1 Introducción

La estacionalidad esta formada por todos los movimientos oscilatorios periódicos o pseudo-periódicos que tardan en completar un ciclo completo un año o menos de un año.Por lo tanto, la estacionalidad sólo esta presente en series temporales periodicidad superior al año, es decir, series temporales trimestrales, mensuales, semanales, diarias,... Al número de observaciones que disponemos por año en una series temporal lo llamamos número de estaciones y lo denotamos por S. En el caso de datos trimestrales $S = 4$, en el caso de datos mensuales $S = 12$.

La forma más habitual de tratar la estacionalidad es suponer que el patrón estacional es estable o fijo a lo largo del año. Mientras que si tratamos la estacionalidad desde un punto de vista estocástico asumimos que el patrón estacional puede ir variando año tras año.

2 Tratamiendo de la estacionalidad determinista

Existen dos formas alternativas y equivalentes de tratar la estacionalidad determinista, el uso de variables ficticias estacionales o la representación triginométrica de la estacionalidad determnista. Tal y como veremos ambas formas son equivalentes, pero en el caso de las variables ficticias estacionales prestamos atención al dominio temporal, mientras que en enfoque trigonométrico prestamos más atención al dominio temporal.

La representación a través de variables ficticias (dummies) estacionales, se puede escribir:

$$\begin{aligned} x_t &= \sum_{s=1}^{S} \delta_s D_t^s + z_t \quad s = 1, 2, \ldots, S \qquad (1) \\ D_t^k &= \begin{cases} 1 & si \quad k = 1 + int\left[(t-1) \bmod S\right] \\ 0 & si \quad k \neq 1 + int\left[(t-1) \bmod S\right] \end{cases} \end{aligned}$$

donde $a \,\mathsf{mod}\, b$ es el operador modulo que devuelve el resto de la division de a entre b. Por lo tanto $1 + int\left[(t-1) \,\mathsf{mod}\, S\right]$ permite calcular para cada observación t a que estacionión pertenece. Finalmente z_t recoge el comportamiento no determinista de x_t. Es decir, el comportamiento estocástico. Una especficación alternativa a (1) cuando se incluye un término independiente (o constante) comun a todas las estaciones, es la siguiente:

$$\begin{aligned} x_t &= \mu + \sum_{s=1}^{S} m_s D_t^s + z_t \quad s = 1, 2, \ldots, S. \\ t &= 1, 2, \ldots, T \end{aligned} \tag{2}$$

Si asumimos que la parte estocástica z_t tiene valor esperado cero $E\left[z_t\right] = 0$, se puede ver que en el caso de (1) δ_s recoge el valor esperado x_t en la estación s. Mientas que para (2) el valor esperado de x_t en la estación s viene dado por $\mu + m_s$ luego:

$$\begin{aligned} E\left[x_t\right] &= \delta_s = \mu + m_s \\ m_s &= \mu - \delta_s \end{aligned}$$

Luego μ recoge el valor medio común a todas las estaciones y m_s el efecto diferencial de la estación s. Para no caer en la trampa de la ficticias es necesario imponer la restricción $\sum_{s=1}^{S} m_s = 0$. O eliminar una de las variables ficticas, entonces:

$$x_t = \mu + \sum_{s=2}^{S} m_s D_t^s + z_t.$$

μ recoge el valor esperado de la estación no incluida y m_s recoge el efecto diferencial para estaciones $s = 2, \ldots, S$.

