

La trimestralización de series temporales generadas bajo un modelo de líneas aéreas trimestral.

Alejandro Rodríguez Caro (alek@empresariales.ulpgc.es)

Delia Dávila Quintana (delia@empresariales.ulpgc.es)

Santiago Rodríguez Feijoo (santi@empresariales.ulpgc.es)

Departamento de Métodos Cuantitativos

Universidad de Las Palmas de G.C.

Resumen: Presentamos un estudio de simulación para estudiar las propiedades de los principales métodos de desagregación temporal derivados del propuesto por Chow y Lin (1971). El ejercicio se desarrolla sobre series trimestrales generadas bajo el modelo más utilizado para describir series económicas como es el modelo de líneas aéreas.

Palabras clave: Desagregación, modelo líneas aéreas, simulación

1.- Introducción

Los métodos de desagregación temporal se han revelado como una solución válida ante la ausencia de información temporal de alta frecuencia. Estos son utilizados por muchos institutos nacionales de estadística para la elaboración de las cuentas nacionales trimestrales. Dentro del amplio conjunto de métodos existentes, los más recurridos en la actualidad son aquellos que utilizan información procedente de variables auxiliares, y dentro de este grupo, el propuesto por Chow y Lin (1971). Sin embargo, este método presenta un problema para su implementación práctica derivado de la utilización de la matriz de covarianzas de la perturbación aleatoria trimestral. Diversos autores han planteado diversas soluciones alternativas, sin quedar clara cual presenta mejores resultados. El presente trabajo aporta evidencia útil para la toma de decisiones sobre que alternativa escoger dado el caso, utilizando para ello un ejercicio de simulación con series trimestrales generadas bajo el modelo más seguido por series económicas, el modelo de líneas aéreas. Este trabajo comienza con una revisión de los métodos más frecuentemente utilizados. En el tercer punto se comenta el ejercicio de simulación para, a continuación, presentar los resultados obtenidos, quedando el último punto para las conclusiones más relevantes.

2.- Los Métodos de Desagregación

Los métodos de desagregación temporal de series temporales se pueden clasificar de forma general en función de la utilización o no de información adicional a la propia serie a desagregar. Aquellos métodos que sí utilizan información procedente de variables auxiliares se denominan métodos de desagregación con indicador. En el desarrollo histórico de este conjunto de métodos tiene una especial relevancia la aportación del método de Chow y Lin en 1971. Este método, que desde el punto de vista teórico es óptimo, presenta limitaciones a la hora de llevarlo a la práctica, ya que se desconoce la matriz de covarianzas de la perturbación aleatoria trimestral, necesaria para su implementación. Este problema ha llevado a diversos autores a proponer distintas soluciones para la estimación de dicha matriz de covarianzas dando lugar a distintos métodos. Comenzamos la exposición de los más importantes por el método de Chow y Lin (1971) para posteriormente ver los métodos desarrollados en torno a esta primera propuesta.

Chow y Lin (1971) desarrollan la forma de obtener el mejor estimador lineal e insesgado de la serie trimestral, mediante la estimación de la relación anual entre los indicadores y la variable objetivo, esto es:

$$Y = B' y = B' x\beta + B'u \quad (1)$$

donde Y es el vector de datos anuales conocidos; y el vector de datos trimestrales desconocidos; x la matriz de datos trimestrales de las variables indicador, y B es la matriz de paso de datos trimestral a anuales, mediante la agregación de los primeros, y que queda definida como sigue en el caso de estudio de variables flujo

$$B' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & . & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & . & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & . & 1 & 1 & 1 & 1 \end{bmatrix} \quad (2)$$

Los autores obtienen la siguiente expresión de trimestralización (3) de los datos anuales, así como la expresión de estimación de los parámetros (4) de la relación anual (1).

$$\hat{y} = x\hat{\beta} + VB(B'VB)^{-1}(Y - B'x\hat{\beta}) \quad (3)$$

$$\hat{\beta} = [x'B(B'VB)^{-1}B'x]^{-1}x'B(B'VB)^{-1}Y \quad (4)$$

en donde V representa a la matriz de covarianzas de la perturbación trimestral desconocida. Se han propuesto distintas formas de esta matriz V, perteneciendo las dos primeras a los propios autores.

La primera solución es la trivial (CL_TRIVIAL) la cuál consiste en suponer que los errores trimestrales tienen un comportamiento ruido blanco, de tal forma que, la estimación a realizar en la expresión (4) corresponde con la de mínimos cuadrados ordinarios (V=I). Esto implica repartir los errores anuales estimados por la regresión en partes iguales entre los correspondientes trimestres. Este reparto puede provocar la presencia de saltos espurios entre el cuarto trimestre de un año y el primero del siguiente.

