

Metodología de estimación de factores estacionales para series de afiliación mensuales y diarias

Enero de 2024

Ministerio de Inclusión, Seguridad Social y Migraciones

Introducción

En esta nota se describe la metodología de estimación de los factores de ajuste estacional para las series mensuales de afiliación en alta laborales totales una vez que se dispone de un año completo de datos tras el periodo de la pandemia y la posterior recuperación del empleo. Además, se introduce una importante innovación metodológica, como es la estimación de factores de ajuste estacional para las series de afiliación con frecuencia diaria y laboral, utilizando para ello un enfoque de modelización basado en la representación de componentes no observados mediante modelos dinámicos, el modelo TBATS. Esta innovación permite el análisis de los datos con frecuencias menores a la mensual y se sitúa en la vanguardia de las mejores prácticas internacionales en el análisis de datos de alta frecuencia.

1. Metodología de las series mensuales

1.1 Naturaleza de los datos

Los datos mensuales de afiliación son medias mensuales del total de afiliados a la Seguridad Social en alta laboral desde enero de 2001. También se publican series desagregadas que distinguen entre asalariados y no asalariados y las principales ramas de actividad económica en el conjunto de regímenes del sistema de Seguridad Social.

A partir de las series temporales anteriores, se obtienen las series ajustadas de estacionalidad, según se indica en el epígrafe correspondiente sobre ajuste estacional.

1.2 Ajuste estacional

La estacionalidad afecta a muchas series temporales económicas, que registran un patrón de variación de periodicidad anual. Éste se manifiesta mediante fluctuaciones del nivel de las series a lo largo del año que se reproducen de forma sistemática todos los años y que son, en parte, independientes de su comportamiento a largo plazo. Las causas profundas de esta regularidad suelen estar relacionadas con el clima, el calendario laboral o las costumbres. Los movimientos

debidos a la estacionalidad son, a menudo, lo suficientemente amplios como para enmascarar otras características de los datos que pudieran resultar de interés, como es la tendencia a largo plazo. El ajuste estacional es un proceso de estimación y eliminación de los efectos debidos a la estacionalidad que, idealmente, resultará en una serie libre de fluctuaciones estacionales (relacionadas con la época del año), que conserva intactas el resto de características. El empleo y la afiliación fluctúan con el momento del año de una forma regular, resultando pertinente la desestacionalización de las series para su mejor interpretación.

El procedimiento de ajuste estacional utilizado para las series de afiliación es TRAMO-SEATS¹ y el software empleado es el programa TSW+, un software libre creado como interfaz para Windows y desarrollado en el Banco de España².

La metodología sobre la que trabaja TSW+ se corresponde con el enfoque de modelización de series temporales ARIMA (p,d,q), donde p hace referencia al orden autorregresivo del proceso, d al orden de integración y q al orden de la media móvil. Este esquema se aplica tanto para la parte regular como para la parte estacional, es decir, obteniéndose un modelo SARIMA de la forma:

$$ARIMA(p, d, q) \times ARIMA(P, D, Q)_{12}$$

El modelo es estimado por Máxima Verosimilitud y el software dispone de opciones automáticas y a elección del usuario, permitiendo también la incorporación de regresores³.

1.3 Aplicación a la serie Total de afiliados en el Sistema de la Seguridad Social

Para realizar el ajuste estacional de la serie Total de afiliados en alta laboral en la Seguridad Social, se han utilizado los datos del período enero.2001 – diciembre.2023. Por tanto, esta actualización de los factores estacionales incorpora, por primera vez, los datos del período de la pandemia y de la recuperación posterior. En la anterior revisión anual del ajuste estacional, la

¹ Victor Gómez and Agustín Maravall (1997), Guide for using the programs TRAMO and SEATS, *Documento de trabajo* 9805, Banco de España

² Gianluca Caporello, Agustín Maravall and Fernando J. Sánchez (2001), *Manual de referencia de TSW*, *Documento de trabajo* 0112, Banco de España.

