

Modelos de estacionalidad

Randall Romero Aguilar, PhD
randall.romero@ucr.ac.cr

EC4301 - Macroeconometría
I Semestre 2020

Última actualización: 25 de mayo de 2020

UCR
UNIVERSIDAD DE COSTA RICA

**ESCUELA de
ECONOMÍA**
UNIVERSIDAD DE COSTA RICA

Tabla de contenidos

1. Introducción
2. Modelo ARIMA estacional
3. Ajuste estacional
4. Raíces unitarias estacionales

1. Introducción

- ▶ En una serie de tiempo, **estacionalidad** es un patrón regular de cambios que se repite cada s períodos.
- ▶ Por ejemplo, una serie mensual es estacional si valores altos tienden a ocurrir en meses particulares y los valores bajos tienen a ocurrir en otros meses particulares. En este caso $s = 12$.
- ▶ Para series trimestrales estacionales, $s = 4$.
- ▶ El primer paso para detectar si una serie es estacional es simplemente graficarla.



¡Estacionalidad no es lo mismo que estacionariedad!

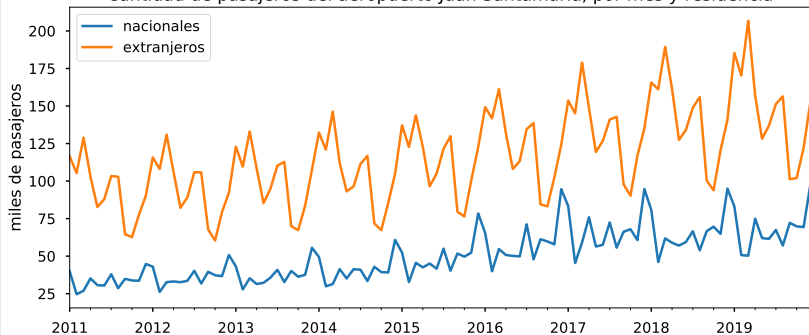
Ejemplo 1:

Pasajeros del aeropuerto Juan Santamaría

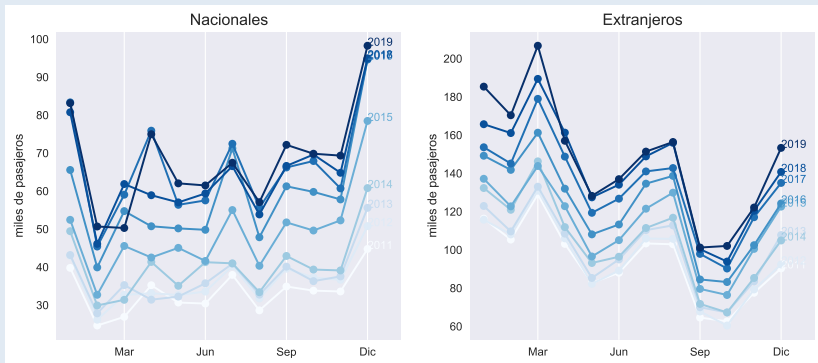


[Datos-ARESEP-Pasajeros-por-aeropuerto.xlsx](#)

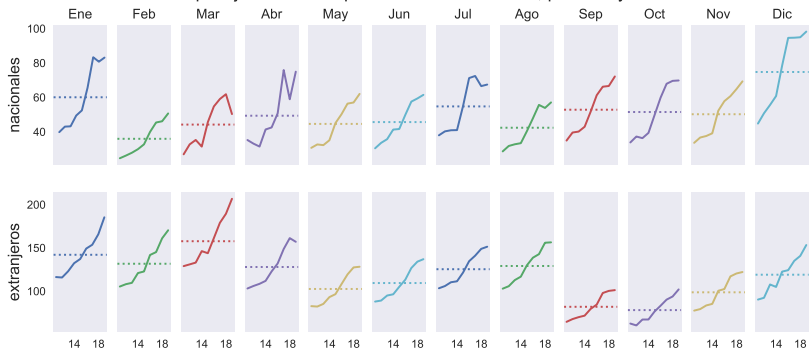
Cantidad de pasajeros del aeropuerto Juan Santamaría, por mes y residencia



- ▶ Un gráfico estacional es similar a un gráfico de tiempo, excepto que los datos se mapean contra la “estación” en la cual se observan.
- ▶ Permiten percibir el patrón estacional subyacente más claramente.
- ▶ Es especialmente útil para identificar años en que cambian los patrones.



