

Sobre el método de desagregación temporal de Guerrero

Enrique M. Quilis¹

**Instituto Nacional de Estadística
Subdirección General de Cuentas Nacionales**

¹ Agradezco a Ana Abad, Silvia Relloso y Eduardo Salazar las discusiones mantenidas así como la colaboración de Leandro Navarro. Las opiniones expresadas corresponden al autor y no reflejan necesariamente las del INE.

Resumen

En este trabajo se expone el procedimiento de desagregación temporal de Guerrero (1990). Dicho método, por estar basado en una modelización ARIMA explícita de los indicadores de alta frecuencia, permite una integración relativamente sencilla de los procedimientos de extracción de señales basados en modelos con los de desagregación temporal. Adicionalmente, este método genera, como caso particular, los mismos resultados que el de Chow-Lin. Finalmente, se presenta una librería Matlab que facilita la aplicación de este procedimiento y se comparan sus resultados con los de Chow-Lin en una aplicación real.

Palabras clave: Desagregación temporal, método de Guerrero, método de Chow-Lin, modelos ARIMA, extracción de señales, indicadores económicos, Contabilidad Nacional Trimestral.

Contenido

1. Introducción
 2. Planteamiento general
 3. Estimador teórico
 4. Estimador factible
 5. Algoritmo
 6. Relación con el método de Chow-Lin
 7. Función Matlab
 8. Aplicación
- Apéndice A: Derivación de los estimadores
- Referencias

1. Introducción

Las técnicas de extracción de señales y de desagregación temporal juegan un papel muy importante en el análisis económico aplicado y, en particular, en la compilación de la Contabilidad Nacional Trimestral (INE, 1992, 2002). No obstante, la práctica habitual consiste en una utilización separada de ambas, a pesar de que existen buenas razones para combinarlas, tanto desde un punto de vista teórico como práctico.

En efecto, desde una perspectiva teórica, ambas técnicas comparten un enfoque basado en el empleo de componentes inobservables para representar los modelos y diseñar los estimadores correspondientes. Por otra parte, desde un punto de vista aplicado, el uso integrado de ambas técnicas permitiría utilizar de modo más eficiente la información disponible, al mismo tiempo que flexibilizaría los métodos empleados, haciéndolos aplicables en contextos cada vez más diversos.

En esta nota se expone el procedimiento de desagregación temporal de Guerrero (1990), que permite una combinación relativamente sencilla de ambos enfoques por estar basado en una representación ARIMA explícita de la información de alta frecuencia, en un contexto general de estimación lineal, insesgada y de varianza mínima (ELIO).

Con el fin de facilitar el uso de este método, se ha programado como una función Matlab que amplía la librería de desagregación temporal descrita en Quilis (2003). El código fuente puede solicitarse en: www.ine.es/infoine

Algunas extensiones del método son las siguientes: en Guerrero y Martínez (1995) se expone cómo aplicar el método para realizar estimaciones concurrentes sin inducir revisiones, en Guerrero y Nieto (1999) se considera el caso multivariante con restricciones transversales y en Guerrero y Peña (2000) se realiza una interpretación del método dentro del marco de la teoría de combinación óptima de estimadores.

El trabajo tiene la siguiente estructura. En la segunda sección se plantea el problema de estimación así como el conjunto de información disponible. Las dos siguientes secciones derivan, respectivamente, el estimador teórico y el factible. La quinta sección resume toda la metodología bajo la forma de un algoritmo. La relación existente entre el método de Guerrero y el de Chow-Lin se examina en la sexta sección y, en la séptima, se presenta la función escrita en Matlab. El trabajo concluye con una aplicación del método y una comparación con los resultados de aplicar el procedimiento de Chow-Lin.

Las líneas de desarrollo de este estudio son numerosas. Entre ellas, cabe citar una comparación detallada con el método de Chow-Lin, tanto a través del estudio pormenorizado de casos reales como mediante simulaciones de Monte Carlo.

2. Planteamiento general

Sea $Y = \{Y_T : T = 1..N, N > 1\}$ una serie temporal de baja frecuencia (p.e., anual). Se considera que esta serie posee una representación inobservable de alta frecuencia (p.e., trimestral), $y = \{y_t : t = 1..n\}$, cuya agregación temporal es, precisamente, la serie observada Y :

$$[2.1] \quad Y = Cy$$

siendo C la matriz de agregación temporal de dimensión $N \times n$:

$$[2.2] \quad C = I_N \otimes c$$

donde N es el número de observaciones de baja frecuencia, \otimes denota el producto tensorial, c es un vector fila de tamaño s que define el tipo de agregación temporal que se desea realizar y s es el número de períodos de alta frecuencia contenidos en uno de baja, siendo $s > 1$. Si $c = [1, 1, \dots, 1]$ se trata de la agregación de un flujo, si $c = [1/s, 1/s, \dots, 1/s]$ se trata del promedio de un índice y , y si $c = [0, 0, \dots, 1]$, se obtiene un problema de interpolación.

Adicionalmente, cuando el tamaño muestral de la serie de alta frecuencia no es conformable con el de la serie de baja frecuencia ($n > sN$), se presenta una situación de extrapolación. Este caso se puede tratar fácilmente mediante una ampliación de la matriz de agregación temporal C , de forma que se consideran nuevas columnas de ceros que no distorsionan la relación de agregación temporal ni, como se verá más adelante, dificultan la incorporación de los últimos $n - sN$ datos en el proceso de estimación de y_t . En consecuencia, la matriz C expandida es:

$$[2.3] \quad C = (I_N \otimes c \mid O_{N, n-sN})$$

Uno de los aspectos fundamentales del método de Guerrero consiste en el uso de la modelización ARIMA como elemento básico para derivar el estimador de la serie de alta frecuencia. En consecuencia, se postula para esta serie un modelo ARIMA de forma general:

$$[2.4] \quad \phi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\theta_q(B^s)a_{y,t}$$

donde $f_p(B)$ y $q_q(B)$ son, respectivamente, polinomios de orden p y q en el operador de desfases B , y $f_p(B^s)$ y $q_q(B^s)$ son polinomios de orden P y Q en B^s . Las expresiones $(1-B)^d$ y $(1-B^s)^D$ son operadores de diferenciación regular y estacional controlados por los parámetros enteros d y D , respectivamente. Por último, $a_{y,t}$ es una secuencia de ruido blanco gaussiano con esperanza nula y varianza constante σ_y^2 .

