UNIVERSIDAD NACIONAL DE INGENIERÍA

ESCUELA CENTRAL DE POSGRADO

Facultad de Ingeniería Económica, Estadística y CCSS

Unidad de Posgrado - FIEECS

Curso: Forecasting for Data Science

Docente: Mg. Omar A. Chincaro Del C.

Forecasting for Data Science

1.- Introducción

2.- Componentes de una serie temporal

3.- Suavizado de un serie

4.- Series estacionarias:

- Modelos Autorregresivos (AR(p))
- Modelos Medias Móviles (MA(q))
- Modelos Mixtos: ARMA(p,q)

5.- Series no estacionarias

- Modelos Mixtos: ARIMA(p,d,q)

6.- Series estacionales: SMA, SAR, SARI, SIMA, SARIMA

7.- Series heterecedásticas: ARCH y GARCH

8.- Pronósticos

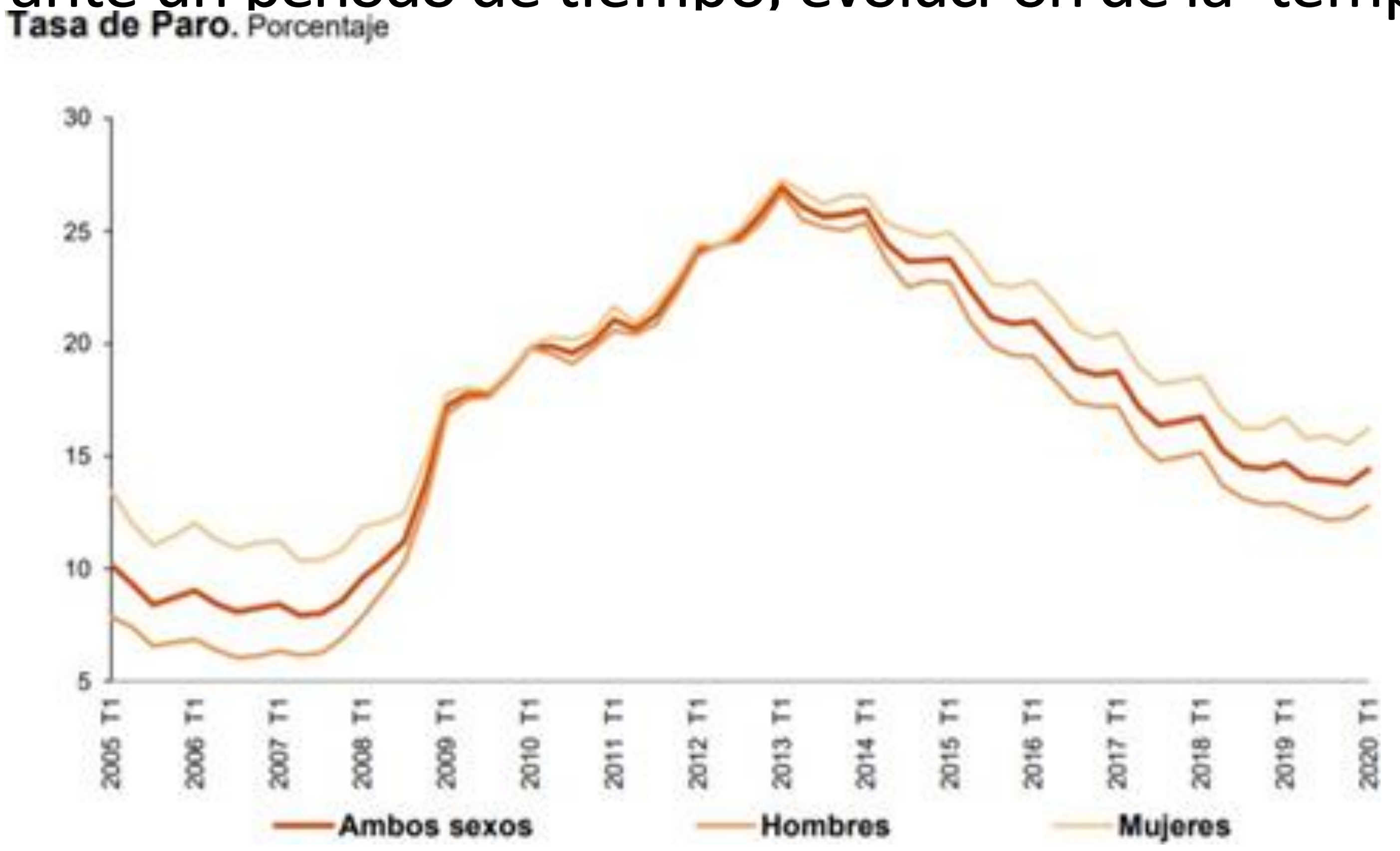
INTRODUCCIÓN

En lo últimos 30 años los modelos de series de tiempo de han convertido en metodologías poderosas para analizar conjuntos de datos, teniendo como propósito exclusivo la de realizar pronósticos. La característica principal de este conjunto de datos, es que se hayan observado puntos equidistantes a lo largo del tiempo y pueden provenir de las distintas área del conocimiento humano. La economía fue el área más privilegiada con el tratamiento de series de tiempo, su aplicación ha revolucionado a sus principales áreas: Macroeconomía, Microeconomía, Gestión, Comercio Exterior, Fianzas entre otras.

Actualmente el estudio de series de tiempo, abarcan desde estructuras mas simples hasta las más complejas y se me mezclan con otras áreas de estudio como en Análisis Bayesiano formándose los Modelos Dinámicos Bayesianos.

Introducción

Las series temporales, Y_t , son observaciones de una variable a lo largo del tiempo. Por ejemplo, datos del paro nacional a lo largo del tiempo, ventas de una empresa durante un período de tiempo, evoluci´on de la temperatura, etc.



COMPONENTES DE UNA SERIE TEMPORAL

