

Desagregación temporal de series económicas con programación lineal

Luis Frank*

Resumen

El artículo presenta un método de desagregación temporal de series de tiempo, que combina la interpolación de datos de baja frecuencia con una o más series relacionadas de alta frecuencia. La serie desagregada es, esencialmente, la solución a un programa lineal que minimiza la suma de desvíos absolutos con la serie de baja frecuencia y con series relacionadas de alta frecuencia. El método es útil para conciliar series de baja frecuencia con otras relacionadas de alta frecuencia, cuando estas últimas presentan valores atípicos, datos faltantes o, incluso, cuando presentan distintas frecuencias. El nuevo método se pone a prueba desagregando la serie trimestral del VAB industrial con esta componente del Estimador Mensual de Actividad Económica.

Palabras clave: desagregación temporal; interpolación; conciliación; Cuentas Nacionales.

Abstract

The article presents a method for temporal disaggregation of time series that combines the interpolation of the low-frequency data with one or more related high-frequency series. The disaggregated series is essentially the solution to a linear program that minimizes the sum of absolute deviations with the low-frequency series and high-frequency benchmark series. The method is useful for reconciling low-frequency series with high-frequency related series when these contain outliers, missing data, or even when they have different frequencies. The new method is tested by disaggregating the quarterly industrial GVA series with the industrial component of the Monthly Estimator of Economic Activity.

Keywords: temporal disaggregation; interpolation; reconciliation; National Accounts.

JEL: C82

Fecha de recepción: 18/11/2018; Fecha de aceptación: 14/05/2019

* Instituto Nacional de Estadística y Censos (INDEC), Dirección Nacional de Cuentas Nacionales; Universidad de Buenos Aires, Facultad de Agronomía. Email: frank@agro.uba.ar

I. Introducción

Supongamos que se nos provee una serie de baja frecuencia y se nos pide construir otra serie de alta frecuencia cuyos valores sumen o promedien por períodos aquellos de la primera. Este es un requerimiento típico en oficinas de Cuentas Nacionales que surge, por ejemplo, cuando la información básica necesaria para calcular cierto agregado macroeconómico se releva con una frecuencia menor que la del agregado macro. En estos casos, el Manual de Cuentas Nacionales (2018) aconseja (i) interpolar la serie de baja frecuencia, o bien (ii) "conciliarla" con una serie de alta frecuencia presumiblemente relacionada. El método de interpolación más difundido en Cuentas Nacionales es el de Boot et al. (1967) (véase también Kroonenberg y Wansbee, 1992), en tanto que para la conciliación de series se utilizan los métodos de Chow y Lin (1971), Denton (1971), y Fernández (1981). En Bee Dagum y Cholette (2006) puede hallarse una revisión exhaustiva de estos y otros métodos más sofisticados para desagregar series de tiempo, y en Infante (2017), una actualización reciente del estado de la cuestión. La desagregación temporal de series económicas con estos métodos presenta, sin embargo, tres dificultades prácticas para las que aún no se ha encontrado una solución convincente.

Primero, los métodos de interpolación y conciliación han tenido desarrollos independientes y, hasta donde sabemos, no se ha intentado combinarlos en un único procedimiento que exima al analista de decidir entre (i) interpolar la serie de baja frecuencia solamente, a riesgo de desaprovechar información útil que pudiera proporcionar una o más series relacionadas de alta frecuencia; o (ii) conciliar la serie de baja frecuencia con otra serie relacionada, aun cuando esta relación fuera débil, desaprovechando la guía que brinda el sendero óptimo de interpolación de los puntos de la serie de baja frecuencia¹.

Segundo, en general, los métodos de conciliación no prevén faltantes de información en las series relacionadas, es decir, no contemplan la posibilidad de que sean discontinuas por tramos, o que esporádicamente les falten valores. En estos casos, es usual descartar las series "defectuosas", reteniendo para la conciliación solamente las series sin discontinuidades, sin reparar en que estas podrían tener una característica común (por ejemplo, misma técnica de muestreo, mismo organismo relevador, mismo financiamiento para el relevamiento, etc.) que no sea compartida por la serie de baja frecuencia, lo cual introducirá un sesgo en la serie desagregada.

Tercero, los métodos estándar de conciliación no son robustos, en sentido estadístico, cuando las series relacionadas son particularmente ruidosas o presentan valores atípicos. Esto se debe a que el criterio de optimización predominante en estos métodos

¹ En este sentido, es ilustrativo recordar que el informe metodológico del Estimador Mensual de Actividad Económica (EMAE) de INDEC (2016) aclara "el criterio de selección del método de desagregación temporal entre los descriptos (Boot et al., 1967; Denton, 1971; Fernández, 1981 y Chow y Lin 1971) para cada componente del EMAE se estableció considerando aquel que minimizara cierta medida de volatilidad" (p. 10); aunque sin aclarar qué medida se utilizó. Es decir, en la misma metodología del EMAE, se reconoce el uso de un criterio *ad-hoc* de selección del método de desagregación temporal ante la imposibilidad de combinar los distintos métodos en un procedimiento general.

es el de mínimos cuadrados, criterio que, como se sabe, otorga más peso a las observaciones más alejadas de los valores esperados. Al igual que el caso mencionado en el párrafo anterior, es usual descartar series especialmente ruidosas y, eventualmente, interpolar la serie de baja frecuencia desaprovechando información potencialmente útil, subyacente al ruido estocástico.

En vista de estos inconvenientes proponemos, a continuación, un método de desagregación temporal de series económicas basado en programación lineal (LP). Este método combina un sistema de restricciones lineales propio de los métodos de interpolación, con otro que vincula una o más series de alta frecuencia con la serie de baja frecuencia, llamado en general "conciliación". El criterio de optimización de este método es la minimización de una suma de los desvíos absolutos en lugar de una suma de cuadrados, lo cual robustece el resultado.

II. El programa lineal

Supongamos que se nos provee una serie $y_1, \dots, y_j, \dots, y_m$ de baja frecuencia a partir de la cual deseamos construir una serie $x_1, \dots, x_i, \dots, x_n$ de alta frecuencia, disponiendo de una o más series relacionadas, o incluso de ninguna. Para que la serie x_i, \dots, x_n sea coherente con la de baja frecuencia, deberá satisfacer un conjunto de condiciones: la primera y principal, que el promedio simple de los $p = n/m$ valores de cada período en que se puede dividir x_i, \dots, x_n coincida con el correspondiente valor de y . Formalmente, para $p = 3$,

$$\frac{1}{3}x_{i-1} + \frac{1}{3}x_i + \frac{1}{3}x_{i+1} = y_j \text{ para todo } j = 1, \dots, m \text{ e } i = jp - 1 \quad (1)$$

Una segunda condición a satisfacer es que las distancias entre los valores extremos de cada uno de los m períodos en los que se puede dividir x_i, \dots, x_n y su valor intermedio, sean aproximadamente iguales. Esta es una condición de monotonía tendiente a preservar el movimiento de la serie de baja frecuencia. En el caso de $p = 3$,

$$(x_{i+1} - x_i) - (x_i - x_{i-1}) - e_j^+ + e_j^- = 0,$$

o bien,

$$x_{i-1} - 2x_i + x_{i+1} - e_j^+ + e_j^- = 0 \text{ para todo } e_j^-, e_j^+ \in \Re^+, \quad (2)$$

donde e^+ y e^- son términos de error que igualan ambos lados de la igualdad y no deben ser interpretados como errores en un sentido estadístico. Para evitar confusiones llamaremos a estos términos "discrepancias". Si la suma de los términos de x es mayor a 0, las discrepancias toman los valores $e^+ > 0$ y $e^- = 0$. En cambio, si la suma de los términos de x es menor que 0, dichas discrepancias toman los valores $e^+ = 0$ y $e^- \geq 0$. Esta descomposición de las discrepancias es similar a la de los modelos de programación de objetivos, descritos en Williams (2013).