El objetivo del método de Guerrero consiste en diseñar un estimador de la serie inobservable de alta frecuencia y de manera que: (i) las estimaciones sean cuantitativamente coherentes con la serie observable de baja frecuencia Y e (ii) que el estimador incorpore la información dinámica contenida en el modelo ARIMA [2.4] postulado para y .

3. Estimador teórico

El predictor de error cuadrático medio mínimo de y condicionado a la información disponible en $t=0$ da lugar al siguiente error de predicción:

$$[3.1] \quad y_t - E_0(y_t) = \sum_{j=0}^{t-1} \psi_j a_{t-j} \quad \psi_0 = 1 \quad t = 1..n$$

Expresando [3.1] en notación matricial:

$$[3.2] \quad y - E_0(y) = Ta$$

siendo T : $n \times n$ una matriz que condensa la expansión de medias móviles (*y-weights*) del modelo ARIMA [2.4]:

$$[3.3] \quad T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi_{n-1} & \psi_{n-2} & \psi_{n-3} & \cdots & 1 \end{bmatrix}$$

Guerrero (1990) demuestra que el estimador lineal de error cuadrático medio mínimo (ELIO) de y que verifica la restricción $Y=Cy$ puede ser representado mediante la siguiente expresión¹:

$$[3.4] \quad \hat{y} = E_0(y) + A[Y - CE_0(y)]$$

siendo:

$$[3.5] \quad A = TT'C'(CTT'C')^{-1}$$

Esta expresión puede interpretarse como sigue: la estimación final \hat{y} se obtiene a partir de una preliminar $E_0(y)$ corregida, de forma ponderada, por las discrepancias entre ésta y la restricción $Y=Cy$. Las ponderaciones, contenidas en la matriz A , se obtienen de forma óptima, en un sentido ELIO.

La correspondiente matriz de varianzas-covarianzas es:

$$[3.6] \quad \Sigma_y = \sigma_y^2 (I_n - AC)TT'$$

De esta manera, resulta posible asociar intervalos de confianza a las estimaciones puntuales obtenidas mediante [3.4].

¹ En el apéndice A se encuentra la demostración detallada.

4. Estimador factible

Puesto que y es inobservable el procedimiento requiere alguna fuente de información que permita hacerlo aplicable. Por esta razón se asume disponible una serie observable $w = \{w_t : t=1..n\}$ cuyo modelo ARIMA es el mismo que el de y , tanto en lo que corresponde al orden de los operadores involucrados como a los valores numéricos de los parámetros:

$$[4.1] \quad \phi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)^D w_t = \theta_q(B)\theta_q(B^s)a_{w,t}$$

Adicionalmente:

$$[4.2] \quad \begin{aligned} \sigma_y^2 &\neq \sigma_w^2 \\ Y &\neq Cw \end{aligned}$$

En consecuencia, la serie w difiere de y en que las varianzas de las innovaciones respectivas no serán iguales en general y que la primera no satisface la condición de agregación temporal mientras que la segunda sí.

¿Cómo obtener w ? En primer lugar, se supone disponible un conjunto de p indicadores de alta frecuencia (p.e., trimestral) $x = \{x_{i,t} : i=1..p, t=1..n, p \geq 1\}$. A continuación, se determina w como una combinación lineal de éstos:

$$[4.3] \quad w = x\hat{\beta}$$

En la sección siguiente se detalla la determinación de las ponderaciones $\hat{\beta}$. A partir de ahí, Guerrero (1990) demuestra que el estimador lineal de error cuadrático medio mínimo de y condicionado a w y que verifica la restricción $Y=Cy$ puede ser representado mediante la siguiente expresión¹:

$$[4.4] \quad \hat{y} = w + A[Y - Cw] = w + A\delta$$

La interpretación es inmediata: \hat{y} es una combinación lineal de un indicador escalado (w) y de una discrepancia ($d=Y-Cw$) temporalmente desagregada. El filtro A que efectúa esta desagregación temporal depende de las propiedades estocásticas de los modelos ARIMA de w y de d , tal y como se condensan en las matrices asociadas a las correspondientes expansiones de medias móviles T y W . El papel de W se detalla en el apéndice A y en la siguiente sección. Las expresiones del filtro son:

$$[4.5a] \quad A = TPT'C'(CTPT'C')^{-1}$$

$$[4.5b] \quad P = T^{-1}\Omega\Omega'T^{-1}$$

¹ En el apéndice A se encuentra la demostración detallada.

La correspondiente matriz de varianzas y covarianzas es:

$$[4.6] \quad \Sigma_y = \sigma^2 (I_n - AC)TPT'$$

Finalmente, la idoneidad de w como *proxy* de $E_0(y)$ puede ser determinada examinando la congruencia de w con Y mediante el siguiente estadístico:

$$[4.7] \quad k = \frac{(Y - Cw)'(CTT'C)^{-1}(Y - Cw)}{\sigma_w^2}$$

Bajo la hipótesis nula de compatibilidad, $Y=CE_0(w)$, el estadístico anterior se distribuye asintóticamente como una χ^2 con $N-p$ grados de libertad.

5. Algoritmo

A continuación se expone un algoritmo completo que resume la metodología propuesta en Guerrero (1990).

5.1. Determinación del indicador escalado

Teniendo en cuenta los resultados anteriores, Guerrero (1990) propone determinar w como una combinación lineal de los p indicadores contenidos en x . Las ponderaciones pueden obtenerse mediante la estimación por mínimos cuadrados ordinarios (MCO) del siguiente modelo de baja frecuencia:

$$[5.1] \quad Y = Cx\beta + U$$

La combinación de los p indicadores se realiza según:

$$[5.2] \quad w = x\hat{\beta}$$

siendo $\hat{\beta} = (x'C'Cx)^{-1} x'C'Y$.

En general, esta estimación será insesgada pero ineficiente¹. No obstante, este último resultado no afecta a la precisión de las estimaciones ya que la matriz de varianzas y covarianzas de $\hat{\beta}$ no interviene en la correspondiente matriz de \hat{y} , véase la expresión [4.6].

