ISSN 1666-0285

Asociación Argentina de Economía Agraria

TITULO:

Suavizado de Series con Programación Lineal

Fecha:
14 de octubre de 2019

Categoría: Comunicación A

Frank, Luis[1]
frank@agro.uba.ar

[1] Universidad de Buenos Aires. Facultad de Agronomía. Av. San Martín 4453, C1417DSE. Buenos Aires, Argentina.

Suavizado de Series con Programación Lineal

Resumen

El artículo presenta un filtro basado en la programación lineal, desarrollado para suavizar series de tiempo ruidosas, entendiendo por ruido cualquier movimiento errático que afecta un sólo un período de tiempo. El artículo presenta también una versión ampliada del filtro, desarrollada para obtener una serie suave y a la vez sintética de múltiples series relacionadas. El desarrollo de este filtro forma parte de un programa de investigación más amplio sobre interpolación, desagregación temporal y equilibrio de matriz con programación lineal.

Abstract

The article presents a filter based on linear programming, developed to smooth noisy time series, understanding noise as any erratic movement that affects only a single time period. The article also presents an expanded version of the filter developed to obtain a smooth and also synthetic series when more than one proxy series are available. The filter is part of a broader research program on interpolation, temporal disaggregation and matrix equilibrium with linear programming.

1 Introducción

Se nos provee una serie de tiempo y se nos pide remover de la misma los movimientos erráticos puntuales, reteniendo aquellos movimientos que prevalecen por más de un período, conocidos en la bibliografía como tendencia, ciclo y estacionalidad. Este requerimiento surge típicamente en oficinas de estadísticas públicas para elaborar agregados macroeconómicos. Por ejemplo, al elaborar indicadores de actividad sectorial es común que al deflactar ventas a precios corrientes se observen en la serie deflactada comportamientos erráticos introducidos por el propio deflactor y que no se relacionan en absoluto con las ventas en unidades físicas. Los mismo ocurre cuando se convierten ventas en moneda extanjera a moneda local con tipos de cambio financieros, o se valorizan inventarios de bienes durables con precios diarios. En todos estos casos es aconsejable remover del resultado la variabilidad introducida espuriamente al procesar los datos originales. En general, esta remoción se llama *filtrado* o *suavizado* de la serie y tiene por objeto exponer aquellos movimientos que abarcan más de un período.

El filtrado y/o suavizado de series tiene una larga tradición en economía, motivada principalmente por la necesidad de extraer la componente cíclica de agregados macro como el PIB. El filtro de Hodrick y Prescott [5, 6] es quizás el más utilizado con para remover tendencias. Otros filtros ampliamente utilizados son el de Henderson [4] (utilizado en el proceso de desestacionalizado X-11 de U.S. Census Bureau [9, § 7.19] para remover la tendencia-ciclo) y el filtro de

Butterworth [7, 8]. Estos tres filtros devuelven una serie suave cuya distancia cuadrática con la serie original es m´ınima. Sin embargo, una desventaja de estos tres filtros (poco tratada en la bibliografía) es que al minimizar formas cuadráticas no permiten filtrar varias series simultáneamente. Esta desventaja es particularmente molesta cuando se dispone de varias series *proxy* de un mismo agregado macro y el analista no encuentra razones para elegir una de ellas como fuente principal y descartar las demás. Ante esta situación, lo ideal sería disponer de un procedimiento que filtre y resuma o sintetice simultáneamente el conjunto de series relacionadas.

El objetivo del trabajo es presentar un método de filtrado/suavizado de series con programación lineal (LP). El trabajo forma parte de un programa de investigación del autor en métodos para interpolar dos o más valores censales [1], desagregar series económicas [3], e incluso para equilibrar y/o desgregar temporalmente matrices [3] insumo-producto. Es decir, para desarrollar un conjunto de técnicas para procesar datos con con LP. El método que presentamos tiene la ventaja de ser apto no solamente para filtrar sino para sintetizar la información contenida en varias series relacionadas.[1]

