, para cubrir la necesidad de desagregar sectorial o geográficamente una serie agregada pueden emplearse procedimientos que hagan uso de variables auxiliares, por ejemplo, para desagregar el PIB nacional trimestral por entidad federativa, es posible utilizar variables auxiliares que sean relevantes en tanto que reflejen las actividades económicas que se realizan en cada

Figura 5

**Consumo privado e índice global de actividad económica trimestral (izquierda) y consumo privado desagregado mensual (derecha). Periodo 2003-2009**



uno de los estados. Por otro lado, también existen procedimientos que pueden emplearse para tener en cuenta la desagregación temporal y contemporánea de varias series de tiempo simultáneamente. Dentro de los trabajos al respecto se encuentran los de di Fonzo (1994) y Guerrero y Nieto (1999), así como las referencias que en ellos se mencionan. Esta situación se considera un problema en esencia de tipo bidimensional y multivariado, ya que aparece cuando se consideran varias series al mismo tiempo; por ejemplo, al desagregar mensualmente varias series trimestrales correspondientes a subsectores de actividad económica, se debe imponer también la restricción contable de que la suma de los subsectores sea igual que el valor agregado trimestral del sector respectivo.

**5. Indicadores de opinión, estimación rápida e índices cíclicos**

Desde la perspectiva del usuario, la oportunidad con la que se publican los datos oficiales es una de las características más importantes para juzgar

su calidad y utilidad. Debido a esto es que el INEGI genera índices de opinión a partir de los datos cualitativos que surgen de manera específica de dos encuestas, la Nacional Sobre Confianza del Consumidor (ENCO) y la Mensual de Opinión Empresarial (EMOE), las cuales permiten obtener datos oportunos de carácter cualitativo que se transforman en índices cuantitativos al aplicarles herramientas estadísticas ya establecidas. La utilidad de estos indicadores radica en mostrar direcciones y cambios de unas cuantas variables macroeconómicas cuantitativas que miden hechos económicos, sin que se puedan asociar magnitudes a los movimientos de dichas variables. La oportunidad de la información transmitida se mide con referencia a una cierta variable cuyo valor se desea conocer antes de que se publique oficialmente la cifra respectiva.

Una primera tarea que debe realizarse para estudiar y establecer la capacidad predictiva de los índices de opinión consiste en determinar las variables cuantitativas que pueden ser anticipadas con ellos. Para esto, conviene hacer un análisis estadístico de carácter exploratorio, como el que se basa en grá-