5.2. Estimación preliminar

Se modeliza w y se determina su modelo ARIMA. Tomando como un dato dicho modelo, se deriva la matriz T y la correspondiente estimación preliminar:

$$[5.3] \quad \hat{y}_1 = w + A_1[Y - Cw]$$

siendo:

$$[5.4] \quad A_1 = TT'C'(CTT'C')^{-1}$$

Esta estimación preliminar sí es cuantitativamente coherente con la serie de baja frecuencia: $Y=Cy_1$.

¹ A menos que U sea ruido blanco.

5.3. Estimación final

Se calcula la discrepancia $d=y_1-w$ y se contrasta si $b=T^{-1}d$ es ruido blanco. En caso afirmativo, $P=I$ y la estimación preliminar coincide con la final. En caso negativo, se identifica, estima y diagnostica un modelo para d , del que se deriva la correspondiente matriz de su expansión de medias móviles W . A partir de ésta y de T se obtiene P :

$$[5.5] \quad P = T^{-1}\Omega\Omega'T^{-1}$$

Una vez determinada P se aplican las fórmulas antes expuestas:

$$[5.6] \quad \hat{y} = w + A[Y - Cw]$$

siendo:

$$[5.7] \quad A = TPT'C'(CTPT'C')^{-1}$$

La correspondiente matriz de varianzas y covarianzas es:

$$[5.8] \quad \Sigma_y = \sigma^2(I_n - AC)TPT'$$

El término escalar que interviene en la anterior matriz de varianzas y covarianzas puede estimarse mediante una expresión de tipo mínimos cuadrados generalizados (MCG):

$$[5.9] \quad \hat{\sigma}^2 = \frac{(y - w)'(TPT')^{-1}(y - w)}{n - r}$$

siendo r el número de parámetros estimados en el modelo ARIMA de d .

6. Relación con el metodo de Chow-Lin

El método de Chow y Lin (1971) asume que existe un modelo del alta frecuencia que relaciona un vector de p indicadores x con la serie inobservable de alta frecuencia y :

$$[6.1] \quad y = x\beta + u$$

donde b es un vector de p parámetros constantes pero desconocidos y u es una perturbación estocástica de media nula y matriz de varianzas y covarianzas v .

Se asume que y satisface la restricción longitudinal:

$$[6.2] \quad Y = Cy$$

Premultiplicando [6.1] por C se obtiene el modelo de baja frecuencia que vincula la serie Y con el indicador temporalmente agregado $X=Cx$. De esta manera, se obtiene un modelo lineal que relaciona variables observables:

$$[6.3] \quad Y = X\beta + U$$

La matriz de varianzas y covarianzas de U es $V=CvC'$. En general, U no será ruido blanco, por lo que la estimación del modelo anterior ha de realizarse por mínimos cuadrados generalizados.

El objetivo del procedimiento de Chow-Lin consiste en definir un estimador lineal para y que satisfaga [6.2] (restricción longitudinal) y que sea al mismo tiempo compatible con [6.1] (modelo de comportamiento). Este estimador lineal tiene la forma:

$$[6.4] \quad \hat{y} = AY$$

La matriz A que determina el estimador se obtiene a partir de dos condiciones: insesgadez y varianza mínima, esto es, se trata de un estimador lineal, insesgado y óptimo¹ (ELIO). A continuación se presentan las ecuaciones principales del método de Chow-Lin. La derivación detallada puede seguirse en di Fonzo (1987) y Quilis (2001), entre otros.

La siguiente expresión indica que la estimación del vector de parámetros se obtiene aplicando el método de los mínimos cuadrados generalizados al modelo de baja frecuencia que relaciona el agregado Y con el indicador X .

$$[6.5] \quad \hat{\beta}_{CL} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$$

Finalmente, la expresión para el estimador ELIO de y es:

$$[6.6] \quad \hat{y} = x\hat{\beta}_{CL} + vC'V^{-1}(Y - Cx\hat{\beta}_{CL}) = x\hat{\beta}_{CL} + L(Y - Cx\hat{\beta}_{CL})$$

¹ En términos de error cuadrático medio mínimo.

La ecuación anterior determina dicho estimador como la agregación de un término relacionado linealmente con el indicador y de un residuo de baja frecuencia temporalmente desagregado. La principal característica del estimador radica en la dependencia del filtro de desagregación temporal L de la forma del modelo de alta frecuencia y, en particular, de la estructura dinámica de sus perturbaciones. Como se verá más adelante, esta es una de las razones de la flexibilidad y generalidad del método de Chow-Lin pero también una de sus limitaciones computacionales.

Los correspondientes intervalos de confianza para las estimaciones de alta frecuencia pueden obtenerse a partir de la correspondiente matriz de varianzas y covarianzas:

$$[6.7] \quad \Sigma_y = (I_n - vC'V^{-1}C)v + (x - LCx)\Sigma_\beta(x - LCx)'$$

Esta última ecuación establece que la incertidumbre que acompaña a las estimaciones de alta frecuencia está asociada con dos fuentes: una vinculada con la variabilidad de las perturbaciones estocásticas de alta frecuencia u y otra asociada con la imprecisión en que se incurre al estimar b .

Las expresiones [6.5] a [6.7], que definen el método de Chow-Lin, requieren para su aplicación el conocimiento de la matriz de varianzas y covarianzas v de las perturbaciones de alta frecuencia u . Se han propuesto diversas especificaciones para dichas perturbaciones. Posiblemente la más habitual consiste en asumir que u obedece a un proceso AR(1) estacionario:

$$[6.8] \quad u_t = \rho u_{t-1} + a_t \quad |\rho| < 1 \quad \forall t$$

siendo a_t un ruido blanco gaussiano con varianza σ_a^2 . Para aplicar el procedimiento de Chow-Lin teniendo en cuenta las dos expresiones anteriores hay que estimar previamente el parámetro ρ . El tratamiento más general es el propuesto por Bournay y Laroque (1979). Estos autores proponen determinar una malla de valores para ρ , con $-1 < \rho < 1$, y determinar los valores de β y σ_a^2 que maximizan la función de verosimilitud del modelo de baja frecuencia, logarítmicamente transformada:

$$[6.9] \quad \ell(\beta, \sigma_a^2 | \bar{\rho}) = -\frac{N}{2} \ln(2\pi\sigma_a^2) - \frac{1}{2} \ln(|Cv(\bar{\rho})C'|) - \frac{1}{2\sigma_a^2} (Y - X\beta)'(Cv(\bar{\rho})C')^{-1}(Y - X\beta)$$

En consecuencia, el enfoque metodológico de los métodos de Guerrero y Chow-Lin es muy similar: derivación de un estimador ELIO asumiendo un determinado modelo para la serie de alta frecuencia. La diferencia radica, precisamente, en el modelo asumido: un modelo ARIMA para el primero y, para el segundo, una relación lineal estática aditivamente perturbada, siendo la perturbación del tipo AR(1) estacionario.

