



Munich Personal RePEc Archive

**Review of models for seasonal
adjustment and projection of EMAE
activity sectors. Year 2021.**

Frank, Luis

2 July 2021

Online at <https://mpra.ub.uni-muenchen.de/112285/>
MPRA Paper No. 112285, posted 08 Mar 2022 18:39 UTC

Revisión de modelos para ajuste estacional y proyección de sectores de actividad del EMAE.

Año 2021

Luis Frank ^{*†}

Resumen

En el artículo se revisan los modelos regARIMA utilizados para ajustar estacionalmente y proyectar el VAB a nivel de sección CIU-3.1. Si bien la revisión estuvo motivada originalmente por el efecto de la pandemia de COVID en el ajuste estacional, también se analizaron en profundidad dificultades previas para reproducir la serie desestacionalizada del EMAE publicada por INDEC. De la revisión se infiere que las discrepancias entre el EMAE desestacionalizado y la serie publicada por INDEC se explican por la exclusión de feriados nacionales del procedimiento de ajuste de INDEC más que por diferencias en las especificaciones de los modelos sectoriales. Además, la discrepancia observada en la tendencia-ciclo agregada se debe a que INDEC sigue el procedimiento propuesto por E. Bee Dagum en lugar de un procedimiento indirecto similar al utilizado para desestacionalizar el EMAE.

Palabras clave: desestacionalización de series, ARIMA estacional, EMAE.

JEL: C820

Abstract

In the paper, regARIMA models used to seasonally adjust and to project the GVA at the section level of ISIC-3.1 are reviewed. Although the review was originally motivated by the effect of the COVID pandemic on seasonal adjustment, previous difficulties in reproducing the seasonally adjusted EMAE published by INDEC were also deeply analysed. From the review, it is inferred that discrepancies between the reproduced seasonally-adjusted EMAE and the series published by INDEC are explained by the exclusion of national holidays from INDEC's adjustment procedure, rather than by differences in the specifications of sectoral models. Besides, the discrepancy observed in the aggregated trend-cycle is due to the fact that INDEC follows the direct procedure proposed by E. Bee Dagum instead of an indirect procedure similar to that used to seasonally adjust the EMAE.

Keywords: seasonal adjustment of series, seasonal ARIMA, EMAE.

JEL: C820

*DNMyP. Secretaría de Política Económica. Ministerio de Economía. Av. Hipólito Yrigoyen 250, C1086AAB. Buenos Aires, Argentina.

†Universidad de Buenos Aires. Facultad de Agronomía. Av. San Martín 4453, C1417DSE. Buenos Aires, Argentina.

1 Introducción

La desestacionalización y proyección de series económicas mediante modelos regARIMA o ARI-MAX estacionales requiere la revisión periódica de especificaciones a fin de incorporar nueva toda nueva información sobre el proceso generador de datos aportada por los nuevos valores que se agregan a las series, o bien por la revisión de cifras consideradas provisorias. Si bien la frecuencia de revisión de modelos no se halla normalizada como en el caso de las revisiones del sistema estadístico, se considera una buena práctica revisarlos al cierre de cada año calendario. Siguiendo esta práctica, revisamos [7] los modelos de proyección de agregados macro trimestrales del SCN y de comercio exterior al cierre del cuarto trimestre de 2020, y en este informe continuaremos el proceso de revisión de modelos con los de las componentes del Estimador Mensual de Actividad Económica (EMAE) elaborado por INDEC.¹

2 Métodos y resultados

Partimos de las componentes del EMAE a nivel de letra de la clasificación CIIU-3 - excepto la letra P (hogares con servicio doméstico) que no se estima mensualmente - incluyendo la serie de impuestos netos de subsidios, en el período 2004-2020. La serie de la letra P se construyó interpolando el valor agregado bruto (VAB) trimestral de esta letra con la serie mensual de la letra K.² Para interpolar la serie trimestral utilizamos el método de [5].³ Todas estas series abarcan el período 2004-2020, siendo los años 2004-2018 definitivos, el año 2019 era provisorio y el 2020 preliminar. Esto significa que todos los datos han sido revisados (trimestralmente) al menos una vez y consistidos con las series trimestrales de VAB. Para chequear la fiabilidad de la estimación de la letra P construimos un índice Laspeyres utilizando como ponderaciones la participación de cada componenete en el PIB de 2004. Las diferencias entre este índice y el EMAE fueron mínimas y atribuibles a redondeos de cifras.