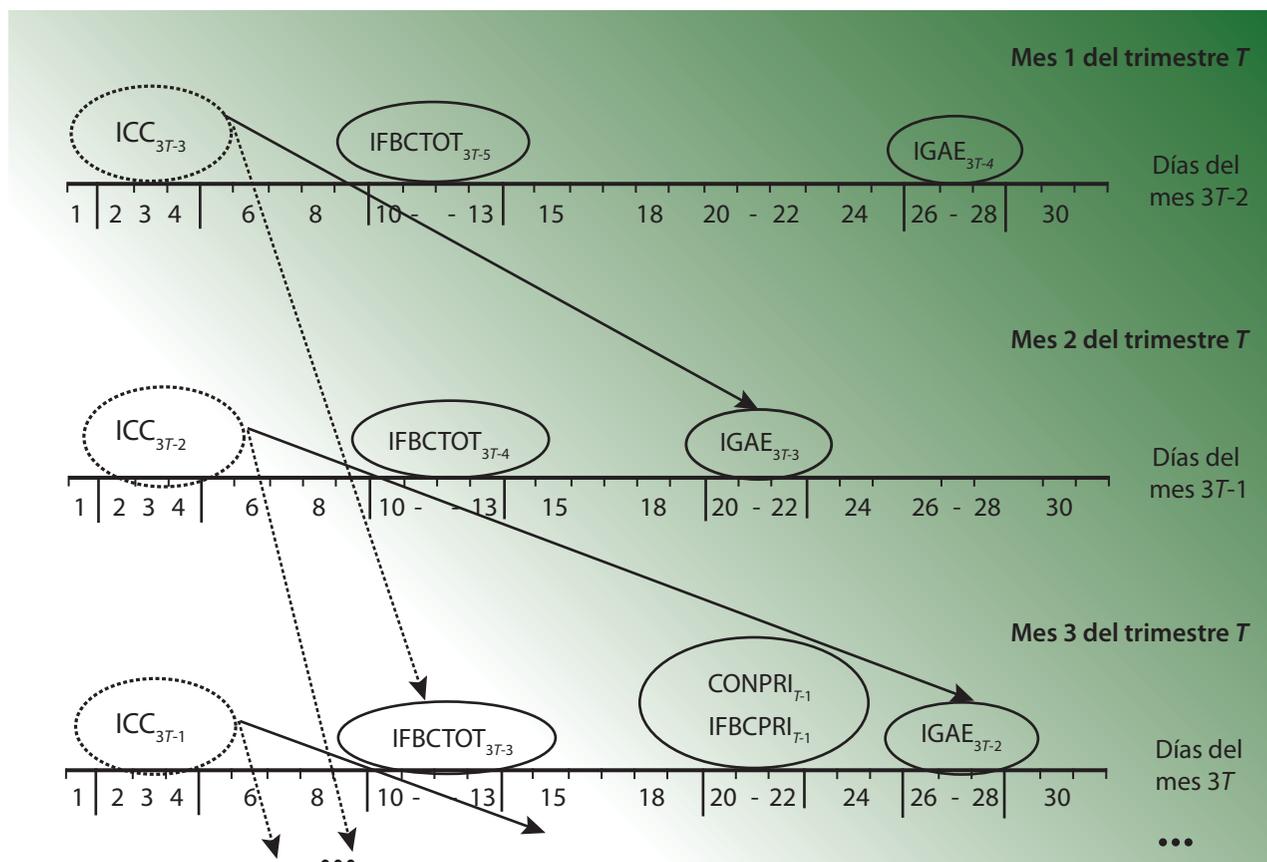
ficas de coincidencias y en el cálculo de correlaciones cruzadas, el cual se efectúa entre los índices de opinión y las variables que se piensa pueden ser anticipadas o que, al menos, se mueven en forma contemporánea con tales índices; por ejemplo, en el trabajo de Guerrero y Sainz (2011) se encontró que el índice de confianza del consumidor (ICC) que surge de la ENCO anticipa el comportamiento de variables como el IGAE, el índice de formación bruta de capital total (IFBCTOT), el índice de formación bruta de capital privado (IFBCPRI) y la CONPRI (estas dos últimas son trimestrales, pero pueden desagregarse por mes, como se indicó en la sección previa). Lo sobresaliente de la oportunidad de publicación puede apreciarse en la figura 6, la cual muestra los días de anticipación del ICC respecto a

las variables mencionadas, que es de un poco más de 50 días para el IGAE, 70 para el IFBCTOT y llega a ser hasta de 80 para la CONPRI y el IFBCPRI.

Otra forma de satisfacer la necesidad expresa de los usuarios acerca de mayor rapidez en la publicación de datos oficiales se puede lograr mediante la aplicación de métodos que permitan completar las series para las que no se cuente con las cifras más recientes. Tales métodos deben: 1) hacer uso de datos oficiales, disponibles sólo en la agencia estadística correspondiente; 2) seguir el procedimiento usual de cálculo de la variable, digamos el PIB trimestral y 3) utilizar sólo modelos estadísticos para obtener la estimación, es decir, que el modelo no surja de una teoría económica que pudiera

Figura 6

**Cronograma de publicación de datos del IGAE e IFBCTOT (mensuales) y de CONPRI e IFBCPRI (trimestrales) en relación con el ICC (mensual), referidos a un cierto trimestre  $T$**



generar controversia y suspicacias sobre los datos que se obtengan. Éstas son las tres principales recomendaciones que se han hecho en las reuniones internacionales sobre el tema de obtención de estimaciones tipo *flash* (e. g. United Nations, 2009). Desde luego, para lograr mayor oportunidad en la publicación de cifras oficiales, el usuario debe aceptar una pérdida en precisión de los datos y tomar conciencia de que las cifras que se publiquen son estimaciones preliminares, de manera que están sujetas a cambios posteriores.