Guerrero (1990) señala que las condiciones de equivalencia de ambos métodos son:

$$[6.10] \quad w = x\hat{\beta}_{CL}$$

y

$$[6.11] \quad \sigma^2 T P T' = v$$

De esta manera, se aprecia cómo el método de Chow-Lin es un caso particular del de Guerrero. Otra importante diferencia entre ambos métodos concierne a la matriz de varianzas y covarianzas ya que, en el caso de Guerrero, se deriva de forma condicionada a w . De esta manera, la expresión del método de Chow-Lin equivalente a la de Guerrero es:

$$[6.12] \quad \Sigma_1 = (I_n - v C' V^{-1} C) v$$

Esta expresión ignora la incertidumbre vinculada con la estimación de los parámetros b y sólo es estrictamente comparable con la matriz de varianzas y covarianzas de Guerrero bajo las condiciones [6.10] y [6.11].

La comparación entre estos procedimientos está más allá del objetivo de este trabajo. No obstante, hay que señalar que ambos métodos efectúan hipótesis fuertes ya que han de resolver un problema de identificación básico: Guerrero asume una equivalencia entre los modelos para y y para w mientras que Chow-Lin supone un modelo estático que liga y con x perturbado por un esquema AR(1) estacionario.

¿Cuál de las dos hipótesis es más verosímil? La respuesta debe tener en cuenta una diferencia de enfoque fundamental entre ambos procedimientos: Guerrero centra la estrategia de estimación en los indicadores mientras que Chow-Lin-Bournay-Laroque sitúan en primer lugar la estimación de un modelo de baja frecuencia por máxima verosimilitud para determinar a continuación los parámetros críticos del procedimiento, definido por las ecuaciones [6.6] y [6.7].

No obstante, el método de Guerrero es más afín a la práctica moderna del análisis de series temporales que el de Chow-Lin, debido al uso explícito de la modelización ARIMA y a que no se emplea una especificación esencialmente estática de la relación entre indicadores x y serie inobservable y . Nótese que, habitualmente, la modelización econométrica de datos de alta frecuencia utiliza modelos explícitamente dinámicos (p.e., funciones de transferencia, vectores de autorregresiones, etc.) y rara vez representaciones estáticas, véase Santos Silva y Cardoso (2001).

Además, el método de Guerrero permite integrar de forma natural la información que se requiere para la extracción de señales con la desagregación temporal, ya que el mismo modelo ARIMA que se usa para la primera puede usarse para la segunda. De esta manera, como ya se ha señalado, este procedimiento constituye una vía sencilla para combinar los aspectos de desagregación temporal y extracción de señales en la práctica del análisis de series temporales, en general, y en la compilación de la Contabilidad Trimestral, en particular.

7. Función MATLAB

Para poder aplicar el método de Guerrero se ha desarrollado una serie de funciones Matlab que forman parte de la librería de desagregación temporal expuesta en Quilis (2003). Su utilización requiere la determinación previa de los modelos ARIMA del indicador escalado w y de la discrepancia d . En la actualidad se está trabajando en la automatización de estas funciones, de manera que se facilite su integración con los procedimientos de modelización ARIMA y de extracción de señales contenidos en los programas TRAMO y SEATS (Gómez y Maravall, 1996).

Computacionalmente, se han empleado algunas funciones Matlab contenidas en dos librerías especializadas: *Signal Processing Toolbox* y *Econometrics Toolbox*, véase Mathworks (1998) y LeSage (1999), respectivamente. De esta manera, se consigue una estimación muy eficiente en modo de producción, esto es, integrada en un esquema general de compilación de datos estadísticos. La transición del modo de investigación al de producción suele ser turbulenta. El uso de *software* de muy alto nivel como Matlab, en el sentido de su abstracción procedural, minimiza los riesgos de la transición, véase Gatheral et al. (1999) para un análisis detallado de estas cuestiones.

A continuación se expone la función principal que permite la aplicación del procedimiento de desagregación temporal expuesto en las secciones anteriores. Asimismo, esta función también permite realizar extrapolaciones. Para ello basta con considerar como *input* un vector de indicadores de dimensión $sN+e$, siendo e el número de extrapolaciones.

La función posee la siguiente estructura:

```
function res=guerrero(Y, x, ta, s, rexw, rexd);
% PURPOSE: ARIMA-based temporal disaggregation: Guerrero method
% -----
% SYNTAX: res=guerrero(Y, x, ta, s, rexw, rexd);
% -----
% OUTPUT: res: a structure
%   res.meth      = 'Guerrero';
%   res.ta        = type of disaggregation
%   res.N         = nobs. of low frequency data
%   res.n         = nobs. of high-frequency data
%   res.pred      = number of extrapolations
%   res.s         = frequency conversion between low and high freq.
%   res.p         = number of regressors (+ intercept)
%   res.Y         = low frequency data
%   res.x         = high frequency indicators
%   res.w         = scaled indicator (preliminary hf estimate)
%   res.y1        = first stage high frequency estimate
%   res.y         = final high frequency estimate
%   res.y_dt      = high frequency estimate: standard deviation
%   res.y_lo      = high frequency estimate: sd - sigma
%   res.y_up      = high frequency estimate: sd + sigma
%   res.delta     = high frequency discrepancy (y1-w)
%   res.u         = high frequency residuals (y-w)
%   res.U         = low frequency residuals (Cu)
%   res.beta      = estimated parameters for scaling x
%   res.k         = statistic to test compatibility
%   res.et        = elapsed time
% -----
% INPUT: Y: Nx1 ---> vector of low frequency data
%   x: nxp ---> matrix of high frequency indicators (without intercept)
%   ta: type of disaggregation
%       ta=1 ---> sum (flow)
%       ta=2 ---> average (index)
%       ta=3 ---> last element (stock) ---> interpolation
%   s: number of high freq. data points for each low freq. data points
%       s= 4 ---> annual to quarterly
%       s=12 ---> annual to monthly
%       s= 3 ---> quarterly to monthly
%   rexw, rexd ---> a structure containing the parameters of ARIMA model
%                   for indicator and discrepancy, respectively (see calT function)
% -----
% LIBRARY: aggreg, calT, numpar, ols
% -----
% SEE ALSO: chowlin, litterman, fernandez, td_print, td_plot
% -----
% REFERENCE: Guerrero, V. (1990) "Temporal disaggregation of time
% series: an ARIMA-based approach", International Statistical
% Review, vol. 58, p. 29-46.
```

