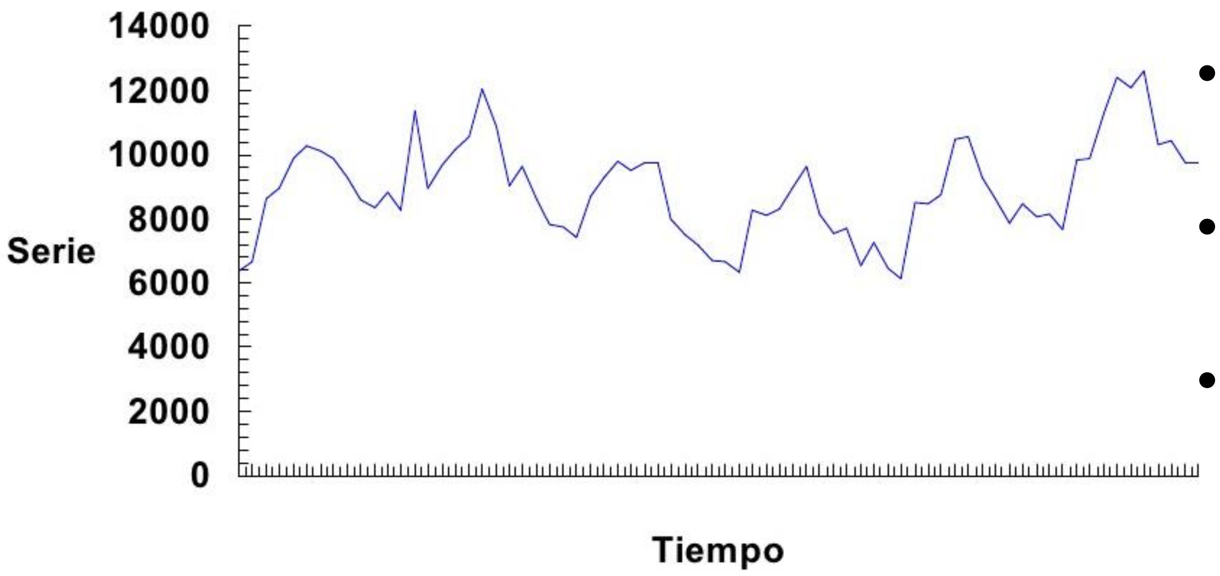
growup
DATA ANALYTICS

SERIES DE TIEMPO CON R

DEFINICIÓN

Una serie temporal se define como una colección de datos sucesivos que se recopilan, observan o registran cronológicamente en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros).



- *Exportaciones por trimestres.*
- *Evolución del PIB entre los años 2000 y 2020.*
- *Venta mensual de teléfonos móviles.*

CARACTERISTICAS

OBJETIVO

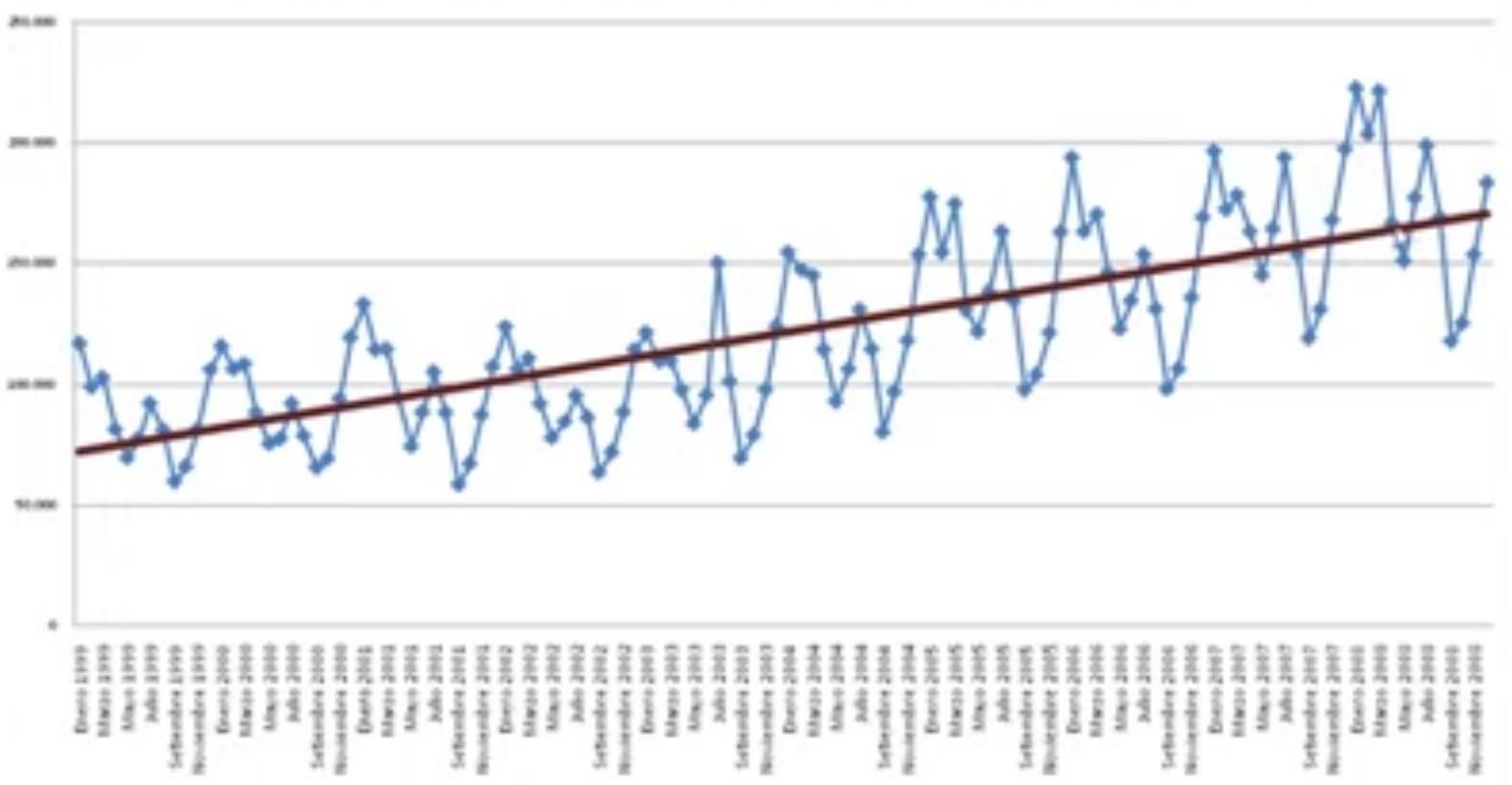
Los cuales servirán para identificar, mediante su comportamiento y evolución, cuál fue el Proceso Generador de los Datos PGD.

Así, se puede realizar pronósticos confiables corto plazo para la toma de decisiones.

COMPONENTES

- ***TENDENCIAL:*** *Se la identifica como un movimiento suave de la serie a largo plazo.*

Generalmente los agregados económicos poseen tendencia y no son estacionarios.