En el trabajo de Guerrero, García y Sainz (2011) se aplicaron los lineamientos anteriores para realizar una estimación rápida del PIB trimestral en México y se logró obtener una con oportunidad de publicación menor a 30 días, mientras que la cifra oficial del PIB se publica alrededor de 52 días después de cerrado el trimestre, tiempos aceptables de acuerdo con estándares internacionales. Una característica fundamental del trabajo mencionado es que se utilizaron fuentes oficiales, así como indicadores de opinión, entre otras variables disponibles con características anticipatorias. Todo esto pudo lograrse mediante el uso de modelos VAR, validados de forma empírica con los datos disponibles. Además, dentro de las variables utilizadas en los modelos VAR que se usaron para obtener estimaciones rápidas del PIB, aparecen los indicadores cíclicos compuestos, coincidente y adelantado que construye el INEGI con el fin de monitorear el estado de la economía y, en particular, para tratar de anticipar sus puntos de giro.

Al respecto de los indicadores cíclicos, conviene notar que la idea que los sustenta se basa en que un índice coincidente debe mostrar el estado agregado de la economía, no sólo de una variable, por más importante que ésta sea —como sucede con el PIB— en el momento actual y con oportunidad, mientras que un indicador adelantado debe señalar el movimiento que seguirá el estado de la economía, sobre todo cuando hay cambio de dirección. Para construir este tipo de indicadores, se cuenta con una gran variedad de procedimientos y no hay consenso internacional sobre el más adecuado. De hecho, en el Sistema de Indicadores Compuestos

Coincidente y Adelantado del INEGI se utilizaba con anterioridad el método del NBER de Estados Unidos de América con enfoque de ciclo clásico, pero en el 2010 se realizó un análisis de los métodos que se emplean en otros países y se encontró que el usado por la OCDE con enfoque de ciclo de crecimiento proporciona resultados que son más adecuados para el caso de la economía mexicana (véase al respecto Guerrero, 2012).<sup>4</sup>

Para llegar a esta conclusión hubo necesidad de decidir entre diversos enfoques de análisis: se empezó por descartar el del ciclo clásico de negocios que antes se aplicaba y se optó por el del ciclo de crecimiento. Esto implicó la necesidad de discriminar entre diversas técnicas para estimar y eliminar tendencias de las series involucradas en los respectivos índices (coincidente y adelantado). Para ello, se realizó un análisis de los métodos disponibles y se aplicó un esquema comparativo semejante al que usaron Nilsson y Gyomai (2008). Después, se decidió entre tres diferentes técnicas para estimar los índices cíclicos (la del NBER, la de la OCDE y la de Stock y Watson). En principio, dado que en México no se cuenta con un organismo que declare oficialmente la entrada a una recesión y la salida de la misma, el INEGI proporcionó una cronología de los periodos recesivos que han afectado al país en los años recientes (ver figura 7) y es fundamental para construir un índice coincidente para la nación, aunque pudiera mejorarse todavía en diversos puntos, en particular si se acepta un tipo de metodología específica como oficial, que podría indicar fechas (ligeramente) distintas que las mostradas por la figura 7 para definir las recesiones.