8. Aplicación

La función expuesta en la sección anterior ha sido aplicada a los datos empleados por Guerrero (1990) para ilustrar su método. Se trata de la serie anual del Producto Interior Bruto (PIB) de México, valorado a pesos constantes de 1980 y del Índice de la Producción Industrial (IPI), base 1980=100, del mismo país. El intervalo temporal abarca desde 1983 hasta 1987. El modelo ARMA para la discrepancia d se ha tomado de Guerrero y Martínez (1995). El modelo ARIMA del indicador escalado w es:

$$[8.1] \quad w_t = \frac{(1-0.40B)(1-0.79B^{12})}{(1-B)(1-0.85B^{12})} a_{w,t} \quad \sigma_w = 4968.716$$

y el de la discrepancia d :

$$[8.2] \quad \delta_t = \frac{1}{(1-0.43B)(1-B)^2(1+0.62B^{12})} a_{\delta,t} \quad \sigma_\delta = 76.95$$

La aplicación se realiza mediante un guión (*script*) de Matlab como el siguiente:

```
Y=load(' PIB. anu' );
x=load(' IPI. men' );
% Type of aggregation
ta=1;
% Frequency conversion
s=12;
% Model for w: (0, 1, 1) (1, 0, 1)
rexw.ar_reg = [1];
rexw.d = 1;
rexw.ma_reg = [1 -0.40];
rexw.ar_sea = [1 0 0 0 0 0 0 0 0 0 0 0 -0.85];
rexw.bd = 0;
rexw.ma_sea = [1 0 0 0 0 0 0 0 0 0 0 0 -0.79];
rexw.sigma = 4968.716^2;
% Model for the discrepancy: (1, 2, 0) (1, 0, 0)
rexd.ar_reg = [1 -0.43];
rexd.d = 2;
rexd.ma_reg = [1];
rexd.ar_sea = [1 0 0 0 0 0 0 0 0 0 0 0 0.62];
rexd.bd = 0;
rexd.ma_sea = [1];
rexd.sigma = 76.95^2;
% Calling the function: output is loaded in a structure called res
res=guerrero(Y, x, ta, s, rexw, rexd);
% Calling printing function
% Name of ASCII file for output
file_sal='guerrero.sal';
output=0; % Do not include series
td_print(res, file_sal, output);
edit guerrero.sal;
% Calling graph function
td_plot(res);
```

La salida impresa es la siguiente:

```
*****
TEMPORAL DIAGGREGATION METHOD: Guerrero
*****
-----
Number of low-frequency observations :    5
Frequency conversion                  :   12
Number of high-frequency observations:   60
Number of extrapolations              :    0
Number of indicators (+ constant)     :    2
-----
Type of disaggregation: sum (flow).
-----
Estimation method: BLUE.
-----
Beta parameters (columnwise):
* Estimate
* Std. deviation
* t-ratios
-----
219988.6766      974531.6756      4.4299
 1723.8723      6174.6540      3.5819
-----
AIC:    7.5245
BIC:    7.3683
-----
Low-frequency correlation (Y, X)
- levels      : 0.9003
- yoy rates   : 0.9973
-----
High-frequency correlation (y, x)
- levels      : 0.9289
- yoy rates   : 0.9835
-----
High-frequency volatility of yoy rates
- estimate    : 3.6623
- indicator   : 6.2899
- ratio       : 0.5823
-----
High-frequency correlation (y, x*beta)
- levels      : 0.9289
- yoy rates   : 0.9832
-----
Compatibility test:
- k :    0.9526
```

ARIMA model for scaled indicator:

(0 1 1) (1 0 1)

- Regular AR operator:

1.0000

- Regular MA operator:

1.0000 -0.4000

- Seasonal AR operator:

1.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 -0.8500

- Seasonal MA operator:

1.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 -0.7900

ARIMA model for discrepancy :

(1 2 0) (1 0 0)

- Regular AR operator:

1.0000 -0.4300

- Regular MA operator:

1.0000

- Seasonal AR operator:

1.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.6200

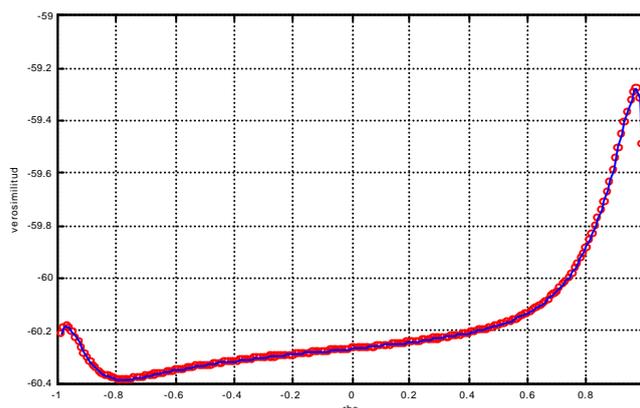
- Seasonal MA operator:

1.0000

Elapsed time: 0.2800

Como ya se ha señalado, la aplicación del método de Chow-Lin requiere la determinación previa del valor del parámetro r que controla la dinámica de la innovación de alta frecuencia. Como se aprecia en el gráfico siguiente, la maximización de la función de verosimilitud [6.9] conduce a un valor de 0.97 para dicho parámetro.

**Gráfico 1: Método de Chow-Lin.
Función de verosimilitud implicada**



Una vez seleccionado el valor de r , se realiza la estimación puntual según Chow-Lin, ecuación [6.6], y Guerrero, ecuación [5.6]. Como se aprecia en la tabla 1, ambos métodos producen una estimación bastante conforme con la información de alta frecuencia contenida en el indicador x . Asimismo, debe señalarse que los resultados de Chow-Lin son bastante robustos frente a r . Así, en la tabla 1 se muestra la similitud de los resultados para dos casos extremos: $\rho=0$ (u se asume ruido blanco) y $\rho=1$ (u se considera un paseo aleatorio)¹.