Por otra parte, si x no presenta discontinuidades, ni saltos abruptos, cabe esperar que la diferencia entre el último valor de cada período y el primero del siguiente sea pequeña. Esta tercera condición se expresa matemáticamente como

$$x_{i+1} + x_{i+2} - e_j^+ + e_j^- = 0, \quad (3)$$

y se aplica a todos los pares (x_{i+1}, x_{i+2}) , a excepción de los pares inicial $\{x_0, x_1\}$ y final $\{x_n, x_{n+1}\}$, porque tanto x_0 como x_{n+1} son valores desconocidos. Es decir, la condición (3) define solo $m-1$ restricciones a las que deberíamos sumar dos restricciones adicionales que vinculen a x_1 con los siguientes valores de la serie y a x_n con sus valores precedentes. Hacemos, a continuación, una breve digresión para hallar dichas restricciones.

Recordemos la condición (2), y sumemos y restemos x_i en el lado izquierdo de la misma. Reemplacemos los sumandos que correspondan por y_j de acuerdo a la condición (1). Omitiendo los términos de discrepancia, al combinar (1) y (2)

$$x_{i-1} + x_i + x_{i+1} - 3x_i = 3y_j - 3x_i = 0.$$

Luego, si (1) y (2) se satisfacen estrictamente, se cumple necesariamente que $x_i = y_j$. A partir de esta igualdad es posible

(a) sustituir x_i por y_j en (1), y x_{i+1} por x_{i+2} (por la tercera condición) de modo que

$$x_{i-1} = 2y_j - x_{i+2},$$

(b) sustituir y_j por x_i en la condición (1), de modo que x_i sea el promedio simple de x_{i-1} y x_{i+1} , y generalizar este resultado para un índice i' cualquiera, ya que el momento en que comienza la serie es completamente arbitrario.

$$x_{i'} = \frac{x_{i'-1} + x_{i'+1}}{2}$$

En (b) podemos reemplazar $x_{i'-1}$ por $x_{i'-2}$, apelando a la condición (3), y llamar $i' - 1 = i + 1$, de manera que x_{i+2} sea el promedio simple entre x_i y x_{i+3} . Y, para concluir el razonamiento, podemos reemplazar x_i por y_j y x_{i+3} por y_{j+1} (estas igualdades se verifican para todo $i = jp - 1, y_j = 1, \dots, m$) de manera que el resultado (a) pueda describirse como

$$x_{i-1} = 2y_j - \frac{y_j + y_{j+1}}{2}$$

Ahora bien, retomando el punto que motivó la digresión, evaluamos la expresión anterior en $i - 1 = 1$ y sumamos los términos de discrepancia para establecer una cuarta condición

$$x_1 - e_j^+ + e_j^- = 2y_1 - \frac{1}{2}(y_1 + y_2) \quad (4)$$

$$= \frac{3}{2}y_1 - \frac{1}{2}y_2$$

Razonando en forma análoga, podemos establecer la quinta condición para x_n ,

$$x_n - e_j^{+''''} + e_j^{-''''} = \frac{3}{2}y_m - \frac{1}{2}y_{m-1} \quad (5)$$

Las condiciones (1) a (5) pueden ser rescritas matricialmente en forma de un sistema lineal del tipo $\mathbf{Ax} = \mathbf{b}$, como el que se muestra a continuación. Para facilitar la lectura, partimos la matriz \mathbf{A} y los vectores \mathbf{x} y \mathbf{b} en bloques, según la siguiente convención: las matrices se escriben en mayúscula negrita y los vectores en minúscula negrita; los escalares se escriben en bastardilla como hasta ahora; todos los vectores son vectores columna; las dimensiones de cada matriz se aclaran como subíndice en forma de producto de filas y por columnas, salvo en matrices cuadradas en las que basta con indicar uno solo de estos números; llamamos \mathbf{I} a la matriz identidad, y 0 y 1 a las matrices de elementos $\mathbf{0}$ o $\mathbf{1}$, respectivamente; los supraíndices entre paréntesis numeran las partes en que se ha dividido cierta matriz. Luego,

$$\begin{bmatrix} \mathbf{I}_m \otimes \mathbf{1}'_p/p & \mathbf{0}_{m \times 2m} & \mathbf{0}_{m \times (2m-2)} & \mathbf{0}_{m \times 2} \\ \mathbf{I}_m \otimes \mathbf{v}' & \mathbf{I}_m \otimes \mathbf{u}' & \mathbf{0}_{m \times (2m-2)} & \mathbf{0}_{m \times 2} \\ \mathbf{C}_{(m-1) \times mp} & \mathbf{0}_{(m-1) \times 2m} & \mathbf{I}_{m-1} \otimes \mathbf{u}' & \mathbf{0}_{(m-1) \times 2} \\ \mathbf{D}_{2 \times mp} & \mathbf{0}_{2 \times 2m} & \mathbf{0}_{m \times (2m-2)} & \mathbf{I}_2 \otimes \mathbf{u}' \end{bmatrix} \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \\ \mathbf{e}^{(3)} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_m \\ \mathbf{0}_{m-1} \\ \mathbf{By} \end{bmatrix} \quad (6)$$

Donde \otimes representa el producto tensorial de Kronecker, $\mathbf{u}' = [-1; 1]$, $\mathbf{v}' = [1; -2; 1]$ (si y sólo si $p = 3$), los vectores $\mathbf{e}^{(c)}$ son vectores de discrepancias y \mathbf{B} , \mathbf{C} y \mathbf{D} son las siguientes matrices

$$\mathbf{B}_{2 \times mp} = \begin{bmatrix} 3/2 & -1/2 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1/2 & 3/2 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{0}'_{p-1} & \mathbf{u}' & \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \dots & \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \mathbf{0}'_{p-1} \\ \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \mathbf{0}'_{p-1} & \mathbf{u}' & \dots & \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \mathbf{0}'_{p-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \mathbf{0}'_{p-1} & \mathbf{0}'_2 & \dots & \mathbf{0}'_{p-1} & \mathbf{u}' & \mathbf{0}'_{p-1} \end{bmatrix} \quad \text{y} \quad \mathbf{D} = \begin{bmatrix} 1 & \mathbf{0}'_{mp-2} & 0 \\ 0 & \mathbf{0}'_{mp-2} & 1 \end{bmatrix}.$$

Podemos normalizar la notación llamando a cada bloque de \mathbf{A} , \mathbf{x} y \mathbf{b} con el nombre de la matriz a la que pertenece y un subíndice que indique su posición, tal como se hace con los elementos de una matriz cualquiera. Por ejemplo, $\mathbf{A}_{31} = \mathbf{C}$, $\mathbf{A}_{41} = \mathbf{D}$, $\mathbf{x}_1 = \mathbf{x}^{(1)}$, $\mathbf{x}_2 = \mathbf{e}^{(1)}$, $\mathbf{b}_1 = \mathbf{y}$, etc. De este modo, podemos formular la desagregación temporal de la serie de alta frecuencia como un problema de optimización que consiste en hallar un vector \mathbf{x}_1 tal que la suma de las discrepancias $\mathbf{x}_2, \mathbf{x}_3$ y \mathbf{x}_4 sea mínima. Dicho de otro modo, se trata de resolver el programa lineal

$$\underset{\mathbf{x}}{\text{mín}\{\mathbf{z}'\mathbf{x}\}} \text{ sujeto a } \mathbf{Ax} = \mathbf{b}, \mathbf{x}_2 \geq \mathbf{0}, \mathbf{x}_3 \geq \mathbf{0}, \mathbf{x}_4 \geq \mathbf{0} \quad (7)$$

donde $\mathbf{z}' = [\mathbf{0}'_{mp}, \mathbf{1}'_{6m-2}]$. Los últimos $6m - 2$ elementos de \mathbf{x} corresponden a los errores del sistema de ecuaciones (6) de modo que la solución al programa planteado (una serie de valores mensuales) es aquella que minimiza la suma de las discrepancias definidas más arriba. Conviene hacer notar que en este programa no se impone la restricción $\mathbf{x}_1 \geq \mathbf{0}$ como es usual en LP.