La segunda solución propuesta por Chow y Lin (1971) consiste en suponer que los errores trimestrales se distribuyen según un proceso autorregresivo de orden uno (CL_AR1). Esta solución pretende evitar la presencia de posibles saltos espurios en la serie trimestral entre el final de un año y el comienzo del siguiente. Para ello se estima la regresión anual a partir de la cual se obtiene un estimador (ϕ_a) del coeficiente autorregresivo de orden uno que se utiliza a su vez para estimar el parámetro autorregresivo trimestral (ϕ) mediante la relación siguiente:

$$\phi_a = \frac{\phi(\phi + 1)(\phi^2 + 1)^2}{2(\phi^2 + \phi + 2)} \quad (5)$$

Sin embargo, el uso de esta solución presenta nuevos problemas debido a que la relación (5) supone la existencia de dos valores trimestrales por cada valor anual, en el caso en que el valor anual sea menor que cero, y no poder estimar valores trimestrales cuando el coeficiente anual sea inferior a -0.1305.

Fernández (1981) trata de añadir una mayor suavidad a la serie trimestral finalmente estimada y propone que la perturbación trimestral siga un camino aleatorio, lo cuál implica que el modelo trimestral propuesto se ajuste a (6)

$$\begin{aligned} y_j &= x_j \beta + u_j \\ u_j &= u_{j-1} + \varepsilon_j \end{aligned} \quad (6)$$

El autor demuestra que la solución coincide con la del método propuesto por Denton (1971) y se expresa como:

$$\begin{aligned} \hat{y} &= x \hat{\beta} + (D' D)^{-1} B (B' (D' D)^{-1} B)^{-1} [Y - B' x \hat{\beta}] \\ \hat{\beta} &= \left[x' B (B' (D' D)^{-1} B)^{-1} B' x \right]^{-1} x' B (B' (D' D)^{-1} B)^{-1} Y \end{aligned} \quad (7)$$

donde D es la matriz de diferenciación siguiente:

$$D = \begin{bmatrix} 1 & 0 & 0 & . & 0 & 0 & 0 \\ -1 & 1 & 0 & . & 0 & 0 & 0 \\ 0 & -1 & 1 & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & -1 & .1 & .0 \\ 0 & 0 & 0 & . & .0 & -1 & 1 \end{bmatrix} \quad (8)$$

Basándose en la propuesta de Fernández (1981), Litterman (1983) añade una penalización de alisado sobre la serie trimestral resultante, como posteriormente demostraron Pinheiro y Coimbra (1983), suponiendo para ello el siguiente modelo trimestral:

$$\begin{aligned} y_j &= x1_j \beta_1 + x2_j \beta_2 + \dots + xk_j \beta_k + u_j \\ u_j &= u_{j-1} + \varepsilon_j \\ \varepsilon_j &= \alpha \varepsilon_{j-1} + e_j \end{aligned} \quad (9)$$

En este caso se añade la dificultad de tener que estimar el valor de α . Para ello se utiliza la expresión (10) en donde q es el coeficiente de correlación de primer orden de la diferencia de los residuos anuales.

$$q = \frac{\alpha^9 + 3\alpha^8 + 7\alpha^7 + 13\alpha^6 + 18\alpha^5 + 22\alpha^4 + 22\alpha^3 + 18\alpha^2 + 14\alpha + 10}{2\alpha^5 + 6\alpha^4 + 14\alpha^3 + 26\alpha^2 + 36\alpha + 44} \quad (10)$$

De igual forma que en (5) la expresión (10) genera nuevas dificultades. En este caso se concreta en el hecho de que no existen valores trimestrales por debajo de valores anuales de q de 0.1666667.

Stram y Wei (1986) y (1990) (SW) desarrollan un método de trimestralización sin indicador, basándose en la obtención del modelo SARIMA trimestral congruente con el correspondiente modelo SARIMA de la serie anual. Proponen solucionar el problema de la matriz V de la solución de Chow y Lin (1971) mediante la aplicación de su método de trimestralización sin indicador sobre la serie de error anual de la relación (1). De esta forma la serie trimestral se obtendría como el producto de los parámetros estimados por los indicadores trimestrales más la serie de error trimestralizada

Pinheiro y Coimbra (1993) (PC) no abordan directamente el problema de la estimación de la matriz de covarianzas de la perturbación trimestral, sino que proponen la estimación de una relación trimestral alternativa, utilizando una transformación especial en logaritmos neperianos para evitar el problema de la no aditividad de los mismos.