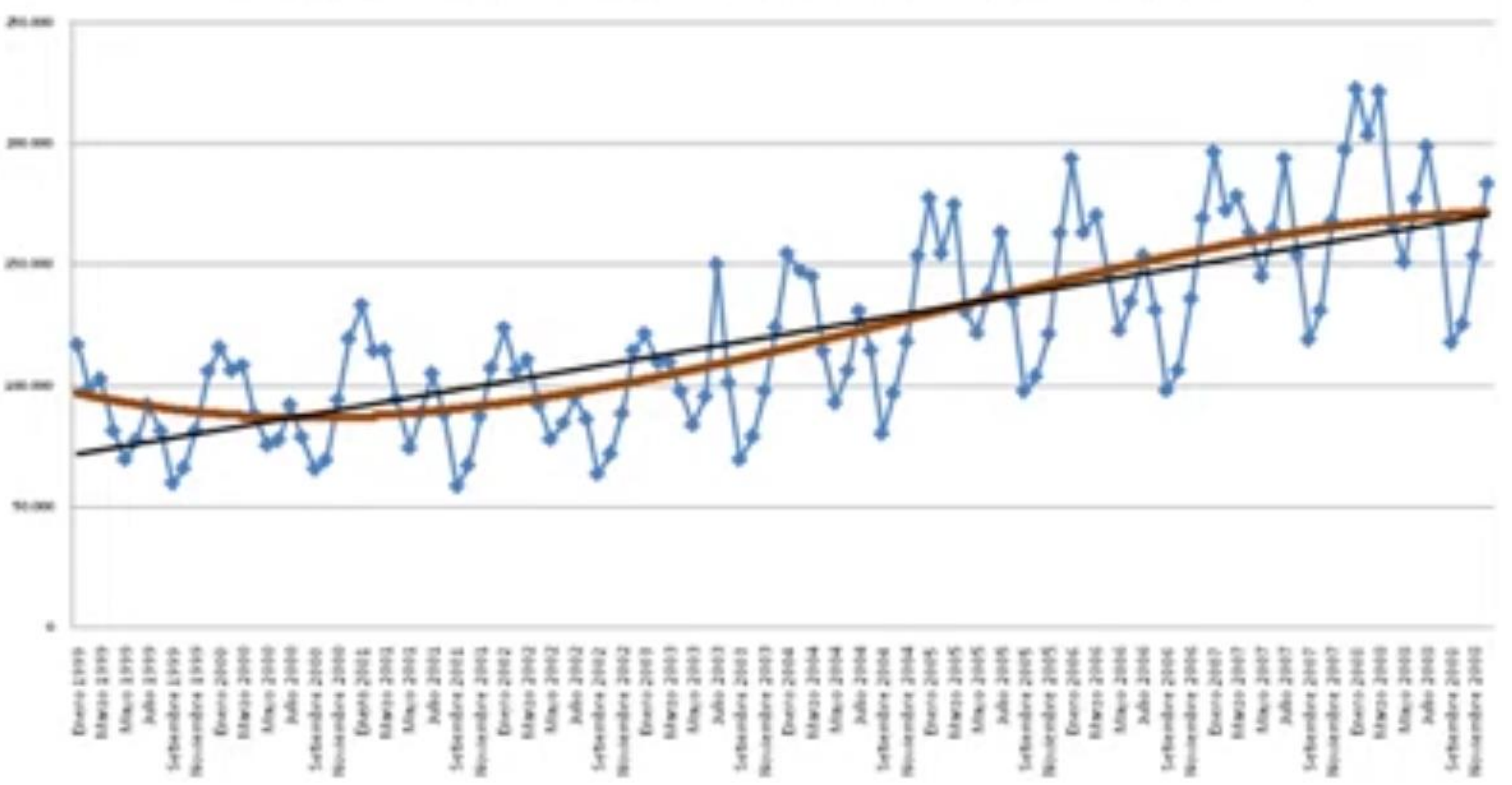


¿Estacionarios?

COMPONENTES

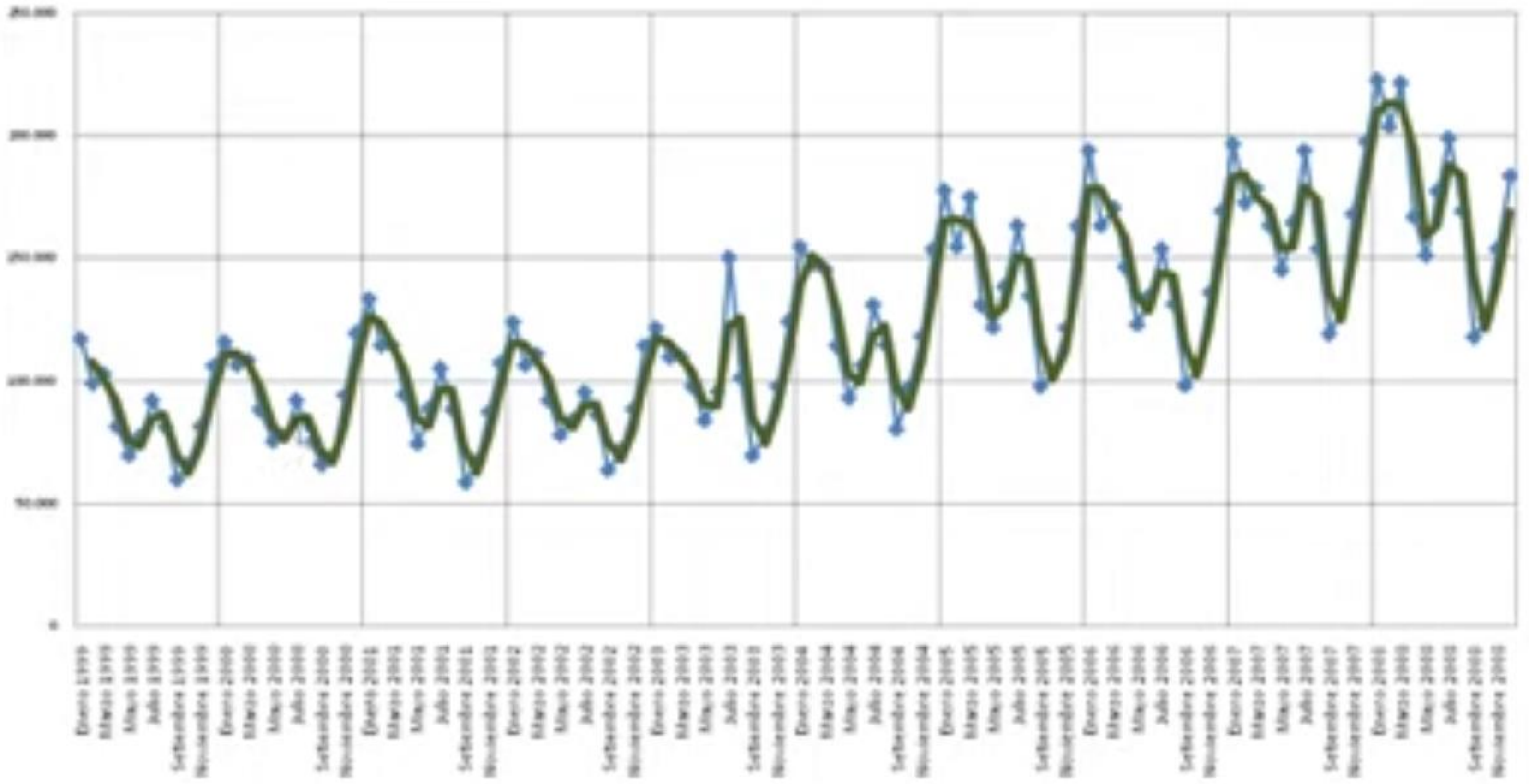
- ***CÍCLICO:*** *Es la fluctuación en forma de onda alrededor de la tendencia.*

