Modelos para series de tiempo con estacionalidad

Curso de Series Cronológicas
Licenciatura de Estadística

Silvia Rodríguez Collazo

Facultad de Ciencias Económicas y de Administración

La estacionalidad es una característica dominante en muchas de las series económicas de frecuencia trimestral, mensual, diaria, horaria por mencionar algunas frecuencias. En este capítulo adoptaremos la perspectiva donde la modelización de los componentes estacionales y no estacionales están estrechamente vinculados.

Diremos que una serie es estacional cuando su valor esperado varía con una pauta cíclica. En concreto si $E(z_t) = E(z_{t+s})$ diremos que la serie tiene una estacionalidad de período s. Intuitivamente, podemos decir que una serie mensual tiene estacionalidad si los valores esperados en distintos meses del año son distintos, pero el valor esperado en el mismo mes en distintos años, es el mismo o similar.

Definición de estacionalidad

Una definición generalmente aceptada es la de Hylleberg (1992): *Estacionalidad son los movimientos sistemáticos, aunque no necesariamente regulares, que se dan dentro del año causados por cambios en el clima, el calendario, o por los tiempos en la toma de decisiones que los agentes económicos realizan ya sea directa o indirectamente y que se vinculan a la producción o el consumo. Estas decisiones están influenciadas por expectativas y preferencias de los agentes así como las técnicas de producción disponibles en la economía*

Esta definición implica que la estacionalidad no necesariamente es fija a través del tiempo, aunque de hecho el calendario no cambie. Por ejemplo el impacto de las fiestas de fin de año, como Navidad, fin de año, año nuevo, pueden modificarse a lo largo del tiempo aunque las fechas de esos eventos no se modifique.

El **período estacional**, s, se define como el número de observaciones que forman el ciclo estacional. Por ejemplo, $s = 12$ para series mensuales, $s = 4$ para series trimestrales, etc.

Supondremos que el valor de s es fijo en la serie. Este supuesto puede no cumplirse, por ejemplo si tenemos datos diarios y el período estacional es la longitud del mes s será aproximadamente 30, pero puede ser 31 o 28.

Una serie puede albergar **estacionalidad múltiples**, si consideramos datos diarios puede estar presente una estacionalidad semanal, con $s = 7$, otra mensual, con $s = 30$ y otra anual, con $s = 365$.

Un modelo simple para representar la estacionalidad en media considera ese patrón estacional como un efecto constante que se suma a los valores de la serie, **estacionalidad determinística**. Donde las medias por estación son constantes a través del tiempo.

Podemos escribir a la serie $S_t^{(s)}$ como suma de un componente estacional y un proceso estacionario η_t, de manera que el modelo para la serie será:

$$Z_t = S_t^{(s)} + \eta_t$$

Esta serie no presenta una única media constante, pues

$$E(z_t) = E(S_t^{(s)}) + \mu$$

donde μ es la media general del proceso η_t.

Por definición el componente estacional no toma el mismo valor en todos los períodos, la serie no tiene una única media constante.

Se pueden elaborar **distintas hipótesis respecto a la forma del patrón estacional** lo que implicará diferentes alternativas de modelización.

Consideremos que $S_t^{(s)}$ es un proceso determinista:

$$S_t^{(s)} = S_{t+ks}^{(s)} \quad k = \pm 1, \pm 2, \dots$$

Una opción para representar las medias estacionales es introducir 12 variables indicatrices, una para cada mes del año. Cuando la estacionalidad se representa de este modo se dice que el proceso presenta estacionalidad determinística.