Como se puede observar, tenemos dos conjuntos de propuestas para solucionar el problema práctico del método de Chow y Lin (1971). Por un lado, están aquellos métodos que proponen fijar a priori la estructura de los errores, buscando un modelo que permita obtener una serie sin cambios bruscos ni saltos espurios entre los distintos años. Por otro lado, el método de Stram y Wei, bajo ciertas condiciones, propone la estimación de la matriz de covarianzas trimestral que sea compatible con la correspondiente anual.

3.- El Proceso de Simulación implementado

Presentamos brevemente las características más importantes del proceso de simulación. Para ello nos basamos en la relación trimestral entre la variable objetivo y la variable indicador representada en (11).

$$y_t = \alpha + \beta x_t + u_t \quad (11)$$

donde y_t es la serie trimestral desconocida objetivo, x_t el indicador trimestral disponible, α y β los parámetros que las relacionan y u_t la perturbación aleatoria trimestral desconocida. Para simular series bajo un modelo de líneas aéreas es necesario definir el valor del parámetro media móvil regular (R) y estacional (S). En el presente trabajo hemos utilizado tres pares de valores correspondientes con ambos parámetros negativos (R=-0.7; S=-0.7), uno positivo y otro negativo (R=0.7; S=-0.7) y ambos parámetros positivos (R=0.7; S=0.7). Otro de los aspectos basados en la relación (1) es la especificación del modelo de la perturbación aleatoria trimestral u_t , aspecto muy importante debido a los supuestos que cada método realiza sobre él. Hemos estudiado tres modelos, ruido blanco (RB), autoregresivo regular positivo ARPO, y autoregresivo regular integrado regularmente de orden uno con parámetro positivo (ARI).

Los parámetros α, β de la relación (11) se han propuesto positivos, mientras que la bondad del ajuste se establece en valores situados entre el 0.90 y 0.95 de coeficiente de determinación, con el fin de estudiar el comportamiento en condiciones óptimas para realizar el proceso, motivo por el cual se ha optado por trimestralizar series de 50 años (200 trimestres).

Para realizar el análisis de los resultados se utilizan en este papel dos estadísticos: de un lado la raíz del error cuadrático medio en porcentaje de media (RECM), como medida del grado de ajuste de la serie trimestral original y la serie trimestralizada por cada uno de los métodos propuestos.

$$RECM(M) = \frac{\sqrt{\frac{1}{4T} \sum_{j=1}^{4T} (y_j^M - y_j^{or})^2}}{\frac{1}{4T} \sum_{j=1}^{4T} y_j^{or}} \quad (12)$$

y de otro, la diferencia de tasas regulares (RT1):

$$RT1(M) = \frac{1}{4T-1} \sum_{j=2}^{4T} |T_1(y_j^{or}) - T_1(y_j^M)| \quad (13)$$

que se basa en calcular las correspondientes tasas de variación regulares de la serie original y de la serie trimestralizada por el método M, esto es:

$$T_1(y_j) = \frac{y_j - y_{j-1}}{y_{j-1}} 100 \quad (14)$$

Este último estadístico nos daría la diferencia media de comportamiento regular que estima la serie trimestralizada con respecto al comportamiento regular de la serie original. Cuanto más cercano a cero esté, más parecida serán las variaciones medias de las serie trimestralizada respecto a las de la serie original

Se han simulado 500 series por cada uno de los 9 escenarios resultantes de todas las posibles combinaciones de las características ya comentadas, esto es, tres modelos ARIMA del error por 3 conjuntos de parámetros del modelo líneas aéreas. Este total de 4500 series se han trimestralizado utilizando cada uno de los métodos de trimestralización mencionados, es decir, la solución trivial (TRIVIAL) de dividir por cuatro la serie anual, la solución trivial de Chow y Lin (1971) (CL_TRIVIAL), utilizar un modelo AR1 para los errores (AR1), Fernández (1981), Litterman (1983), Stram y Wei (SW) (1986 y 1990), y Pinheiro y Coimbra (PC) (1993). Por último, se han comparado las series resultantes con las series originales a través del RECM.

4.- Los Resultados Obtenidos

Los resultados obtenidos tanto globalmente como para cada uno de los nueve escenarios propuestos se presentan en la tabla 1. Los resultados globales, en las últimas filas, muestran la mejora que supone utilizar un método de desagregación, contra la alternativa de utilizar la solución trivial consistente en dividir por cuatro la serie anual. Tanto de forma global, como para cada uno de los modelos de los residuos, el comportamiento del método trivial está, como mínimo, un punto de RMSPE por encima del resto de métodos propuestos.