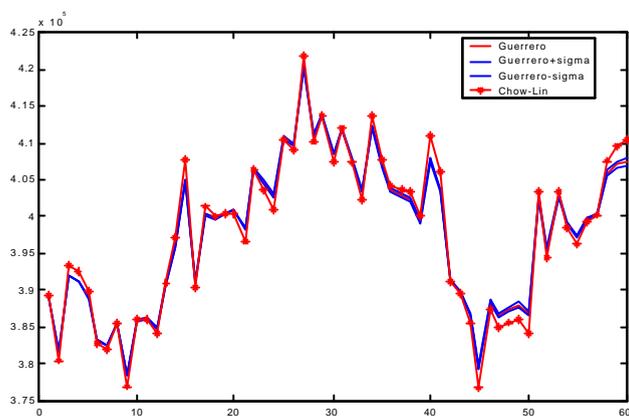
**Tabla 1: Métodos de Guerrero y Chow-Lin.
Relación con el indicador**

Método	Correlación		Ratio de varianzas
	Niveles	Tasas interanuales	
Guerrero	0.93	0.98	0.58
Chow Lin	$\rho = 0.97$	0.93	0.99
	$\rho = 0.00$	0.93	0.98
	$\rho = 1.00$	0.93	0.99

En el gráfico 2 se compara la estimación puntual de Chow-Lin con la efectuada mediante el método de Guerrero. Aún cuando la semejanza es notable, con una correlación de 0.99 tanto en niveles como en tasas interanuales, no son equivalentes, ya que la estimación puntual de Chow-Lin se sitúa sistemáticamente fuera del intervalo $\hat{y} \pm \sigma$ de la de Guerrero.

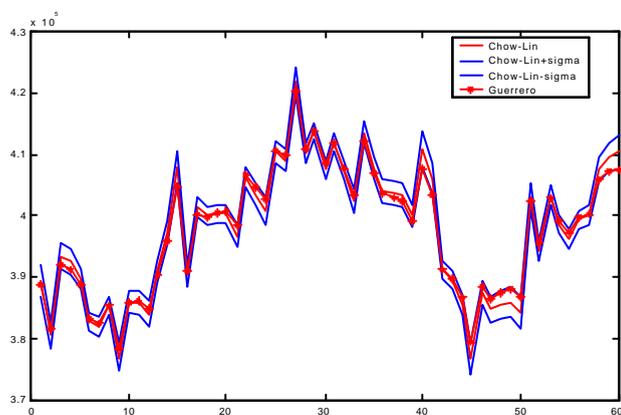
¹ Este caso límite equivale al método de Fernández (1981).

**Gráfico 2: Método de Guerrero.
Comparación con Chow-Lin**



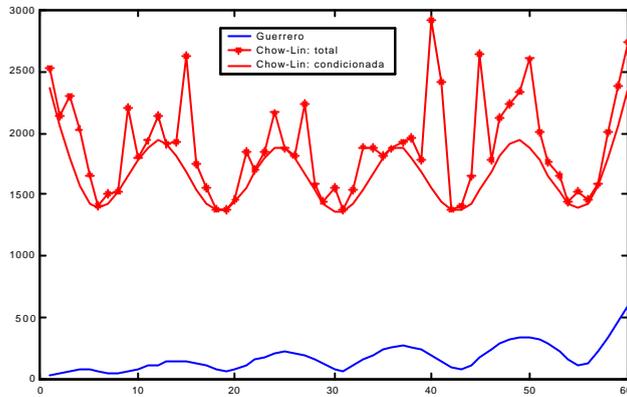
Si se realiza la comparación inversa, la estimación de Guerrero sí que parece más compatible con la de Chow-Lin, como se aprecia en el siguiente gráfico:

**Gráfico 3: Método de Chow-Lin.
Comparación con Guerrero**



Esta asimetría se debe a la notable diferencia que existe entre las varianzas generadas por ambos métodos. Ya se ha señalado que la comparación resulta delicada debido a las diferencias metodológicas básicas entre ambos procedimientos. No obstante, desde un punto de vista aplicado, resulta ilustrativa la comparación. En el gráfico 4 se presentan las series de desviaciones típicas de los dos procedimientos detallando, en el caso de Chow-Lin, la estimación total (ecuación [5.8]) y el componente condicionado (ecuación [6.12]). La diferencia de precisión entre ambos métodos es sustancial.

**Gráfico 4: Métodos de Guerrero y Chow-Lin.
Incertidumbre de las estimaciones puntuales**

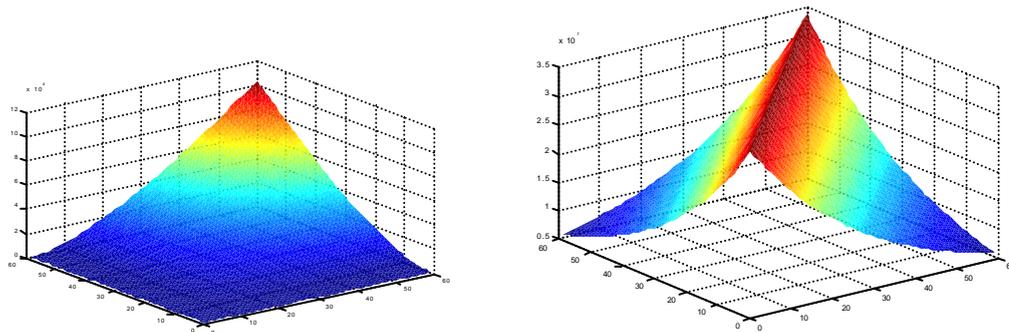


Por otra parte, una de las condiciones de equivalencia entre ambos métodos viene dada por la condición:

$$[6.11] \quad \sigma^2 TPT' = v$$

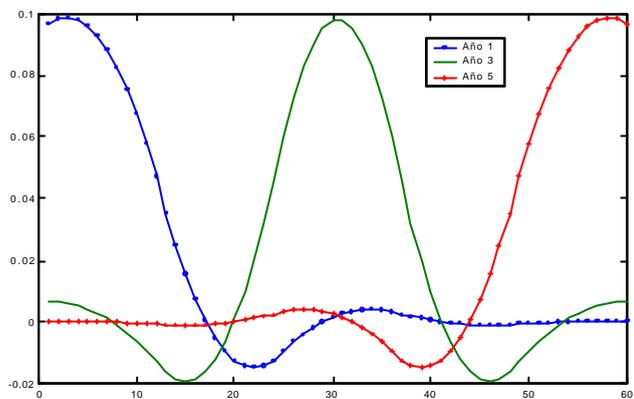
Como se observa en el gráfico 5, la diferencia entre ambas matrices es notable, tanto en lo que se refiere a la forma como a los valores numéricos (10^4 frente a 10^7).

