

Detalle del ajuste estacional de la secuencia de posts en Nada es Gratis,

por Jesús Ruiz y Luis Puch

Una forma muy general de desestacionalizar una serie temporal de forma simple, entendible para una audiencia amplia con conocimientos básicos de econometría de series temporales, en presencia de un fuerte efecto continuado sobre un periodo largo (como los efectos que la pandemia ha producido sobre la serie temporal de afiliados), es como sigue:

- 1) El primer paso sería identificar y estimar un modelo de regresión ARIMA (“regARIMA”) como el siguiente:

$$(1-L)^d(1-L^s)^D y_t = \sum_{j=1}^n \beta_j (1-L)^d (1-L^s)^D \xi_{j,t} + N_t$$
$$N_t = \frac{(1 + \theta_1 L + \dots + \theta_q L^q)(1 + \Theta_1 L^s + \dots + \Theta_Q L^{sQ})}{(1 - \phi_1 L - \dots - \phi_p L^p)(1 - \Phi_1 L^s - \dots - \Phi_P L^{sP})} a_t, \quad a_t \underset{iid}{\sim} N(0, \sigma_a^2) \quad (1)$$

es decir, $N_t \sim ARMA(p, q) \times ARMA(P, Q)_s$

donde L es el operador retardo, d el número de diferencias regulares, D el número de diferencias estacionales, p y P son los órdenes de los polinomios regular y estacional del componente autorregresivo de N_t , q y Q son los órdenes de los polinomios regular y estacional del componente de medias móviles de N_t , y s es el periodo estacional (igual a 12 para series mensuales o igual a 4 para series trimestrales). Las variables $\xi_{j,t}$, $j=1, \dots, n$ son variables fijas que suelen recoger efectos tipo impulso o escalón de atípicos a lo largo de la muestra. Por tanto, la variable N_t es la serie temporal estacional neta de efectos de valores atípicos.

- 2) El segundo paso sería aplicar algún tipo de filtro que desestacionalice la serie temporal N_t . Por ejemplo, X13ARIMA-SEATS o TRAMO-SEATS, que, básicamente, lo que hacen tales filtros es identificar y estimar los componentes típicos de una serie temporal económica: tendencial, cíclico, irregular y estacional; descontando de N_t el componente estacional, daría como resultado una serie temporal, neta de efectos de atípicos, desestacionalizada. Llamemos a esa serie $N_{sa,t}$.

3) Por último, si sumamos a la serie $N_{sa,t}$ obtenida en el paso 3 los efectos de los atípicos estimados en el paso 1, podemos obtener la serie temporal desestacionalizada:

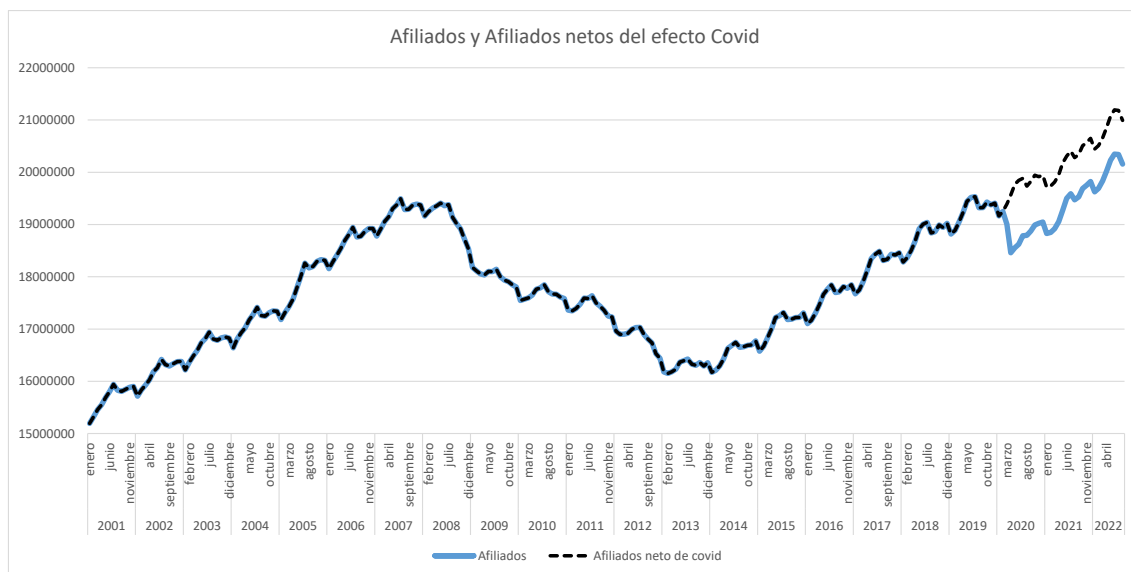
$$(1-L)^d(1-L^s)^D y_{sa,t} = \sum_{j=1}^n \beta_j (1-L)^d(1-L^s)^D \xi_{j,t} + N_{sa,t}$$