Miles de pasajeros del aeropuerto Juan Santamaría, por mes y residencia



- ▶ La temporada “alta” de viajes de extranjeros ocurre entre enero y abril, mientras que la “baja” sucede en septiembre-octubre.
- ▶ La temporada “alta” de viajes de nacionales es en diciembre y enero, la baja en “febrero”.

2. Modelo ARIMA estacional

- ▶ En un modelo ARIMA estacional, términos AR y MA predicen y_t usando valores de datos y errores con rezagos que son múltiplos de s
- ▶ Por ejemplo, con datos trimestrales ($s=4$),
 - ▶ un modelo autorregresivo estacional de primer orden usa y_{t-4} para predecir y_t , mientras que uno de segundo orden usa y_{t-4}, y_{t-8} para ello.
 - ▶ un modelo de media móvil estacional de primer orden usa ϵ_{t-4} para predecir y_t , mientras que uno de segundo orden utiliza $\epsilon_{t-4}, \epsilon_{t-8}$.

- ▶ Por ejemplo, si quisiéramos pronosticar el número de pasajeros extranjeros que viajarán por el SJO en agosto 2020, tiene mucho sentido modelar ese valor en función del número de pasajeros extranjeros que viajaron por SJO en agosto de años anteriores.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
2011	116	105	129	103	83	88	103	103	65	63	78	90
2012	116	108	131	106	82	89	106	106	68	60	79	92
2013	123	110	133	108	85	95	110	113	70	67	84	108
2014	132	121	146	112	93	96	111	117	72	67	85	105
2015	137	123	144	123	97	105	122	130	80	76	101	123
2016	149	142	161	132	108	113	135	139	85	83	102	124
2017	154	145	179	149	119	127	141	143	98	90	117	135
2018	166	161	189	161	127	134	149	156	100	94	120	141
2019	185	170	207	157	128	137	151	156	101	102	122	153

Un modelo estacional puro

	May	Jun	Jul	Ago
2016			y_{t-36-1}	$y_{t-36} = 138.7$
2017			y_{t-24-1}	$y_{t-24} = 142.8$
2018			y_{t-12-1}	$y_{t-12} = 156.0$
2019	$y_{t-3} = 128.2$	$y_{t-2} = 137.0$	$y_{t-1} = 151.4$	$y_t = 156.4$

- ▶ En general, un modelo puramente estacional puede representarse por

$$y_t = \varphi_1 y_{t-s} + \varphi_2 y_{t-2s} + \cdots + \varphi_P y_{t-Ps} + \varepsilon_t + \vartheta_1 \varepsilon_{t-s} + \vartheta_2 \varepsilon_{t-2s} + \cdots + \vartheta_Q \varepsilon_{t-Qs}$$

- ▶ o bien, en términos de polinomios de rezagos

$$\underbrace{(1 - \varphi_1 L^s - \cdots - \varphi_P L^{Ps})}_{\Phi(L^s)} y_t = \underbrace{(1 + \vartheta_1 L^s + \cdots + \vartheta_Q L^{Qs})}_{\Theta(L^s)} \varepsilon_t$$

Diferenciación estacional

- ▶ La diferenciación estacional se define como la diferencia entre un valor y_t y un valor rezagado un múltiplo de s períodos.