Componentes de una serie temporal

Una serie temporal se puede descomponer en cuatro tipos de componentes:

- ►**Efecto estacional (EE):** se define como una conducta repetitiva a lo largo del tiempo, provocada por factores que se presentan de forma peri´odica (trimestralmente, semestralmente...) y que influyen en el comportamiento de la serie. La periodicidad del efecto estacional es inferior a un año. Por ejemplo, el consumo de energía presenta picos más altos en verano e invierno debido al consumo de los aires acondicionados y calefacciones.
- ►**Componente cicilica (CC):** algunas series muestran conductas repetitivas sin periodo fijo, debido a las fluctuaciones de la actividad social y económica. Por ejemplo, el resultado de la sucesión de las fases expansivas y recesivas de la econom´ıa. La periodicidad del efecto c´ıclico es superior a un año.

Componentes de una serie temporal

- **Tendencia (T):** esta componente recoge el comportamiento de la serie a lo largo del tiempo. Es importante que la serie cuente con un gran número de observaciones para poder detectar la tendencia. El comportamiento de la tendencia puede ser lineal, exponencial, logıstico... Por tanto, una serie tiene tendencia cuando hay una variación significativa en el valor medio de la serie. Mediante la tendencia se puede estudiar si la serie es **estacionaria** o **evolutiva**. Una serie es **estacionaria** si los datos fluctúan en torno a un valor constante, es decir, si no hay tendencia en la serie.

Componentes de una serie temporal

- **Otras variaciones irregulares (VI):** fluctuaciones presentes en los datos no explicadas por la componente estacional ni por la tendencia. Por ejemplo, las huelgas, inundaciones son factores fortuitos que afectan de forma aislada y no permanente en una serie.