³ Para un mayor detalle sobre este procedimiento puede consultarse el Manual de referencia de TSW: <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeridas/DocumentosTrabajo/01/Fic/dt0112e.pdf>

incorporación de los años 2020 y 2021 al período muestral afectaba a la estimación, de modo que los factores mensuales resultantes presentaban una estacionalidad cambiante, que evolucionaba con rapidez. Al disponer de un año como 2022, apenas afectado por las restricciones para hacer frente a las olas de la pandemia, se ha podido examinar si las pautas estacionales habituales iban recuperándose, permitiendo estimar el ciclo estacional con mayor fiabilidad. Ello ha requerido estimar previamente el efecto del confinamiento y la desescalada mediante una regresión con tres variables *dummy*, que recogen este impacto. La primera de ellas recoge el efecto del confinamiento sobre el nivel de la afiliación y las otras dos, el proceso de recuperación en 2020 -interrumpido por la ola de invierno- y la notable recuperación de la afiliación a partir del verano de 2021. El ajuste estacional con el procedimiento TRAMO-SEATS se aplica a la serie resultante de “depurar” el efecto de la pandemia, obteniéndose, así, la estimación de factores estacionales.

En la Tabla 1, se presentan los factores estacionales de la serie de afiliación del total del sistema para el año 2024 y los dos años anteriores.

Tabla 1. Factores estacionales para la serie de Afiliación- Total Sistema

	2022		2023		2024
ene-22	98,8976	ene-23	98,7434	ene-24	98,6757
feb-22	98,9851	feb-23	98,8566	feb-24	98,8241
mar-22	99,4012	mar-23	99,3890	mar-24	99,3786
abr-22	100,0141	abr-23	100,1083	abr-24	100,1337
may-22	100,7054	may-23	100,8575	may-24	100,8798
jun-22	100,9811	jun-23	101,0445	jun-24	101,0664
jul-22	100,9308	jul-23	101,0104	jul-24	101,0138
ago-22	99,9241	ago-23	99,9437	ago-24	99,9562
sep-22	99,8825	sep-23	99,8824	sep-24	99,8931
oct-22	100,2080	oct-23	100,1986	oct-24	100,2059
nov-22	100,0539	nov-23	99,9986	nov-24	100,0032
dic-22	100,0161	dic-23	99,9668	dic-24	99,9694

2. Metodología adoptada en la serie diaria de afiliación

2.1 Ajuste estacional

El ajuste estacional de series temporales diarias reviste mayor complejidad que el de las series mensuales, por la posible presencia de ciclos estacionales múltiples. Además del ciclo anual, que se aprecia en el distinto nivel de cada mes en relación con la media del año, pueden existir

patrones intra-mensuales que se repiten todos los meses, o semanales. La propia naturaleza de los datos diarios de afiliación, solo disponibles para días laborables, induce una estacionalidad de periodicidad semanal caracterizada por incrementos de afiliación en los lunes y descensos en los viernes. También está presente una regularidad en torno a los últimos días de cada mes, que presentan un descenso más intenso que cualquier otro día del mes. Sin embargo, estas regularidades no tienen una periodicidad fija ya que los meses tienen distinta duración (28, 30 y 31 días), hay años bisiestos y fiestas móviles como la Semana Santa, etc. En consecuencia, los ciclos estacionales tienen una periodicidad media que no es un número entero, como ocurre con los datos trimestrales (la periodicidad en un año es 4) o los mensuales (la periodicidad en un año es 12), sino fraccionario. Así, la periodicidad del ciclo anual con datos diarios es 365,24 y la del ciclo mensual es 30,44. Los métodos de ajuste estacional usuales para series mensuales, como X13-ARIMA o TRAMO-SEATS, no pueden aplicarse en el caso de periodicidades fraccionarias, si bien se está avanzando en su adaptación.

Para modelizar la serie diaria de afiliación Total del Sistema se ha usado la metodología basada en modelos estructurales de componentes no observables desarrollada en De Livera et al. (2011)⁴.

El modelo, cuyo acrónimo es TBATS (T: Trigonometric, estacionalidad aproximada por funciones trigonométricas, B: Box-Cox transformación, A: modelo ARMA para el componente irregular, T: Tendencia, S: Seasonal, componente estacional), se basa en un modelo estructural del tipo $z_t = p_t + s_t + u_t$ donde la serie (z) es el resultado de la agregación de tres componentes: tendencia-ciclo (p), estacionalidad (s) y componente irregular (u), todos ellos dependientes de un shock único a_t .