La representación triginométrica equivalente a (1) es:

$$\begin{aligned} x_t &= \mu + \sum_{k=1}^{\left[\frac{S}{2}\right]} \left[a_k \cos\left(\frac{2\pi k t}{S}\right) + b_k \sin\left(\frac{2\pi k t}{S}\right)\right] + z_t \qquad (3) \\ t &= 1, 2, \ldots, T. \end{aligned}$$

A partir de (1) y (3) y teniendo en cuenta que la funciones coseno y seno son periódicas, que existe la siguiente relación entre los coeficientes de ambas especificaciones con S par:

$$\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \vdots \\ \delta_S \end{bmatrix} = \mathbf{T} \times \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_{S/2} \end{bmatrix}$$

donde $\mathbf{T}$ es una matriz de order $S \times S$ con elementos genéricos en la fija iesima:

$$\begin{bmatrix} 1 & \cos\left(\frac{2\pi t}{S}\right) & \sin\left(\frac{2\pi t}{S}\right) & \dots \\ \cos\left(\frac{2\pi t[(S/2)-1]}{S}\right) & \sin\left(\frac{2\pi t[(S/2)-1]}{S}\right) & \cos(\pi t) \end{bmatrix}.$$

En el caso de S impar:

$$\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \vdots \\ \delta_{S-1} \\ \delta_S \end{bmatrix} = \mathbf{T} \times \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_{(S-1)/2} \\ \beta_{(S-1)/2} \end{bmatrix}$$

y los elementos genéricos en la fila t-esima de $\mathbf{T}$ serán :

$$\begin{bmatrix} 1 & \cos\left(\frac{2\pi t}{S}\right) & \sin\left(\frac{2\pi t}{S}\right) & \ldots \\ \cos\left(\frac{2\pi t[(S-1)/2]}{S}\right) & \sin\left(\frac{2\pi t[(S-1)/2]}{S}\right) \end{bmatrix}.$$

Las columnas de $\mathbf{T}$ están muy relacionas con las transformaciones que se utilizan en el contraste HEGY de raíces unitarias estacionales.

3 Estacionalidad estocástica

En la metodología Box-Jenkins, la estacionalidad está asociada a procesos SAR(P):

$$\begin{aligned} \Phi_P\left(L^S\right)x_t &= \varepsilon_t \\ \Phi_P\left(L^S\right) &= \left(1-\Phi_1 L^S - \cdots - \Phi_P L^{PS}\right) \end{aligned}$$

SMA(Q):

$$\begin{aligned} x_t &= \Theta_Q\left(L^S\right)\varepsilon_t \\ \Theta_Q\left(L^S\right) &= \left(1 - \Theta_1 L^S - \cdots - \Theta_Q L^{QS}\right) \end{aligned}$$

SARMA(P,Q):

$$\Phi_P\left(L^S\right)x_t = \Theta_Q\left(L^S\right)\varepsilon_t$$

y SARIMA(P,D,Q):

$$\begin{aligned} \Phi_P\left(L^S\right)\Delta_S^D x_t &= \Theta_Q\left(L^S\right)\varepsilon_t \\ \Delta_S^D &= \left(1 - L^S\right)^D. \end{aligned}$$

Los Procesos estacionales habitualmente son como máximo de orden 1. Se identifican de forma similar a los procesos AR, MA y ARMA, pero mirando en los retardos estaciones de la FAS y la FAP, es decir los múltiplos de S. Es decir S, $2S$ y $3S$. Es importante tener presentre que por ejemplo paar un SAR(1) tendremos:

$$\begin{aligned}
\Phi_1\left(L^S\right) &= \left(1-\Phi_1 L^S\right)\\
\left(1-\Phi_1 L^S\right) &= \left(1-\Phi_1^{1/S}L\right)\left(1+\Phi_1^{1/S}L\right)\times\\
&\quad \prod_{j=1}^{(S-1)/2}\left(1-e^{-i\omega_j}\Phi_1^{1/S}L\right)\left(1-e^{i\omega_j}\Phi_1^{1/S}L\right)\\
&= \left(1-\Phi_1^{1/S}L\right)\left(1+\Phi_1^{1/S}L\right)\times\\
&\quad \prod_{j=1}^{(S-1)/2}\left(1-2\cos\left(\omega_j\right)\Phi_1^{1/S}L+\Phi_1^{2/S}L^2\right)\\
\omega_j &= \frac{2\pi j}{S} \quad j=1,2,\ldots,(S-1)/2
\end{aligned}$$