2 El programa lineal

Llamemos $y_1, \ldots, y_i, \ldots, y_n$ a la serie original y $x_1, \ldots, x_i, \ldots, x_n$ a la serie suavizada que pretendemos hallar, es decir, libre de comportamientos erráticos. La relación entre cada valor y_i y el correspondiente valor x_i, así como entre este último y sus valores más próximos, puede representarse mediante un sistema de tres ecuaciones como el que sigue

$$\begin{aligned} 0\,x_{i-1} + x_i + 0\,x_{i+1} - e_i^- + e_i^+ &= y_i \\ -x_{i-1} + x_i + 0\,x_{i+1} - e_{i'}^- + e_i^+ &= 0 \\ 0\,x_{i-1} - x_i + x_{i+1} - e_{i''}^- + e_{i''}^+ &= 0. \end{aligned} \tag{1}$$

En este sistema, la primera ecuación representa la relación entre x_i e y_i, la segunda ecuación representa la relación entre x_i y su valor inmediato anterior, y la tercera entre x_i y su valor inmediato posterior. La discrepancia entre ambos lados de cada ecuación (a la que llamaremos e) aparece descompuesta en dos términos, de manera que $y_i - x_i = e_i^-$ si $x_i > y_i$ e $y_i - x_i = e^+$ si $y_i > x_i$ (ver [10, pp. 32-34]). Los términos $e_i \in \mathbb{R}^+$ son pequeñas discrepancias debidas a novedades que sólo afectan al período de referencia y no repercuten más allá de éste. En cambio, los términos $e_{i'}$ y $e_{i''}$ (también números reales positivos) son diferencias entre valores consecutivos de x atribuibles a factores que abarcan más de un período. La descomposición de las discrepancias en dos términos tiene la finalidad de garantizar la no-negatividad de la solución del programa lineal que expondremos en breve.

Si bien planteamos tres ecuaciones para describir la relación entre cada valor de x y los demás puntos relacionados, el lector notará que la tercera ecuación es en verdad idéntica a la segunda pero evaluada en $i+1$, de manera que al componer todas las ecuaciones en un solo sistema de n per´ıodos tendremos n ecuaciones que vinculan a x_i con y_i, pero $n-1$ ecuaciones que vinculan a x_i con x_{i-1}. El sistema de ecuaciones para los n periódos puede expresarse en forma matricial

[1] El lector notará que a lo largo del trabajo evitaremos utilizar términos tales como *error*, *tendencia*, *estacionalidad*, etc. a fin de no adjudicarle involuntariamente a los resultados propiedades análogas a las de conceptos estad´ısticos del mismo nombre.

como

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \otimes \mathbf{u}' & \mathbf{0}_{n\times 2(n-1)} \\ \mathbf{D} & \mathbf{0}_{(n-1)\times 2n} & \mathbf{I}_{n-1} \otimes \mathbf{u}' \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{n-1} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{e}^{(1)} \\ \mathbf{e}^{(2)} \end{bmatrix} \geq \mathbf{0}_{4n-2}. \tag{2}$$

La primera fila de bloques representa la primera ecuación de (1) para n períodos de la serie, mientras que la segunda fila representa la segunda ecuación, que como dijimos es idéntica a la tercera del período anterior. La notación utilizada en (2) es la siguiente: $\mathbf{I}_n$ es una matriz identidad de dimensión $n \times n$; $\mathbf{u}'$ es el vector $[-1, 1]$; el símbolo $\otimes$ representa el producto Kronecker entre $\mathbf{I}$ y $\mathbf{u}'$; y $\mathbf{x}$ e $\mathbf{y}$ son vectores columna con los n elementos de x e y, respectivamente. Llamamos $\mathbf{e}^{(1)}$ al vector de n pares $[e_i^-, e_i^+]'$ y $\mathbf{e}^{(2)}$ al vector de $n-1$ pares $[e_i^-, e_i^+]'$. Los subíndices al pie de cada matriz indican las dimensiones de la misma en forma de filas por columnas, salvo en el caso de matrices cuadradas cuyas dimensiones resumimos en un solo sub´ındice tanto para filas como para columnas. Los supraíndices entre paréntesis numeran vectores con el mismo nombre. El lector no deber´ıa confundir matrices y vectores ya que las primeras se escriben con mayúscula y los segundos con minúscula. Todos los vectores son vectores columna salvo indicación en contrario. La forma expandida del producto $\mathbf{I} \otimes \mathbf{u}'$ y de la matriz $\mathbf{D}$ son

$$\mathbf{I} \otimes \mathbf{u}' = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & -1 & 1 & \dots & \dots & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & -1 & 1 & 0 \end{bmatrix} \quad \text{y} \quad \mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & \dots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 1 & 1 & & \ddots & \vdots \\ 0 & \dots & \dots & 0 & -1 & 1 \end{bmatrix}$$