A continuación revisamos los modelos de desestacionalizado del año 2020. Para ello seguimos el procedimiento usual de dos etapas: uno de búsqueda y selección, y otros de desestacionalizado propiamente dicho y conciliación con los totales anuales de las series originales. En la primera etapa las series fueron ajustadas al modelo $(0, 1, 1)(0, 1, 1)$ transformadas en logaritmos y sin transformar a fin de determinar - mediante el criterio de información de Akaike o AICC - si correspondía transformarlas. Seleccionada la versión de menor criterio de información realizamos una búsqueda exhaustiva del modelo “óptimo” ajustando las 576 especificaciones $\text{regARIMA}(p,d,q)(P,D,Q)_{12}$ que surgen de combinar los parámetros $p, q = \{0, 1, 2, 3\}$ y $P, Q = \{0, 1, 2\}$ hasta un orden de integración máximo de 1, y reteniendo para cada serie la especificación que presentó el menor error porcentual absoluto medio o MAPE en los últimos tres años de la serie, siempre que éste no superara el 15%. En el único caso en que se superó este límite se retuvo el modelo de menor MAPE aunque sabiendo que excedía la tolerancia

¹Tomamos como referencia metodológica de esta revisión las notas de [6] del año 2020.

²Recordemos que de acuerdo a la metodología publicada del EMAE [8, p. 9]: “No se dispone de indicador mensual específico para esta actividad [letra P]. Su evolución se condice con la del valor agregado a precios básicos deducidos [de] los *servicios inmobiliarios con bienes propios y arrendados*”. Pero los servicios inmobiliarios son una componente (50% del VAB, aproximadamente) de la letra K (Actividades Inmobiliarias, Empresariales y de Alquiler) que sólo se publica trimestralmente, al igual que la letra P. En consecuencia, al no publicarse la letra K abierta en los comunicados de EMAE, la única posibilidad para reproducir aproximadamente el EMAE es construir una serie *proxy* interpolando la serie trimestral de la letra P con la serie mensual de la letra K, lo cual determinará que haya inevitablemente una pequeña discrepancia entre nuestro agregado y el EMAE desestacionalizado.

³El método propuesto por R. Fernández es una variante del método de Denton utilizado por la DNCN. La interpolación se realizó con un programa propio escrito en el lenguaje matricial Euler Math Toolbox

máxima admitida por el programa X-13 ARIMA-SEATS, la que actualmente se considera un estándar para el desestacionalizado de series. Todos los modelos probados incluyeron variables exógenas de efecto calendario y feriados o huelgas generales, así como una cantidad de variables indicativas de valores atípicos o *outliers*. Las variables de efecto calendario se refirieron básicamente a la cantidad relativa de días laborables y al año bisiesto. Los feriados y huelgas generales, se incorporaron en cantidad de días por mes sin distinción del día de la semana en que cayeran. Los detalles de codificación de todas estas variables pueden hallarse en [4, pp. 27-28]. Cabe aclarar que no surge del último informe metodológico del EMAE [8] el uso de variables de efecto calendario, lo cual agrega una fuente de discrepancia adicional entre nuestro indicador mensual desestacionalizado y la serie publicada de EMAE desestacionalizado.