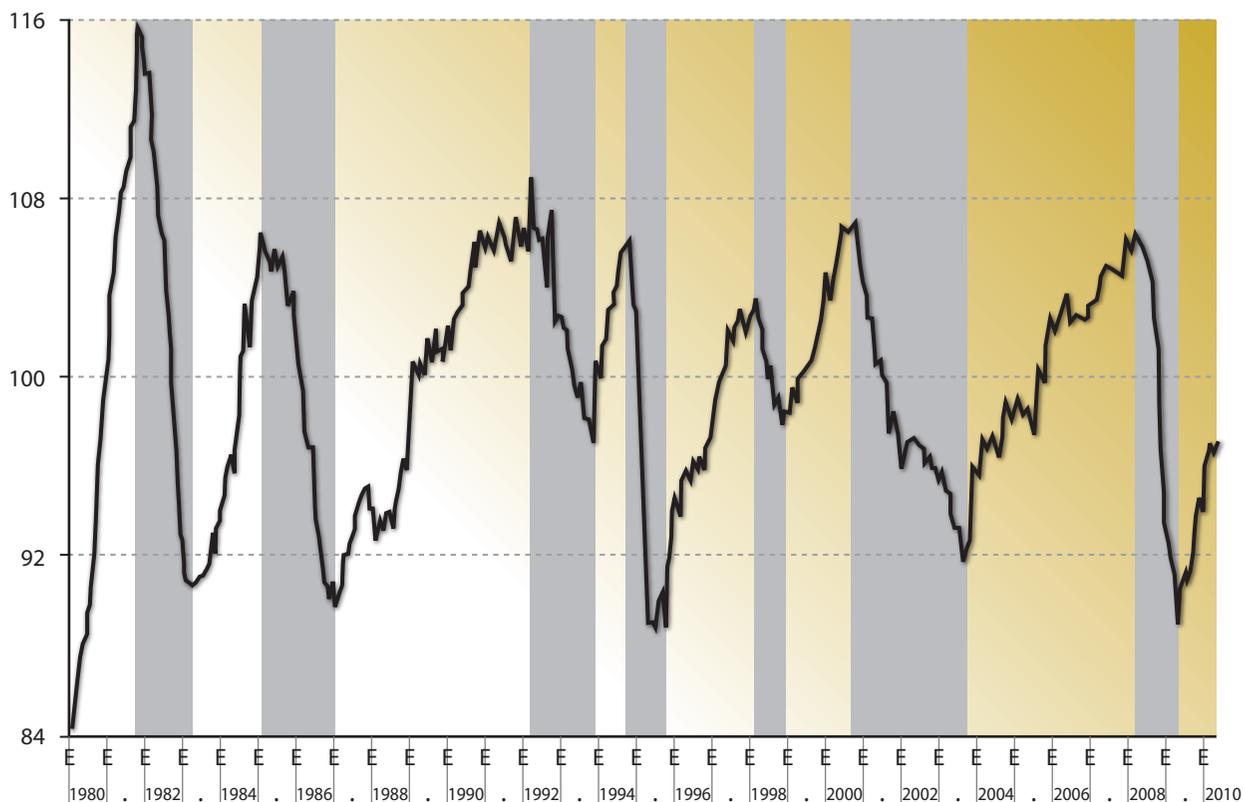
## 6. Conclusiones y recomendaciones

Una de las primeras es que los institutos de estadística oficial tienen muchas posibilidades de realizar análisis de los datos antes de darlos a conocer al público, lo cual debería considerarse como una

<sup>4</sup> En la actualidad, el INEGI continúa realizando el cálculo y la difusión de los índices que se obtienen con el método del NBER, para que los usuarios cuenten con la información de ambos enfoques.

Figura 7

**Serie del índice coincidente del estado de la economía, obtenida con enfoque de ciclo de crecimiento y metodología del NBER**



obligación y enfrentarla con responsabilidad. La mira tendría que estar puesta en brindar al usuario la mayor cantidad de información a través de los datos que se publiquen, lo cual implica que éstos sean claros, interpretables, de buena calidad, en forma oportuna, con la mayor cobertura posible y acerca de las variables más relevantes. De esta manera, los usuarios (incluidas las instituciones gubernamentales) podrían considerarse bien informados acerca de lo que ocurre con el sistema económico para hacer el diagnóstico pertinente y tomar decisiones adecuadas. Además, ya que la adherencia de los institutos de estadística oficial a los métodos y convenciones internacionales establecidos respalda la calidad de la información que proporcionan, se requiere que estén al tanto de lo que se acuerda en el plano internacional y se comprometan a implementarlo en el ámbito de su incumbencia. Para evitar los rezagos en su aplicación, el personal de

estas instituciones debería participar —no sólo asistir para escuchar— de manera regular en los diversos seminarios internacionales especializados —que son organizados por Eurostat, la OCDE, la ONU, etc.—, donde se discuten y buscan consensos sobre temas estadísticos de actualidad.

Esto último conduciría a una mayor participación en áreas de investigación sobre metodologías de estadística oficial y la difusión de resultados de manera visible y generalizada, no sólo para uso interno del instituto respectivo. El propósito debería ser no sólo mantenerse en la frontera del conocimiento sino, de ser posible, tratar de mover dicha frontera hacia adelante.

Desde luego, debe ser claro que los temas considerados en este artículo son meros ejemplos del análisis que podría hacerse en las instituciones de

estadística, pues hay muchos otros temas que no fueron considerados aquí y que no por ello deberían menospreciarse; por ejemplo, una situación interesante que surge al trabajar con series de tiempo es que, conforme pasan los días, los datos publicados de algunas variables cambian, ya sea debido a la acumulación de otros nuevos —porque algunos eran en realidad preliminares y fueron modificados—, o bien porque se realizó algún ajuste a las cifras para compatibilizarlas con otras de la misma variable, que son obtenidas con mayor detalle y precisión (digamos provenientes de una encuesta con mayor cobertura). Así es como surge la necesidad de realizar revisiones de datos y de analizar los efectos de dichas revisiones.