El método de suavizado que proponemos es básicamente la solución del programa lineal que minimiza la suma de discrepancias de $\mathbf{1}'_{2n}\mathbf{e}^{(1)} + \mathbf{1}'_{2(n-1)}\mathbf{e}^{(2)}$ del sistema (2). Es decir, si representamos este sistema en el formato típico de sistemas lineales $\mathbf{Ax} = \mathbf{b}$, el programa que proponemos es aquel que resuelve el problema de minimización

$$\min_{\mathbf{x}} \{\mathbf{z}_1'\mathbf{x}_1 + \mathbf{z}_2'\mathbf{x}_2\} \quad \text{sujeto a} \quad \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \quad \text{para todo} \quad \mathbf{x}_2 \geq \mathbf{0}, \tag{3}$$

donde $\mathbf{z}_1 = \mathbf{0}_n$, $\mathbf{z}_2 = \mathbf{1}_{4n-2}$, $\mathbf{x}_1$ es la serie suavizada que deseamos hallar (antes llamada $\mathbf{x}$ a secas) y $\mathbf{x}_2$ es el vector de discrepancias compuesto por $\mathbf{e}^{(1)}$ y $\mathbf{e}^{(2)}$. No debería confundir el hecho de que en esta última expresión los subíndices de bloques no indican dimensiones sino posiciones relativas dentro de la matriz. Entonces, la función objetivo es simplemente la suma de las discrepancias en valor absoluto. En esta nueva notación el bloque $\mathbf{A}_{11}$ es la matriz identidad $\mathbf{I}_n$, $\mathbf{A}_{21}$ es la matriz que antes llamamos $\mathbf{D}$ y el lado derecho de $\mathbf{A}$, representado por $\mathbf{A}_{12}$ y $\mathbf{A}_{22}$, es el producto Kronecker $\mathbf{I}_{2n-1} \otimes \mathbf{u}'$. Del lado derecho del sistema de restricciones, $\mathbf{b}_1 = \mathbf{y}$ y $\mathbf{b}_2 = \mathbf{0}_{n-1}$. En este programa solamente los elementos del vector de discrepancias aparecen restringido al campo de los números reales positivos.

2.1 Múltiples series y graduado del suavizado

El programa (3) puede expandirse fácilmente para filtrar o suavizar varias series simultáneamente. Esta posibilidad es útil cuando se dispone de varias series *proxy* del mismo agregado macro como ocurre, por ejemplo, con los índices de producción industrial (IPI) elaborados por INDEC e instituciones privadas, o la producción agrícola estimada por MINAGRI y USDA.

Para ello, podemos expandir el sistema (2) para incorporar múltiples series agregando una fila de bloques $\mathbf{A}_{11}$ y $\mathbf{A}_{12}$ por cada serie adicional, con sus respectivas series *proxy* del lado derecho de la igualdad. Es decir, si se nos provee una matriz $\mathbf{Y}$ de n per´ıodos por m series, necesitamos vectorizar $\mathbf{Y}$ y plantear el sistema de restricciones

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \otimes \mathbf{u}' & \dots & \mathbf{0}_{n\times 2n} & \mathbf{0}_{n\times 2(n-1)} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{I}_n & \mathbf{0}_{n\times 2n} & \dots & \mathbf{I}_n \otimes \mathbf{u}' & \mathbf{0}_{n\times 2(n-1)} \\ \mathbf{D} & \mathbf{0}_{(n-1)\times 2n} & \dots & \dots & \mathbf{I}_{n-1} \otimes \mathbf{u}' \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(m+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(m)} \\ \mathbf{0}_{n-1} \end{bmatrix}, \quad \text{vec}(\mathbf{Y}) = \begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(m)} \end{bmatrix}. \tag{4}$$