**Gráfico 5: Métodos de Guerrero y Chow-Lin.
Matrices TPT' y $S^{-2}v$**



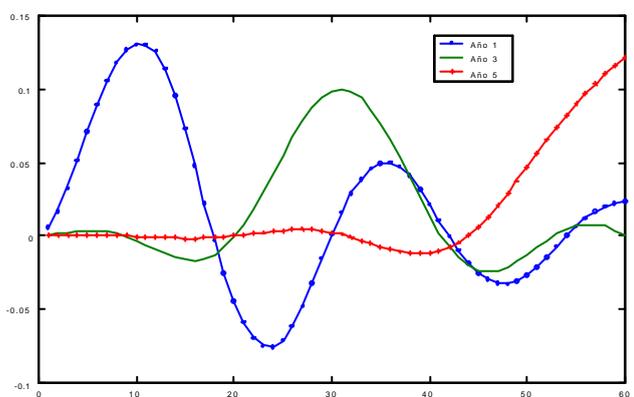
Otra diferencia importante entre ambos procedimientos reside en las distintas matrices de desagregación temporal de las discrepancias o residuos de baja frecuencia. En el gráfico 6 se muestran las columnas 1, 3 y 5 de la matriz L del método de Chow-Lin, de forma que se representan las contribuciones de los años primero, tercero y quinto a la generación de los $n=60$ valores mensuales. Dichas contribuciones poseen una forma simétrica entre ellas tomando $t=30$ como eje.

Gráfico 6: Método de Chow-Lin: matriz L
Dependencia de los meses respecto a los años



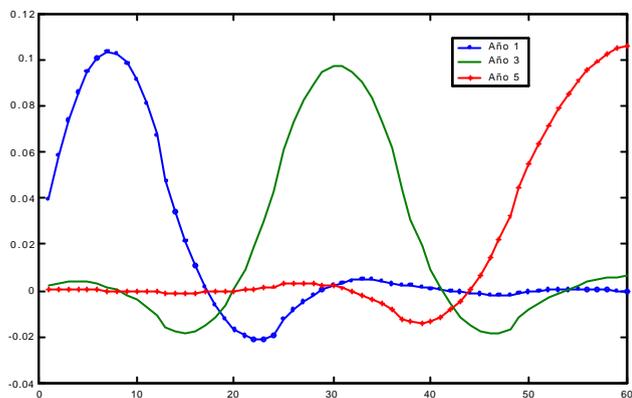
Esta simetría no se observa en el caso de la matriz A del método de Guerrero, como se aprecia en el gráfico siguiente:

Gráfico 7: Método de Guerrero: matriz A
Dependencia de los meses respecto a los años



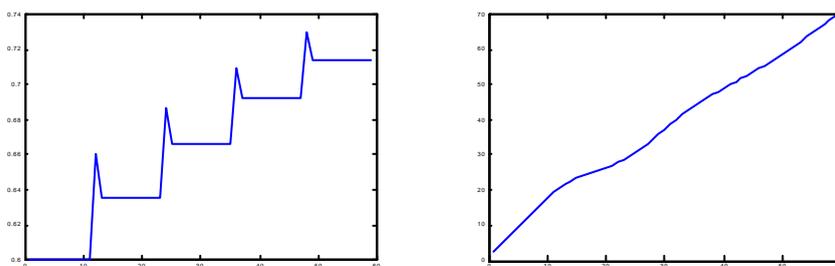
Este resultado es debido, esencialmente, al importante papel del modelo de la discrepancia \mathbf{d} y de la correspondiente matriz \mathbf{W} en la estimación de y , ya que la matriz A_1 asociada a la estimación preliminar sí es más simétrica:

**Gráfico 6: Método de Guerrero: matriz A_1
Dependencia de los meses respecto a los años**



Los modelos ARIMA empleados para caracterizar al indicador escalado w y a la discrepancia d , $(011)(101)_{12}$ y $(120)(100)_{12}$, respectivamente, son radicalmente distintos tanto cualitativa como cuantitativamente, como puede apreciarse en sus respectivas funciones de respuesta al impulso:

**Gráfico 8: Funciones de respuesta al impulso
Matrices T y W**



Como estas funciones determinan la forma de las matrices T y W , se explica fácilmente la diferencia entre las matrices A_1 y A . La selección del modelo ARIMA para d aparece, en consecuencia, como un elemento importante del método de Guerrero. La modelización ARIMA de d es una tarea complicada debido a su naturaleza artificial, derivada de la desagregación temporal de una discrepancia anual preliminar (recuérdese el punto 3 del algoritmo), por lo que es uno de los aspectos que deberá examinarse con detalle en las aplicaciones prácticas del procedimiento.

Apéndice A: Derivación de los estimadores

A continuación se derivan de forma detallada los estimadores propuestos, tanto de tipo teórico como factible.

Se propone un estimador lineal de y de la forma general:

$$[A.1] \quad \hat{y} = AY$$

Se desea determinar A de manera que el estimador presentado en [A.1] sea insesgado y de varianza mínima.