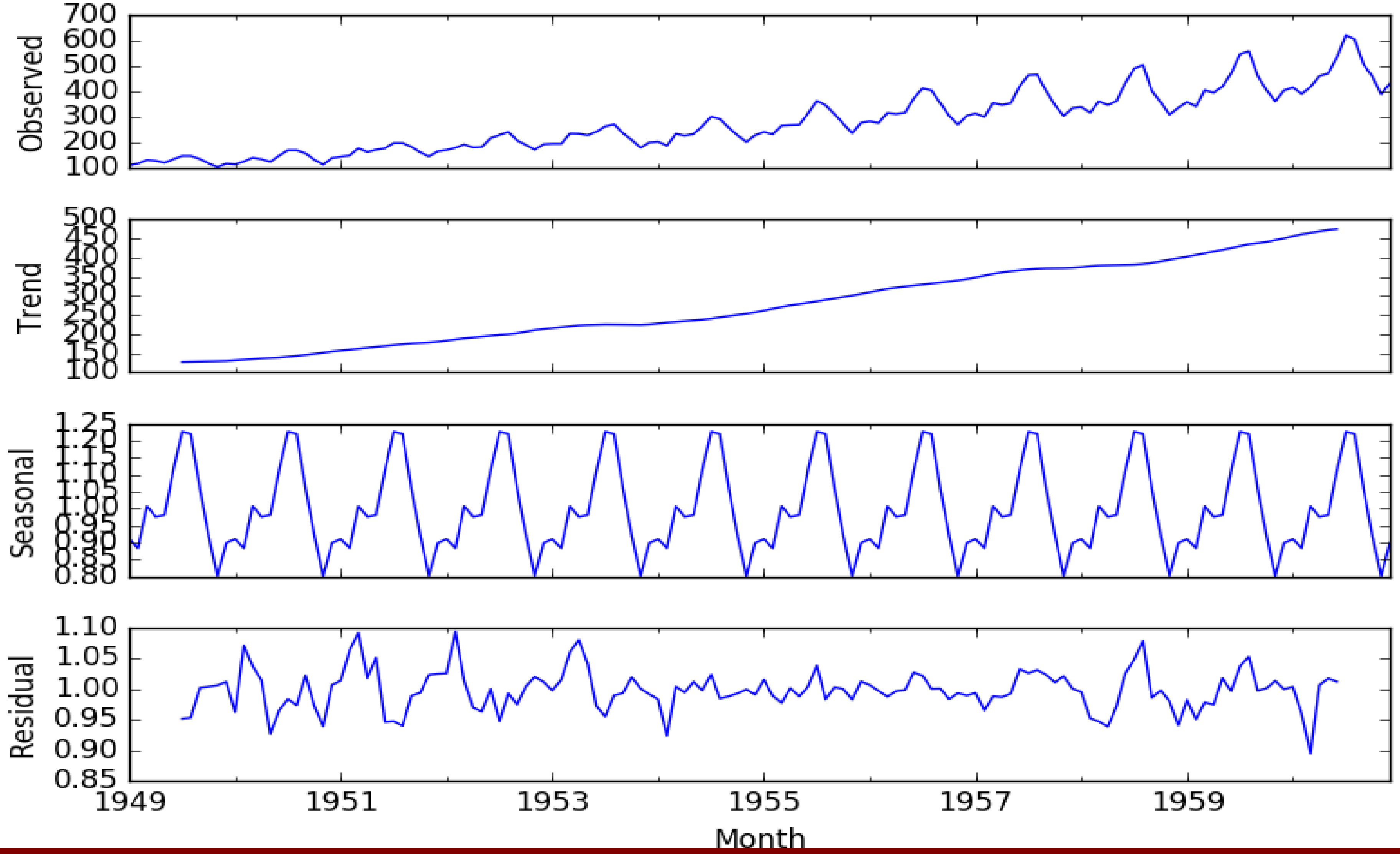
En resumen, una serie temporal se podr´a representar como

$$Y_t = f\,(\mathrm{EE}_t, \mathrm{CC}_t, \mathrm{T}_t, \mathrm{VI}_t), \qquad t = 1, 2, \ldots, n.$$

Dependiendo de la forma que tome f , los modelos de descomposici´on m´as frecuentes son:

- Modelo multiplicativo: $Y_t = \mathrm{EE}_t \cdot \mathrm{CC}_t \cdot \mathrm{T}_t \cdot \mathrm{VI}_t$.
- Modelo aditivo: $Y_t = \mathrm{EE}_t + \mathrm{CC}_t + \mathrm{T}_t + \mathrm{VI}_t$.
- Modelo mixto: $Y_t = \mathrm{EE}_t \cdot \mathrm{CC}_t \cdot \mathrm{T}_t + \mathrm{VI}_t$.