$$\Delta_s y_t = (1 - L^s) y_t = y_t - y_{t-s}$$

- ▶ Por ejemplo:

trimestral $\Delta_4 y_t = (1 - L^4) y_t = y_t - y_{t-4}$

mensual $\Delta_{12} y_t = (1 - L^{12}) y_t = y_t - y_{t-12}$

- ▶ Nótese que siguiendo esta notación $\Delta y_t = \Delta_1 y_t$. Pero

$$\Delta_2 y_t \neq \Delta^2 y_t$$



$$(1 - L^2) y_t \neq (1 - L)^2 y_t$$

$$(1 - L^2) y_t \neq (1 - 2L + L^2) y_t$$

$$y_t - y_{t-2} \neq y_t - 2y_{t-1} + y_{t-2}$$

Un modelo ARIMA para observaciones de la misma estación

- ▶ En general, podríamos plantear un modelo ARIMA para observaciones de una sola estación (mes, trimestre)

$$\Phi(L^s)\Delta_s^D y_t = \Theta(L^s)\varepsilon_t$$

- ▶ Es usualmente razonable asumir que esta misma relación la cumplen las observaciones de la estación anterior

$$\Phi(L^s)\Delta_s^D y_{t-1} = \Theta(L^s)\varepsilon_{t-1}$$

- ▶ En general, los errores de estas relaciones $\varepsilon_t, \varepsilon_{t-1}$ podrían estar correlacionadas, por lo que en principio podemos plantear el modelo ARIMA:

$$\Phi(L)\Delta^d \varepsilon_t = \Theta(L)\varepsilon_t$$

donde ε_t es ruido blanco.

- ▶ Asumiendo que el proceso ARIMA de cada estación es invertible obtenemos

$$\begin{aligned}\Phi(L^s)\Delta_s^D y_t &= \Theta(L^s)\varepsilon_t \\ \Theta^{-1}(L^s)\Phi(L^s)\Delta_s^D y_t &= \varepsilon_t \\ \Phi(L)\Delta^d\Theta^{-1}(L^s)\Phi(L^s)\Delta_s^D y_t &= \Phi(L)\Delta^d\varepsilon_t \\ &= \Theta(L)\varepsilon_t \\ \Phi(L)\Phi(L^s)\Delta^d\Delta_s^D y_t &= \Theta(L)\Theta(L^s)\varepsilon_t \\ &\quad \underset{p}{} \quad \underset{P}{} \quad \underset{d}{} \quad \underset{D}{} \quad \underset{q}{} \quad \underset{Q}{}\end{aligned}$$

- ▶ Este modelo se denomina **SARIMA** $(p, d, q) \times (P, D, Q)_s$.

- ▶ SARIMA(0, 0, 0) \times (0, 1, 0)₄ = SARIMA(0, 1, 0)₄ =

$$\Delta_4 y_t = \epsilon_t$$

$$(1 - L^4)y_t = \epsilon_t$$

$$y_t - y_{t-4} = \epsilon_t$$

- ▶ SARIMA(2, 0, 0)₄ = SAR(2)₄

$$(1 - \varphi_1 L^4 - \varphi_2 L^8) y_t = \epsilon_t$$

$$y_t - \varphi_1 y_{t-4} - \varphi_2 y_{t-8} = \epsilon_t$$

- ▶ SARIMA(0, 0, 3)₄ = SMA(3)₄

$$y_t = (1 + \vartheta_1 L^4 + \vartheta_2 L^8 + \vartheta_3 L^{12})\epsilon_t$$

$$= \epsilon_t + \vartheta_1 \epsilon_{t-4} + \vartheta_2 \epsilon_{t-8} + \vartheta_3 \epsilon_{t-12}$$

- ▶ SARIMA(0, 1, 1) × (0, 1, 1)₄

$$\Delta\Delta_4 y_t = (1 + \theta L)(1 + \vartheta L^4)\epsilon_t$$

$$(1 - L)(1 - L^4)y_t = (1 + \theta L)(1 + \vartheta L^4)\epsilon_t$$

$$(1 - L - L^4 + L^5)y_t = (1 + \theta L + \vartheta L^4 + \theta\vartheta L^5)\epsilon_t$$

$$y_t - y_{t-1} - y_{t-4} + y_{t-5} = \epsilon_t + \theta\epsilon_{t-1} + \vartheta\epsilon_{t-4} + \theta\vartheta\epsilon_{t-5}$$