Se caracteriza porque su duración es irregular



COMPONENTES

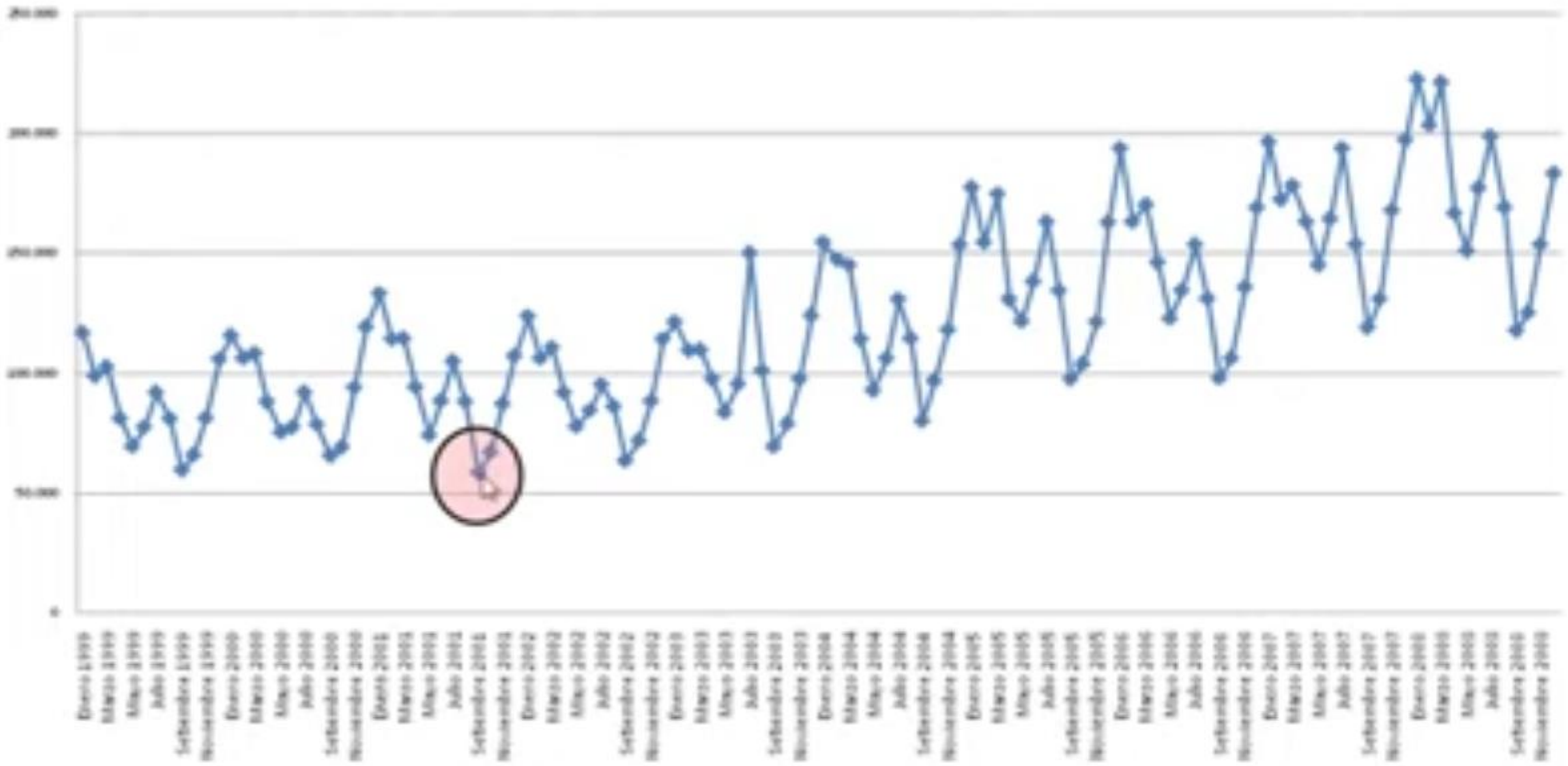
- ***ESTACIONAL:*** *Es el comportamiento recursivo o patrón que responde a las mismas fechas en el tiempo, durante algunos periodos.*



COMPONENTES

growup
DATA ANALYTICS

- ***IRREGULAR:*** *Se conoce como la parte* ***Estocástica o Aleatoria****.*



Esta componente no responde a ningún patrón o comportamiento sistemático sino que es el resultado de cuestiones fortuitas implícitas de la serie.

PROCESOS ESTOCÁSTICOS

Un proceso estocástico $(\boldsymbol{x_t})$ *es una sucesión de variables aleatorias ordenadas en el tiempo (en el caso de series temporales).*

Por lo que, las series temporales se definen como un caso particular de los procesos estocásticos.

Lo ideal es tener una serie de tiempo con media y varianza (más o menos) constante.

CLASIFICACIÓN

- ***ESTACIONARIA:***

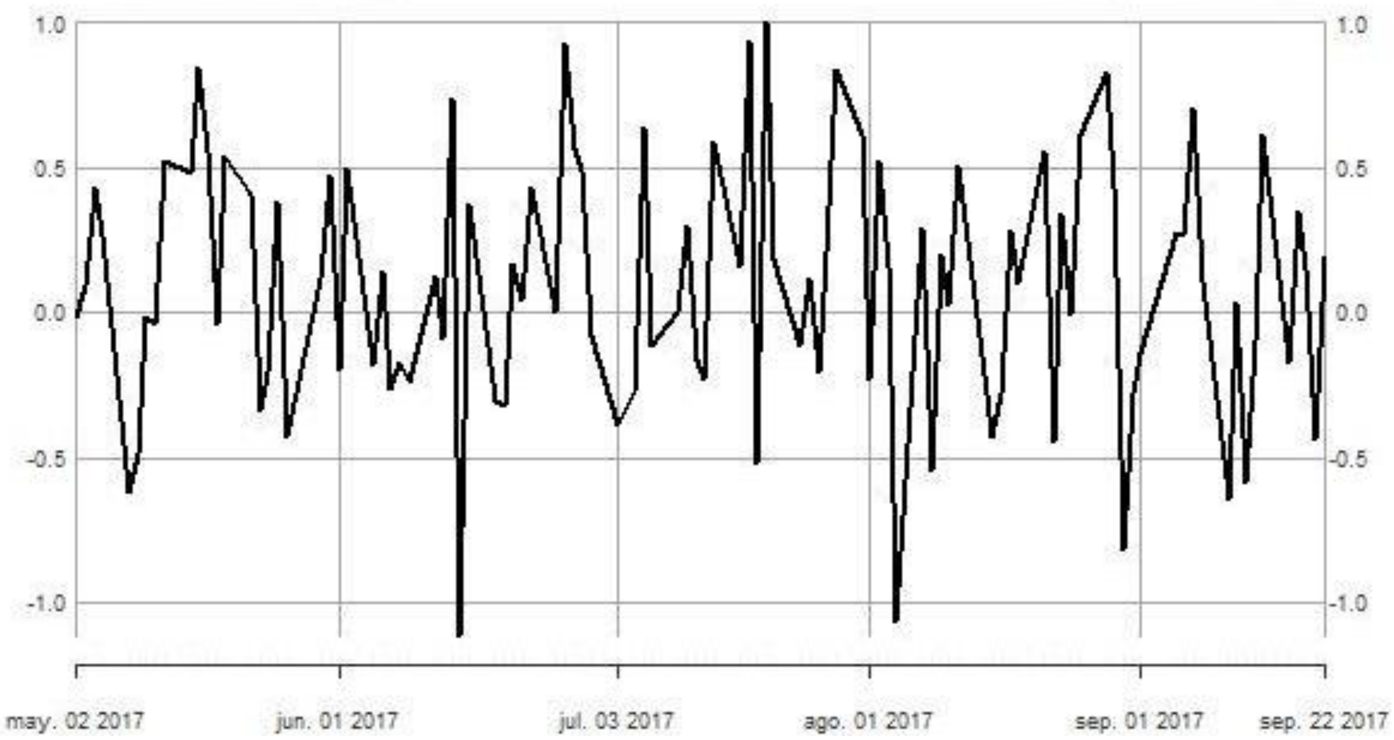
Cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo.