La expresión (6) describe un sistema de restricciones que es específico para desagregar una serie trimestral en una serie mensual. Para desagregar una serie anual en trimestres, la única modificación que se necesita realizar en la matriz \mathbf{A} es redefinir el vector \mathbf{v}' como $\mathbf{v}' = [1, -1, -1, 1]$, ya que, contrariamente a lo que sugiere la intuición, las matrices \mathbf{B} y \mathbf{D} no cambian porque las condiciones (1) a (3) no cambian. Para comprobarlo, basta llamar \bar{x} al promedio simple $(x_2 + x_3) / 2$, reemplazar x_i por \bar{x} en (2) y seguir el razonamiento que condujo a las condiciones (4) y (5) que definen \mathbf{B} y \mathbf{D} .

El programa (7) es solo un programa minimizador que devuelve una serie de alta frecuencia compatible con otra de baja frecuencia sin más información que un conjunto de restricciones lineales que garanticen una cierta coherencia entre ambas. No obstante, en ocasiones se dispone de una o más series de alta frecuencia presumiblemente relacionadas con la de baja frecuencia. En estos casos, el programa (7) puede ampliarse agregando k ecuaciones adicionales del tipo

$$\mathbf{I}_{mp}\mathbf{x}_1 + \mathbf{0}_{m \times 2m}\mathbf{x}_2 + \mathbf{0}_{m \times (2m-2)}\mathbf{x}_3 + \mathbf{0}_{m \times 2}\mathbf{x}_4 + (\mathbf{I}_{mp} \otimes \mathbf{u}')\mathbf{x}_h = \mathbf{b}_h, \forall h = 5, \dots, k + 4 \quad (8)$$

donde \mathbf{x}_h es un vector de discrepancias análogo a \mathbf{x}_2 , \mathbf{x}_3 y \mathbf{x}_4 , sujeto a las mismas restricciones de no-negatividad; y \mathbf{b}_h es un vector de dimensión $mp \times 1$ que representa cada una de las k series relacionadas. Los k sistemas adicionales se agregan al final de las matrices \mathbf{A} , \mathbf{x} y \mathbf{b} para conformar un programa lineal análogo a (7). Lógicamente, el vector \mathbf{z} también se extiende agregando tantos vectores $\mathbf{1}'_{mp}$ como series relacionadas haya. En lo sucesivo, llamaremos a esta forma de desagregar series temporales el método LP. Si se tratara solo de interpolar una serie, hablaremos de interpolación LP; o de conciliación LP si además se aprovecharan una o más series de alta frecuencia para desagregar \mathbf{y} . A continuación, ejemplificamos el uso del método LP desagregando una serie simulada y otra real.

III. Desagregación temporal de datos simulados

Desagregamos primero una serie corta, de apenas cuatro trimestres, con una serie relacionada de 12 meses. Ambas series son ficticias y emulan números índices con base 100 en algún período fuera del intervalo. La primera columna de la Tabla 1, a continuación de la numeración de meses, contiene la serie trimestral original, y la columna siguiente, la serie mensual relacionada. La tercera, cuarta y quinta columna contienen, respectivamente, (a) la serie mensual interpolada (es decir, ignorando la existencia de la serie relacionada) con el método LP; (b) la serie conciliada por LP; y

(c) la serie conciliada por LP a partir de una serie relacionada defectuosa en la que faltan el séptimo y octavo mes (elementos b_{20} y b_{21} de \mathbf{b}). Más adelante, ampliaremos sobre la forma de operar con datos faltantes. Para conciliar la serie mensual con la trimestral eliminamos las restricciones sobre el primero y el último valor de la serie (bloques \mathbf{A}_{41} a \mathbf{A}_{44} de la matriz \mathbf{A} , y los vectores \mathbf{x}_4 y \mathbf{b}_4), para no oscurecer el efecto de la serie relacionada en dichos puntos. Y, tanto para interpolar como para conciliar, rescalamos las series trimestral y mensual originales, rescalamiento que consiste en centrar las series restándoles sus respectivas medias, y sumar al resultado la media de y . Formalmente, conciliamos las series

$$\mathbf{y}^* = \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) \mathbf{y} \quad \text{y} \quad \mathbf{b}_h^* = \left(\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right) \mathbf{b}_h,$$

donde $\mathbf{J} = \mathbf{1}\mathbf{1}'$, y calculamos $\mathbf{x} = \mathbf{x}^* + \mathbf{1}_n \bar{y}$.

La simple inspección de la tabla, permite apreciar que (i) la condición (1) se satisface para todos los trimestres, tanto en la serie interpolada como en las dos conciliaciones; (ii) en las tres series, la segunda condición también se satisface estrictamente, aunque no la hayamos planteado como una restricción de igualdad exacta; (iii) la condición (3), $x_{i+1} = x_{i+2}$, se satisface estrictamente en todos los trimestres de las series conciliadas, pero solo en dos de tres de la serie interpolada; (iv) las series interpolada y conciliada sin datos faltantes coinciden en general, aunque en algunos períodos se mueven en sentidos opuestos, como puede observarse en los primeros cuatro meses y en el último mes; las series conciliadas se mueven prácticamente igual; (v) los valores inicial y final de la serie interpolada satisfacen las condiciones (4) y (5), aunque no hayamos planteado una igualdad estricta; del mismo modo, el valor intermedio de cada trimestre coincide con el promedio trimestral, aunque tampoco hayamos planteado una igualdad estricta. En síntesis, en este primer ejercicio del programa (7), hallamos una solución única que satisfizo estrictamente casi todas las restricciones. Todos los cálculos del ejemplo se realizaron con un programa informático escrito por el autor en el lenguaje matricial de Euler Math Toolbox². Para resolver el programa lineal, recurrimos al algoritmo simplex incluido en Euler Math Toolbox. A pesar de no ser el algoritmo computacional más eficiente, los tiempos de cálculo fueron prácticamente despreciables, tanto en este ejercicio como en el que desarrollaremos en la próxima sección. En el apéndice adjuntamos el código Euler utilizado para desagregar las series.

IV. Desagregación temporal del valor agregado industrial

Es sabido que las Cuentas Nacionales de Argentina son trimestrales, aunque INDEC elabora indicadores mensuales para anticipar la evolución de algunos agregados

² En este sentido, es ilustrativo recordar que el informe metodológico del Estimador Mensual de Actividad Económica (EMAE) de INDEC (2016) aclara "el criterio de selección del método de desagregación temporal entre los descriptos (Boot et al., 1967; Denton, 1971; Fernández, 1981 y Chow y Lin 1971) para cada componente del EMAE se estableció considerando aquel que minimizara cierta medida de volatilidad" (p. 10); aunque sin aclarar qué medida se utilizó. Es decir, en la misma metodología del EMAE, se reconoce el uso de un criterio *ad-hoc* de selección del método de desagregación temporal ante la imposibilidad de combinar los distintos métodos en un procedimiento general.