Este componente irregular obedece a un esquema ARMA estacionario e invertible tal que

$$(1 - \rho_1 B - \dots - \rho_p B^p)u_t = (1 - \theta_1 B - \dots - \theta_q B^q)a_t$$

$$a_t \sim iid N(0, v)$$

La tendencia sigue un modelo muy general (tendencia determinística, o estocástica con un orden de integración de orden 1 o 2). El componente estacional (de carácter múltiple) admite periodos fraccionales, como por ejemplo 21,74 que es el que corresponde a los días laborables

⁴ De Livera, A., Hyndman, R. J. and Snyder, R. D. (2011): "Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing", *Journal of the American Statistical Association*, 106(496), 1513-1527

en un mes. Este componente se aproxima mediante ciclos derivados de funciones trigonométricas que permiten perfilar patrones de diversa periodicidad. En el caso de la serie de afiliación el componente estacional está formado por un ciclo semanal, un ciclo intra-mensual y el ciclo anual. A su vez, TBATS considera que cada sub-componente estacional puede estar formado por oscilaciones de menor periodicidad (o mayor frecuencia), denominados armónicos. El número de armónicos se determina mediante una estimación preliminar.

El modelo TBATS se estima utilizando el filtro de Kalman y el software libre de las librerías en R y la interfaz RStudio. Previamente a la estimación del modelo TBATS, es preciso realizar un pre-tratamiento para “limpiar” la serie diaria de efectos calendario ligados a festividades de carácter móvil, como la Semana Santa u otras. La validación del modelo, como es usual, se apoya en diagnósticos sobre los residuos: normalidad, homogeneidad de media y varianza, así como la existencia de valores atípicos.

2.2 Elementos específicos de la serie diaria de Total de afiliados al Sistema de la Seguridad Social

Los datos proceden de la base diaria del número de afiliados en el total del Sistema de la Seguridad Social. Esta base de datos, que está disponible de forma homogénea desde 2011, ofrece datos solo para los días laborables (excluidos el día 24 y 31 de diciembre). Es, por tanto, una serie diaria con semanas de 5 días y sin dato en los festivos que caen entre lunes y viernes, en cuyo caso se acorta la longitud de la semana afectada. De cara a obtener periodicidades exactas en el componente semanal, se interpolan linealmente los datos de los festivos que caen de lunes a viernes⁵.

Previamente a estimar el modelo TBATS, se realiza una regresión con variables calendario para recoger el efecto del último día del mes, que tiene carácter móvil, ya que puede caer en cualquier día de la semana. Después de valorar diversos regresores, el efecto del fin de mes se capta con una única variable que identifica el fin de mes, con valor “1” y es “0” el resto de días del mes, siempre que el último día del mes sea laborable o esté interpolado. Cuando el último día de mes cae en fin de semana esta variable toma el valor “0”. La regresión incluye, además,

⁵ Por tanto, las periodicidades de los componentes semanal, intra-mensual y anual son: 5, 21,7411 y 260,8929 respectivamente, teniendo en cuenta la distinta longitud de los meses y los años bisiestos.

una variable para controlar el efecto de la pandemia sobre la afiliación, similar a la empleada en Cuevas y Quilis (2022)⁶.

Por último, los factores estacionales diarios obtenidos se han compatibilizado con los factores mensuales estimados con la metodología TRAMO-SEATS, con el fin de que las estimaciones de datos ajustados en series de alta frecuencia que se publiquen sean consistentes con los datos ajustados de estacionalidad de periodicidad mensual. Para ello se multiplica el factor del modelo mensual por la ratio del factor diario respecto al promedio, en cada mes, que resulta del ajuste con datos diarios, obteniéndose un factor diario compatible.

⁶ Cuevas, A. y E. Quilis (2022), "The COVID-19 shock through the lens of the Spanish sales daily data: Modelling and seasonal adjustment challenges", OECD Workshop, septiembre, mimeo.