Componentes de una serie temporal



SUAVIZADO DE UNA SERIE TEMPORAL

MEDIAS MOVILES

Métodos de suavizados de la serie

Medias Móviles (MA)

Un proceso que se encuentra dentro de todos los métodos de descomposición es el **suavizado de datos**. El suavizado de datos es una transformación de los datos para que la serie resultante contenga menos fluctuaciones que la original. El método más popular es el de medias móviles.

Este método consiste en promediar la serie. Se promedian los valores de la serie para periodos de tiempo fijos que se encuentran a lo largo de toda la serie.

Índice	Datos	q=3	q=5	q=4
1	28			
2	20	26,00		
3	30	26,00	25,40	25,63
4	28	26,33	24,40	25,13
5	21	24,00	24,00	24,00
6	23	20,67	23,00	22,13
7	18	22,00	22,20	22,13
8	25	22,33	24,00	23,38
9	24	26,33	24,60	25,25
10	30	26,67	26,60	26,63
11	26	28,00	25,60	26,50
12	28	24,67	24,20	24,38
13	20	21,67	22,20	22,00
14	17	19,00	21,00	20,25
15	20	19,00	19,20	19,13
16	20	19,67	18,60	19,00
17	19	18,67	19,40	19,13
18	17	19,00	21,00	20,25
19	21	22,00		
20	28			

EXPONENCIAL

Métodos de suavizados de la serie – Suavización exponencial

Los métodos de suavizado exponencial describen la serie temporal basándose en el estudio de las fluctuaciones que sufren las misma, sin crear un modelo matemático a partir de los datos de la serie. Aunque no se basan en ningún modelo, implícitamente sí que dependen de algún modelo sencillo debido a la naturaleza de los datos.

El objetivo que persiguen es eliminar o "suavizar" las fluctuaciones o movimientos aleatorios en los datos, aprovechando el comportamiento de la serie para predecir valores futuros.

Las predicciones que ofrece este método son medias de valores pasados, ponderando los datos según un conjunto desigual de pesos. Por ello, el nombre del método "suavizado exponencial" es porque, en general, los pesos decaen de forma exponencial desde el dato más reciente a las antiguo.

Métodos de suavizados de la serie – Suavización exponencial

Los métodos de suavizado exponencial son:

- **Suavizado exponencial simple.** Se estima 1 parámetro. Adecuado para predecir valores futuros de series que no presentan ni componente estacional ni tendencia.

- **Suavizado exponencial doble:** método lineal de Holt. Se estiman 2 parámetros. Es adecuado para predecir valores futuros de series que presentan tendencia lineal.

- **Suavizado exponencial triple:** método de Holt-Winters. Se estiman 3 parámetros. Es adecuado para predecir valores futuros de series que presentan componente estacional.

Métodos de suavizados de la serie – Suavización exponencial simple

Es un método iterativo que ofrece una serie suavizada $\hat{Y}_t$ de la serie original : Y_t

$$\hat{Y}_{t+1} = \alpha Y_{t+1} + (1-\alpha)\hat{Y}_t \quad t = 1 \ldots n$$

donde α (parámetro de suavizado) es la constante con valores comprendidos entre 0 y 1.

Cuanto más cercano a 1 sea su valor, ...

Para obtener los pesos que se asignan a cada valor pasado de la serie, se debe de aplicar de forma repetitiva la ecuación anterior.

Por ello, la predicción en *t+1* es una media ponderada de todos los valores pasados.

Métodos de suavizados de la serie – Suavización exponencial doble

Tendencia	Estacionalidad	**Suavización Exponencial**
No	No	Simple Un Parámetro (alfa)
Si	No	Doble (Brown, Holt) Dos Parámetros (alfa, beta)
Si	Si	Holt Winters Tres parámetros (alfa, beta, gamma)

SERIES ESTACIONARIAS

¿Qué es la estacionariedad en serie temporal? ¿Por qué es estacionaria una serie?