macroeconómicos fundamentales. Por ejemplo, para anticipar la evolución del PIB, INDEC (2016) elabora el Estimador Mensual de Actividad Económica (EMAE), al que define como "un indicador provisorio de la evolución del PIB a precios constantes de 2004 [...] difundido con un rezago de 50 a 60 días de concluido el mes de referencia" (INDEC, p. 4). Sin embargo, el mismo Instituto advierte: "dado que la estimación del PIB trimestral reúne mayor cantidad de datos, por cerrarse y difundirse unos 30 días después que el EMAE, es habitual que se observen diferencias entre las variaciones trimestrales de ambos indicadores" (INDEC, p. 4). Estas diferencias justifican la política de INDEC de conciliar la serie de EMAE con la del PIB, a fin de presentar al público series de distinta frecuencia, pero perfectamente coherentes entre sí. El método de conciliación del EMAE con el PIB es indirecto, es decir, la conciliación se realiza a nivel de indicadores sectoriales, y estos se agregan en un índice general. Cada componente del EMAE se concilia mediante "aquel [método] que minimiza[ra] cierta medida de volatilidad" (INDEC, p. 10). En el caso de la Industria Manufacturera, se utiliza el método de Denton (DNCN, comunicación personal) y, como indicador sectorial, un índice que se basa en el Estimador Mensual Industrial (EMI) de INDEC. Este indicador, al que llamaremos EMAE-D, es el que se concilia con el VAB trimestral de la letra D (clasificación CIIU rev. 3.1) de la cuenta de producción del Sistema de Cuentas Nacionales.

Tabla 1. Interpolación de una serie trimestral ficticia y conciliación de la misma con una serie mensual también ficticia, por LP.

Mes	Serie trimestral	Serie mensual	Interpolación LP	Conciliación LP	Conciliación LP sin b_{20} y b_{21}
1		97,0	110,5	113,1	112,9
2		101,0	112,0	112,0	112,0
3	112,0	103,0	113,5	110,9	111,1
4		102,0	108,5	110,9	111,1
5		105,0	115,0	115,0	115,0
6	115,0	111,0	121,5	119,1	118,9
7		112,0	121,5	119,1	118,9
8		115,0	119,0	119,0	119,0
9	119,0	114,0	116,5	118,9	119,1
10		112,0	116,5	118,9	119,1
11		107,0	114,0	114,0	114,0
12	114,0	106,0	111,5	109,1	108,9

El informe metodológico del EMI (INDEC, 2006) define a este indicador como "un indicador de coyuntura que mide la evolución de la producción industrial con periodicidad mensual" (INDEC, p. 9), elaborado a partir de un relevamiento a empresas líderes, cámaras empresariales y organismos públicos. Se trata, básicamente, de un índice del tipo Laspeyre de 12 bloques industriales, cada uno de ellos conformado por una cantidad de rubros (no especificada en el informe metodológico) que combinan indicadores de producción directos e indirectos³. Los ponderadores del EMI son los VAB del Censo Nacional Económico 2004/2005, por lo cual, la base del índice se estableció en 100 para el año 2004. Aparte de INDEC, el sector privado también elabora indicadores de coyuntura de la Industria Manufacturera. Entre los cinco índices privados que conocemos, seleccionamos dos por su trayectoria y acceso público: el índice de Producción Industrial (IPI) de la Fundación de Investigaciones Económicas Latinoamericanas (FIEL) y el IPI de Orlando J. Ferreres & Asociados (OJF). El primero (Martín Lucero, 2001) se compone de 78 series (provenientes de cámaras y asociaciones empresariales, organismos oficiales e informantes propios), agrupadas en 15 capítulos, de las cuales se publican regularmente 10. Los ponderadores del IPI-FIEL son básicamente los VAB del Censo Nacional Económico 1994 (INDEC, 1997) (con datos de 1993), por lo cual la base del índice es 100 en el año '93. El segundo índice, el IPI-OJF, es metodológicamente similar al anterior, tiene el mismo año base y una apertura similar en 12 capítulos⁴.

Por otra parte, conviene recordar que tanto el EMAE como el PIB son series provisorias que se actualizan siguiendo un protocolo de revisión, según el cual, al cierre de cada período, se corrige el período anterior. Es decir que, en el informe de abril de 2019, por ejemplo, la serie de Industria Manufacturera del EMAE presentará una primera estimación del mes de febrero, junto con una revisión del mes de enero, en tanto que los meses anteriores serán el VAB trimestral conciliado con el EMAE-D. Además, en abril de 2019 el VAB trimestral tendrá una primera estimación del cuarto trimestre de 2018 (que cierra provisoriamente ese año), una revisión del tercer trimestre, y una revisión del año 2017 completo, siendo los valores anteriores a 2017 ya definitivos. Los IPI de FIEL y OJF, en cambio, siguen esquemas de revisión continuos. FIEL, por ejemplo, estima que, al analizar cada mes, solo dispone de un 40% de la información (contabilizada en términos de las ponderaciones de las series básicas) requerida para cerrar el cálculo del período, de un 60% para el mes inmediato anterior, y de un 70 y 80% de la información para los meses siguientes. A su vez, Orlando J. Ferreres & Asociados advierte en todos sus comunicados de prensa, que los últimos cuatro datos de la serie están sujetos a revisión, en tanto que el último dato es solo una estimación preliminar. Las distintas fuentes de información, así como las distintas políticas de revisión, explican las diferencias entre EMAE-D, IPI-FIEL e IPI-OJF.

³ En referencia al EMI, el informe metodológico de FIEL (Martín Lucero, 2001) afirma que en la primera edición de este índice se utilizaban 100 series para su elaboración.

⁴ El informe metodológico del IPI-OJF no es de acceso público por lo cual no conocemos más detalles que los mencionados.

La multiplicidad de índices disponibles para estimar la evolución de la producción industrial, nos permite conciliar y comparar el método LP con el método de Denton utilizando series reales. Para ello, realizamos tres contrastes estadísticos. En el primero, comparamos la serie de VAB industrial a precios básicos de 2004 (letra D del cuadro 4 del Informe de Avance del Nivel de Actividad de INDEC) conciliada con EMI mediante el programa (7) con el agregado (8), con la serie EMAE-D publicada por INDEC. Se trata de una comparación justa, porque esta componente del EMAE es básicamente EMI. Además, al ser este último un índice de acceso público, los resultados son perfectamente reproducibles por cualquier investigador. En el segundo, comparamos el VAB industrial conciliado con EMI, IPI-FIEL e IPI-OJF a través del método LP contra el VAB industrial conciliado solo con EMI. En el tercer contraste, comparamos el VAB industrial conciliado con las tres series relacionadas con EMAE-D publicado. Las tres pruebas se focalizaron en las discrepancias de la serie desagregada con la serie trimestral como se explicará más abajo. El primer contraste (LP con EMI versus Denton con EMAE-D), pone a prueba una hipótesis de que el método LP produce series menos ruidosas que el método de Denton porque utiliza un criterio de optimización más robusto (pondera menos las observaciones más extremas) y/o porque combina interpolación y conciliación en un solo proceso, lo cual conduce a una serie desagregada más suave. El segundo contraste (LP con tres series relacionadas versus LP con EMI), pone a prueba la hipótesis que sugiere que la desagregación será más suave cuantas más series se incorporen al programa lineal. Si esta hipótesis fuera cierta, la exclusión de series alternativas en un proceso de desagregación tendría consecuencias potencialmente graves, ya que incrementaría injustificadamente la variabilidad del ajuste. El último contraste (LP con tres series versus Denton con EMAE-D), tiene la finalidad de confirmar las conclusiones de los contrastes anteriores, en particular las del segundo contraste, porque, si bien la serie EMAED se basa principalmente en EMI, incluye también otras series cuya influencia es incierta. El protocolo de las tres pruebas es el siguiente:

1. Se desagregan las series por LP y se calculan las discrepancias entre ellas y las cifras correspondientes del VAB trimestral, elevadas al cuadrado; la serie EMAE-D conciliada se descarga del sitio web de INDEC.
2. Se transforma la serie de discrepancias cuadráticas en rangos; en casos de empates se asigna, como es usual, a los valores repetidos el rango intermedio entre ellos.
3. Se plantea un modelo lineal que clasifica los rangos en función del método de desagregación, el trimestre y el año, siguiendo un esquema de diseño factorial sin interacción (Toutenburg y Shalabh, 2009) como el siguiente:

$$R(d_{ijk}^2) = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2) \quad (9)$$

donde α_i es el efecto del i -ésimo método de desagregación $i = \{1,2\}$; β_j es el efecto del j -ésimo trimestre, $j = \{1, \dots, 4\}$ y γ_k es el efecto del k -ésimo año, $k = f\{1, \dots, 15\}$ y $R(\cdot)$ es el rango de la discrepancia $d_{ijk} = x_{ijk} - y_{jk}$ entre el valor mensual x_{ijk} y el trimestral y_{jk} elevada al cuadrado. El supuesto de normalidad se sustenta en el tamaño de la

muestra (360 observaciones), la que, a priori, consideramos lo "suficientemente grande" como para invocar el Teorema Central del Límite.