Cuando se modeliza de este modo la estacionalidad (12 indicatrices de la estación) no es posible diferenciar la media incondicional general de las estacionales. Por tanto si se desea explicitar estas diferentes medias, se puede especificar de la siguiente forma:

$$\mu_t = \mu + \sum_{s=1}^{s=s-1} \delta_s D_s$$

Para evitar multicolinealidad exacta se incluyen solo $(s-1)$ indicatrices.
Otra forma de modelizar el patrón estacional es mediante una representación trigonométrica:

$$\mu_t = \mu + \sum_{j=1}^{j=s*} \gamma_j cos\lambda_{jt} + \gamma_j^* sen\lambda_{jt} + Z_t$$

Con $s^* =$ parte entera de $(s/2)$ y $\lambda_{jt} = 2\pi jt/s$ con $j = 1, 2, \dots\dots s^*$

Modelos para representar la estacionalidad estacionaria y no estacionaria

Hay series cuya pauta estacional evoluciona también como otras propiedades de la serie a lo largo del tiempo. En ese caso la forma de modelar la **estacionalidad** es suponer que esta evolución es **estacionaria**.

Los factores estacionales no son constantes pero siguen un proceso estacionario en torno a la media estacional:

$$S_t^{(s)} = \mu^{(s)} + \upsilon_t$$

$\mu^{(s)}$ es una constante que depende del mes y representa el efecto determinista de la estacionalidad y υ_t es un proceso estacionario de media cero.

La **estacionalidad puede ser cambiante en el tiempo**, sin evolucionar en torno a un valor medio fijo. Es este caso la estacionalidad sigue un proceso no estacionario, siendo υ_t un proceso estacionario de media cero. Y la representación sería:

$$\begin{gathered} S_t^{(s)} = S_{t-s}^{(s)} + \upsilon_t \\ S_t^{(s)} - S_{t-s}^{(s)} = \upsilon_t \\ \Delta^s S_t^{(s)} = \upsilon_t \end{gathered}$$

Como se puede observar el proceso, $S_t^{(s)}$ es no estacionario, tiene raíces unitarias, tiene s raíces de módulo 1, las llamaremos raíces unitarias en las frecuencias estacionales.$S_t^{(s)}$ es un **random walk estacional**.

Transformación estacionaria de una serie con raíces unitarias estacionales

Podemos aplicar una transformación estacionaria a la serie. En caso de que la serie contenga raíces de módulo 1 en el polinomio autoregresivo estacional, la transformación adecuada es una diferencia estacional.

$$\Delta_s = (1 - L^s)$$
$$\Delta_s Z_t = Z_t - Z_{t-s}$$

Esta transformación implica que hay s raíces de módulo uno. Se puede ver si factorizamos el polinomio $(1 - L^s)$:

$$\Delta_s = (1 - L^s) = (1 - L)(1 + L + L^2 + L^3 + ...L^{s-1})$$

La raíz $+1$ se relaciona con la frecuencia 0, el largo plazo, período infinito, mientras que las s-1 raíces remanentes, representan raíces unitarias en las frecuencias estacionales.

La aplicación de este filtro implica de hecho suponer que existencia de $s - 1$ raíces unitarias estacionales y una raíz unitaria regular, , en total s raíces. Por tanto, si aplicamos :

$$\Delta_s Z_t = \Delta_s S_t^{(s)} + \Delta_s \eta_t$$

la serie $\Delta_s Z_t$ sería estacionaria.

Cuando exista dependencia estacional podemos generalizar el modelo ARMA para series estacionarias incorporando además de la dependencia regular, la dependencia estacional que es la asociada a observaciones separadas por s períodos.

La dependencia regular se refiere a la dependencia entre observaciones consecutivas, mientras que la estacional a las separadas por s períodos.

La primera solución es incorporar la dependencia estacional a la regular ya estudiada, añadiendo a los operadores AR o MA en el operador L, términos del tipo L^s , para representar la dependencia entre observaciones separadas por s períodos.

El inconveniente de esta formulación es que llevará a polinomios de órdenes altos en la parte AR y MA.
Por ejemplo, con datos mensuales, $s = 12$, si un mes está relacionado con el mismo mes en tres años anteriores necesitamos un AR o MA de orden 36 para representar esta dependencia.
Box y Jenkins propusieron un enfoque más simple, es modelar de forma separada la dependencia regular y la estacional, y construir el modelo incorporando ambas.