Una serie estacionaria cuando la media y varianza de la serie se mantienen constante a lo largo del tiempo, es decir tanto la media como la varianza no dependen del tiempo.

MODELOS PARA UNA SERIE DE TIEMPO

A) FORMA DE MEDIA MOVIL INFINITO

Esta forma se fundamenta en el hecho de que toda la serie se origina como una suma infinita de choques aleatorios:

$$Z_t = \theta_0 + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \theta_4\varepsilon_{t-4} + \ldots\ldots\ldots\ldots\ldots.$$

$$Z_t = \theta_0 + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}$$

$$Z_t = \Theta_0 + \theta(B)\,\epsilon_t$$

Donde θ_j son los parámetros del modelo, ε_t las perturbaciones del modelo, B es operador de desface y t es el tiempo.

¿Qué es la estacionariedad en serie temporal? ¿Por qué es estacionaria una serie?

B) FORMA AUTOREGRESIVA E INVERTIDA

Es cuando el modelo se puede expresarse en su forma autorregresiva, es decir, como una suma ponderada de sus valores pasados mas un choque aleatorio:

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \phi_4 Z_{t-4} + \ldots\ldots\ldots\ldots\ldots.$$

$$Z_t = \phi_0 + \sum_{j=1}^{\infty} \phi_j Z_{t-j}$$

$$Z_t = \phi_0 + \phi(B)\, Z_t$$

Puede observarse que los polinomios $\theta(B)$ **y** $\phi(B)$ son inversos y determinan la invertibilidad y estacionariedad de los procesos respectivos.

CONDICIONES DE ESTACIONARIEDAD E INVERTIBILIDAD

1.- ESTACIONARIEDAD

Una serie de tiempo es estacionaria si tienen medias y varianzas finitas, no dependan del tiempo. Es decir cuando las raíces del polinomio que la generan tengas raíces mayores que la unidad.

2.- INVERTIBILIDAD

Una serie de tiempo es invertible cuando es posible expresar en su forma autorregresiva, implicando que la ecuación característica que lo generan tenga raíces mayores que la unidad

3.- CORRELOGRAMA

El análisis estadístico y la inspección gráfica visual de los valores muestrales tanto de las funciones de autocorrelación como los funciones de autocorrelación parcial constituyen lo que se llama “Análisis del correlograma”

3.1.- FUNCION DE AUTOCORRELACIÓN (FAC)

Esta función es la más usada que se utiliza para describir la estacionariedad o no estacionariedad de una serie de tiempo, consecuentemente en ella se determinan el orden de las Medias Móviles. Se definen de la siguiente manera:

$$\text{CORR}(K) = \frac{COV(K)}{COV(0)}$$

3.2.- FUNCION DE AUTOCORRELACIÓN PARCIAL (FACP)

La función de autocorrelación parcial se define como:

$\Phi(kk) = \text{CORR}[(Z(t) - E(Z(t) / Z(t-1),Z(t-2), \ldots\ldots,Z(t-k+1)) ; Z(t-k) - E(Z(t-k) / Z(t-1),Z(t-2), \ldots\ldots,Z(t-k+1))]$

y sirve para determinar el orden de la componente autoregresiva de la serie temporal.

¿Por qué es importante que la serie de tiempos sea estacionaria?

En la mayoría de los modelo de series temporales es indispensable que la serie sea estacionario

¿Cómo conseguimos la estacionariedad en varianza?: Aplicando log neperiano a la serie y/o haciendo uso de las transformaciones de box – cox.

¿Cómo conseguimos la estacionariedad en media? Tomando la primera diferencia de la serie

¿Raíces unitarias?

Una raíz unitaria es una tendencia estocástica de la serie temporal. Existen dos formas de remover la tendencia de una serie: 1) Hacer las primeras diferencias y 2) Regresión tendencial.