Insensgadez: $E_0(\hat{y} - y) = 0$

Teniendo en cuenta que $y = E_0(y) + Ta$ y que $Y = Cy$ se obtiene:

$$[A.2] \quad \hat{y} = AY = ACE_0(y) + ACTa$$

En consecuencia, $E_0(\hat{y} - y) = 0$ implica:

$$[A.3] \quad (AC - I_n)E_0(y) = 0$$

La expresión para el estimador es:

$$[A.4] \quad \hat{y} - E_0(y) = ACTa$$

Sumando y restando y en la anterior ecuación:

$$[A.5] \quad \hat{y} - y = (AC - I_n)Ta$$

Varianza mínima

La expresión de la matriz de varianzas y covarianzas de \hat{y} es:

$$[A.6] \quad \Sigma_{\hat{y}} = E[(\hat{y} - y)(\hat{y} - y)'] = \sigma^2(I_n - AC)TT'(I_n - AC)'$$

La determinación de A se obtiene resolviendo el siguiente programa de minimización:

$$[A.7] \quad \text{MIN}_A \quad \Sigma_{\hat{y}} \quad \Rightarrow \quad \frac{\partial \Sigma_{\hat{y}}}{\partial A} = 0$$

Expandiendo la matriz varianzas y covarianzas se tiene:

$$[A.8] \quad \frac{\partial}{\partial A} (TT' - ACTT' - TT'C'A' + ACTT'C'A') = 0$$

Tras algunas manipulaciones algebraicas se obtiene la expresión óptima para A :

$$[A.9] \quad \hat{A} = (TT' C')(CTT' C')^{-1}$$

Premultiplicando $y - E_0(y) = Ta$ por $\hat{A}C$ y teniendo en cuenta [A.4] se obtiene la expresión final del estimador:

$$[A.10] \quad \hat{y} = E_0(y) + \hat{A}(Y - CE_0(y))$$

Sustituyendo [A.9] en la versión expandida de [A.6] se deriva la expresión de la matriz de varianzas y covarianzas de \hat{y} :

$$[A.11] \quad \Sigma_{\hat{y}} = \sigma^2 (I_n - \hat{A}C)TT'$$

Asumiendo que w es observable y que posee una representación ARIMA idéntica a la de y , excepto por un factor de escala, se puede condicionar el error de predicción $y - E_0(y)$ a w para obtener:

$$[A.12] \quad w - E_0(y) = T E(a | w)$$

Naturalmente, el elemento crítico es $E(y/w) = w$ que resulta factible gracias a la hipótesis referida a la equivalencia entre la estructura dinámica de y y la de w . Sustrayendo [A.12] de $y - E_0(y)$ se obtiene:

$$[A.13] \quad y - w = T(a - E(a | w)) = Ta^*$$

De esta forma, la distribución de a^* condicionada a w es $N(0, \sigma^2 P)$, siendo P una matriz definida positiva. Nótese que, si W es la matriz que contiene la expansión de medias móviles de $d=y-w$ se verifica:

$$[A.14] \quad \Omega\Omega' = TPT'$$

de donde se deduce:

$$[A.15] \quad P = T^{-1}\Omega\Omega'T^{-1}$$

Las expresiones correspondientes al estimador factible se derivan siguiendo el mismo razonamiento que en el caso teórico, teniendo en cuenta que $y - w = Ta^*$ juega el mismo papel que $y - E_0(y) = Ta$. En consecuencia, las citadas expresiones son:

$$[A.16] \quad \hat{y} = w + A[Y - Cw]$$

siendo:

$$[A.17] \quad A = TPT'C'(CTPT'C')^{-1}$$

La correspondiente matriz de varianzas y covarianzas es:

$$[A.18] \quad \Sigma_y = \sigma^2 (I_n - AC)TPT'$$

Referencias

Bournay, J. y Laroque, G. (1979) "Réflexions sur la méthode d'elaboration des comptes trimestriels", *Annales de l'INSEE*, n. 36, p. 3-30.

Chow, G. y Lin, A.L. (1971) "Best linear unbiased distribution and extrapolation of economic time series by related series", *Review of Economic and Statistics*, vol. 53, n. 4, p. 372-375.

di Fonzo, T. (1987) *La stima indiretta di serie economiche trimestrali*, Cleup Editore, Padua, Italia.

Fernández, R.B. (1981) "Methodological note on the estimation of time series", *Review of Economic and Statistics*, vol. 63, n. 3, p. 471-478.

Gatheral, J., Epelbaum, Y., Han, J., Laud, K., Lubovitsky, O., Kant. E. y Randall, C. (1999) "Implementing option-pricing models using software synthesis", *IEEE Computing in Science and Engineering*, november/december, p. 54-64.

Gómez, V. y Maravall, A. (1996) "Programs TRAMO and SEATS", Documento de Trabajo n. 9628, Banco de España <<http://www.bde.es>>.

Guerrero, V. (1990) "Temporal disaggregation of time series: an ARIMA-based approach", *International Statistical Review*, volumen 58, no. 1, pag. 29-46.

Guerrero, V. y Martínez, J. (1995) "A recursive ARIMA-based procedure for disaggregating a time series variable using concurrent data", *Test*, vol. 4, n. 2, p. 359-376.

Guerrero, V. y Nieto, F.H. (1999) "Temporal and contemporaneous disaggregation of multiple time series", *Test*, vol. 8, n. 2, p. 459-489.

Guerrero, V. y Peña, D. (2000) "Linear combination of restrictions and forecasts in time series analysis", *Journal of Forecasting*, vol. 19, p. 103-122.

INE (1993) *Contabilidad Nacional Trimestral de España. Metodología y serie trimestral 1970-1992*, Instituto Nacional de Estadística, Madrid, España.

INE (2002) "Ajuste estacional y extracción de señales en la Contabilidad Nacional Trimestral", Instituto Nacional de Estadística, *Boletín Trimestral de Coyuntura*, n. 84, p. 129-151 <<http://www.ine.es>>.

INE (2002) "Ajuste estacional y extracción de señales en la Contabilidad Nacional Trimestral", Banco de España, Documento de Trabajo n. 0210 <<http://www.bde.es>>.

LeSage, J.P. (1999) "Applied econometrics using MATLAB", University of Toledo (Ohio), Department of Economics, Documento Interno <<http://www.econ.utoledo.edu/~lesage>>.

MathWorks (1998) *Signal processing toolbox*, The MathWorks Co., Natick, Massachusetts, United States of America.

Quilis, E.M. (2001) "Notas sobre desagregación temporal de series económicas", Instituto de Estudios Fiscales, Papeles de Trabajo n. 1/01 <<http://www.minhac.es/ief>>.

Quilis, E.M. (2003) "A Matlab library of temporal disaggregation methods", Instituto Nacional de Estadística, Documento Interno <<http://www.ine.es>>.

Santos Silva, J.M.C. y Cardoso, F. (2001) "The Chow-Lin method using dynamic models", *Economic Modelling*, vol. 18, p. 269-280.