- ***NO ESTACIONARIA:***

Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

PROCESOS ESTOCÁSTICOS ESTACIONARIOS

Un proceso estocástico (x_t) *es estacionario en sentido débil cuando su distribución de probabilidad varía de forma más o menos constante a lo largo de cierto periodo de tiempo.*



PROCESOS ESTOCÁSTICOS ESTACIONARIOS

En otras palabras:

- *su media o primer momento:*

$$E(x_t) = \mu, \qquad |\mu| < \infty$$

- *su varianza o segundo momento:*

$$E(x_t - \mu)^2 = \sigma^2 < \infty$$

- *su covarianza*

$$E(x_t - \mu)(x_{t+\tau} - \mu_{t+\tau}) = \gamma_\tau, \qquad |\gamma_\tau| < \infty$$

Se mantienen (mas o menos) constante o invariante a través del tiempo.

PROCESOS ESTOCÁSTICOS ESTACIONARIOS

- *Si el proceso es estacionario, estos momentos son finitos y no dependen de "t".*
- *Si un proceso no es estacionario, tanto* $\boldsymbol{E}(\boldsymbol{x_t})$ *como* $\boldsymbol{Var}(\boldsymbol{x_t})$ *pueden depender de "t".*

EVALUACIÓN LA ESTACIONARIEDAD DE UNA SERIE

1. *Método gráfico*

2. *Test de estacionariedad / Raíz Unitaria*
    - *Dickey Fuller*
    - *Aumented Dickey Fuller*
    - *KPSS*
    - *Phillips Perron*
    - *Hegy*

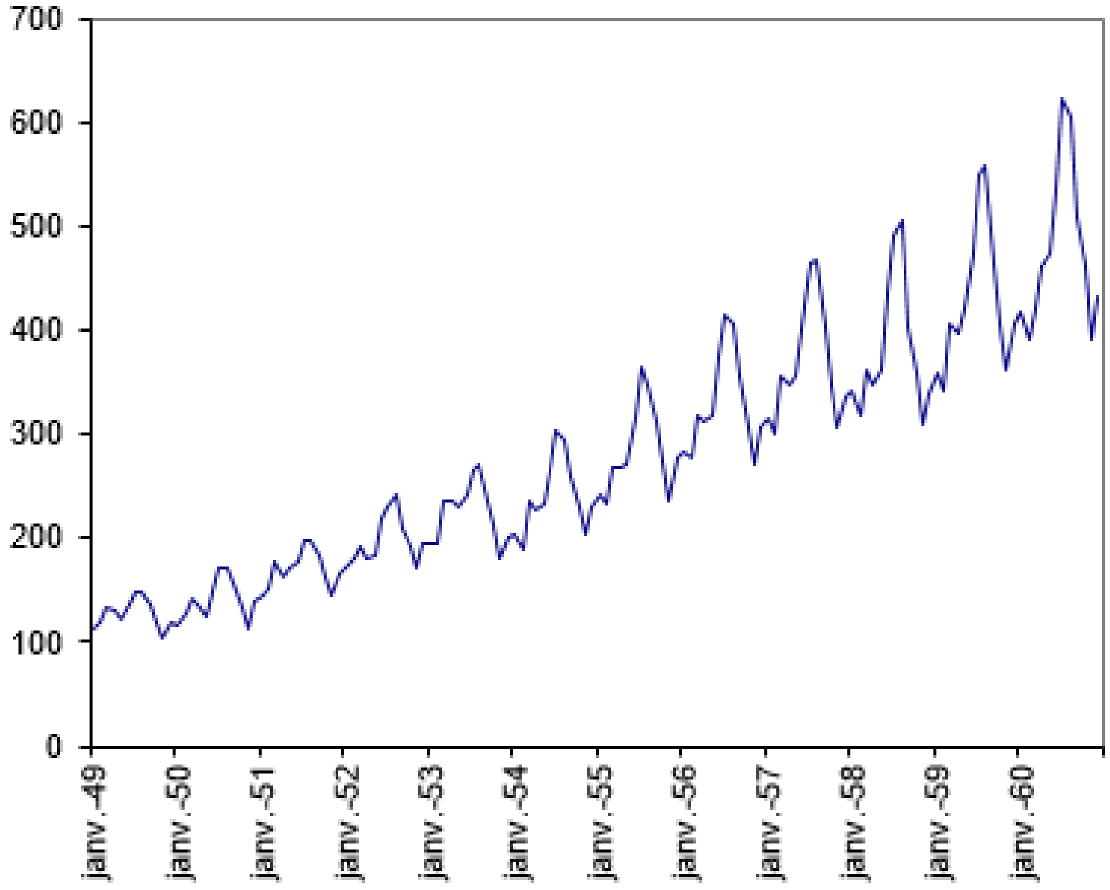
3. *Análisis de la Función de Autocorrelación y la Función Autocorrelación Parcial*

MÉTODO GRÁFICO

- *Se grafica la serie y se analiza su comportamiento*

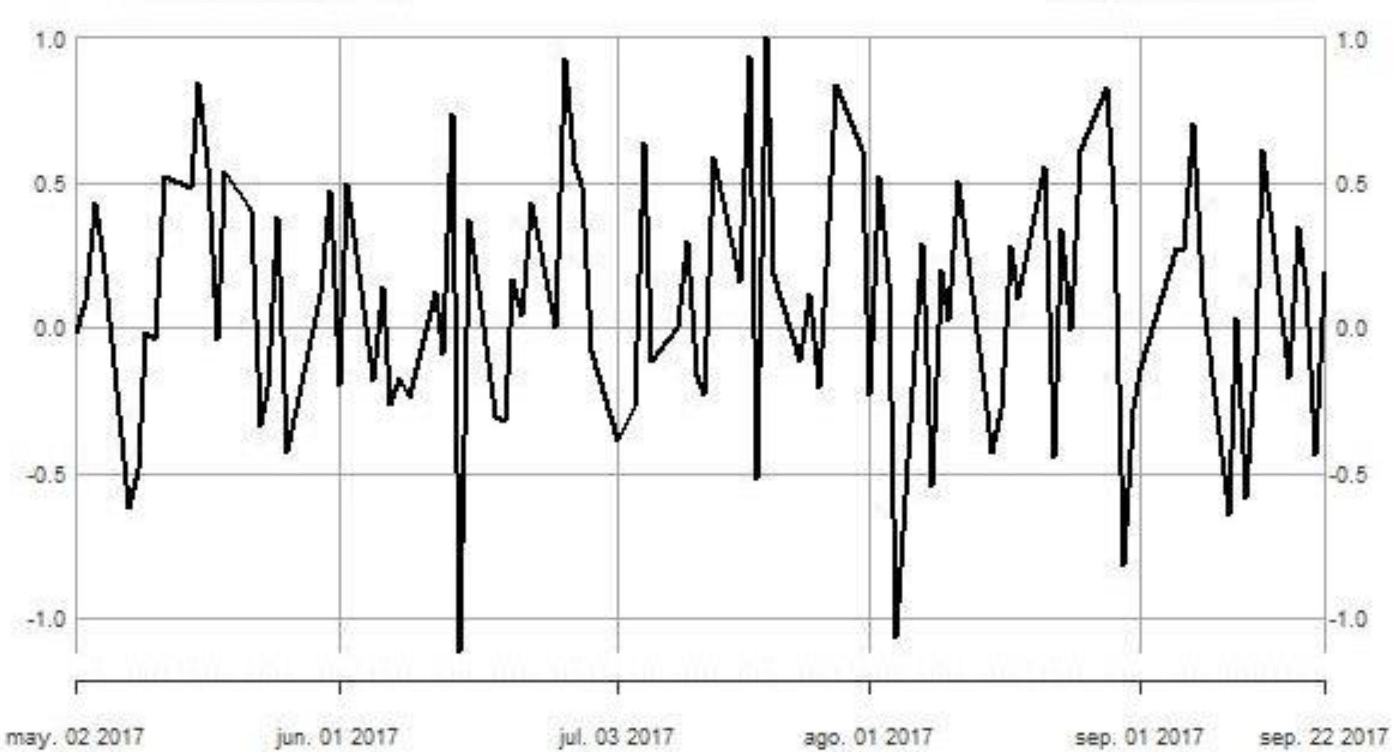
- *Si su comportamiento presenta una tendencia, entonces la serie no es estacionaria*