Se obtiene así el modelo ARIMA estacional multiplicativo, SARIMA $(p,d,q)(P,D,Q)s$ que se especifica de la siguiente forma:

$$\Phi_p(L^s)\varphi_{p1}(L)\Delta_s^D\Delta^d Z_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

con:

Polinomio autoreg. estacional:
$\Phi_p(L^s) = (1 - \Phi_1 L^s - \Phi_2 L^{2s} - ... - \Phi_p L^{sp})$

Polinomio autoreg. regular:
$\phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 ... - \phi_p L^p)$

Polinomio de medias móviles estacional:
$\Theta_Q(L^s) = (1 - \Theta_1 L^2 - \Theta_2 L^{2s} - ... - \Theta_Q L^{sQ})$

Polinomio de medias móviles regular:
$\theta_q(L) = (1 - \theta_1 L - \theta_2 L^2 .... - \theta_Q L^q)$

Cuando la serie es no estacionaria, esto es, tiene raíces unitarias regulares y estacionales, las variaciones anuales y entre estaciones adyacentes cobran importancia.

En los modelos donde se aplica la diferencia regular y la estacional, desaparecen los componentes determinísticos (como las indicatrices estacionales o una tendencia determinística).

El filtro $(1 - L^s)$ captura la tendencia para los valores de la serie en una estación particular, que está altamente correlacionado con el valor en igual estación un año antes y el filtro $(1 - L)$ captura la no estacionariedad regular, el componente de tendencia estocástico.

Raíces unitarias estacionales

Mencionamos el proceso random walk estacional. Veámoslo para el caso de las series trimestrales.

$$Y_t = Y_{t-4} + \varepsilon_t$$

siendo ε_t un proceso Ruido Blanco.
Pero este proceso no es un proceso integrado $I(1)$ convencional, pues contiene 4 raíces unitarias.

$$(1-L^4) = (1-L)(1+L+L^2+L^3) =$$
$$(1-L)(1+L)(1-L^2) = (1-L)(1+L)(1+iL)(1-iL)$$

$(1-L)$ corresponde a la raíz $+1$, $(1+L)$ corresponde a la raíz -1 y $(1+iL)(1-iL)$ corresponden a las raíces imaginarias de módulo 1.

La primera remueve la raíz regular, la segunda remueve la raíz -1, que es una raíz asociada a una frecuencia estacional, al igual que las raíces $+$i y -i.
Es importante tener presente que las raíces $+$i y -i no pueden ser separadas, por tanto la transformación $(1+iL)(1-iL)$ se aplica en forma conjunta, no se separa.

La transformación estacionaria adecuada para el random walk estacional es la diferencia estacional, $(1-L^4)$, si sólo se aplica la regular, sólo se filtra la raíz $+1$.
Por ello algunos autores hablan de procesos $SI(1)$

En procesos mensuales $\Delta_{12} Y_t = (1 - L^{12}) Y_t = Y_t - Y_{t-12}$.

Este operador de diferencias estacionales está vinculado a la estacionalidad no estacionaria. Para el caso de las series mensuales:

$$\Delta_{12} = \Delta(1 + L + L^2 + ... + L^{11}) = \Delta U_{11}(L)$$
$$(1 + L^{12}) = (1 - L)(1 + L)(1 + L^2 + ... + L^{10}) =$$
$$(1 - L)(1 + L)(1 - \sqrt{3}L + L^2)(1 - L + L^2)(1 + L^2)(1 + L + L^2)(1 + \sqrt{3}L + L^2)$$

La diferencia estacional es la diferencia regular aplicada a un polinomio suma $U_{11}(L)$ que afecta a los últimos 12 valores observados en cada momento.