Aunque nuestro interés primario es evaluar el efecto del método de conciliación α_i sobre la dispersión de la serie conciliada con respecto de la serie trimestral, incluimos variables clasificatorias de trimestre y año para controlar por efectos estacionales y posibles shocks anuales. Como es usual, en la estimación de este tipo de modelos, cada factor fue reparametrizado, restando la variable clasificatoria del último nivel a las demás variables del mismo factor, lo cual equivale a introducir la restricción lineal de que todos los efectos del mismo factor suman 0. La reparametrización del modelo es fundamental para que el modelo sea estimable. La siguiente expresión representa el modelo anterior en forma matricial para el primer contraste. Advertimos al lector que las matrices **A** y **B** en (10) no deben confundirse con otras matrices del mismo nombre en otras secciones del trabajo.

$$R \left(\begin{bmatrix} \tilde{\mathbf{d}}_{LP-EMI} \\ \tilde{\mathbf{d}}_{EMAE-D} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{1}_{180} & \mathbf{1}_{180} & (\mathbf{1}_{15} \otimes \mathbf{A}) & \mathbf{B} \\ \mathbf{1}_{180} & -\mathbf{1}_{180} & (\mathbf{1}_{15} \otimes \mathbf{A}) & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta^* \\ \gamma^* \end{bmatrix} + \epsilon, \quad \epsilon \sim N(\mathbf{0}_{360}, \sigma^2 \mathbf{I}_{360}) \quad (10)$$

Donde

$$\mathbf{A}_{12 \times 3} = \begin{bmatrix} \mathbf{I}_3 \otimes \mathbf{1}_3 \\ -\mathbf{1}'_3 \otimes \mathbf{1}_3 \end{bmatrix} \text{ y } \mathbf{B}_{180 \times 14} = \begin{bmatrix} \mathbf{I}_{14} \otimes \mathbf{1}_{12} \\ -\mathbf{1}'_{14} \otimes \mathbf{1}_{12} \end{bmatrix}.$$

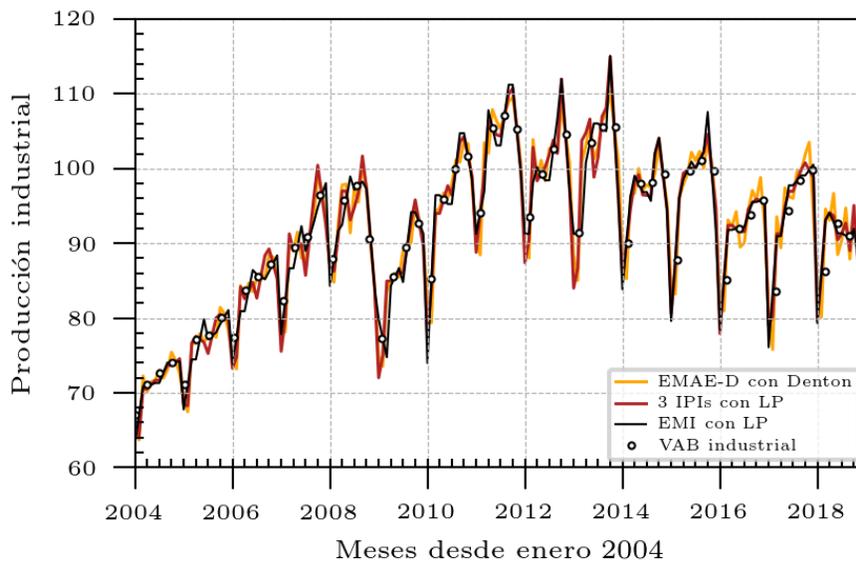
Los vectores $\tilde{\mathbf{d}}_{LP-EMI}$ y $\tilde{\mathbf{d}}_{EMAE-D}$ en (10) corresponden a los rangos $R(\cdot)$ de la serie de VAB conciliada con EMI (siguiendo el método LP) y la serie EMAE-D publicada por INDEC, respectivamente. Para los otros contrastes, el modelo es similar en todo, salvo que los vectores $\tilde{\mathbf{d}}$ se reemplazan por las alternativas que correspondan. Los vectores $\mathbf{1}_h$ simbolizan columnas de dimensión $h \times 1$; \mathbf{I}_{2h} es una matriz identidad de tamaño $2h \times 2h$, y el símbolo \otimes representa el producto Kronecker entre matrices. Nótese que, por la forma en que se reparametrizó el modelo, si el signo del parámetro asociado al método de conciliación fuera negativo, significaría que el método LP reduce la variabilidad respecto a las cifras trimestrales. Los resultados del análisis de regresión se presentan en el Apéndice A.

En la Figura 1 se muestra el VAB trimestral desagregado con LP y la serie de EMI solamente, con LP y las series de EMI, IPI-FIEL e IPI-OJF, y con Denton y la serie EMAE-D (publicada por INDEC) en el período 2004-2018. Con fines comparativos, superpusimos en el primer mes de cada trimestre las cifras trimestrales del VAB. Todas las series fueron rescaladas a 100 en el año 2012, para facilitar la comparación. Se puede observar que las series conciliadas con LP ajustan mejor a la serie trimestral

que aquella conciliada con Denton, pero entre las primeras, la serie conciliada solo con EMI es más suave que aquella conciliada con tres indicadores. Estos efectos son particularmente notables en los últimos años.

La simple inspección de las tablas del Apéndice A revela que, en los modelos que comparan el método LP con Denton (Tablas A2 y A4), el cociente de regresión asociado al método de conciliación es negativo y estadísticamente significativo. En cambio, en la comparación de la serie conciliada con tres IPIs versus la serie conciliada con EMI solo (Tabla A3), el signo es positivo y significativo, revelando que el agregado de series no necesariamente mejora el ajuste con la serie de baja frecuencia. En las tres pruebas los efectos de trimestre resultaron significativos, no así los efectos de año en general. Esto revela un comportamiento sistemático en la varianza de las series conciliadas, introducido por el proceso de conciliación, un punto que parece haber sido pasado por alto en la bibliografía. En otras palabras, hallamos evidencia de que el proceso de conciliación, sea Denton o LP, introduce propiedades espurias en la serie conciliada, las cuales deberían ser consideradas al modelar la serie conciliada con fines de pronóstico. Todos estos resultados, no obstante, deben tomarse con precaución, ya que en dos de las tres regresiones no se verificó el supuesto de normalidad del error que subyace a los estadísticos de prueba mencionados.

Figura 1. Índices de Producción Industrial conciliados por el VAB industrial por Denton y LP.