Una vez seleccionado un modelo regARIMA adecuado para cada serie, calculamos sendas series desestacionalizadas y conciliadas con los totales anuales de cada una de las series originales a través del método de Denton. Luego agregamos estas series en un indicador general equivalente a la serie desestacionalizada del EMAE. La agregación de series se realizó mediante un índice Laspeyres en base 2004=100 al igual que en el EMAE. Este indicador, sin embargo, no necesariamente coincidirá con la serie publicada por INDEC por varias razones. La primera, es el uso, en nuestro cálculo, de una serie *proxy* para la letra P, que si bien creemos que reproduce la construcción de INDEC, en rigor es imposible saberlo consultado solamente el informe metodológico del EMAE. La segunda, la incorporación en nuestros modelos de una componente calendario que pareciera no estar en el EMAE. La tercera razón, el desconocimiento de las especificaciones utilizadas por INDEC para modelar las componentes del EMAE, así como la periodicidad de las revisiones de estos modelos, visto que la metodología actual [8] y los comunicados mensuales no aclaran estos puntos. En caso que INDEC siguiera una estrategia de selección automática de modelos, aprovechando facilidades del programa X-13 ARIMA-SEATS, y revisara las especificaciones con una frecuencia distinta a la nuestra caben esperar resultados levemente distintos.

Para el cálculo la tendencia-ciclo general construímos un índice Laspeyres de tendencia-ciclo considerando y sin considerar en las componentes saltos de nivel o *level shifts*. El índice Laspeyres fue rescalado a 100 en el año 2004 para normalizar todas las series en dicha base. INDEC, sin embargo, no calcula la tendencia-ciclo general de este modo sino por suavizando de la series desestacionalizada de EMAE con el filtro de Henderson de 13 términos, siguiendo un método propuesto por Bee Dagum (ver [2] y [3]) a fines de los años '90. Hasta donde tenemos conocimiento este método no es estándar sino una solución particular que perduró en el cuerpo metodológico de INDEC desde las primeras publicaciones de series desestacionalizadas.⁴ En el cuadro 1 detallamos las especificaciones seleccionadas para desestacionalizar cada componente del EMAE. Nótese que, a diferencia de la revisión anterior, en ésta omitimos la inclusión de la variable climática ONI (“anomalías de temperaturas superficiales en el Pacífico Sur) porque solamente resultó significativa para la letra A en la modelación mensual, pero no en el modelo trimestral, y porque incluso en el modelo mensual su aporte a la reducción del MAPE fue exiguo.

3 Discusión

Revisamos los modelos utilizados para desestacionar las componentes del EMAE. Las nuevas especificaciones se muestran en el cuadro 1. La simple inspección del cuadro permite apreciar que las nuevas especificaciones presentan elevados niveles de ajuste - evaluados a través del

⁴En [3] se mencionan dos consultas oficiales realizadas en 1999 a la Dra. Estela Bee Dagum de Statistics Canada por este tema, así como otras sugerencias realizadas por el Dr. Jacob Ryten de la misma institución

MAPE - salvo la serie de Pesca (letra B) cuyo MAPE supera la tolerancia del 15% aceptada en oficinas de estadísticas públicas para rechazar modelos candidato. Además, la calidad de ajuste estacional - evaluada a través del índice Q sin M_2 - también exhibió elevados niveles calidad, salvo en Pesca y por Hogares con Servicio Doméstico. En este último caso, atribuimos la mala calidad de ajuste estacional a la manipulación de datos a la que recurrimos para obtener una serie *proxy* de la letra P, manipulación que altera las propiedades estadísticas de la verdadera serie mensual suyacente a los datos trimestrales.