Por otra parte, la minimización de discrepancias entre valores consecutivos de x puede extenderse a una cantidad arbitraria de períodos hacia adelante y hacia atrás. Si llamamos k a una cierta cantidad de per´ıodos, el sistema de restricciones puede ampliarse agregando sendas matrices $\mathbf{D}^{(k)}$, $k = 1, \dots, p$, análogas a $\mathbf{D} = \mathbf{D}^{(1)}$ en $\mathbf{A}$. Lógicamente cuantas más restricciones de este tipo se agreguen, mayor será el suavizado. La siguiente expresión muestra el sistema expandido para minimizar la distancia entre cada valor x_i y los p valores anteriores y posteriores.

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \otimes \mathbf{u}' & \mathbf{0}_{n\times 2(n-1)} & \dots & \mathbf{0}_{n\times 2(n-p)} \\ \mathbf{D}^{(1)} & \mathbf{0}_{(n-1)\times 2n} & \mathbf{I}_{n-1} \otimes \mathbf{u}' & \dots & \mathbf{0}_{n\times 2(n-p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}^{(p)} & \mathbf{0}_{(n-p)\times 2n} & \dots & & \mathbf{I}_{n-p} \otimes \mathbf{u}' \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(p)} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{n-1} \\ \vdots \\ \mathbf{0}_{n-p} \end{bmatrix}, \tag{5}$$

donde, por ejemplo, $\mathbf{D}^{(2)}$ es igual a

$$\mathbf{D}^{(2)} = \begin{bmatrix} -1 & 0 & 1 & 0 & \dots & 0 \\ 0 & -1 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -1 & 0 & 1 \end{bmatrix}.$$

Como advertimos al describir el programa para una sola serie, los primeros y los últimos valores de $\mathbf{x}$ estarán sujetos a menos restricciones que los valores centrales de la serie. Este problema se potencia al considerar p valores por lo cual aconsejamos considerar los extremos de la serie con elevados niveles de suavizado con precaución. Nótese que el valor de p opera como un *parámetro* de suavizado fijado de antemano por el analista. Cuanto mayor sea p, mayor será el suavizado. Sin embargo, aconsejamos prudencia en la elección de p a fin de evitar sobresuavizados de la serie. Como regla práctica, valores de p de 3 o 4 devuelven la tendencia-ciclo en series mensuales mientras que en series trimestrales la misma se obtendr´ıa con valores de $p = 1$. Un alternativa a explorar sería la ponderación de las discrepancias en función de la distancia k. Es decir, redefinir la función objetivo como

$$\min_{\mathbf{x}} \{\mathbf{z}_1' \mathbf{x}_1 + \mathbf{z}_2^{(1)'} \mathbf{x}_2^{(1)} + \cdots + \mathbf{z}_2^{(p)'} \mathbf{x}_2^{(p)}\} \quad \text{donde} \quad \mathbf{z}_1 = \mathbf{0} \quad \text{y} \quad \mathbf{z}_2^{(k)} = \mathbf{f}(k)$$

Por el momento no profundizaremos en esta posibilidad para evitar dispersarnos en la búsqueda en el análisis de funciones que podrían ser adecuadas para este fin.

2.2 Forzado de subtotales

Con frecuencia se requiere que la suma o el promedio de todos los valores de la serie suavizada, o de ciertos tramos de la misma, coincida con la correspondiente suma o promedio de la serie original. Este requisito tiende a evitar confundir al público presentando series supuestamente equivalentes pero con distintos subtotales. Para asegurarlo, se pueden agregar al sistema $\mathbf{Ax} = \mathbf{b}$ las siguientes restricciones

$$\begin{bmatrix} \mathbf{I}_r \otimes \mathbf{1}_s^J & \mathbf{0}_{r\times 2n} & \dots & \mathbf{0}_{r\times 2(n-1)} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \vdots \end{bmatrix} = (\mathbf{I}_r \otimes \mathbf{1}_s^J)\,\mathbf{y}. \tag{6}$$