- ▶ SARIMA(1, 0, 0) × (0, 1, 1)₄

$$(1 - \phi L)\Delta_4 y_t = (1 + \vartheta L^4)\epsilon_t$$

$$(1 - \phi L)(1 - L^4)y_t = (1 + \vartheta L^4)\epsilon_t$$

$$(1 - \phi L - L^4 + \phi L^5)y_t = (1 + \vartheta L^4)\epsilon_t$$

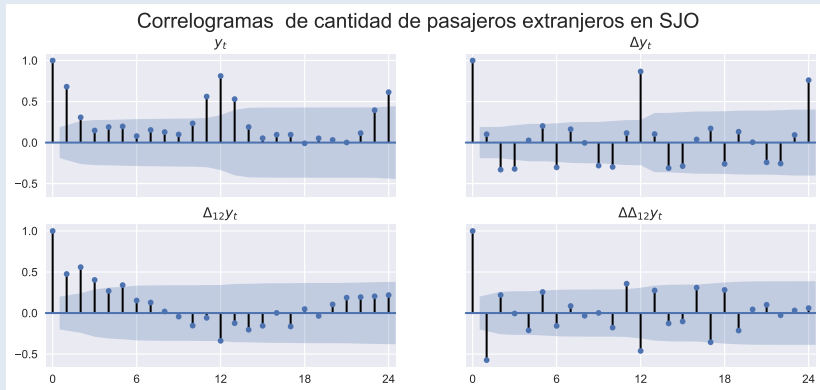
$$y_t - \phi y_{t-1} - y_{t-4} + \phi y_{t-5} = \epsilon_t + \vartheta\epsilon_{t-4}$$

- ▶ La identificación de modelos estacionales es más difícil que la identificación de modelos no estacionales por dos razones:
 1. Muchas series estacionales exhiben también patrones no estacionales y por lo tanto las FAC y las FACP estimadas contienen ambos patrones.
 2. No hay muchas correlaciones en valores k múltiplos de s . Por ejemplo, en una serie mensual podríamos contar únicamente con $k = 12$, $k = 24$ y $k = 36$
- ▶ En la práctica, cuando se tienen dudas, se utilizan herramientas que automatizan esta selección de parámetros a partir de criterios de selección.

Ejemplo 2:

Estimación de un modelo SARIMA

- ▶ El correlograma de la serie de movimientos de pasajeros extranjeros en SJO sugiere que la serie tiene un componente estacional.
- ▶ Al usar una herramienta de selección de modelos (`statsmodels.tsa.x13.x13_arima_select_order` en Python) se selecciona el modelo $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$.



Dep. Variable:	0	No. Observations:	108
Model:	SARIMAX(0, 1, 1)x(0, 1, 1, 12)	Log Likelihood	194.101
Date:	Sun, 24 May 2020	AIC	-382.202
Time:	22:03:25	BIC	-374.540
Sample:	01-31-2011 - 12-31-2019	HQIC	-379.106

	coef	std err	z	P > z	[0.025	0.975]
ma.L1	-0.6110	0.087	-7.030	0.000	-0.781	-0.441
ma.S.L12	-0.7646	0.117	-6.523	0.000	-0.994	-0.535
sigma2	0.0009	0.000	6.348	0.000	0.001	0.001
Ljung-Box (Q):		47.51		Jarque-Bera (JB):	0.23	
Prob(Q):		0.19		Prob(JB):	0.89	
Heteroskedasticity (H):		1.56		Skew:	0.09	
Prob(H) (two-sided):		0.21		Kurtosis:	3.16	