A lo largo del proceso de revisión comparamos resultados parciales con los de INDEC a fin de aclarar cuáles serían las causas de las discrepancias que venimos observando entre nuestras series y las publicadas por el instituto. En principio, concluimos que las diferencias entre indicadores desestacionalizados se deben a la forma en que se incorporan los días feriados en nuestros modelos y, en menor medida, a la frecuencia y oportunidad de revisión de modelos de cada una de las oficinas. En nuestro caso, todos los feriados que no caigan en domingo se suman y se incorporan a la matriz de variables exógenas como variable separada. En el caso de INDEC, como ya dijimos, el informe metodológico de EMAE [8] no es explícito acerca de la incorporación de feriados. Sin embargo, el informe metodológico de las cuentas trimestrales [9] menciona que “el programa considera el efecto de las fiestas de Pascua, de los días de comercio, de los años bisiestos y el número de días en cada mes” por lo cual podemos conjeturar (i) que los únicos feriados considerados son los de Pascua, o bien (ii) que todos los feriados, excepto éstos últimos, se descuentan de la cuenta relativa de días laborables (“días de comercio”) respecto a la cantidad de domingos. Las diferencias entre los indicadores de tendencia-ciclo (TC) en cambio son principalmente metodológicas. Mientras que nosotros construimos el indicador general agregando las TC de cada componente, de manera análoga al cálculo del nivel de actividad general desestacionalizado, INDEC suaviza la serie del EMAE desestacionalizada con el filtro de Henderson de 13 términos, previa extensión con el modelo de la aerolínea para evitar el conocido problema del “punto final”.⁵ Es decir, mientras que nuestro método es indirecto, el de INDEC es directo pero sobre una serie desestacionalizada indirectamente. Hasta donde sabemos, no hay un estándar metodológico que aconseje un camino en particular.⁶

⁵El cálculo indirecto de la TC es también una práctica común en estadísticas públicas. Ver, por ejemplo, [1].

⁶Ver [10] para un panorama de la práctica corriente en varias oficinas de estadísticas públicas.

Referencias

- [1] ALADI, s/f. Metodología de cálculo de los indicadores de tendencia-ciclo del comercio de la ALADI. Disponible *on-line* en: <http://www2.aladi.org/sitioAladi/documentos/GraficosTenci/Metodologia.pdf>
- [2] Bee Dagum E., 1996. A New Method to Reduce Unwanted Ripples and Revisions in Trend-Cycle Estimates From X-11-ARIMA. *Survey Methodology* 22(1): 77-83.
- [3] Instituto Nacional de Estadística y Censos - INDEC, 2000. Síntesis metodológica. Estimador Mensual Industrial. Ajuste Estacional y Estimador EMI Ajustado. Disponible *on-line* en: https://biblioteca.indec.gob.ar/bases/minde/2mi409_4.pdf
- [4] Felipe H., Correa V., Luna L. y F. Ruiz 2002. Desestacionalización de series económicas: el procedimiento usado por el Banco Central de Chile. Documentos de Trabajo Nro. 177. Banco Central de Chile.
- [5] Fernández R. 1981. A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics* 63(3): 471-476.
- [6] Frank L. 2020. Revisión de modelos para la desestacionalización de series mensuales y trimestrales de actividad económica. MPRA Paper No. 111423. Disponible en <https://mpra.ub.uni-muenchen.de/111423/>
- [7] Frank L. 2021. Revisión de modelos para la desestacionalización y proyección de series macroeconómicas trimestrales. MPRA Paper No. 112278. Disponible en <https://mpra.ub.uni-muenchen.de/112278/>
- [8] Instituto Nacional de Estadística y Censos - INDEC, 2016. Estimador Mensual de Actividad Económica (EMAE), Base año 2004. Metodología INDEC Nro. 20. ISBN 978-950-896-484-7. Disponible *on-line* en: <https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-9-48>
- [9] Instituto Nacional de Estadística y Censos - INDEC, 2016. Cuentas Nacionales. Metodología de estimación. Base 2004 y serie a precios constantes y corrientes. Metodología INDEC Nro. 21. ISSN 2545-7179. Disponible *on-line* en: <https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-9-47>
- [10] Mc Laren C. y X. Zhang, 2010. The Importance of Trend-Cycle Analysis for National Statistics Institutes. *Estudios de Economía Aplicada* 28(3): 607-624.