- *Si su comportamiento es muy volátil a través del tiempo, entonces no es estacionaria*



MÉTODO GRÁFICO

- *Si su comportamiento se mantiene alrededor de la media con un comportamiento estable, entonces es estacionaria.*



RAÍZ UNITARIA

- *Es una tendencia estocástica en la serie temporal.*

- *Algunas veces se le llama "paseo aleatorio con deriva o caminata aleatoria".*

- *Así, la no estacionariedad, caminata aleatoria, raíz unitaria son sinónimos.*

- *Por tanto, si la serie tiene una raíz unitaria, ésta presenta un patrón sistemático* que es impredecible. D*e manera que no es estacionaria y los pronósticos no serán confiables.*

AUTOCORRELACIÓN

- En una serie de tiempo los valores que toma una variable en el tiempo no son independientes entre sí.

- Sino que un valor determinado depende de los valores anteriores, existen dos formas de medir esta dependencia de las variables

FUNCIÓN DE AUTOCORRELACIÓN (ACF)

ACF: *Mide la relación estadística entre las observaciones de una serie de tiempo.*

Es decir, la correlación entre la variable $(\boldsymbol{x_t})$ *vs* $(\boldsymbol{x_{t-1}})$*, es decir la variable en* $(\boldsymbol{x_t})$ *contra la variable* $(\boldsymbol{x_t})$ *rezagada un período.*

FUNCIÓN DE AUTOCORRELACIÓN (ACF)

En otras palabras, la Autocorrelación mide la correlación entre dos variables separadas por k periodos:

$$\boldsymbol{\rho_j = corr\left(x_j, x_{j-k}\right) = \frac{cov\left(x_j, x_{j-k}\right)}{\sqrt{Var\left(x_j\right)}\sqrt{Var\left(x_{j-k}\right)}}}$$

La función de Autocorrelación simple tiene las siguientes propiedades:

$$\boldsymbol{\rho_0 = 1}$$

$$\boldsymbol{-1 < \rho_j < 1}$$

$$\boldsymbol{\rho_j = \rho_{-j}}$$

FUNCIÓN AUTOCORRELACIÓN PARCIAL (PACF)

PACF: *Se define como la correlación de las observaciones de la serie de tiempo, libre de influencias de otras observaciones retardadas de la misma serie de tiempo.*

FUNCIÓN AUTOCORRELACIÓN PARCIAL (PACF)

En otras palabras, La Autocorrelación Parcial mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas.

$$\boldsymbol{\pi_j = corr\left(x_j, x_{j-k} \middle| x_{j-1}, \ldots, x_{j-k+1}\right) = \frac{cov\left(x_j - \hat{x}_j, x_{j-k} - \hat{x}_{j-k}\right)}{\sqrt{Var\left(x_j - \hat{x}_j\right)}\sqrt{Var\left(x_{j-k} - \hat{x}_{j-k}\right)}}}$$

PROCESOS LINEALES ESTACIONARIOS

DEF: *Sea* $(\boldsymbol{x_t})$ *centrado, de segundo orden. Se dice que el proceso es lineal si se puede expresar como:*

$$x_t = \sum_{j=0}^{\infty} \varphi_j u_{t-j}$$

Donde

$$\varphi_0 = 1$$

$$\sum_{j=0}^{\infty} |\varphi_j|^2 < \infty$$

$$(u_t):\ ruido\ blanco$$

RUIDO BLANCO

Sea $(\boldsymbol{u_t})$ *un proceso aleatorio real de segundo orden* $\left(\boldsymbol{E u_t^2} < \infty\right)$ *tal que*

$$\boldsymbol{E[u_t] = 0}$$
$$\boldsymbol{Var[u_t] = \sigma^2 > 0}$$
$$\boldsymbol{Cov[u_t, u_s] = 0, t \neq s}$$

Este proceso se conoce como Ruido Blanco.

Es decir, un proceso estrictamente estacionario es un Ruido Blanco.

PROCESO AUTOREGRESIVO AR(p)

AR(p): *Sea* $(\boldsymbol{x_t})$ *un proceso estacionario se define un autoregresivo como:*

$$\boldsymbol{x_t = \sum_{j=0}^{p} \alpha_j x_{t-j} + u_t}$$

donde $(\boldsymbol{u_t})$ *es un ruido blanco.*

Condición para Estacionariedad es que

$$|\boldsymbol{\alpha_0 + \alpha_1 + \cdots + \alpha_p}| < \mathbf{1}$$

PROCESO AUTOREGRESIVO AR(p)

La idea es que el valor actual de la serie, $(\boldsymbol{x_t})$ puede explicarse en función de valores pasados $\boldsymbol{x_{t-1}}$, $\boldsymbol{x_{t-2}}$,…, $\boldsymbol{x_{t-p}}$ donde determina el número de rezagos necesarios para pronosticar un valor actual.

$$\boldsymbol{x_t = \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + u_t}$$

donde $(\boldsymbol{u_t})$ es un ruido blanco.

*Se observa mediante **ACFP***

PROCESO MEDIA MOVIL MA(q)

MA(q): *Sea* $(\boldsymbol{x_t})$ *un proceso estacionario se define una media móvil como:*

$$\boldsymbol{x_t = u_t - \sum_{j=0}^{q} \theta_j u_{t-j}}$$

donde $(\boldsymbol{u_t})$ *es un ruido blanco.*

$$\boldsymbol{\theta_j} < \infty$$

Un proceso MA(q) es centrado, estacionario y tiene una representación única.