Entonces se puede ver que tendremos el factor $\left(1-\Phi_1^{1/S}L\right)$ asociado a la frecuencia zero, el factor $\left(1+\Phi_1^{1/S}L\right)$ asociado a la frequencia π (ciclo cada dos periodos) y los factores $\left(1-2\cos\left(\omega_j\right)\Phi_1^{1/S}L+\Phi_1^{2/S}L^2\right)$ asociados a los pares de frecuencias harmónicas conjugadas $\omega_j = 2\pi j/S$ y $2\pi-\omega_j$ con $j=1,2,\ldots,(S-1)/2$ que tardan en completar un ciclo $2\pi/\omega_j = S/j$ periodos. Una situación equivalente ocurre con el operador diferencia estacional:

$$\begin{aligned}\left(1-L^S\right) &= (1-L)(1+L)\times \\ &\quad \prod_{j=1}^{(S-1)/2}\left(1-e^{-i\omega_j}L\right)\left(1-e^{i\omega_j}L\right) \\ &= (1-L)(1+L)\times \\ &\quad \prod_{j=1}^{(S-1)/2}\left(1-2\cos\left(\omega_j\right)L+L^2\right) \\ \omega_j &= \frac{2\pi j}{S} \quad j=1,2,\ldots,(S-1)/2.\end{aligned}$$

Para trabajar con series mensuales o trimestrales se utiliza el modelo ARIMA(p,d,q)×SARIMA(P,D,Q):

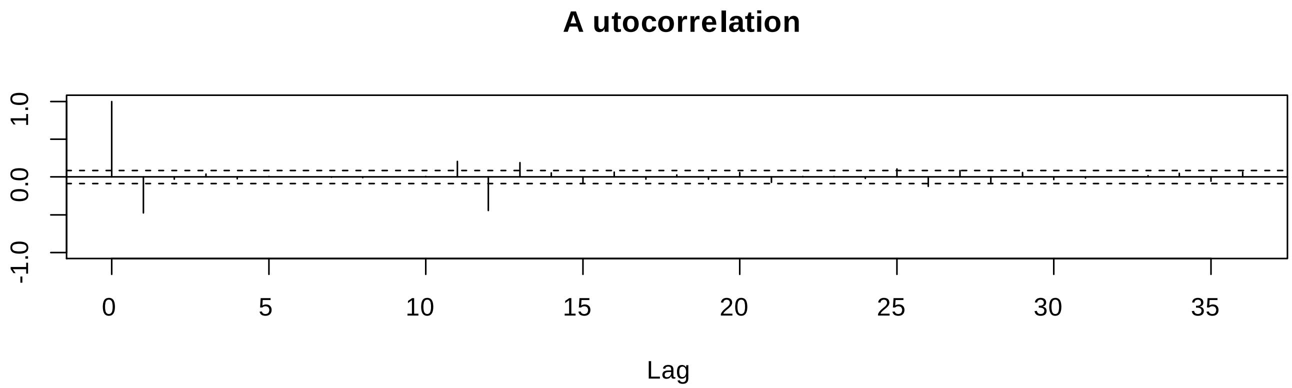
$$\Phi_P\left(L^S\right)\phi_p(L)\Delta_S^D\Delta^d x_t=\Theta_Q\left(L^S\right)\theta_q(L)\varepsilon_t.$$