Si comenzamos a estudiar los resultados por el estadístico RECM, éstos se muestran sensibles a cambios en la estructura de los errores. El primer cambio se encuentra al pasar de un modelo sin estructura, como el ruido blanco que de forma global obtiene un valor de 3.04%, a un modelo que sí la tiene, como el autorregresivo con valor de 1.73%. Además, se pone de manifiesto la diferencia existente en los resultados provocada por la existencia de orden de integración en el modelo de la perturbación, obteniendo el

modelo ARI un valor de 0.80 % global, casi un punto por debajo del obtenido por el modelo no integrado (ARPO).

Tabla 1: Resultados obtenidos

	LINEAS AEREAS	MODELO ERROR	METODO DE TRIMESTRALIZACIÓN							
			TRIVIAL	FERNANDEZ	CL-TRIVIAL	AR1	SW	PC	LITTERMAN	Total
RAIZ DEL ERROR CUADRÁTICO MEDIO (RECM)	R=-0.7; S=-0.7	RB	5.76%	2.81%	2.57%	2.57%	2.57%	2.80%		3.18%
		ARPO	5.27%	1.61%	1.43%	1.29%	1.30%	1.91%	1.54%	2.05%
		ARI	5.05%	.43%	.38%	.58%	.24%	.84%	.22%	1.10%
		Total	5.36%	1.62%	1.46%	1.48%	1.37%	1.85%	.88%	2.06%
	R=0.7; S=-0.7	RB	3.01%	2.85%	2.78%	2.85%	2.78%	2.84%		2.85%
		ARPO	1.90%	1.40%	1.54%	1.37%	1.40%	1.73%	1.31%	1.52%
		ARI	1.08%	.22%	.37%	1.27%	.20%	.47%	.15%	.54%
		Total	2.00%	1.49%	1.56%	1.83%	1.46%	1.68%	.73%	1.58%
	R=0.7; S=0.7	RB	3.21%	3.08%	3.01%	3.05%	3.01%	3.09%		3.08%
		ARPO	1.99%	1.51%	1.68%	1.49%	1.52%	1.90%	1.28%	1.63%
		ARI	1.05%	.28%	.40%	2.75%	.21%	.51%	.19%	.77%
		Total	2.08%	1.62%	1.70%	2.43%	1.58%	1.83%	.74%	1.76%
	Total	RB	3.99%	2.91%	2.79%	2.83%	2.79%	2.91%		3.04%
		ARPO	3.06%	1.51%	1.55%	1.39%	1.41%	1.85%	1.38%	1.73%
		ARI	2.39%	.31%	.38%	1.53%	.21%	.61%	.19%	.80%
		Total	3.15%	1.58%	1.57%	1.91%	1.47%	1.79%	.78%	1.80%
RESTA DE TASAS REGULARES (RT1)	R=-0.7; S=-0.7	RB	7.75%	3.66%	3.35%	3.37%	3.35%	3.66%		4.19%
		ARPO	7.11%	1.94%	1.62%	1.45%	1.46%	2.32%	1.86%	2.53%
		ARI	6.93%	.52%	.43%	.55%	.27%	1.13%	.27%	1.44%
		Total	7.26%	2.04%	1.80%	1.79%	1.69%	2.37%	1.06%	2.65%
	R=0.7; S=-0.7	RB	3.81%	3.63%	3.56%	3.68%	3.56%	3.64%		3.65%
		ARPO	2.14%	1.53%	1.71%	1.51%	1.54%	1.95%	1.45%	1.69%
		ARI	1.24%	.17%	.39%	.51%	.16%	.51%	.14%	.45%
		Total	2.40%	1.78%	1.89%	1.90%	1.75%	2.04%	.80%	1.84%
	R=0.7; S=0.7	RB	4.12%	3.96%	3.89%	3.96%	3.89%	4.00%		3.97%
		ARPO	2.26%	1.66%	1.87%	1.64%	1.68%	2.15%	1.42%	1.81%
		ARI	1.24%	.24%	.43%	.66%	.16%	.56%	.19%	.50%
		Total	2.54%	1.95%	2.06%	2.09%	1.91%	2.24%	.81%	2.00%
	Total	RB	5.23%	3.75%	3.60%	3.67%	3.60%	3.77%		3.94%
		ARPO	3.84%	1.71%	1.73%	1.53%	1.56%	2.14%	1.58%	2.01%
		ARI	3.14%	.31%	.42%	.57%	.20%	.73%	.20%	.80%
		Total	4.07%	1.92%	1.92%	1.92%	1.79%	2.21%	.89%	2.16%