Para el caso que nos ocupa, la serie temporal mensual de afiliados sigue un modelo de regresión-ARIMA como el siguiente (muestra 2001:01-2022:08):

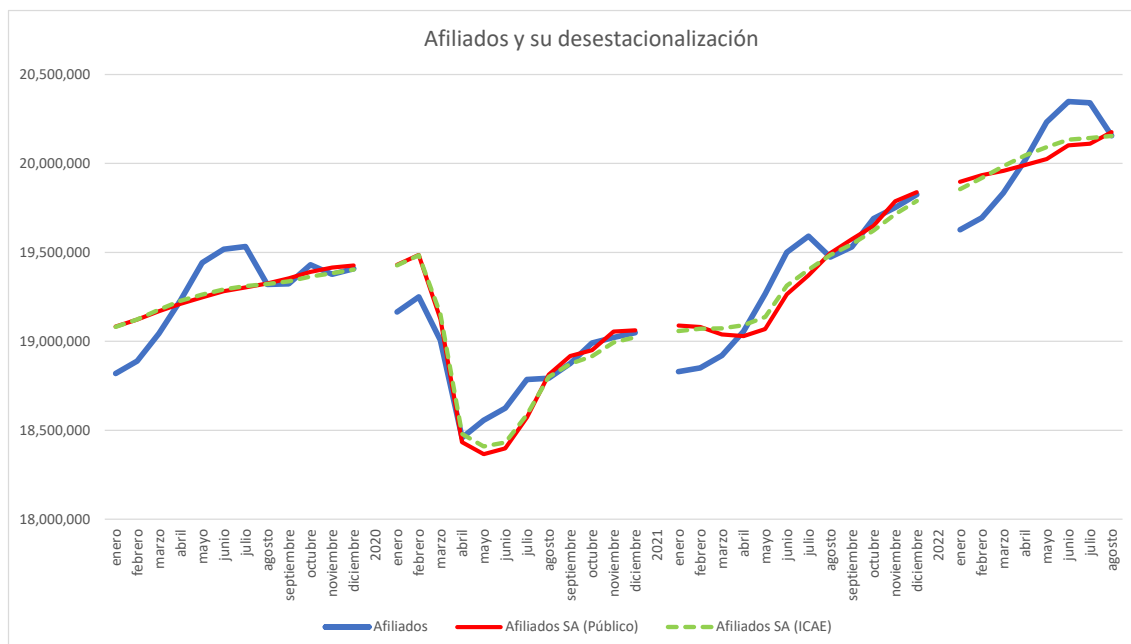
$$\begin{aligned} \nabla_{12} \nabla \ln(\text{afiliados}_t) = & -0.020033 \cdot \nabla_{12} \xi_t^{\text{impulso_mar20}} \\ & -0.038251 \cdot \nabla_{12} \xi_t^{\text{impulso_apr20}} \\ & -0.005202 \cdot \nabla_{12} \xi_t^{\text{impulso_may20}} \\ & +0.008547 \cdot \nabla_{12} \xi_t^{\text{impulso_jul20}} \\ & +0.007918 \cdot \nabla_{12} \xi_t^{\text{impulso_aug20}} \\ & +0.003006 \cdot \nabla_{12} \xi_t^{\text{impulso_nov20}} + \\ & +0.005346 \cdot \nabla_{12} \xi_t^{\text{impulso_jun21}} + N_t \\ N_t = & \frac{1-0.422355 L^{12}}{1-0.694337 L-0.123692 L^4} a_t; \quad a_t \sim N(0, 0.00186) \end{aligned}$$

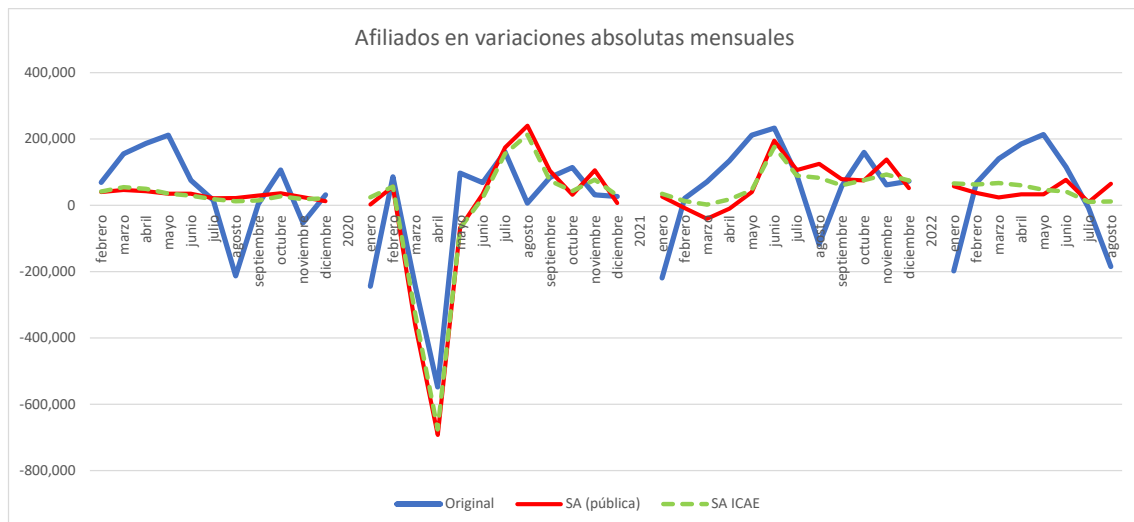
[Nótese que las variables dummy incluidas son impulsos en primeras diferencias regulares, es decir, escalones en niveles].