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

► El modelo estimado es

$$(1 - L)(1 - L^1 2)y_t = (1 - 0.611 L)(1 - 0.765 L^1 2)\epsilon_t$$

$$y_t - y_{t-1} - y_{t-12} + y_{t-13} = \epsilon_t - 0.611\epsilon_{t-1} - 0.765\epsilon_{t-12} + 0.467\epsilon_{t-13}$$

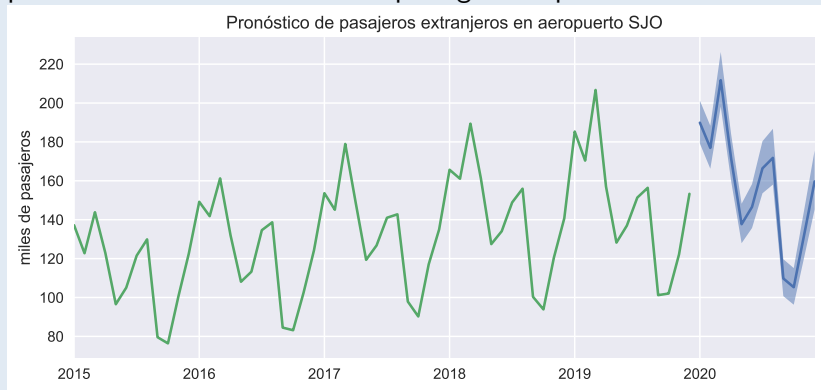
Los resultados de la tabla anterior muestran que:

- ▶ Los coeficiente estimados son significativos.
- ▶ Los residuos del modelo...
 - ▶ parecen no estar autocorrelacionados: el valor p del estadístico Q de Lung-Box es 0.19.
 - ▶ parecen ser normales: la asimetría es 0.09, la kurtosis es 3.16, y el valor p de la prueba de Jarque-Bera es 0.89.
- ▶ Por ello, podemos pensar que el modelo estimado es una buena representación de los datos.

La expresión

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \epsilon_t - 0.611\epsilon_{t-1} - 0.765\epsilon_{t-4} + 0.467\epsilon_{t-5}$$

puede utilizarse recursivamente para generar pronósticos.



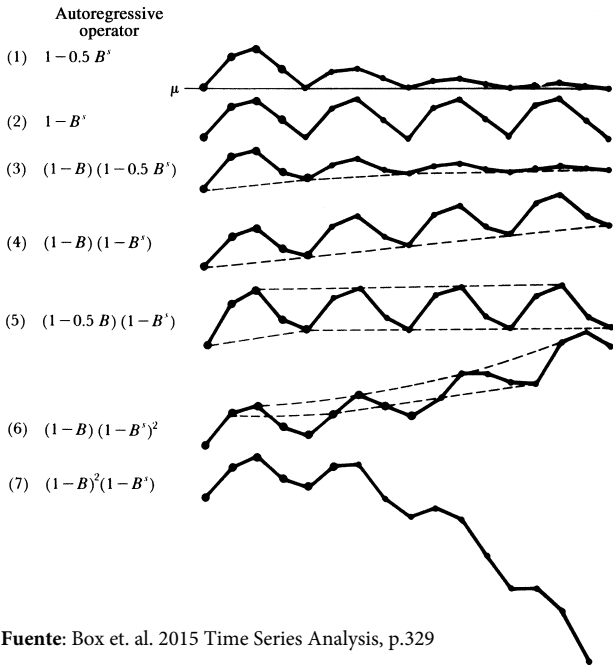
- ▶ Si expandimos los polinomios del proceso SARIMA(p, d, q) \times (P, D, Q) $_s$

$$\underbrace{\Phi(L)}_p \underbrace{\Phi(L^s)}_P \underbrace{\Delta^d}_{d} \underbrace{\Delta_s^D}_{D} y_t = \underbrace{\Theta(L)}_q \underbrace{\Theta(L^s)}_Q \epsilon_t$$

el resultado será un polinomio de grado $p + Ps + d + D$ del lado izquierdo y uno de grado $q + Qs$ del lado derecho.