PROCESO MEDIA MOVIL MA(q)

Estos modelos suponen linealidad, el valor actual de la serie, $(\boldsymbol{x_t})$ *está influenciado por los valores de la fuente externa.*

$$\boldsymbol{x_t = \theta_0 - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q} + u_t}$$

donde $(\boldsymbol{u_t})$ *es un ruido blanco.*

Se observa mediante la ***ACF***

PROCESO ARMA(p , q)

Se pueden expresar como:

$$\boldsymbol{x_t = \sum_{j=0}^{p} \alpha_j x_{t-j} - \sum_{j=0}^{q} \theta_j u_{t-j} + u_t}$$

donde $(\boldsymbol{u_t})$ *es un ruido blanco.*

PROCESO LINEAL NO ESATCIONARIO ARIMA(p, d, q)

- *Muchas series de tiempo y en especial las series económicas no son estacionarias a este tipo de proceso se les considera procesos integrados.*

- *Por lo que, se debe diferenciar una serie de tiempo **(d)** veces para hacerla estacionaria*

- *Luego aplicarla a esta serie diferenciada un modelo, es decir, una serie de tiempo autoregresiva integrada de media móvil **ARIMA(p, d, q)**.*

PROCESO LINEAL NO ESATCIONARIO ARIMA(p, d, q)

Donde

- ***(p)*** *denota el número de términos autoregresivos*
- ***(d)*** *el número de veces que la serie debe ser diferenciada para hacerla estacionaria y*
- ***(q)*** *el número de términos de la media móvil invertible.*

$$\boldsymbol{x_t^d = \sum_{j=0}^{p} \alpha_j x_{t-j}^d - \sum_{j=0}^{q} \theta_j u_{t-j}^d + u_t}$$

donde $(\boldsymbol{u_t})$ *es un ruido blanco.*

PROCESO ARIMA Estacionales (SARIMA)(P,D,Q)

Los modelos ARIMA estacionales se representan por ARIMAs(P , D ,Q) o bien SARISMA(P , D ,Q), donde

- ***(P)*** *es el orden de la parte autoregresiva.*
- ***(D)*** *es el número de diferencias estacionales*
- ***(Q)*** *es el orden de la parte de medias móviles.*

$$x_t^D = \sum_{j=0}^{P} \alpha_j x_{t-jest}^D - \sum_{j=0}^{Q} \theta_j u_{t-jest}^D + u_t$$

donde (u_t) *es un ruido blanco.*

PROCESO ARIMA Estacionales (SARIMA)(P,D,Q)

Los modelos SAR (P), donde

- ***(P)*** *es el orden de la parte autoregresiva.*

$$\boldsymbol{x_t = \alpha_0 + \alpha_1 x_{t-12} + \alpha_2 x_{t-24} + \cdots + u_t}$$

Los modelos SMA (Q), donde

- ***(Q)*** *es el orden de la parte de medias móviles.*

$$\boldsymbol{x_t = \theta_0 - \theta_1 u_{t-12} - \theta_2 u_{t-24} + \cdots + u_t}$$

tal que $(\boldsymbol{u_t})$ *es un ruido blanco.*

PROCESO ARIMA Estacionales (SARIMA)(p,d,q)(P,D,Q)

Diferencias Ordinal

$$\Delta x_t = x_t - x_{t-j}$$

Diferencia Estacional

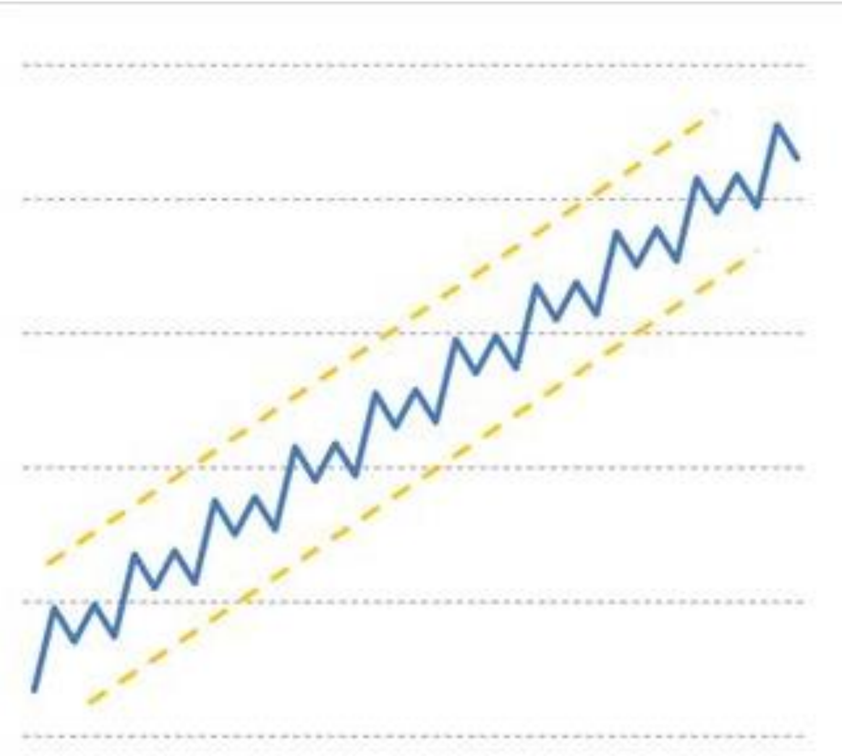
$$\Delta x_t = x_t - x_{t-jest}$$

Modelo

$$x_t^{d,D} = \sum_{j=0}^{p} \alpha_j x_{t-j}^{d} - \sum_{j=0}^{q} \theta_j u_{t-j}^{d} + \sum_{j=0}^{P} \alpha_j x_{t-jest}^{D} - \sum_{j=0}^{Q} \theta_j u_{t-jest}^{D} + u_t$$

MODELOS DE AISLAMIENTO EXPONENCIAL

MODELO ADITIVO



Se usa cuando la serie tiene una tendencia al menos localmente, y un patrón estacional constante.

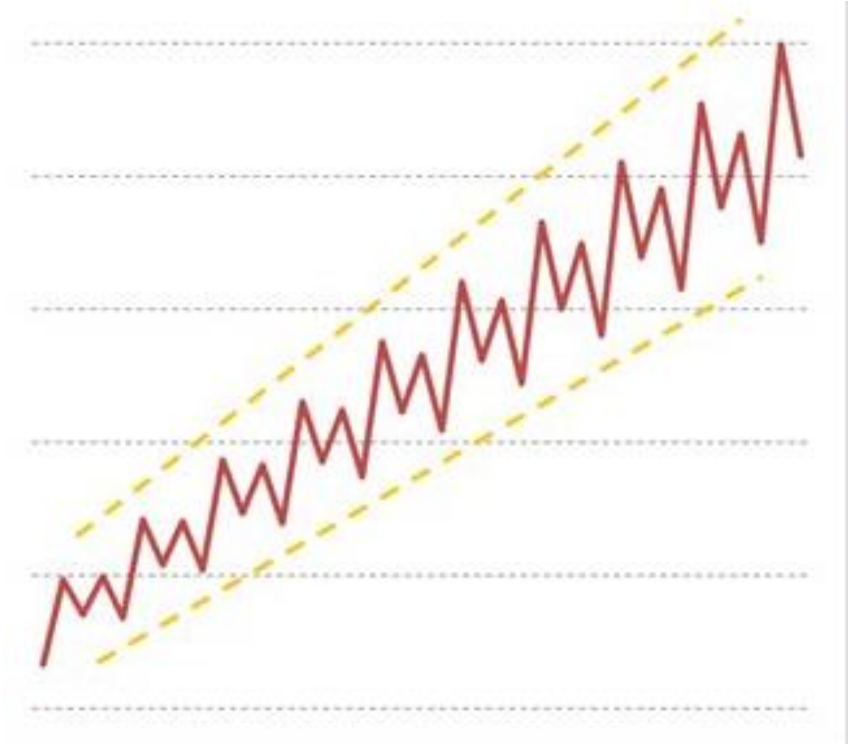
Al modelo Holt, se resta el factor.

MODELOS DE AISLAMIENTO EXPONENCIAL

MODELO MULTIPLICATIVO

Se usa cuando la serie tiene una tendencia, al menos localmente, y un patrón estacional creciente.