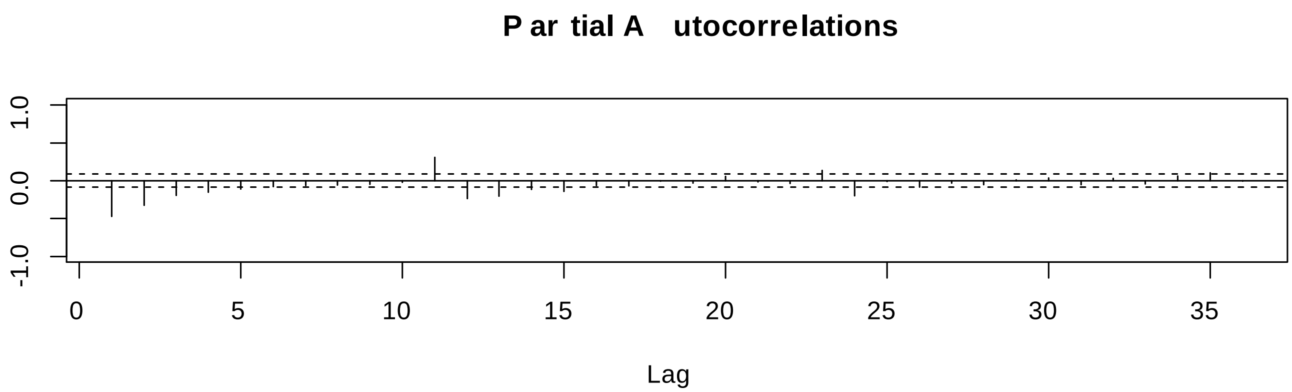
Es decir tenemos la parte regular ARIMA(p,d,q) que se identifica en los primeros retardos de las FAS y la FAP y la parte estacional que se identifica en los retardos múltiplos de S, es decir S, $2S$, $3S$, ...

En este tipo de modelos se da lo que se conoce como efectos satelites. Para ilustrarlo utilizaremos un ejemplo, supongamos que la serie sigue un proceso $ARMA(0,1) \times SARMA(0,1)$ una vez que ha sido transformado en estacionario:

$$
\begin{aligned}
z_t &= \left(1 - \Theta_1 L^S\right)(1 - \theta_1 L)\,\varepsilon_t \\
&= \left(1 - \theta_1 L - \Theta_1 L^S + \theta_1 \Theta_1 L^{S+1}\right)\varepsilon_t \\
&= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-S} + \theta_1 \Theta_1 \varepsilon_{t-(S+1)}.
\end{aligned}
$$

Se puede apreciar que aparece el término $\theta_1 \Theta_1 \varepsilon_{t-(S+1)}$ que se reflejará en los coeficientes de la FAS y la FAP asociados los retardos próximos a los multiplos de S, pero esto no significa que tengamos procesos adicionales a un MA(1) y un SMA(1):





4 Análisis de intervención y Outliers

Las series temporales pueden verse afectadas por susceso puntuales y/o permanentes conocidos o no, que pueden afectar su evolución con mayor o menor intensidad.

- Hablaremos de análisis de intervención cuando los sucesos o hechos que afectan a las series son conocidos. Nos encontrariamos ante situaciones tales como una huelga, un cambio legal, un accidente, un cambio de festividad, cambio de base, cambio de definición en una variable, etc...

- Cuando desconocemos los hechos que afectan a la variable son desconocidos, hablamos de valores atípicos. Nos encontrariamos en situaciones tales como errores de media, errores de transcripción o cualquiera de las situaciones citadas para el caso del análisis de intervención pero si la causa es desconocida.

En todas las situaciones anteriores utilizaremos variables ficticias para poder tratar de forma satisfactoria estos efectos en en el tratamiento de series temporales.

4.1 Análisis de intervención.

La principal característica del análisis de intervención es que el analista conoce la existencia de un hecho o fenómeno que afecta a la evolución de la serie temporal. Además tambíen se conoce la fecha de ocurrencia del hecho o fenómeno.