Cuadro 1: Especificaciones para la desestacionalización de componentes del EMAE. Letras A a P de la clasificación CIIU-3 e impuestos netos de subsidios. Revisión 2021 con datos del período 2004-2020.

Letra	Descripción	Transf.	Modelo	Const.	Días lab.	Feriatos y paros	Año bis.	MAPE	Q s/M_2	<i>Outliers</i>
A	Agricultura	log	(3 0 1) (1 1 1)	si	si	no	no	6,51	0,56	LS2004.Jul TC2009.Mar TC2009.Apr TC2018.Apr
B	Pesca	log	(0 1 1) (1 0 0)	no	si	no	no	25,9	1,07	LS2005.Nov
C	Minería	–	(1 0 2) (2 1 1)	no	si	no	si	2,61	0,43	AO2005.Oct AO2008.May AO2009.Aug AO2010.Dec TC2011.Apr LS2020.Apr
D	Industria	–	(3 0 2) (2 0 0)	si	si	si	si	4,48	0,30	TC2020.Mar TC2020.Apr
E	Agua y energía	log	(3 0 1) (2 0 2)	si	si	no	no	2,44	0,53	TC2020.Apr
F	Construcción	–	(3 0 1) (0 0 2)	si	si	si	no	6,77	0,48	LS2011.Jan AO2016.Aug TC2018.Nov TC2019.Aug TC2020.Mar TC2020.Apr AO2020.Sep
G	Comercio	–	(2 0 3) (1 0 0)	si	si	si	no	4,75	0,42	TC2020.Mar TC2020.Apr
H	Hoteles y restaur.	–	(3 0 3) (2 0 0)	no	no	si	no	3,07	0,38	TC2009.Jul TC2020.Mar AO2020.Apr LS2020.Apr AO2020.Jun AO2020.Jul
I	Transporte y comunic.	–	(0 0 3) (1 0 1)	si	si	si	si	2,36	0,26	LS2006.Nov AO2008.Jun LS2010.May TC2020.Mar LS2020.Apr
J	Interm. financ.	log	(2 0 0) (2 0 2)	si	si	si	no	5,53	0,43	LS2018.Oct
K	Actividades empresarias	–	(2 0 3) (1 0 1)	si	si	si	si	1,08	0,19	AO2007.Dec LS2010.Jan TC2014.Jun AO2015.Jul AO2020.Mar TC2020.Apr
L	Admin. pública	log	(3 0 2) (2 0 1)	si	si	no	si	0,79	0,35	LS2006.Jan TC2007.May LS2007.Nov TC2008.Jan LS2009.Jan LS2009.Jul AO2013.Feb LS2020.Apr LS2020.Jul LS2020.Oct
M	Educación	log	(2 0 2) (2 0 1)	si	no	si	no	0,87	0,20	LS2006.Jan AO2006.Apr TC2007.May AO2007.Nov AO2008.Jan LS2009.Jan LS2009.Jul LS2015.Apr TC2020.Apr LS2020.Jun AO2020.Jul
N	Serv. soc. y de salud	–	(2 0 1) (1 0 0)	no	no	si	no	0,70	0,20	AO2005.Mar TC2005.Jun LS2006.Jan AO2009.Jun TC2010.Jan AO2010.Jul LS2012.Jan LS2020.Jan AO2020.Mar TC2020.Apr AO2020.May AO2020.Jul TC2020.Oct
O	Servicios personales	–	(2 1 3) (0 1 1)	si	si	si	no	2,46	0,46	TC2020.Mar LS2020.Apr LS2020.Aug LS2020.Oct
P	Hogares c/ serv. dom.	–	(1 0 3) (2 0 0)	si	si	si	no	2,82	1,01	TC2007.Jan LS2020.Mar TC2020.Apr AO2020.May TC2020.May
–	Imp. netos	–	(2 0 3) (2 0 0)	si	si	si	si	3,68	0,66	AO2020.Apr