Al modelo Holt, se divide por el factor estacional



ENFOQUE DE SUAVISAMIENTO

Los métodos de suavizado se basan en modelos paramétricos ***deterministas*** *que se ajustan a la evolución de la serie.*

Son técnicas de tipo predictivo más que descriptivo (resultan más adecuados para pronosticar). Estos modelos se pueden emplear en:

1. ***Series temporales sin tendencia ni estacionalidad***
2. ***Series temporales no estacionales con tendencia***
3. ***Series temporales con tendencia y estacionalidad***

Series temporales sin tendencia ni estacionalidad

Este tipo de series tienen un comportamiento más o menos estable que sigue un patrón subyacente salvo fluctuaciones aleatorias (comportamiento estacionario), a este tipo de series se le pueden aplicar:

MODELOS "NAIVE"

Según la importancia que se le de a las observaciones se tiene:

- *Se otorga la misma importancia a todas las observaciones a la hora de predecir, de esta forma la previsión vendrá dada por la media de las observaciones.*

- *Se da importancia únicamente a la última de las observaciones ignorando el resto, de forma que el ajuste de la serie es su "sombra", es la misma serie pero retardada en una unidad de periodo.*

MODELOS DE SUAVIZADO EXPONENCIAL SIMPLE

- *Consisten en dar importancia a todos los datos anteriores*
- *Pero concediéndoles diferentes pesos, ya que los datos más relevantes a la hora de efectuar una previsión son los últimos de los que se dispone, disminuyendo la importancia conforme nos alejamos de ellos.*
- *De esta manera se sustituye cada dato de la serie por una media ponderada de las observaciones anteriores, considerando que los pesos de las mismas decaen de forma exponencial conforme éstas se alejan en el tiempo (la fórmula del ajuste es recursiva).*

Series temporales no estacionales con tendencia

En el caso de series temporales con tendencia lineal (creciente o decreciente) pero sin comportamiento estacional, el modelo clásico que más se suele aplicar es el de Holt:

MODELOS DE SUAVIZADO EXPONENCIAL DE HOLT

- *Se aplican un suavizado exponencial simple de manera doble*
- *Por lo que también son conocidos como suavizado exponencial doble*
- *En un principio se alisa directamente a la variable objeto de estudio, mientras que en la segunda operación se procede a alisar a la variable alisada previamente obtenida.*

Series temporales con tendencia y estacionalidad

En el caso de series temporales con tendencia lineal (creciente o decreciente) y comportamiento estacional, el modelo clásico que se aplica es el de Holt-Winters que es una extensión del modelo de Holt, solo que además considera estacionalidad.

PRONÓSTICO

Un pronóstico usando suavizamiento exponencial es un promedio ponderado de las observaciones pasadas, en el que las observaciones más recientes tienen mayor peso y la ponderación disminuye de manera exponencial a medida que los registros se alejan del tiempo.

En este sentido, el parámetro común a estimar en todas las variantes existentes de esta metodología es el peso que tendrá la observación más reciente.

PRONÓSTICO

Dependiendo de la característica de la serie de tiempo a pronosticar se debe elegir entre métodos de suavizado exponencial que van desde

- *versiones simples, en las que solo se estima el peso de la observación más recientes,*
- *Versiones en los que además se estiman componentes como: tendencia (lineal o no), estacionalidad (aditiva o multiplicativa) y el tipo de intervalo de pronóstico.*

Para elegir la variante adecuada de suavizamiento exponencial sin tener que detenernos a hacer pruebas específicas a la serie de tiempo, utilizaremos la función

ETS(Errors, Trend, Seasonal)

Esta función se encarga de analizar las características de las series y aplicar la variante de suavizamiento que mejor le va.

PRONÓSTICO

El método de pronóstico más simple es utilizar la observación más reciente; esto se denomina pronóstico "NAIVE" o ingenuo y se puede implementar en una función homónima.

Proporciona un punto de referencia útil para otros métodos de pronóstico.

Para los datos estacionales, una idea relacionada es utilizar la temporada correspondiente del último año de datos.

Por ejemplo, si desea pronosticar el volumen de ventas para el próximo mes de marzo, utilizaría el volumen de ventas del mes de marzo anterior. Esto se implementa en la función

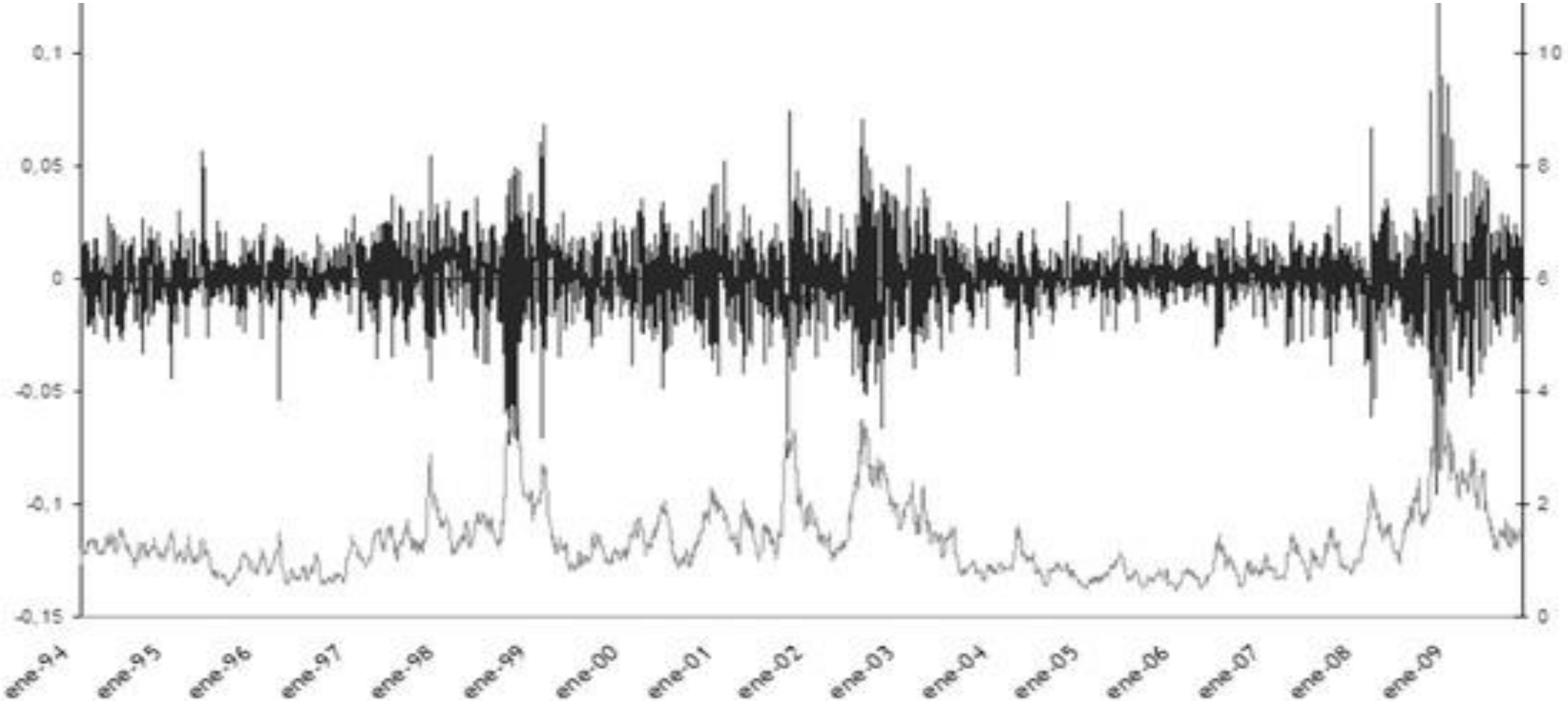
snaive()

es decir, naive estacional .