Los dos principales herramientas que se utilizarán para el análisis de intervención son las variables ficticias tipo impulso y escalón. La variable ficticia tipo impulso de define de la siguiente manera:

$$I_t^{(h)} = \begin{cases} 0 & t \neq h \\ 1 & t = h \end{cases}$$

Las variables ficticias tipo impulso son útiles para modelizar fenómenos que tienen un efecto en un sólo período. Supongamos un modelo ARIMA:

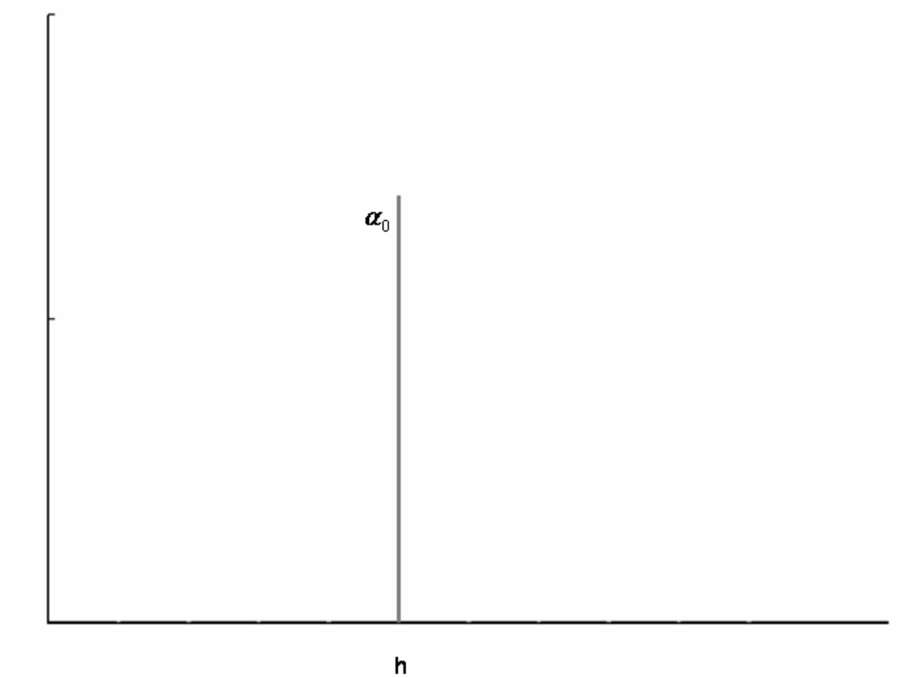
$$
\begin{aligned}
\Phi_P(L)\,\phi_p(L)\,\Delta_S^D\Delta^d x_t &= \Theta_Q(L)\,\theta_q(L)\,\varepsilon_t \\
\psi(L) &= \frac{\Theta_Q(L)\,\theta_q(L)}{\Phi_P(L)\,\phi_p(L)\,\Delta_S^D\Delta^d} \\
x_t &= \psi(L)\,\varepsilon_t.
\end{aligned}
$$

Y ahora supongamos que la serie se ve afectada por un evento que sucede en el período h resultando en que la variable se ve modificada como:

$$
\begin{aligned}
y_h &= \alpha_0 + x_h \\
y_t &= I_t^{(h)}\alpha_0 + \psi(L)\,\varepsilon_t.
\end{aligned}
$$

Es decir:

$$
y_t = \begin{cases} x_t & t \neq h \\ \alpha_0 + x_t & t = h \end{cases}
$$



Esta especificación es útil cuando el hecho o evento afecta a un único período. Pero existen situaciones en las que el efecto del evento puede afectar a varios períodos. Por ejemplo supongamos una huelga que afecta a una serie durante varios períodos, pero a cada punto de ellos con un efecto diferente es decir:

$$
\begin{aligned}
y_t &= \begin{cases} x_t \quad t < h \\ \alpha_j + x_{h+j} \quad j = 0, 1, \ldots, m \\ x_{h+j} \quad j > m \end{cases} \\
con &: \\
x_t &= \psi(L)\,\varepsilon_t \\
entonces &: \\
y_t &= \sum_{j=0}^{m} I_t^{(h+j)} \alpha_j + \psi(L)\,\varepsilon_t \qquad (4) \\
donde &: \\
I_t^{(h+j)} &= \begin{cases} 0 \quad t \neq h + j \\ 1 \quad t = h + j \end{cases}
\end{aligned}
$$