Nota: Los VAB trimestrales se ubicaron en el primer mes del trimestre)

V. Conclusión

En secciones anteriores presentamos un método de desagregación de temporal basado en programación lineal. Como ya dijimos, este método combina restricciones lineales propias de un método de interpolación, con otras propias de un método de conciliación, ya sea con una o con más series relacionadas. Utilizamos este método para conciliar la componente industrial del EMAE con el VAB trimestral de la industria manufacturera, obteniendo resultados promisorios en términos de ajuste de la serie conciliada con los datos trimestrales. Podemos analizar las implicancias de estos resultados desde dos perspectivas, una relacionada al procesamiento de datos en oficinas de estadísticas públicas, y otra econométrica. Desde la primera:

(a) Al combinar interpolación y conciliación en un solo proceso de desagregación temporal, se exige al analista de decidir a cuál técnica recurrir con cada juego de datos. Esta ventaja, de por sí valiosa, no resuelve, sin embargo, el problema de decidir cuáles ni cuántas series del conjunto de series relacionadas son suficientemente informativas como para incluirlas en el proceso de conciliación. Creemos que la forma adecuada de resolver este problema sería incluir todas las series disponibles, pero penalizando las menos informativas a través de la ponderación que reciban sus discrepancias en la función objetivo. Es decir, en vez de establecer $z_j = 1$ para todo $j > mp$, pensamos que sería más adecuado introducir ponderadores que reduzcan la influencia de las series ruidosas, o poco relacionadas, con la serie de baja frecuencia. Lamentablemente, no pudimos avanzar en este sentido simplemente porque no hallamos un criterio de ponderación que sopesa adecuadamente la cantidad de información aportada por cada serie, asumiendo que esta pudiera ser conocida de antemano.

(b) En segundo lugar, nuestro método permite utilizar series relacionadas incluso con datos faltantes, como mostramos en la simulación. Este problema aparece frecuentemente al trabajar con estadísticas públicas, principalmente por discontinuidades en las series o por retrasos en la publicación. En presencia de datos faltantes, basta con reemplazar los 1 por 0 en las filas del bloque \mathbf{A}_{h1} que correspondan a los elementos faltantes en (6), y reemplazar por 0 los elementos faltantes de \mathbf{b}_h . Estos reemplazos equivalen, en la práctica, a eliminar del sistema de restricciones $\mathbf{Ax} = \mathbf{b}$ las filas con datos faltantes, pero facilitan el manipuleo de matrices porque evitan redefinir las dimensiones de \mathbf{z} , \mathbf{A} y \mathbf{b} en cada oportunidad, aunque quizás a costa de alargar el tiempo de ejecución del algoritmo simplex. Lógicamente, en caso que los datos faltantes, sean el primero y el último valor de la serie relacionada \mathbf{b}_h , se deben reponer los bloques \mathbf{A}_{41} a \mathbf{A}_{44} y los vectores \mathbf{x}_4 y \mathbf{b}_4 en el sistema de restricciones para contar con valores de referencia para x_1 y x_n .

(c) Con el nuevo método, también es posible conciliar la serie de baja frecuencia con series relacionadas de distintas frecuencias. Para ello, se debe reemplazar la matriz identidad del primer término de (8) por una matriz que emule el bloque \mathbf{A}_{11} de la matriz de restricciones, y redefinir la cantidad de filas de los otros términos de la expresión. Por ejemplo, si en el ejercicio disimulación hubiéramos incluido, además de

la serie relacionada mensual una serie semestral, el primer término de (8) debería reemplazarse por $(\mathbf{I}_2 \otimes \mathbf{1}'/6) \mathbf{x}_1$. Los métodos de reconciliación tradicionales no poseen esta flexibilidad y, por lo tanto, tienden a desaprovechar información útil, simplemente porque se la provee con una frecuencia distinta a la deseada.

(d) El método LP constituye una alternativa simple para corroborar los resultados obtenidos por los métodos tradicionales. Este punto no es menor, ya que en la práctica es casi imposible confrontar las interpolaciones obtenidas por el método de Boot et al. (1967), o las conciliaciones obtenidas por Denton (1971), con métodos alternativos, principalmente por la falta de software capaz de manipular grandes cantidades de series.

Desde la perspectiva econométrica, el método de conciliación seleccionado definirá el proceso generador de datos de series, que serán modeladas posteriormente. En este sentido, cabe aclarar que:

(e) A lo largo del trabajo evitamos equiparar conceptos del modelo (7) con conceptos estadísticos. Por ejemplo, en todo momento nos referimos a discrepancias en lugar de errores, o a la solución en lugar de la estimación. Tuvimos este cuidado porque no realizamos supuesto alguno sobre la naturaleza estadística de la serie a desagregar (\mathbf{y}) ni de las series relacionadas \mathbf{b}_h . Por lo tanto, el método descrito en el trabajo debe entenderse como un método de ajuste que garantiza una solución única (siempre que los datos no presenten alguna patología), pero no se trata de una estimación en el sentido estadístico de la verdadera (aunque desconocida) serie de alta frecuencia que originó la de baja frecuencia.

(f) Los resultados del ejercicio de desagregación del VAB industrial revelaron que sumar más series al proceso de conciliación no necesariamente mejora el ajuste a la serie de baja frecuencia. De hecho, al incorporar al proceso de conciliación, los IPI de FIEL y OJF, la variabilidad de la serie conciliada respecto del VAB trimestral fue superior a la obtenida utilizando solo el EMI. Esta conclusión, no obstante, debe tomarse con precaución, en vista de que el efecto es más notable en años más alejados de los años base. Insistimos en que este efecto podría atenuarse a través de la especificación de ponderadores adecuados en la función objetivo, como propusimos más arriba.

(g) Por último, destacamos que, a través del estudio de las discrepancias de la serie conciliada con la serie trimestral, hallamos evidencia de un patrón de heteroscedasticidad espuria introducida por el propio proceso de conciliación, tanto con el método LP como con el de Denton. En el caso del método LP, este patrón podría deberse a la interacción entre las restricciones (2) y (3). Esta observación es de suma importancia para la modelación posterior de series conciliadas, ya que sugiere que debieran probar modelos ARCH o GARCH estacionales (Lütkepohl, 2007). Estudios sobre series desagregadas (mensuales) del PBI del Reino Unido (Mitchell y Weale, 2005b) y la Eurozona (Mitchell y Weale, 2005a) no hallaron evidencia de una estructura ARCH(4) en el error, y un estudio sobre la serie de PIB trimestral de Argentina (Camacho, Dal Bianco y Martínez Martín, 2013) rechazó un estructura ARCH(1), aunque

a decir verdad, como ninguno de estos estudios buscó patrones de heteroscedasticidad estacionales, en general consideraron una cantidad insuficiente de rezagos⁵. Por otra parte, el patrón de estacionalidad (fija) observado en nuestro estudio se basa en una transformación de las discrepancias y no en las discrepancias en sí, lo cual limita también los alcances de nuestra presunción sobre la estructura del error de la serie conciliada. Investigar este punto excede ampliamente los objetivos del trabajo.

⁵ Al respecto, Camacho et al. (2013) exponen: "However, it is important to note that in spite of its high volatility, the Argentine GDP growth does not follow an ARCH process. We carry out ARCH tests for the residual from an AR(1) specification of Argentine real GDP growth and were not able to reject the null of no-presence of ARCH in the residuals at usual significance levels" (p. 2)

VI. Referencias

Bee Dagum, E. y P. A. Cholette (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. New York: Springer.

Boot, J.C.G., Feibes, W. y J.H.C. Lisman (1967). "Further methods of derivation of quarterly figures from annual data", *Applied Statistics*, 16 (1), 65-75.