- ▶ Para horizontes de más allá de $q + Qs$ períodos, la dinámica de estos pronósticos estará gobernada únicamente por la ecuación en diferencia homogénea

$$\Phi(L)\Phi(L^s)\Delta^d\Delta_s^D\hat{y}_{t+h} = 0$$



Fuente: Box et. al. 2015 Time Series Analysis, p.329

3. Ajuste estacional

- ▶ Estos métodos surgieron a mitad del siglo XX para resolver dos problemas:
 1. Remover la tendencia de una serie, con el fin de estudiar apropiadamente su autocorrelación.
 2. Separar las variaciones estacionales de los demás componentes de las series.
- ▶ Al inicio, se utilizaban métodos determinísticos, pero luego se fueron abandonando en favor de métodos empíricos basados en promedios móviles

¿Para qué ajustar una serie por estacionalidad?

- ▶ El objetivo de los métodos de ajuste estacional es extraer una señal clara de y_t , que permita identificar con facilidad la verdadera evolución de la serie
- ▶ Si una serie de tiempo tiene mucho ruido (variabilidad) y un componente estacional fuerte, este ruido afectará a las variaciones inter-periodos o a las variaciones interanuales.
- ▶ El ajuste estacional permite hacer comparaciones de un período respecto al anterior, removiendo el ruido que introduciría la presencia de la estacionalidad.

- Tendencia (T)** Es un movimiento de larga duración que se mantiene durante todo el período de observación.
- Movimientos cíclicos (C)** Son oscilaciones alrededor de la tendencia producidos por periodos alternativos de prosperidad y depresión.
- Variación estacional (S)** Son los movimientos que se producen dentro del año y que se repiten de un año a otro.
- Movimientos irregulares (ϵ)** Son las oscilaciones erráticas o accidentales que obedecen a variadas causas. No siguen ningún patrón específico de comportamiento y por tanto son impredecibles.

Multiplicativa

$$y_t = T_t \times C_t \times S_t \times \epsilon_t$$

- ▶ Se utiliza cuando se observa que la amplitud del componente estacional varía en forma proporcional al valor de la tendencia.
- ▶ En este caso el componente estacional se expresa como un índice y se interpreta como una diferencia relativa. Lo mismo ocurre para el componente irregular.
- ▶ La tendencia siempre se mide en las mismas unidades que Y_t

Aditiva

$$y_t = T_t + C_t + S_t + \epsilon_t$$

- ▶ Se utiliza cuando el componente estacional permanece constante, aún si existen cambios en el nivel de tendencia.
- ▶ En este caso el componente estacional se expresa en las mismas unidades que Y_t , al igual que ocurre con el componente irregular.
- ▶ La tendencia siempre se mide en las mismas unidades que Y_t

Serie desestacionalizada

Se obtiene cuando a la serie original se le elimina el componente estacional:

$$y_t^{s.a.} = \frac{T_t \times C_t \times S_t \times \epsilon_t}{S_t} = T_t \times C_t \times \epsilon_t$$

Serie tendencia-ciclo

Se obtiene cuando se elimina el componente estacional y el irregular de la serie original:

$$y_t^{t.c.} = \frac{T_t \times C_t \times S_t \times \epsilon_t}{S_t \times \epsilon_t} = T_t \times C_t$$

Basados en medias móviles

- ▶ Son empíricos
- ▶ La estimación de los componentes es local a través de promedio móviles centrados de distintos tamaños
- ▶ Las revisiones están sujetas a la bondad del pronóstico.

Basados en modelos

- ▶ Proveen medidas de bondad de ajuste sobre la estimación de los componentes.
- ▶ Permite obtener pronósticos de los componentes.
- ▶ Minimiza las revisiones.
- ▶ Minimiza el riesgo de inducción de propiedades espurias.