Estudiando los resultados de los método de trimestralización, se posiciona como el mejor método el de Litterman (1983), aunque cabría hacer algunas consideraciones sobre el mismo. Dadas las condiciones en las que se utiliza el método, no se han trimestralizado series cuyo escenario esté formado por errores ruido blanco, ya que las series no pasaban las condiciones de aplicabilidad del mismo, mientras que el número de series en escenarios con errores ARPO trimestralizadas es muy bajo como para tener en cuenta los resultados del mismo. Solamente en el caso de errores ARI el número de series se acercan a las 500 de partida, por lo que consideramos estos como los únicos resultados relevantes del método de Litterman (1983). En estos escenarios sus resultados se muestran mejores que el resto, con un valor de 0.18%, aunque este valor no es significativamente diferente del que obtiene el método de Stram y Wei (1986-1990), que se cifra en 0.21%. Por otro lado, utilizar el método AR1 conlleva una pérdida superior a un punto porcentual en RECM con respecto a los métodos de

Litterman (1983) y Stram y Wei (1986-1990), como por otro lado era de esperar, dados los supuestos sobre los que se desarrollan.

Los escenarios con errores no integrados muestran una situación similar, puntualmente los mejores resultados los obtiene el método AR1 con 1.38%, aunque el método de Stram y Wei (1986-1990) no es significativamente distinto a este con un valor de 1.40%. Por último, en escenarios con errores ruido blanco, todos los métodos basados en modelos quedan por detrás de optar por la solución trivial, como era de esperar ya que en estos casos se ha demostrado que la mejor opción es dividir por cuatro la serie, (en este caso Stram y Wei (1986-1990) se reduce a la solución trivial, es por ello la igualdad de valores)

Las tasas de variación regular nos llevan a las mismas conclusiones que las obtenidas con el estadístico RECM. Así, se cuantifica en torno a un punto y medio la diferencia mínima entre el método trivial y el resto de métodos propuestos. Esto se puede observar en la tabla 1, en donde para el modelo de los errores RB el método trivial obtiene un valor del 5.23% mientras que el método PC presenta un valor del 3.77%. De igual forma podemos observar el cambio que se produce al utilizar otro modelo de los errores, fundamentalmente ante cambios en el orden de integración del mismo, por ejemplo, el método de Stram y Wei (1986-1990) pasa del 1.56% en el modelo ARPO al 0.20% en el modelo ARI,

Conclusiones

Bajo las condiciones en las que se ha llevado a cabo el experimento de simulación, es el método de Stram y Wei (1986-1990) el que presenta mejores resultados generales en cada uno de los modelos del error que se han estudiado, mientras que el resto de métodos se comportan bien cuando el modelo del error bajo los que se desarrollan coinciden con el que presenta la serie trimestral. En cuanto a los elementos que definen el modelo de líneas aéreas, se pone claramente de manifiesto la influencia sobre los resultados finales del cambio de signo en uno o ambos parámetros del modelo. Cabe comentar por último, la mayor dificultad a la hora de llevar a la práctica el método de Stram y Wei (1986-1990) con respecto a los modelos basados en estructuras prefijadas, ya que estos últimos utilizan fórmulas fijas con las que trimestralizar todas las series,

frente a la necesidad de estimar una estructura para cada uno de los casos en función de la estructura de los datos anuales.

Bibliografía

- Chow G.C. and Lin A.L., 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 quaterly series to annuals totals: An approach based on quadratic minimization, *Journal of the American Statistical Association* 66, 99-102.
- Fernandez R.B., 1981, A methodological note on the estimation of time series, *The Review of Economics and Statistics* 63, 471-476.
- Litterman R.B., 1983, A random walk, Markov model for the distribution of time series, *Journal of Business & Economic Statistics* 1, 169-173.
- Pinheiro M. and Coimbra C., 1993, Distribution and extrapolation of time series by related eries using logarithms and smoothing penalties, *Economía*, 359-374.
- Stram D.O. and Wei W.W.S., 1986, A methodological note on the disaggregation of time series totals, *Journal of Time Series Analysis* 7, 293-302.
- Wei W.W.S. and Stram D.O., 1990, Disaggregation of time series models, *Journal of the Royal Statistical Society. Series B.* 52, 453-467.