La matriz $\mathbf{I}_r \otimes \mathbf{1}_s^J$ se ubicaría como bloque $\mathbf{A}_{31}$ en el programa (3), y las matrices nulas en lugar de $\mathbf{A}_{32}$. El vector $(\mathbf{I}_r \otimes \mathbf{1}_s^J)\mathbf{y}$ se ubicaría como vector $\mathbf{y}_3$ en el mismo programa. Si se desea que coincidan los promedios de ambas series, cada elemento del vector unitario se divide por s. Los subíndices r y s se refieren, respectivamente, a la cantidad de tramos en que las sumas de la serie original y la serie suavizada deberían coincidir, y a la cantidad de períodos de cada uno de estos tramos. Por ejemplo, si la serie a suavizar es una serie trimestral y abarca 5 años completos, $r = 5$ y $s = 4$. Lógicamente, si $n > rs$, la restricción no alcanzaría a los últimos valores de la serie. En la expresión (5) asuminos que $n = rs$ para que las dimensiones de los bloques sean conformables. Si rs hubiera sido menor a n, las dimensiones del bloque $\mathbf{I}_r \otimes \mathbf{1}_s^J$ podrían extenderse hasta $r \times n$ simplemente completando con $n - rs$ columnas $\mathbf{0}_r$ al final del bloque.

3 Ejemplo de aplicación

Presentamos a continuación un ejemplo de aplicación en el que combinamos tres índices de producción industrial, el IPI de Orlando Ferreres & Asociados y el de la Fundación de Investigaciones Económicas Latinoamericanas (FIEL), y el Estimador Mensual Industrial (EMI) de INDEC, para obtener una serie suavizada sintética de la producción industrial en el período 2004-2018.[2] Como las dos primeras series se publican con base 100 en 1993, mientras que la tercera (EMI) se publica con base 100 en el año 2004, normalizamos las tres en base 2012 = 100. La figura 1 muestra la serie suavizada con el programa (3) y el sistema de restricciones (4), superpuesta con los mencionados IPI y EMI.

La simple inspección del gráfico permite apreciar que la serie suavizada reproduce el comportamiento de las series originales pero sin tanta variabilidad, tal como cabía esperar. Se puede apreciar también que los valores de la serie suavizada no pasan necesariamente por los puntos medios de las tres series en cada período.

El gráfico que sigue muestra cuatro niveles de suavizado desde $p = 1$ hasta $p = 4$. En el mismo se aprecia el efecto distorsivo que genera la pérdida de restricciones en los extremos de la serie. Este problema, conocido en la bibligrafía como el problema del punto final, es común a casi todos los filtros y no tiene una solución simple. En nuestro caso sugerimos dos alternativas: (i) estimar mediante algún método estándar un valor hacia adelante y hacia atrás de la serie y proceder como si éstos fueran parte de la serie; o (ii) omitir de la presentación de resultados los puntos extremos de la serie suavizada, es decir, presentar sólo los $n - 2$ puntos intermedios.

[2] Recientemente, INDEC amplió la cobertura de EMI y llamó al nuevo indicador Índice de Producción Industrial Manufacturero.