VOLATILIDAD

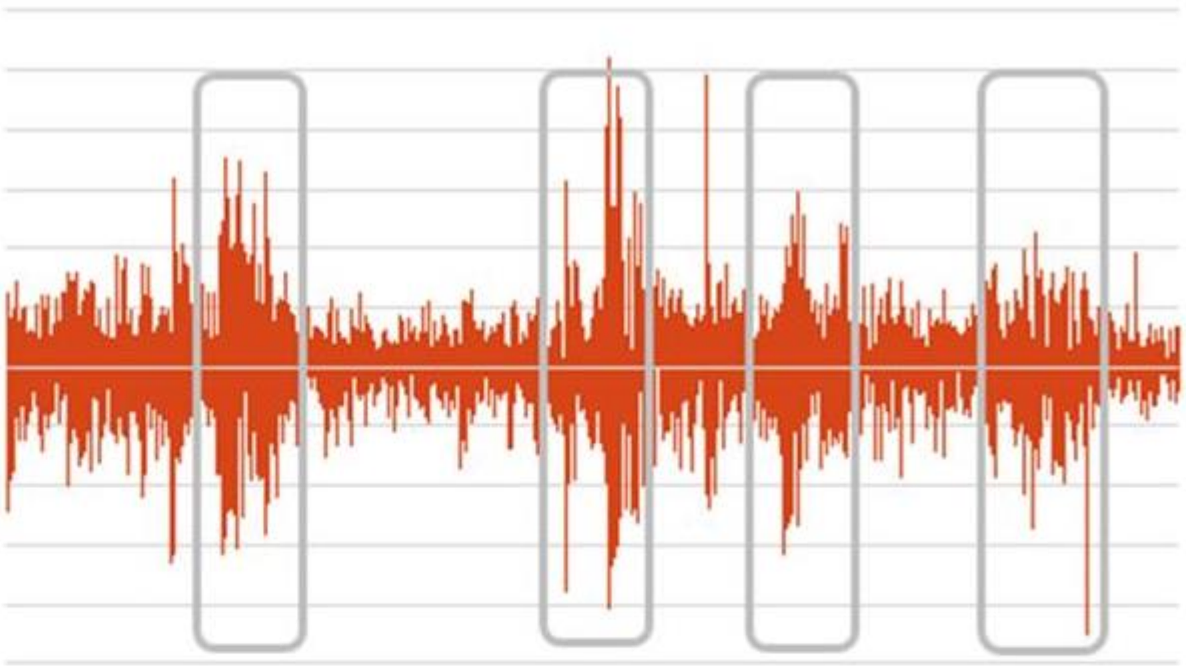
DEF: *Es la varianza condicional de la serie subyacente.*

En el caso de las series de tiempo financieras, se modela la volatilidad de los retornos.



VOLATILIDAD

Es de anotar que, aunque la serie sea estacionaria y tenga, por tanto, varianza constante, puede presentar oscilaciones a corto plazo es lo que recoge la varianza condicional para el estudio de la volatilidad cuyo conocimiento es de interés, en particular, para hacer predicciones a corto plazo.



En las series financieras se presenta períodos largos de alta volatilidad seguidos por períodos de baja volatilidad, lo que indica la presencia de **heterocedasticidad.**

MODELO ARCH

Engle (1982) propuso el modelo ARCH, que significa ***modelo auto regresivo condicionalmente heterocedástico****, el cual hace parte de la familia de modelos adecuados para modelar la volatilidad de una serie.*

Se estudia :

$$\boldsymbol{Magnitud\ de\ los\ Residuos}:\ \boldsymbol{\varepsilon_t}$$

Se usa:

$$\boldsymbol{Residuos\ al\ cuadrado}:\ \boldsymbol{\varepsilon_t^2}$$

Resuelve:

- *Problema de signos positivos y negativos*
- *Penaliza más las diferencias más altas entre valores reales y predicciones*

MODELO ARCH(q)

ARCH(q): *Sea* (x_t) *un proceso estacionario se define la varianza condicionada como:*

$$Var(x_t|x_{t-1},\dots,x_{t-q}) = \sum_{j=0}^{q} \alpha_j \varepsilon_{t-j}^2$$

Es decir,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-q}^2$$

Se observa mediante la ***PACF*** *de los residuos y de los residuos al cuadrado*

Condiciones que deben satisfacer los modelos ARCH(q)

No negatividad: *La varianza condicional, su valor siempre debe ser estrictamente positivo (recuerde que la varianza condicional es el cuadrado de los errores).*

Confirmar que hay efectos ARCH en las series: *En esta prueba, la hipótesis nula es que hay efectos ARCH, es decir, que los q rezagos de los errores al cuadrado son significativos (o que son distintos de 0) en la serie.*

La sumatoria de los parámetros no puede ser mayor a 1: *si la suma de los valores que reportan los parámetros del modelo es mayor uno, la volatilidad de la serie explota con el tiempo, en otras palabras, el modelo es inestable.*

Limitaciones de los modelos ARCH

1. *No hay una forma precisa de calcular el número de rezagos óptimos (q) para el modelo ARCH.*

2. El valor de (q), es decir, el número de rezagos del error al cuadrado que es requerido para capturar toda la dependencia en la varianza condicional, puede llegar a ser muy largo.

3. Muchos rezagos (q) puede llegar a ocasionar que uno de los coeficientes se vuelva negativo, lo cual no tendría sentido en la interpretación.

MODELO GARCH

*Significa **modelo auto regresivo condicionalmente heterocedástico generalizado**.*

Intuición:

$$\boldsymbol{Volatilidad = Varianza}$$

En el modelo ARCH sólo se incluye valores de residuos al cuadrado de períodos. Ahora, añadir valores pasados y **¿cuales tendrían que ser esos valores?**

*Al incluir valores pasados de los retornos en este caso **no tendría sentido** ya que generalmente contamos con ellos en la ecuación de la media.*

Incluir esas varianzas condicionales pasadas $\boldsymbol{\sigma_{t-1}^2}$ para ayudarnos a explicar las varianzas condicionales actuales.

MODELO GARCH

Se usa :

$$\textit{Residuos al cuadrado del pasado}: \varepsilon_{t-j}^2$$
$$\textit{Varianza condicional del periodo anterior}: \sigma_{t-j}^2$$

Por ejemplo:

$$Var(x_t|x_{t-1}) = \Omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

Ahora, una varianza condicional pasada en realidad ya incluiría dentro de sí misma los efectos de los residuos al cuadrado.

$$\sigma_{t-1}^2 = \Omega + \alpha_1\varepsilon_{t-2}^2$$

Esto haría que agregar más residuos al cuadrado sea redundante si estamos agregando la varianza de periodos anteriores.

Es decir, todos los efectos de la varianza condicional de hace dos días estarían contenidos en la varianza condicional de ayer.

MODELO GARCH(p, q)

GARCH(p , q): *Sea* (x_t) *un proceso estacionario se define la varianza condicionada como:*

$$Var\left(x_t \middle| x_{t-1}, \ldots, x_{t-p}\right) = \Omega + \sum_{j=0}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=0}^{p} \beta_j \sigma_{t-j}^2$$

Es decir,

$$\sigma_t^2 = \Omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

MODELO GARCH(p, q)

- *Se ha demostrado matemáticamente que ningún modelo GARCH de orden superior supera al modelo **GARCH(1,1)** cuando se trata de la variación de los retornos de precios del mercado.*

- *Se debe a la naturaleza recursiva en la que se calculan las variaciones condicionales pasadas.*