De igual manera se presentan problemas cuando ocurre un cambio de año base en una serie de índices (con o sin modificaciones en las ponderaciones que se utilicen), o bien por la unión de series con diferentes periodos de observación para variables semejantes; asimismo, aparece la necesidad de efectuar retroproyección de series para completar bases de datos históricos. También, se requiere aplicar análisis estadístico cuando se presenta algún cambio de definición de conceptos o variables lo cual, por ejemplo, ocurrió en México con los datos de desempleo que se obtenían anteriormente con la Encuesta Nacional de Empleo Urbano y que se captan a partir del 2005 mediante la Encuesta Nacional de Ocupación y Empleo. Otras oportunidades de análisis ocurren por los cambios de necesidades, como ocurrió en 1987 cuando el Banco de México decidió generar datos quincenales para complementar los de la inflación mensual debido a la hiperinflación que prevalecía en ese tiempo; o al igual que por la incorporación de datos de variables de reciente aparición en los análisis ya establecidos, como sucede con el impuesto empresarial a tasa única (IETU) que se aplica en México a partir del 2009, etcétera.

Por otro lado, se debe tomar conciencia de que en la actualidad las empresas y otras organizaciones están generando grandes bases de datos, lo cual se denomina ya un *diluvio de datos*, y que debería ser aprovechado por los institutos oficiales de estadís-

tica para anticipar necesidades de información de los usuarios. Para ello, se requiere ampliar la visión y la capacidad de recolección de datos, porque éstos pueden provenir de registros bancarios, redes sociales, Internet y puntos de venta, entre otras fuentes. No sólo se trata de recopilarlos y almacenarlos, sino de validarlos, clasificarlos, darles una forma presentable y analizarlos con, entre otras, las herramientas mencionadas en este artículo, para después difundirlos como verdadera información. En el reporte de Manyika *et al.* (2011) se da cuenta de esta situación, la cual puede ser aprovechada de manera favorable si los institutos de estadística se preparan para ello.

Además, para que estas instituciones oficiales se beneficien de la opinión de los usuarios y conozcan sus necesidades de información, deben admitir críticas y estar abiertos a recibir comentarios. De hecho, deberían considerar las dudas y preguntas que surjan sobre los procedimientos usados en el respectivo instituto, no como una amenaza o crítica negativa, sino como una manifestación de interés y de reconocimiento a la importancia que tiene la elaboración de los datos. Más aún, debería considerarse como una posibilidad de mejora en los procedimientos actuales y de una colaboración en potencia con el individuo u organismo que hizo la crítica o comentario correspondiente. En adición, tendría que distinguirse con claridad entre: 1) las dudas sobre los procedimientos y 2) las dudas sobre la integridad y capacidad de la gente que los aplica en el instituto, para que el personal del mismo no se sienta agredido por cada comentario o duda que surja.

Desde luego, se vuelve cada vez más evidente que los productores de datos oficiales deben tener mejor preparación técnica. Antes, lo fundamental era que el personal pudiera transcribir, clasificar, registrar y archivar datos eficaz y eficientemente, mientras que ahora eso se puede hacer de manera casi automática y lo que se requiere es mayor involucramiento del personal en las tareas de análisis de datos para producir información. Si los institutos de estadística no lo hacen, los usuarios tenderán a restarles importancia y harán los análisis por ellos mismos, si tal posibilidad está a su alcance.

Por último, cabe recordar que los temas tratados aquí se refieren sólo a datos económicos y que algo semejante, aunque con adecuaciones específicas, podría decirse respecto a los sociodemográficos, de medioambiente y de otra índole que también son generados por los institutos de estadística.