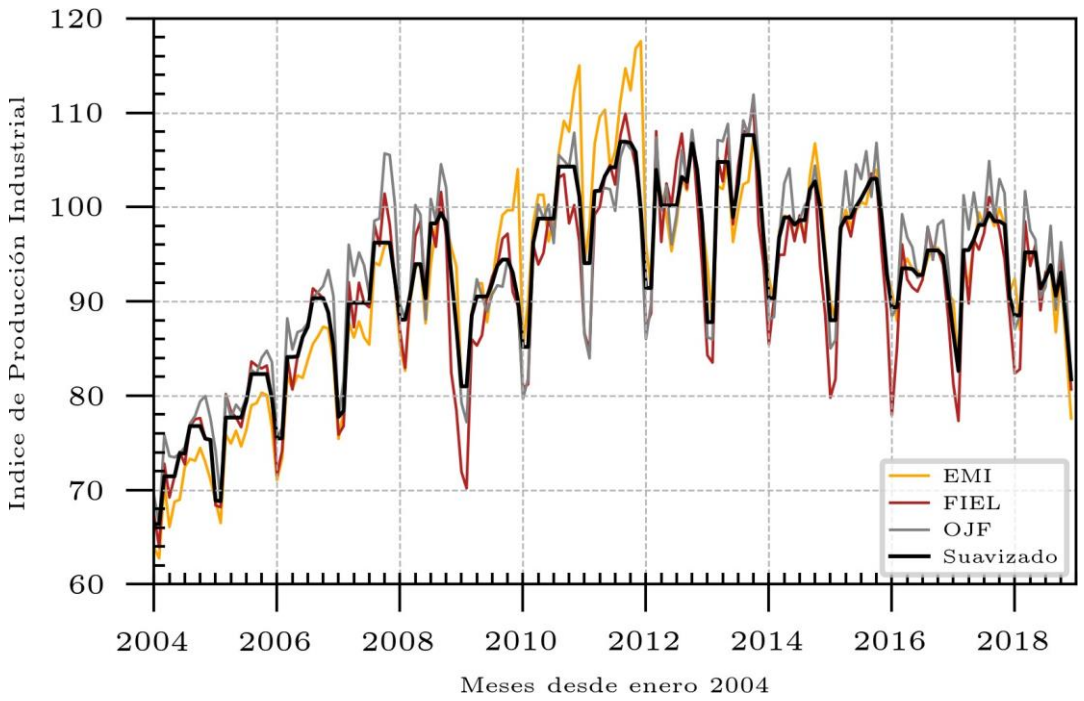

Figura 1: Índices de Producción Industrial suavizados con LP

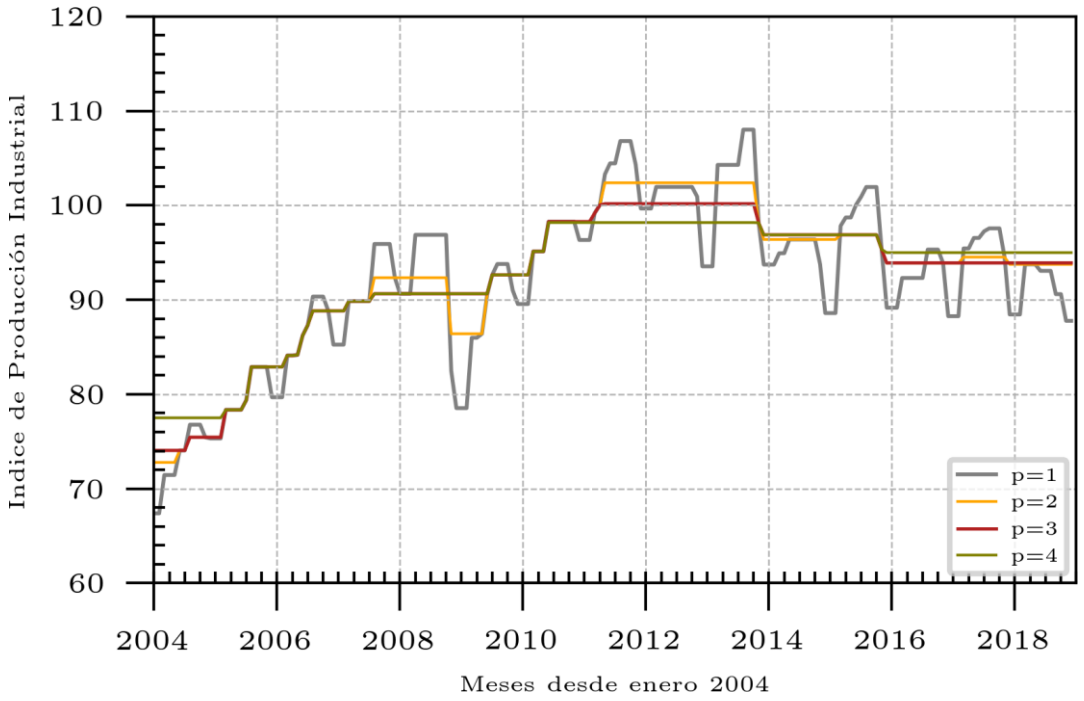

Figura 2: Índices de Producción Industrial suavizados con LP

4 Conclusión

A lo largo del trabajo desarrollamos un método de filtrado/suavizado de series de tiempo con LP. Este método se suma a otras técnicas propuestas por el autor para interpolar, desagregar y equilibrar agregados macro económicos. Básicamente, el suavizado LP halla una serie $\mathbf{x}$ cuyos puntos guardan una distancia o discrepancia m´ınima con aquellos de la serie observada y con puntos vecinos dentro de la misma serie. El nivel de suavizado puede ser graduado por el analista mediante el *parámetro p* que es simplemente la cantidad de períodos hacia adelante y hacia atrás considerados en la minimización de discrepancias entre puntos consecutivos. El método LP tiene la ventaja frente a los métodos de suavizado basados en formas cuadráticas de ser extensible al filtrado simultáneo de varias series relacionadas. Esta posibilidad exime al analista de tener que elegir una fuente de información entre las disponibles evitando sesgos de selección en el resultado final. En este sentido, el método LP es exhaustivo en el uso de información. A futuro prevemos perfeccionar el método para ponderar las series observadas según su valor informativo. Ello se lograría utilizando distintos vectores $\mathbf{z}_2^{(1)}, \ldots, \mathbf{z}_2^{(m)}$ en la función objetivo asociada al sistema (4) de acuerdo al peso informativo que se le quiera dar a cada una de las series observadas.