Línea de tiempo: programas de ajuste estacional

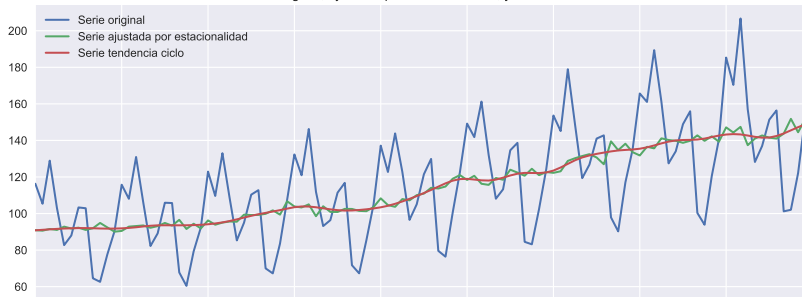
- 1967 X11, filtros ad-hoc, Julius Shiskin, US Bureau of the Census
 - 1980 X-11 ARIMA, Statistics Canada
 - 1997 X-12 ARIMA, US Bureau of Census
 - 2012 X-13 ARIMA-SEATS, US Bureau of Census [▶ x13as](#)
-
- 1992 TRAMO y SEATS (prueba), métodos estocásticos
 - 1998 TRAMO y SEATS, Banco de España
 - 2001 TSW, Banco de España
 - 2012 JDEMETRA+ Eurostat [▶ jdemetra+](#)
 - 2014 TSW+, Banco de España

Ejemplo 3:

Descomposición de la serie pasajeros extranjeros en SJO, usando X13

Descomposición multiplicativa de la serie de pasajeros extranjeros en SJO, X13as

Series original, ajustada por estacionalidad, y tendencia-ciclo



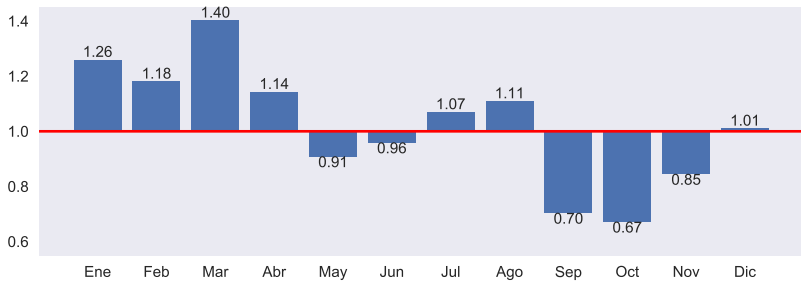
Factor estacional



Componente irregular



Factores estacionales de 2019

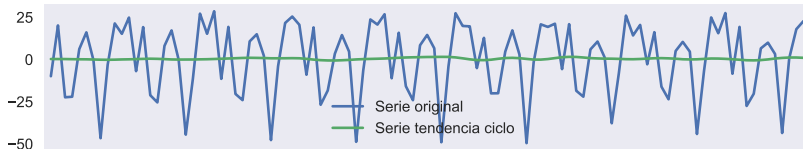


- ▶ La temporada “alta” ocurre entre enero y abril.
- ▶ En marzo, el tráfico de pasajeros es 40% mayor que el valor de tendencia-ciclo.
- ▶ La temporada “baja” es entre septiembre y noviembre.
- ▶ En octubre, el tráfico de pasajeros es 33% menor que el valor de tendencia-ciclo.