El papel de esta media de 12 meses provocada por el operador $U_{11}(L)$ es el de eliminar la oscilación estacional en el nivel de la serie, ya que promedia los valores a lo largo de un ciclo estacional completo.

$$Y_t = U_{11}(L) Y_t = (Y_t + Y_{t-1} + ... + Y_{t-11})$$

$$\Delta_{12} Y_t = \Delta U_{11}(L) Y_t$$

Raíces unitarias estacionales mensuales II

El operador Δ_{12} es lo mismo que el operador Δ ampliado con $U_{11}(L)$, la diferencia regular se aplica a $U_{11}(L)Y_t$ que ha suavizado posibles oscilaciones estacionales que pudiera presentar el nivel de la serie.

$$\Delta_s = (1-L)(1+L+L^2+...+L^{s-1}) = (1-L)U_{s-1}(L)$$

Este polinomio tiene S raíces de módulo 1. La raíz de $(1-L)$ es unitaria positiva y las $S-1$ raíes de $U_{s-1}(L)$ se componen de $(S-2)/2$ pares de complejas conjugadas y una real negativa, si S es par.

En cuanto a las raíces unitarias de $(1-L^s)$, la real positiva está relacionada con el nivel no estacional de la serie y las otras $(S-1)$ con las oscilaciones estacionales.

La **función de autocorrelación simple** de este proceso es una mezcla de las funciones de autocorrelación correspondientes a las partes regular y estacional.

Puede demostrarse (Peña, 1984) que si llamamos r_i a los coeficientes de autocorrelación del proceso ARMA (p,q) regular y R_{si} a los coeficientes de autocorrelación en los retardos $s, 2_s, 3_s, \ldots$ del proceso ARMA (P,Q) estacional, y ρ_j a los coeficientes de autocorrelación del proceso completo, se verifica:

$$\rho_j = \frac{r_j + \sum_{i=1}^{\infty} R_{si}(r_{si+j} + r_{si-j})}{1 + 2\sum_{i=i}^{\infty} r_{si} R_{si}}$$

Si suponemos $s = 12$ y admitimos que $r_i \cong 0$ para retardos altos (por ejemplo, para $j \geq 8$):

- En los retardos bajos $(j = 1, \ldots, 6)$ se observará únicamente la parte regular, es decir: $\rho_j \simeq r_j \qquad j = 1, 2, 3, \ldots.$
- En los retardos estacionales se observará básicamente la parte estacional básica: $\rho_{12i} \simeq R_{12i}(r_{24i} + r_o) + R_{24i}(r_{36i} + r_{12i})$ Suponiendo que $\gamma_{12i} \cong 0$ para $i >= 1$ y $r_0 = 1$, entonces $\rho_{12i} \simeq R_{12i} \qquad i = 1, 2, \ldots\ldots$
- Alrededor de los retardos estacionales observaremos la interacción entre la parte regular y estacional.

Ejemplo de la modelización de una serie con estacionalidad

A continuación se muestra un ejemplo clásico para el análisis de series temporales con estacionalidad.

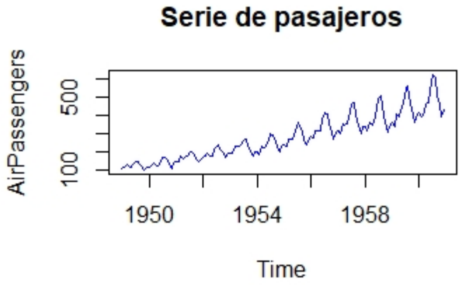

Figura: Serie de Pasajeros

Observar las características de la serie.

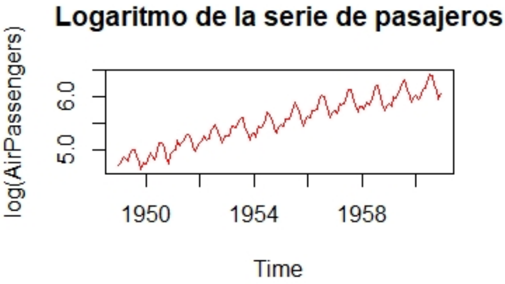

Figura: Logaritmos de la serie de Pasajeros

Al aplicar la transformación logarítmica, que se modificó?