Los test de raíces unitarias puede ser utilizado para determinar si una serie tienen tendencias:

a) **El test de Dicker – Fuller Aumentado (ADF),** elimina la autocorrelación e indica si una serie es estacionario o no
b) **El test de Phillips - Perron (PP),** corrige la autocorrelación y heterocedasticidad de los errores.

PROCESO DE MODELACIÓN

METODOLOGÍA DE BOX JENKINS

ESPECIFICACIÓN

Analizar la gráfica de la serie (análisis situacional y estadístico).
Identificar tendencias y volatilidades en la serie.

IDENTIFICACIÓN

Analizar la estacionariedad en varianza y en media
Usar test de raíz unitaria
Transformación de Box Cox
Usar las FAC y FCP para identificar posibles modelos.

ESTIMACIÓN

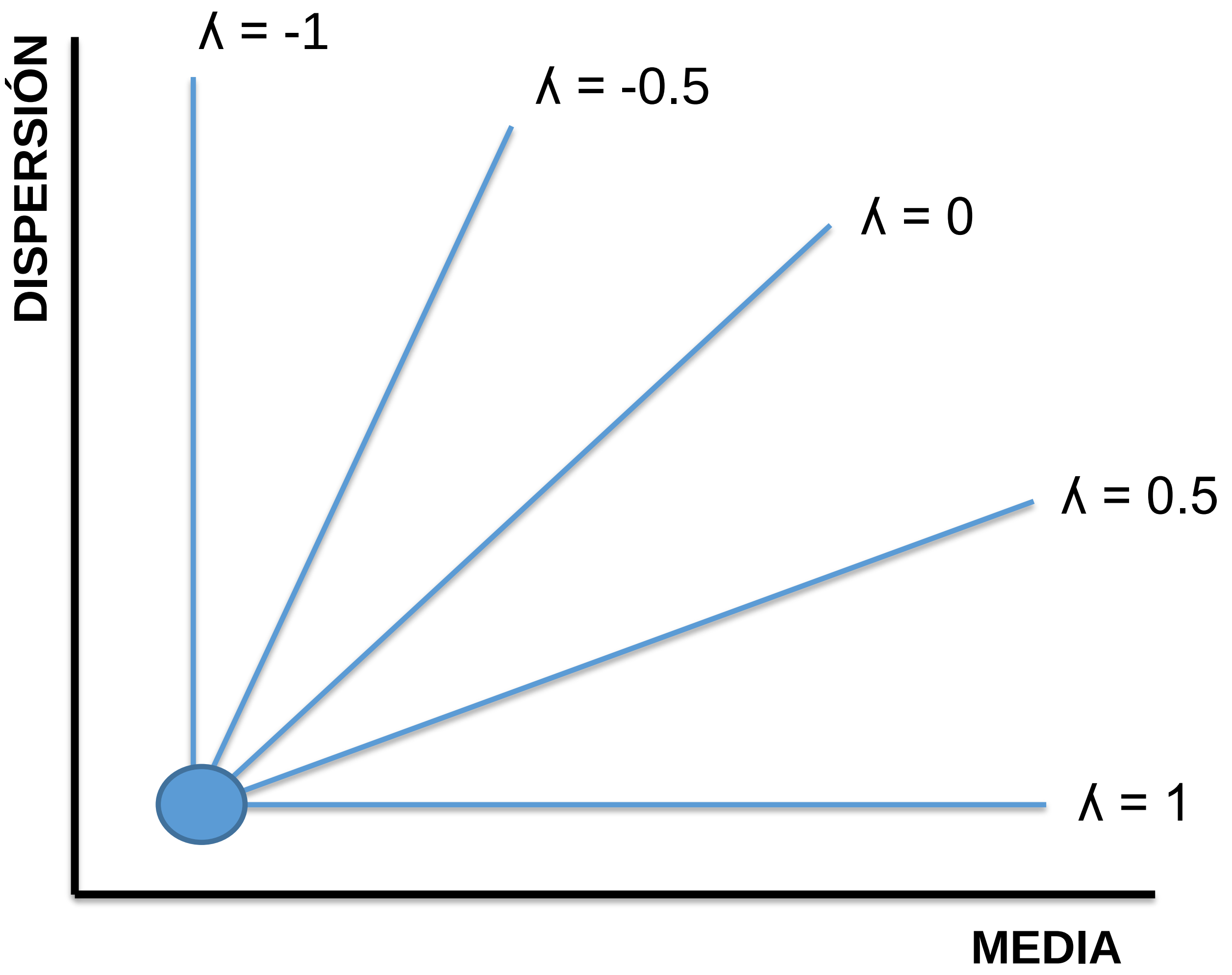
Estimar los parámetros del modelo con métodos estadísticos.
Analizar la significancia de los parámetros estimados.