Aplicando los tres pasos anteriores, hemos obtenido la serie de afiliados desestacionalizada por nosotros. En el siguiente gráfico tenemos representadas las series de afiliados y la de afiliados sin el efecto covid (algo así como el contrafactual); es esta última serie la que desestacionalizamos y, una vez realizado tal ajuste estacional, volvemos a agregar el efecto covid para obtener la serie de afiliados desestacionalizada.



En los dos gráficos que siguen a continuación se muestra que la serie desestacionalizada por nosotros sigue una evolución muy parecida a la serie estimada por el ministerio público.





Sin embargo, nuestro modelo no prevé una variación absoluta en agosto como la del ministerio: +25000 frente a los +65000 anunciados por el ministro. Escribá a mediados de agosto.

Pero, ¿es esto una diferencia sustancial?; ¿fue una diferencia sustancial que la variación absoluta en julio fuera para el ministerio de +9000, para nosotros fuera de +11000 y para otros profesionales llegase a ser de una disminución de más de 20000?

Veamos qué podemos decir a todo esto. La serie desestacionalizada de afiliados es una estimación (de una serie temporal que no es una estimación, es la contabilización real de los afiliados en cada instante temporal), lo cual significa que tiene un modelo de probabilidad asociada cuyos parámetros han sido estimados. Esto significa que podríamos obtener muchas sendas para la serie desestacionalizada perturbando adecuadamente los parámetros estimados del modelo (por ejemplo, mediante bootstrapping, es decir, remuestreando los residuos de los modelos estimados y re-estimando los parámetros de los modelos identificados). ¿Cómo hacer esto? Para ello, es importante que tengamos en cuenta dónde están las fuentes de incertidumbre que afectan a la serie desestacionalizada; esencialmente, tenemos dos fuentes de incertidumbre en los parámetros de los modelos usados en la desestacionalización: i) la incertidumbre asociada a la estimación del modelo que nos sirve para eliminar el efecto covid

(modelo (1)); y ii) la incertidumbre asociada al filtro utilizado para estimar el ajuste estacional de la serie neta de covid.

Nosotros, por sencillez y por no entrar en los métodos de desestacionalización, sólo vamos a medir la primera fuente de incertidumbre, lo cual significa que posiblemente midamos un mínimo de la incertidumbre asociada a la estimación de un ajuste estacional típico.

A través de la metodología bootstrapping, remuestreamos con reemplazamiento los residuos a_t del modelo (1). A partir de ese re-muestreo obtenemos una nueva serie de N_t , y con esta nueva serie, obtenemos una nueva serie de afiliados; con esta nueva serie, volvemos a estimar el modelo (1), obteniendo una nueva estimación para los parámetros. Si repetimos esto muchas veces, obtendremos muchos valores estimados de los parámetros. Con todos estos valores paramétricos estimados podremos obtener muchas series netas de covid, susceptibles de desestacionalizarse. Para cada una de estas series desestacionalizadas obtenidas volvemos a agregarles el efecto covid y obtendremos muchas series temporales de afiliados desestacionalizadas.

Así, en términos medios, nosotros obtenemos los valores de la serie desestacionalizada para junio, julio y agosto: 20.105.613, 20.117.739, y 20.141.215, mientras que el ministerio público, para esos mismos periodos dio los valores 20.102.037, 20.111.141 y 20.176.141; sólo se aprecia una diferencia importante en agosto; sin embargo, para tales periodos las series desestacionalizadas usando bootstrapping tenían una desviación típica de 11.050, 10.990 y 10.613 para junio, julio y agosto respectivamente, lo cual significa que las bandas de confianza estarían entre 20.127.713 y 20.083.513 para junio, entre 20.139.719 y 20.095.759 para julio, y entre 20.162.441 y 20.119.989. Salvo agosto, los otros valores están dentro de bandas, y el dato de agosto del ministerio no está muy lejos del límite superior de la banda de confianza (al 5%). Con todo, las diferencias entre los límites superiores e inferiores de las bandas de confianza están en un entorno de 43.000 afiliados, como para tenerlo en cuenta cuando se discute entre estimaciones puntuales.

Nótese que no hemos tenido en cuenta la otra fuente de incertidumbre, la asociada a la metodología de filtrado para el ajuste estacional. Cabría esperar unas bandas de confianza aún mayores, por lo que la discusión entre estimaciones puntuales quedaría fuera de lugar, ya que las diferencias dadas por cualquier institución o investigador apenas son significativas estadísticamente.