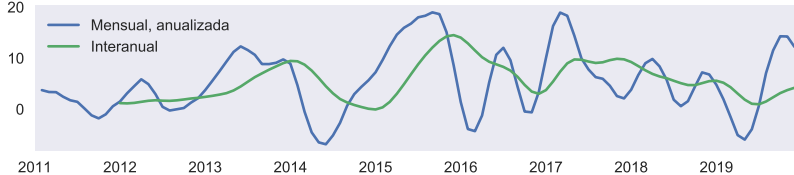
Tasa de crecimiento interanual



Tasa de crecimiento mensual



Tasa de crecimiento de la serie tendencia-ciclo



- ▶ La tasa de crecimiento interanual no tiene el ruido que presenta la tasa de crecimiento interanual de la serie original.
- ▶ Dado el fuerte componente estacional, la tasa de crecimiento mensual de la serie original es un pobre indicador del estado de la actividad.
- ▶ Al realizar cálculos con la serie tendencia ciclo encontramos:
 - ▶ la tasa de crecimiento interanual es de magnitud similar a la tasa de crecimiento mensual anualizada.
 - ▶ la tasa interanual presenta un desfase respecto a la mensual anualizada.
 - ▶ este desfase se debe a que la interanual es una media móvil de la mensual y sus últimos 11 rezagos.

4. Raíces unitarias estacionales

Diferenciación estacional y raíces unitarias

- ▶ Si para modelar una serie estacional es necesario aplicar el operador $\Delta_s = (1 - L^s)$, entonces la serie original tiene una raíz unitaria estacional: $1 - 1^s = 0$.
- ▶ Pero 1 no es la única raíz de $1 - L^s$: al ser este un polinomio de grado s , tiene exactamente s raíces (no necesariamente todas reales):

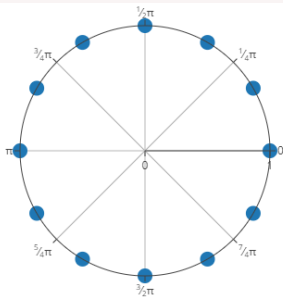
$$z_k = e^{\frac{2k\pi}{n}i}, \quad k = 0, 1, \dots, s$$

o bien, usando la fórmula de Euler

$$= \cos\left(\frac{2k\pi}{n}\right) + i \sin\left(\frac{2k\pi}{n}\right)$$

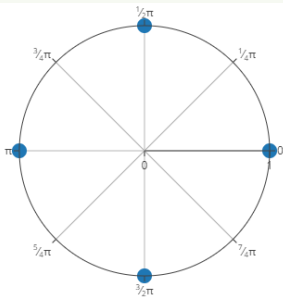
- ▶ Todas estas raíces están equidistantes en la circunferencia unitaria.

Mensuales



$$\cos\left(\frac{k\pi}{6}\right) + i \sin\left(\frac{k\pi}{6}\right)$$

Trimestrales



$$\cos\left(\frac{k\pi}{2}\right) + i \sin\left(\frac{k\pi}{2}\right)$$

- ▶ La literatura provee varias pruebas para detectar raíces unitarias.
- ▶ Una de ellas es la prueba de Dickey, Hasza y Fuller (DHF, 1984), la cual es una extensión de la prueba Dickey Fuller:

$$y_t = \rho y_{t-s} + \epsilon_t$$
$$y_t - y_{t-s} = (\rho - 1)y_{t-s} + \epsilon_t$$
$$\Delta_s y_t = \gamma y_{t-s} + \epsilon_t$$

- ▶ Al igual que en la prueba DF, la hipótesis nula es $\gamma = 0$ contra la alternativa $\gamma < 0$, aunque los valores críticos son distintos.
- ▶ Una limitación de esta prueba es que no es apropiada para diagnosticar la presencia de las otras raíces unitarias estacionales.

- ▶ Existen otras pruebas de raíces unitarias estacionales
 - ▶ La prueba HEGY de Hylleberg, Engle, Granger, y Yoo (1990)
 - ▶ La prueba de Kunst (1997)
 - ▶ La prueba OCSB, de Osborn, Chui, Smith y Birchenhall (1988)
 - ▶ La prueba CH, de Canova y Hansen (1995)
- ▶ Sin embargo, no estudiaremos estas pruebas en este curso.



Box, George E. P. y col. (2016). *Time Series Analysis*. 5ª ed. Wiley. ISBN: 978-1-118-67502-1.