Teniendo en cuenta que el operador de retardos (L) funciona de la siguiente manera:

$$
LI_t^{(h)} = I_{t-1}^{(h)} = I_t^{(h+1)}.
$$

Entonces podemos expresar (4) como:

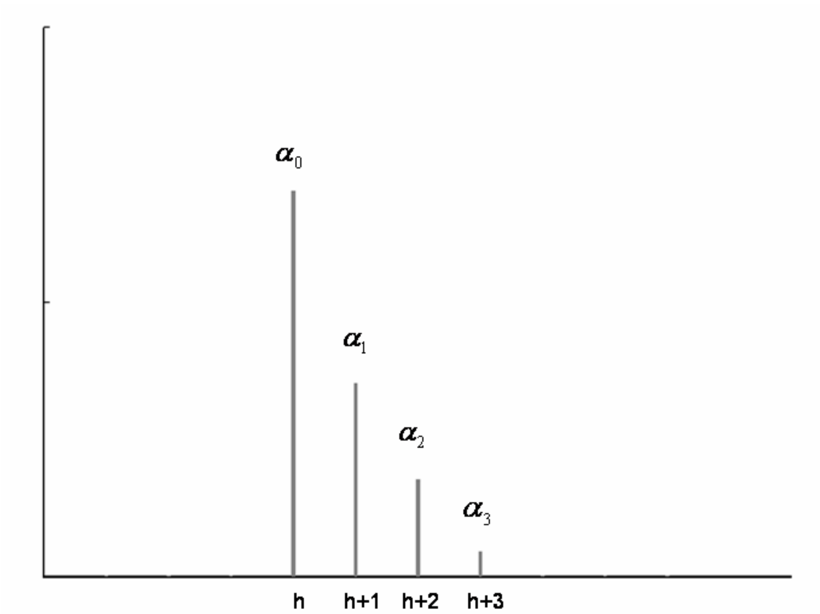
$$
\begin{aligned}
y_t &= \alpha(L) I_t^{(h)} + \psi(L) \varepsilon_t \\
con &: \\
\alpha(L) &= (\alpha_0 + \alpha_1 L + \cdots + \alpha_m L^m).
\end{aligned}
$$

Si el número de períodos afectados por el evento es largo, la representación anterior puede obligar a estimar un número elevado de parámetros Una forma sencilla de recudir el número de parámetros consiste en suponer que el efecto de del evento en el hecho decae exponencialmente a medida que nos alejamos del período h, según el esquema $\alpha_j = \alpha_0 \lambda^j$ donde $0 < \lambda < 1$:

$$
\begin{aligned}
\alpha(L) &= \left(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots\right) \\
&= \left(\alpha_0 + \alpha_0 \lambda L + \alpha_0 \lambda^2 L^2 + \cdots\right) \\
&= \frac{\alpha_0}{(1 - \lambda L)}
\end{aligned}
$$

Entonces tenemos:

$$y_t = \frac{\alpha_0}{(1-\lambda L)} I_t^{(h)} + \psi(L)\,\varepsilon_t$$



La variable ficticia tipo escalón se define de la siguiente manera:

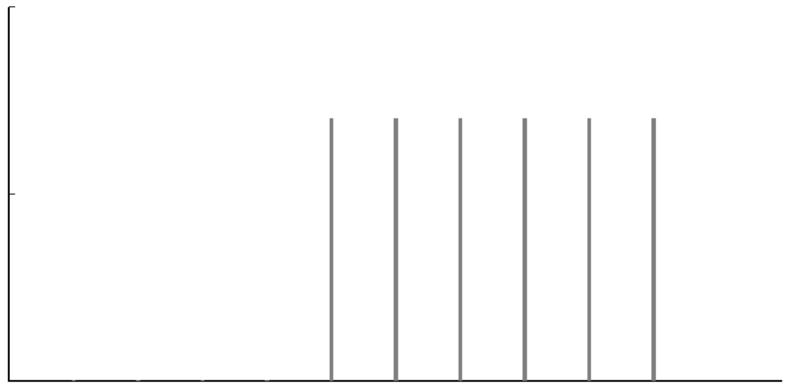
$$S_t^{(h)} = \begin{cases} 0 & t < h \\ 1 & t \geq h \end{cases}$$