VERIFICACIÓN

Análisis de residuales del modelo: Aleatoriedad y Normalidad.
Validar los supuestos del modelo.

PRONÓSTICOS

TRANSFORMACIÓN DE BOX COX



$$y(\lambda) = \begin{cases} \frac{y^{(\lambda)}-1}{\lambda}, \lambda \neq 0 \\ ln(y), \lambda = 0 \end{cases}$$

LAMBDA	TRANSFOMRACIÓN
1	No existe transfomación a la serie
0.5	Raíz cuadrada a la serie
0.125	Raíz cuarta
0	Logaritmo neperiano
-0.125	Inversa de la raíz cuarta
-0.5	Inversa de la raiz cuadrada
-1	Inversa de la serie

MODELOS AR(P)

DEFINICIÓN

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \phi_4 Z_{t-4} + \ldots\ldots\ldots\ldots\ldots + \phi_p Z_{t-p} + \varepsilon_t$$

ε_t ~ (0,Var(error)) Var(error) = cte ε_t = Aleatorios independientes

ESPERANZA

$$E(Z_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \ldots\ldots\ldots - \phi_P}$$

VARIANZA

$$VAR(Z_t) = \frac{Var(Error)}{1 - \phi_1\rho_1 + \phi_2\rho_2 + \ldots\ldots\ldots + \phi_p\rho_P}$$

ρ_p Autocorrelación de la serie

CONDICIONES DE ESTACIONARIEDAD E INVERTIBILIDAD

1.- ESTACIONARIEDAD

Una serie de tiempo es estacionaria si tienen medias y varianzas finitas, no dependan del tiempo. Es decir cuando las raíces del polinomio que la generan tengas raíces mayores que la unidad.

2.- INVERTIBILIDAD

Una serie de tiempo AR(p) es invertible cuando es posible expresar en su forma autorregresiva

3.- CORRELOGRAMA

El análisis estadístico y la inspección gráfica visual de los valores muestrales tanto de las funciones de autocorrelación como los funciones de autocorrelación parcial constituyen lo que se llama “Análisis del correlograma”

MODELOS MA(q)

DEFINICIÓN

$$Z_t = \theta_0 + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \theta_4\varepsilon_{t-4} + \ldots\ldots\ldots\ldots\ldots + \varepsilon_{t-q} + \varepsilon_t$$

ε_t ~ (0,Var(error)) Var(error) = cte ε_t = Aleatorios independientes

ESPERANZA

$$E(Z_t) = \theta_0$$

VARIANZA

$$VAR(Z_t) = Var(Error) * (1 + \theta_1^2 + \theta_2^2 + \ldots\ldots\ldots\ldots.. + \theta_q^2)$$

CONDICIONES DE ESTACIONARIEDAD E INVERTIBILIDAD

1.- ESTACIONARIEDAD

La serie MA(q) es estacionario dado que no existen tendencia siempre y cuando la V(Error) sea constante en el tiempo.