El filtro LP puede relacionarse con otros filtros más conocidos como el de Hodrick y Prescott (HP) mencionado en la introducción. Para ello, es necesario representar este último como programa lineal, es decir, como un problema de optimización en la norma unitaria en vez del cuadrado de la norma euclídea (ver apéndice A). A continuación transcribimos el filtro HP en forma de programa lineal. Esta versión del filtro es un caso particular del programa (3) en el que $\mathbf{z}_2^J = [\mathbf{1}_{2n}^J, \lambda \mathbf{1}_{2(n-2)n}^J]$ y $\mathbf{x}_2^J = [\mathbf{c}^J, \mathbf{e}^J]$, y los bloques $\mathbf{A}_{11} = \quad \mathbf{I}_n$ y $\mathbf{A}_{21} = \mathbf{D}_1 \quad \mathbf{D}_2$. El vector $\mathbf{c}$ y las matrices $\mathbf{D}_1$ y $\mathbf{D}_2$ son las homónimas del apéndice A.

$$\min_{\mathbf{x}} \{\mathbf{0}_n^J \mathbf{x} + \mathbf{1}_{2n}^J \mathbf{c} + \lambda \mathbf{1}_{2(n-2)}^J \mathbf{e}\} \quad \text{s.a} \quad \begin{bmatrix} -\mathbf{I}_n & \mathbf{I}_n \otimes \mathbf{u}^J & \mathbf{0}_{n\times 2(n-2)} \\ \mathbf{D}_1 - \mathbf{D}_2 & \mathbf{0}_{(n-2)\times 2n} & \mathbf{I}_{n-2} \otimes \mathbf{u}^J \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{c} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{n-2} \end{bmatrix}.$$

Expresado de este modo resulta evidente la imposibilidad de separar en el filtro HP el parámetro de suavizado λ de la ponderación informativa que se le quiera dar a cada punto, o incluso a cada serie si este filtro se expandiera a múltiples series como en (5), ya que el parámetro λ aparecería confundido con los ponderadores $\mathbf{z}_2^{(\cdot)}$. En conclusión, el filtro LP y sus variantes representan verdaderamente una familia distinta de filtros que merece ser estudiada en profundidad, tanto por sus ventajas conceptuales como por su buen comportamiento empírico evidenciado a través del filtrado simultánea de tres IPI.

Referencias

[1] Frank L., 2019. Interpolating Data with Linear Programming. Statistical Journal of the IAOS. Enviado.