Las variables ficticias escalón son útiles para modelizar fenómenos en los que hecho que afecta a la variable tiene un efecto permanente sobre la serie temporal a partir del período h. De forma equivalente a lo que tedriamos para la variable impulso partimos de un modelo ARIMA para x_t:

$$x_t = \psi(L)\,\varepsilon_t.$$

Y ahora supongamos que la serie se ve afectada por un evento que sucede en el período h de forma tal que a partir de este período la variable se ve modificada como:

$$\begin{aligned} y_t &= x_t \quad t < h \\ y_{h+j} &= \alpha_0 + x_{h+j} \quad j \geq 0 \\ y_t &= S_t^{(h)}\alpha_0 + \psi(L)\,\varepsilon_t. \end{aligned}$$



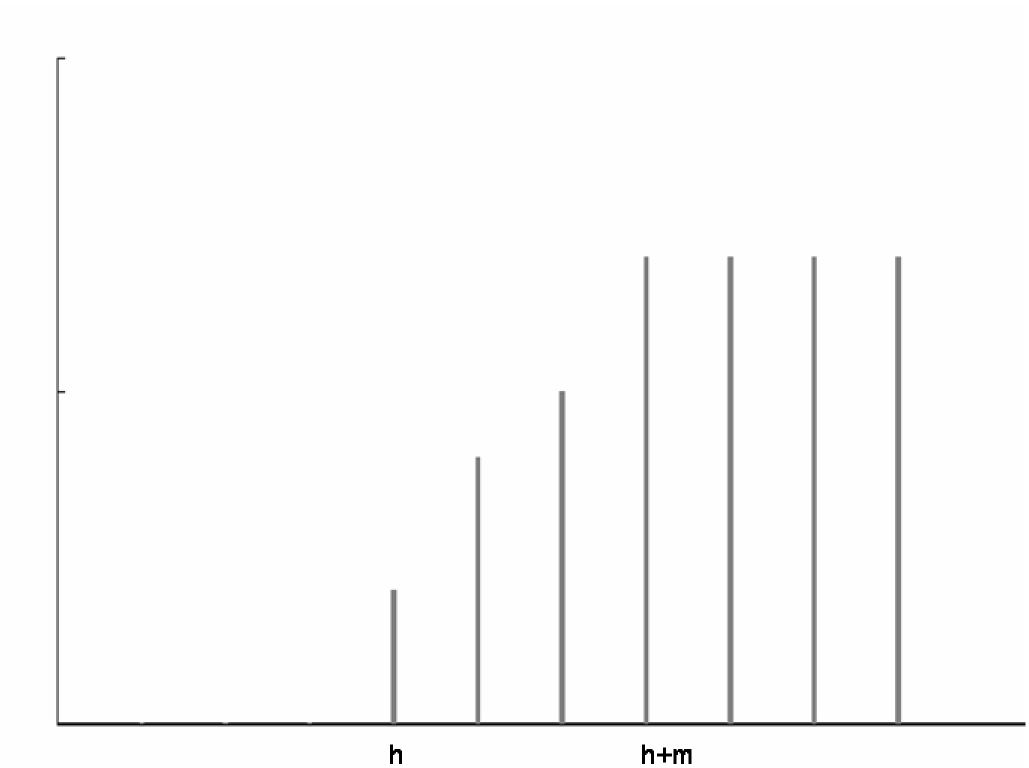
Que implica que todos los valores posteriores a h se ven afectados por una cantidad constante α_0.