Análisis exploratorio de una serie con estacionalidad I

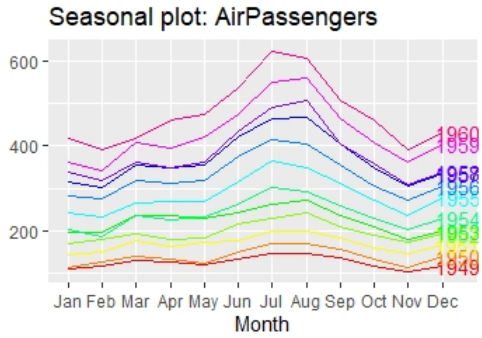

Figura: Estacionalidad Serie de Pasajeros

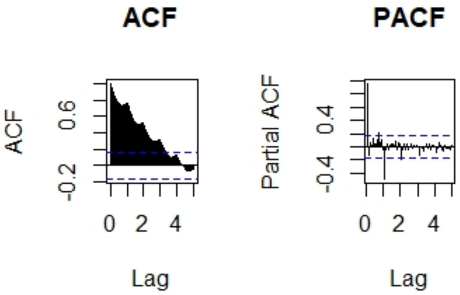

Figura: ACF y PACF logaritmos de Serie de Pasajeros

Efectos especiales, efecto calendario

El componente Efecto Calendario engloba todos aquellos efectos determinísticos producto de la composición del calendario de un país. Este efecto influye particularmente en series mensuales que resultan del agregado diario de los datos de la variable considerada en cada caso.

- Los meses no tienen el mismo número de días
- Un mismo mes a medida que transcurren los años presenta cambios en el número específico de días de la semana: cambian por ejemplo la cantidad de lunes que hay en un mismo mes.
- Hay feriados fijos, que se conmemoran en ese día y feriados móviles, que se conmemoran otro día de a semana de ese mes .

Efectos especiales, efecto calendario II

Otros aspectos a incluir en la modelización de estos efectos implica considerar por ejemplo:

- en series mensuales por ejemplo, el número total de días hábiles (o no) del mes
- el largo del mes
- considerar la existencia de años bisiestos
- el número y composición de días de la Semana (Lunes, Martes, ..,Viernes)
- el efecto de los feriados móviles como Turismo y Carnaval.

A su vez es importante considerar si el evento tiene efectos dinámicos que abarcan otros períodos de tiempo posteriores o anteriores a la ocurrencia del suceso a modelizar. Esta dinámica puede tener una forma específica para cada variable que habrá que explorar.

Muchas series pueden presentar anomalías o quiebres en su comportamiento estacional. Es necesario indorporar estos eventos en el modelo.
Ejemplo: cierre de los puentes entre Argentina y Uruguay en el período 2008-2010 Ejemplo: Pandemia , cierre de fronteras Afecta variables como movimiento turístico, especialmente en sus picos estacionales.

Se podría incorporar un AO en cada mes o trimestre que se quiere intervenir, pero el modelo no sería parsimonioso.

$$\eta(L) = \frac{1-\alpha L}{1-L^s} \tag{1}$$

Cuando $\alpha = 1$, $\eta(L) = \frac{1-L}{1-L^s}$

Cuando $\alpha = 0$, $\eta(L) = \frac{1}{1-L^s}$

Cuando la serie está sujeta a la presencia de este tipo de outliers, la detección de outliers (LS,TC,AO) estandar no presenta buen desempeño.

La etapa de diagnóstico se ve afectada como consecuencia de ignorar la presencia de este tipo de outliers (SLS), especialmente las correlaciones $\rho_1 2$ y el estadístico de Ljung-Box.