## Referencias

- Caporello, G. y A. Maravall. "A tool for quality control of time series data", en: *Program TERROR*, 2002. Disponible en el sitio del Banco de España ([www.bde.es/servicio/software/econom.htm](http://www.bde.es/servicio/software/econom.htm)).
- Christiano, L. J. y T. J. Fitzgerald. "The band pass filter", en: *International Economic Review*. Vol. 44, 2003, pp. 435-465.
- Deville, J. C. y E. Malinvaud. "Data Analysis in Official Socio-economic Statistics", en: *Journal of the Royal Statistical Society, Series A*. 146, 1983, pp. 355-361.
- Di Fonzo, T. "Temporal disaggregation of a system of time series when the aggregate is known", en: *INSEE-Eurostat Workshop on Quarterly National Accounts*. París, diciembre de 1994.
- Gómez, V. y A. Maravall. *Programs TRAMO and SEATS*. 1996. Disponible en el sitio del Banco de España ([www.bde.es/servicio/software/econom.htm](http://www.bde.es/servicio/software/econom.htm)).
- Guerrero, V. M. "Desestacionalización de series de tiempo económicas: introducción a la metodología", en: *Comercio Exterior*. Vol. 40, 1990, pp. 1035-1046.
- \_\_\_\_\_. "Monthly disaggregation of a quarterly time series and forecasts of its unobservable monthly values", en: *Journal of Official Statistics*. 19, 2003, pp. 215-235.
- \_\_\_\_\_. "Estimating Trends with Percentage of Smoothness Chosen by the User", en: *International Statistical Review*. 76, 2008, pp. 187-202.
- \_\_\_\_\_. "Capacidad predictiva de los índices cíclicos compuestos para los puntos de giro de la economía mexicana". Por aparecer publicado en *Economía Mexicana*. 2013.
- Guerrero, G. V. M. "Medición de la tendencia y el ciclo de una serie de tiempo económica desde una perspectiva estadística", en: *Realidad, Datos y Espacio. Revista Internacional de Estadística y Geografía*. Vol. 2, núm. 2, 2011, pp. 50-73.
- Guerrero, V. M. y B. I. Gaspar. "Edition and Imputation of Multiple Time Series Data Generated by Repetitive Surveys", en: *Journal of Data Science*. Vol. 8, 2010, pp. 555-577.
- Guerrero, V. M. y F. H. Nieto. "Temporal and contemporaneous disaggregation of multiple time series", en: *Test*, Vol. 8, 1999, pp. 459-489.
- Guerrero, V. M. y E. Sainz. "Capacidad predictiva del Índice de Confianza del Consumidor en México", en: *Documento de Trabajo*. DE-C11.6. México, ITAM, 2011.
- Guerrero, V. M., A. C. García y E. Sainz. "Rapid Estimates of Mexico's Quarterly GDP", en: *Documento de Trabajo*. DE-C11.5. México, ITAM, 2011.
- Hodrick, R. J. y E. C. Prescott. "Postwar U.S. business cycles: An empirical investigation", en: *Journal of Money, Credit and Banking*. Vol. 29, 1997, pp. 1-16.
- INEGI. *Ley del Sistema Nacional de Información Estadística y Geográfica*. 2.ª edición. Aguascalientes, INEGI, 2008.
- Kaiser, R. y A. Maravall. *Measuring Business Cycles in Economic Time Series*. New York, Springer-Verlag, 2001.
- Ladiray, D. y B. Quenneville. *Seasonal Adjustment with the X-11-Method*. New York, Springer-Verlag, 2001.
- Little, R. J. A. y D. B. Rubin. *Statistical Analysis with Missing Data*. New York, John Wiley, 1987.
- Manyika, J. et al. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011. Disponible en el sitio de McKinsey Global Institute ([www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation)).
- Nilsson, R. y G. Gyomai. "Cycle Extraction. A comparison of the Phase-Average Trend method, the Hodrick-Prescott and Christiano-Fitzgerald filters", en: *Technical Report*. París, OECD, 2008.
- OECD. *Handbook on Constructing Composite Indicators. METHODOLOGY AND USER GUIDE*. París, OECD, 2008. Disponible en: [www.oecd.org/dataoecd/37/42/42495745.pdf](http://www.oecd.org/dataoecd/37/42/42495745.pdf)
- Pfefferman, D. y G. Nathan. "Imputation for Wave Nonresponse: Existing Methods and a Time Series Approach", en: *Survey Nonresponse* (editado por Groves, R. M., D. A. Dillman, J. L. Eltinge y R. J. A. Little). New York, John Wiley, pp. 417-429, 2002.
- Quilis, E. M. *MATLAB Temporal Disaggregation and Interpolation Library*. 2009. Disponible en: [www.mathworks.com/matlabcentral/fileexchange/24438](http://www.mathworks.com/matlabcentral/fileexchange/24438)
- United Nations. Statistical Division. "International Seminar on Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends". Ottawa, 27-29 May, 2009. Disponible en: <http://unstats.un.org/unsd/nationalaccount/workshops/2009/ottawa/ac188-2.asp>
- Zhang, P. "Multiple imputation: Theory and methods", en: *International Statistical Review*. Vol. 71, 2003, pp. 581-592.