[2] Frank L., 2019a. Desagregación temporal de series económicas con programación lineal. Revista Ensayos de Política Económica. Aceptado, 14 de mayo de 2019.

[3] Frank L., 2019b. Reconstructing Matrices with Linear Programming. 2019 Joint Statistical Meeting, Denver, Colorado. Disponible en: `https://www.amstat.org/`

[4] Henderson R., 1916. Note on graduation by adjusted average. Trans. Amer. Math. Soc. 17:43-48.

[5] Hodrick R. J. y E.C. Prescott, 1980. Postwar US business cycles: an empirical investigation. Carnegie Mellon University discussion paper, 451.

[6] Hodrick R. J. y E.C. Prescott, 1997. Postwar US business cycles: an empirical investigation. Journal of Money, Credit, and Banking, (1)16.

[7] Pollock D.S.G., 2000. Trend Estimation and De-Trending via Rational Square Wave Filters. Journal of Econometrics 99:317-334.

[8] Pollock D.S.G., 2014. Econometric Filters. Working Paper No. 14/07 Department of Economics. University of Leicester. UK

[9] U.S. Census Bureau, Center for Statistical Research and Methodology, 2017. X-13ARIMA-SEATS Reference Manual. Disponible en: http://www.census.gov/srd/www/x13as/

[10] Williams H.P., 2013. Model Building in Mathematical Programming. 5th ed. John Wiley & Sons Ltd. West Sussex. 432 p.

A El filtro de Hodrick-Prescott

El filtro de Hodrick y Prescott [5, 6] es un método ampliamente utilizado para remover la componente de tendencia de una serie de tiempo. Para deducirlo, supongamos que la serie bajo estudio y_i puede interpretarse como la suma de dos componentes, un de tendencia (a la que llamaremos x_i) y otra c´ıclica (que llamaremos c_i). Formalmente,

$$y_i = x_i + c_i$$

Hodrick y Prescott definen la siguiente función del ciclo y la tendencia, condicionada al parámetro λ

$$L(c, x|\lambda, y_i) = \sum_{i=1}^{n} c_i^2 + \lambda \sum_{i=3}^{n-1} [(x_{i+1} - x_i) - (x_i - x_{i-1})]^2, \qquad \text{para todo} \quad \lambda \in \mathsf{R}^+,$$

de modo que x es aquella serie de valores que minimiza L respecto de c y x, es decir,

$$\min_x L(x|\lambda, y) \quad \text{sujeto a} \quad y_i = x_i + c_i.$$

El parámetro λ es un factor de ponderación que regula la suavidad de la tendencia. La simple inspección de la fórmula permite apreciar que si $\lambda \to \infty$, la tendencia tiende a ser lineal. Expresado en forma matricial, el problema y la solución son

$$L(\mathbf{x}|\boldsymbol{\lambda}, \mathbf{y}) = (\mathbf{y} - \mathbf{x})'(\mathbf{y} - \mathbf{x}) + \lambda \mathbf{x}'(\mathbf{D}_1 - \mathbf{D}_2)'(\mathbf{D}_1 - \mathbf{D}_2)\mathbf{x} \tag{7}$$

donde $\mathbf{D}_1 = [\mathbf{0}_{n-2}, \mathbf{D}]$ y $\mathbf{D}_2 = [\mathbf{D}, \mathbf{0}_{n-2}]$ y $\mathbf{D}$ es una matriz análoga a la matriz homónima el texto pero de dimensión $(n-2) \times (n-1)$.

$$\mathbf{D}_1 - \mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & \dots & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 & -2 & 1 \end{bmatrix}.$$

Luego, resolviendo las condiciones de primer orden, se obtiene la solución

$$\hat{\mathbf{x}} = [\mathbf{I}_n + \lambda\, (\mathbf{D}_1 - \mathbf{D}_2)'(\mathbf{D}_1 - \mathbf{D}_2)]^{-1}\, \mathbf{y} \tag{8}$$

El parámetros de suavizado es arbitrario, pero la bibliografía sugiere utilizar $\lambda = 100$ con datos anuales, $\lambda = 1600$ con datos trimestrales y $\lambda = 14400$ con datos mensuales.