Camacho, M., Dal Bianco, M. y J. Martínez-Martín (2013). *Short-Run Forecasting of Argentine GDP Growth*. Working Paper Number 13/14. BBVA Research. Madrid. Disponible online en: <https://www.bbvaresearch.com>

Chow, G.C. y A.L. Lin (1971). "Best linear unbiased interpolation, distribution and extrapolation of time series by related series". *The Review of Economics and Statistics*, 53, 372-375.

Denton, F.T. (1971). "Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization". *Journal of the American Statistical Association*, 66 (333), 99-102.

Fernández, R. B. (1981). "A methodological note on the estimation of time series". *The Review of Economics and Statistics*, 63 (3), 471-476.

Martín Lucero, L. E. (2001). *Reseña: Índice de Producción Industrial y sus ciclos*. Documento de Trabajo Nro. 70. Fundación de Investigaciones Económicas Latinoamericanas (FIEL).

International Monetary Fund. Statistics Department (2018). *Quarterly National Accounts Manual 2017 Edition*. Washington, DC. Disponible on-line en: <https://www.imf.org/external/pubs/ft/qna/>

Greene W. (2008). *Econometric Analysis*. New Jersey: Pearson-Prentice Hall.

Instituto Nacional de Estadística y Censos (INDEC) (1997). *Censo Nacional Económico 1994*. Resultados definitivos, versión revisada. Total del país: industria, comercio, servicios, petróleo y gas, otras explotaciones mineras, intermediación financiera. (Serie A). Buenos Aires.

Instituto Nacional de Estadística y Censos (INDEC) (2006). *Estimador Mensual Industrial (EMI)*, Año base 2004=100. Anexo metodológico. Buenos Aires. ISSN 0327-7968. Disponible on-line en: <https://www.indec.gob.ar/>

Instituto Nacional de Estadística y Censos (INDEC) (2016). *Estimador Mensual de Actividad Económica*, EMAE: base año 2004. Metodología INDEC nro. 20. Primera edición. Ciudad Autónoma de Buenos Aires. 12 p. Disponible on-line en: <https://www.indec.gob.ar/>

Infante, E. (2017). *Two-Step Reconciliation of Time Series. New Formulation and Validation*. Tesis. Università degli Studi di Napoli. Dipartimento di Scienze Economiche e Statistiche. 198 p. Disponible en: <http://www.fedoa.unina.it/11731/>

Jacobs J., Kroonenberg S. y T. Wansbee (1992). *Dividing by 4. An efficient algorithm for the optimal disaggregation of annual data into quarterly data*. Disponible on-line en:

<http://www.eco.rug.nl/medewerk/JACOBS/jjdownload/graytex.pdf>

Jarque, C y Bera, A. (1987). "A test for normality of observations and regression residuals", *International Statistical Review*, 55,163-172.

Lütkepohl, H. (2007) *New Introduction to Multiple Time Series Analysis*. Heidelberg: Springer-Verlag.

Mitchell, J. y M. Weale (2005a). *Estimates of monthly GDP for the Euro area: an application of the regression approach to interpolation*. Working Papers and Studies. Luxembourg: Office for Official Publications of the European Communities. ISSN 1725-4825. ISBN 92-79-01307-6.

Mitchell, J. y M. Weale (2005b). "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth", *The Economic Journal*, 115 (501),F108-F129.

Toutenburg H. y Shalabh (2009). *Statistical Analysis of Designed Experiments.*, New York: Springer.

Williams, H.P. (2013). *Model Building in Mathematical Programming.*, West Sussex: John Wiley & Sons Ltd

VII. Apéndices

VII.1. Comparación de los métodos LP y Denton

Las Tablas A1 a A3 presentan los parámetros estimados, desvíos estándar, estadísticos de prueba y valor- p de los tres modelos utilizados para evaluar el ajuste de las series conciliadas con programación lineal versus la serie EMAE-D conciliada con el método de Denton (serie oficial de INDEC). Se puede observar por inspección de los estadísticos t que todos los coeficientes asociados al método de conciliación y al trimestre resultaron significativos al 5 %, al igual que los estadísticos F de ajuste global del modelo (incluso al 1 %) aunque la variabilidad total explicada por el modelo, medida a través del R^2 , apenas superó el 30% en los tres modelos. La verificación de supuestos de las regresiones arrojó los siguientes resultados:

(a) No se observaron indicios de multicolinealidad entre los regresores. Este resultado era esperable debido a la reparametrización del modelo. Los factores de inflación de varianzas (VIF) de todas las variables resultaron menores a 2, en los 3 modelos (Greene, 2008).

(b) No se hallaron indicios de heteroscedasticidad en el error en ningún caso a través de la prueba de White (Greene, 2008). No obstante, con el *test* de Golfeld-Quandt hallamos que en los dos modelos que involucraban la serie EMAE-D, la variable asociada al método de conciliación explicaba significativamente la variabilidad del error. Este resultado no debe sorprender ya que es sabido que el *test* de White es muy general, pero tiene baja potencia.

(c) Hallamos evidencia de autocorrelación en el error en los tres modelos con el *test* de Breusch-Godfrey (Greene, 2008) (corrido hasta un orden $p = 4$), razón por la cual recalculamos los coeficientes de regresión con el estimador iterativo de Cochran-Orcutt. En las tablas de resultados señalamos con el supraíndice a aquellos coeficientes que resultaron significativos con el estimador OLS y con b aquellos que resultaron significativos con el estimador de Cochran-Orcutt.

(d) Las pruebas de normalidad de los residuos de Jarque y Bera (1987) arrojaron resultados mixtos. En el modelo comparativo del método LP con EMI versus Denton con EMAE-D no rechazamos la hipótesis de normalidad. En cambio, en los otros dos modelos rechazamos dicha hipótesis con probabilidad de error tipo II menor al 1 %.

Tabla A1. Modelo explicativo de las discrepancias entre la serie conciliada por LP (con EMI) y por Denton (con EMAE-D) y el VAB trimestral de la Industria Manufacturera.

Variable	Coefficiente	Desv. Est.	Estadístico <i>t</i>	Valor- <i>p</i>
Ordenada	180,5000	4,8342	37,3384	< 0, 0001 ^{ab}
LP _{EMI} - DEN _{EMAE}	-22,3500	4,8342	-4,6233	< 0, 0001 ^{ab}
Trim. I - Trim. IV	65,3667	8,3730	7,8068	< 0, 0001 ^{ab}
Trim. II - Trim. IV	-32,4667	8,3730	-3,8775	0, 0002 ^{ab}
Trim. III - Trim. IV	-39,5667	8,3730	-4,7255	< 0, 0001 ^{ab}
Año 2004-2018	-48,1667	18,0878	-2,6629	0, 0040 ^{ab}
Año 2005-2018	-38,2292	18,0878	-2,1135	0, 0165 ^{ab}
Año 2006-2018	-31,8125	18,0878	-1,7588	0,0379
Año 2007-2018	8,7708	18,0878	0,4849	0,3169
Año 2008-2018	18,7500	18,0878	1,0366	0,1505
Año 2009-2018	-23,3333	18,0878	-1,2900	0,0974
Año 2010-2018	12,8542	18,0878	0,7107	0,2424
Año 2011-2018	24,1042	18,0878	1,3326	0,0901
Año 2012-2018	12,9583	18,0878	0,7164	0,2406
Año 2013-2018	3,2292	18,0878	0,1785	0,4263
Año 2014-2018	12,4375	18,0878	0,6876	0,2503
Año 2015-2018	14,8750	18,0878	0,8224	0,2083
Año 2016-2018	-11,7500	18,0878	-0,6496	0,2618
Año 2017-2018	20,3333	18,0878	1,1242	0,1303

Notas: La variable dependiente es el rango del cuadrado de las discrepancias. Los supraíndices a y b indican significatividad al 5% de los coeficientes ajustados por OLS y Cochran-Orcutt.