Tambíen es posible que es escalón o cambio permanente tenga un efecto gradual, es decir que a partir del período h se empiezen a dar efectos acumulativos hasta un período $h + m$ en el que se consolidan:

$$
\begin{aligned}
y_t &= \alpha(L) S_t^{(h)} + \psi(L) \varepsilon_t \\
con &: \\
\alpha(L) &= (\alpha_0 + \alpha_1 L + \cdots + \alpha_m L^m) \\
L S_t^{(h)} &= S_{t-1}^{(h)} = S_t^{(h+1)}.
\end{aligned}
$$

Entonces tenemos una situación en la que:

$$
\begin{aligned}
y_t &= x_t \;\; t < h \\
y_h &= \alpha_0 + x_h \\
y_{h+1} &= \alpha_0 + \alpha_1 + x_{h+1} \\
y_{h+j} &= \sum_{i=0}^{j} \alpha_i + x_{h+j} \;\; j \geq 0 \\
y_{h+j} &= \sum_{i=0}^{m} \alpha_i + x_{h+j} \;\; j \geq m \\
con &: \\
x_t &= \psi(L)\, \varepsilon_t.
\end{aligned}
$$



- Es importante tener en cuenta que las variables impulso y escalón están relacionadas entre si por:

$$
\begin{aligned}
I_t^{(h)} &= S_t^{(h)} - S_{t-1}^{(h)} = (1 - L)\, S_t^{(h)} \\
&= \Delta S_t^{(h)}.
\end{aligned}
$$

Finalmente comentar que en el análisis de intervención se conoce la existencia de un hecho o suceso que afectará a la serie analizada y que tambíen se conoce en período (h) en que se transmite el efecto sobre la variable. Dependiendo de si el hecho tiene un efecto transitorio o permanente sobre la variable objeto de estudio utilizaremos variables ficticias de tipo impulso o escalón.

4.2 Datos atípicos

La diferencia fundamental entre el análisis de intervención y los datos atípicos es que para el análisis de intervención el hecho o suceso que afecta a la serie temporal es conocido y el período en el que transmite tambíen. Mientras que en el caso de los datos atípicos los hechos que provocan la aparición del dato atípico es desconocido y en período de tiempo en el que se manifiesta en la serie temporal tambíen.

Se han propuesto cuatro tipos de valores atípicos. Que reciben los siguientes nombres Outlier Aditivo (Aditive Outlier, AO) ,Outlier Inovacional (Innovational Outlier, IO), Cambio de Nivel (Level Shift, LS) y Cambio Temporal (Temporary Change, TC).

- **Outlier Aditivo (AO):** Diremos que ha ocurrido un outlier o atípico aditivo en una serie temporal en el instante h si el valor de las serie se genera en ese instante de manera distinta al resto. Por ejemplo si existe un error de medida que afecta únicamente al instante h y esta situación no afecta al resto de observaciones:

$$y_t = \begin{cases} x_t & t \neq h \\ \alpha_{AO} + x_t & t = h \end{cases}$$

Entonces en modelo para la serie observada es:

$$y_t = I_t^{(h)} \alpha_{AO} + \psi(L) \varepsilon_t.$$

Donde $I_t^{(h)}$ es una variable ficticia impulso$I_t^{(h)} = 0$ para $t \neq h$ y $I_t^{(h)} = 1$ para $t = h$. Este modelo es idéntico al de intervención con un impulso, en cuanto a formulación, pero se diferencian en que la variable $I_t^{(h)}$ se supone conocida en el análisis de intervención y en el caso del outlier aditivo el período h es desconocido y es necesario identificarlo.

- Outlier Innovacional (IO): Diremos que ha ocurrido un outlier innovacional en una serie temporal en el período h cuando la innovación (el ruido blanco) ε_t en ese punto o período se ve afectada por una cantidad desconocida debida a un suceso imprevisto. El modelo responde a la siguiente especificación:

$$y_t = \psi(L)\left(I_t^{(h)}\alpha_{IO} + \varepsilon_t\right).$$

El efecto final del outlier innovacional sobre la serie dependera del filtro $\psi(L)$, es decir del modelo ARIMA que siga la serie.

Existen otros tipos de outliers pero no están implementado en R.