Tabla A2. Modelo explicativo de las discrepancias entre la serie conciliada por LP con 3 IPIs y por LP con EMI.

Variable	Coefficiente	Desv. Est.	Estadístico <i>t</i>	Valor- <i>p</i>
Ordenada	180,5000	5,0664	35,6268	< 0, 0001 ^{ab}
LP3-IPI - LPEMI	11,6056	5,0664	2,2907	0, 0106 ^{ab}
Trim. I - Trim. IV	56,1778	8,7753	6,4018	< 0, 0001 ^{ab}
Trim. II - Trim. IV	-38,1000	8,7753	-4,3418	0, 0001 ^{ab}
Trim. III - Trim. IV	-15,5389	8,7753	-1,7708	0, 0369 ^b
Año 2004-2018	-49,3958	18,9568	-2,6057	0, 0047 ^{ab}
Año 2005-2018	-31,2500	18,9568	-1,6485	0,0481
Año 2006-2018	-13,8125	18,9568	-0,7286	0,2368
Año 2007-2018	7,5417	18,9568	0,3978	0,3468
Año 2008-2018	16,3750	18,9568	0,8638	0,1963
Año 2009-2018	-2,3125	18,9568	-0,1220	0,4493
Año 2010-2018	15,2500	18,9568	0,8045	0,2137
Año 2011-2018	8,1458	18,9568	0,4297	0,3358
Año 2012-2018	22,1875	18,9568	1,1704	0,1204
Año 2013-2018	31,5833	18,9568	1,6661	0,0463
Año 2014-2018	21,3333	18,9568	1,1254	0,1301
Año 2015-2018	3,6667	18,9568	0,1934	0,4205
Año 2016-2018	-36,9583	18,9568	-1,9496	0, 0245 ^{ab}
Año 2017-2018	-2,9167	18,9568	-0,1539	0,4362

Notas: La variable dependiente es el rango del cuadrado de las discrepancias. Los supraíndices a y b indican significatividad al 5% de los coeficientes ajustados por OLS y Cochran-Orcutt.

Tabla A3. Modelo explicativo de las discrepancias entre la serie conciliada por LP con 3 IPIs y por Denton con EMAE-D.

Variable	Coefficiente	Desv.Est.	Estadístico <i>t</i>	Valor- <i>p</i>
Ordenada	180,5000	4,4869	40,2283	< 0, 0001 ^{ab}
LP3-IPI - DENEMA	-9,0139	4,4869	-2,0089	0, 0213 ^{ab}
Trim. I - Trim. IV	88,5944	7,7715	11,3999	< 0, 0001 ^{ab}
Trim. II - Trim. IV	-51,7278	7,7715	-6,6561	< 0, 0001 ^{ab}
Trim. III - Trim. IV	-37,1389	7,7715	-4,7788	< 0, 0001 ^{ab}
Año 2004-2018	-58,6875	16,7884	-3,4957	0, 0004 ^{ab}
Año 2005-2018	-49,4792	16,7884	-2,9472	0, 0019 ^{ab}
Año 2006-2018	-23,8958	16,7884	-1,4234	0,0759
Año 2007-2018	16,3125	16,7884	0,9717	0,1666
Año 2008-2018	25,6458	16,7884	1,5276	0,0617
Año 2009-2018	-21,6875	16,7884	-1,2918	0,0971
Año 2010-2018	6,9792	16,7884	0,4157	0,3407
Año 2011-2018	9,1042	16,7884	0,5423	0,2974
Año 2012-2018	0,1875	16,7884	0,0112	0,4938
Año 2013-2018	52,8750	16,7884	3,1495	0, 0011 ^a
Año 2014-2018	-0,8333	16,7884	-0,0496	0,4797
Año 2015-2018	-1,3333	16,7884	-0,0794	0,4673
Año 2016-2018	-15,8125	16,7884	-0,9419	0,1746
Año 2017-2018	24,6667	16,7884	1,4693	0,0693

Notas: La variable dependiente es el rango del cuadrado de las discrepancias. Los supraíndices a y b indican significatividad al 5% de los coeficientes ajustados por OLS y Cochran-Orcutt.

VII.2. B Programa informático para desagregación de series con LP

El siguiente programa devuelve una serie mensual x conciliada con la serie trimestral y , a partir de k series mensuales relacionadas, ubicadas en las columnas de \mathbf{X} . Las dimensiones de \mathbf{X} e y son, respectivamente, $mp \times k$ y $m \times 1$. Si no se dispone de series relacionadas, la matriz \mathbf{X} se reemplaza por la matriz vacía $[\]$ en el argumento de la función. No es necesario centrar las series, ya que el programa lo realiza automáticamente. El programa es específico para conciliar series trimestrales con series mensuales y no admite datos faltantes. Para conciliar series anuales con series

trimestrales, se debe definir $p = 4$ en la primera fila del código y reemplazar el vector $[1,-2,1]$ en la fila 13 del código por $[1,-1,-1,1]$.

```
function LPrecon(X,y)

p:=3;

m:=rows(y);

k:=cols(X);

n:=m*p;

mu:=mean(y');

if isempty(X)==0

    X:=(eye(n)-ones(n,n)/n).X;

endif;

y:=(eye(m)-ones(m,m)/m).y;

A:=zeros(3*m-1,m*p);

for i=1 to m

    A[i,(i-1)*p+1:i*p]:=ones(1,p)/p;

    A[m+i,(i-1)*p+1:i*p]:=[1,-2,1];

end;

for i=1 to m-1

    A[2*m+i,i*p:i*p+1]:=[-1,1];

end;

B:=zeros(3*m-1,4*m-2);

for i=m+1 to 3*m-1

    B[i,2*(i-m)-1:2*(i-m)]:=[-1,1];

end;

A:=A|B;

b:=y_zeros(2*m-1,1);

c:=zeros(1,m*p)|ones(1,4*m-2);

eq:=zeros(3*m-1,1);
```

```
restrict:=c;
if isempty(X)==0
    A:=A[:,1:m*p];
    for j=1 to k
        A:=A_eye(m*p);
    end;
    B:=zeros((k*p+3)*m-1,2*(2+k*p)*m-2);
    for i=m+1 to (k*p+3)*m-1
        B[i,2*(i-m)-1:2*(i-m)]:=[-1,1];
    end;
    A:=A|B;
    b:=y_zeros(2*m-1,1);
    for j=1 to k
        b:=b_X[:,j];
    end;
    c:=zeros(1,m*p)|ones(1,2*(2+k*p)*m-2);
    eq:=zeros((k*p+3)*m-1,1);
    restrict:=c;
endif;
if isempty(X)==1
    A:=A|zeros(3*m-1,4);
    B:=zeros(2,(p+4)*m+2);
    B[1,1]=1; B[1,(p+4)*m-1:(p+4)*m]:=[-1,1];
    B[2,m*p]=1; B[2,(p+4)*m+1:(p+4)*m+2]:=[-1,1];
    A:=A_B;
    b:=b_(3*y[1]/2-y[2]/2);
    b:=b_(3*y[m]/2-y[m-1]/2);
```

```
c=c|ones(1,4);  
eq=eq_zeros(2,1);  
restrict=c;  
endif;  
{x,r,i}:=simplex(A,b,c,eq,restrict);  
x[1:m*p]:=x[1:m*p]+mean(mu);  
return x[1:m*p];  
endfunction;
```

El siguiente ejemplo aclara cómo invocar la función LPrecon(X,y) desde la línea de comandos del Euler Math Toolbox.

```
>load LPrecon.e  
>y:=[112,115,119,114]';  
>X:=[97,101,103,102,105,111,112,115,114,112,107,106]';  
>LPrecon(X,y)  
113.083  
112  
110.917  
110.917  
115  
119.083  
119.083  
119  
118.917  
118.917  
114  
109.083
```