2.- INVERTIBILIDAD

Una serie de tiempo MA(q) es invertible cuando $|\theta_j| < 1$

3.- CORRELOGRAMA

El análisis estadístico y la inspección gráfica visual de los valores muestrales tanto de las funciones de autocorrelación como los funciones de autocorrelación parcial constituyen lo que se llama "Análisis del correlograma"

MODELOS ARMA(p,q)

DEFINICIÓN

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots\ldots. + \phi_p Z_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \varepsilon_t$$

ε_t ~ (0,Var(error)) Var(error) = cte ε_t = Aleatorios independientes

ESPERANZA

$$E(Z_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \ldots\ldots\ldots - \phi_P}$$

SERIES NO ESTACIONARIAS

INTRODUCCIÓN

También llamados Procesos Arima, presentan en su mayoría series de tiempo con un grado de no estacionariedad en media y/o varianza. Consiste en realizar diferencia simples y/o realizar transformaciones sobre la serie observada, con la finalidad de obtener una serie estacionaria. Dentro de esta clase de Modelos se encuentran: ARIMA(p,d,q), ARI(p,d,0) y los IMA(0,d,q).

PROCESOS ARIMA(p,d,q)

$$\Delta^d Z_t = \varphi(B) Z_t + \omega(B)\varepsilon_t$$

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots \ldots . + \phi_p Z_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \varepsilon_t$$

ESPECIFICACIÓN DE UN MODELO ARIMA

PASO 1: No estacionariedad en varianza

a) **Gráfica de la serie:** Se da cuando en la gráfica de la serie existe mucha volatilidad con respecto a su valor promedio. Si esto ocurre hay que hacer transformaciones de BOX – COX
b) **Grafica de dispersión vs media:** Esta grafica es importante por que dá una idea que tipo de transformación hay que realizar a la serie.
c) **Determinar el valor optimo de (ʎ):** Con la finalidad de encontrara la transformación de BOX – COX más optima.

ESPECIFICACIÓN DE UN MODELO ARIMA

PASO 2: No estacionariedad en media

a) **Gráfica de la serie:** Se da cuando en la gráfica de la serie existe tendencia. Si esto ocurre hay que tomar tantas diferencias sea necesaria a la serie, hasta obtener un grafica de una serie sin tendencias.
b) **Gráfica de correlograma:** Cuando no hay un decaimiento rápido en las funciones de autocorrección parcia (facp) y autocorrelación (fac), entonces las serie no es estacionaria en media. Sin embargo el correlograma de la serie diferenciada si existe un decaimiento rápido, si esto ocurre la serie es estacionaria en media.

PASO 3: Determinación de los valores de "p" y "q"

a) **Valores de p:** Aquí hay que ver el decaimiento rápido de la facp del correlograma
b) **Valores de q:** Aquí hay que ver el decaimiento rápido de la fac del correlograma

SERIES ESTACIONALES

INTRODUCCIÓN

También llamados Procesos Sarima, presentan en su mayoría series de tiempo con un grado de no estacionariedad en media y/o varianza en patrones de intervalos de observación menores a un año. Estas variaciones sistemáticas inferiores a un año son llamados "factores estacionales". Consiste en realizar diferencia simples y/o realizar transformaciones sobre los intervalos de observación inferiores a un año, con la finalidad de obtener una serie no estacional.

PROCESOS SMA(Q) -- SIMA(dq)(D,Q)

$$Z_t = \phi_0 + \theta_1 \varepsilon_{t-1s} + \theta_2 \varepsilon_{t-2s} + \cdots + \varepsilon_t$$

PROCESOS SAR(P) -- SARI(pd)(PD)

$$Z_t = \phi_0 + \phi_1 Z_{t-1s} + \phi_2 Z_{t-2s} + \ldots\ldots + \varepsilon_t$$

PROCESOS SARMA(P,Q) -- SARIMA(pdq)(PDQ)

$$Z_t = \phi_0 + \phi_1 Z_{t-1s} + \phi_2 Z_{t-2s} + \ldots\ldots + \phi_p Z_{t-ps} + \theta_1 \varepsilon_{t-1s} + \theta_2 \varepsilon_{t-2s} + \cdots + \varepsilon_t$$

RESUMEN

		NO ESTACIONALES (SIN ESTACIÓN)	ESTACIONALES (CON ESTACIÓN)
ESTACIONARIO (SIN TENDENCIA)	AUTOREGRESIVOS	AR(p)	SAR(P)
	MEDIAS MOVILES	MA(q)	SMA(Q)
	MIXTOS	ARMA(p,q)	SARMA(P,Q)
NO ESTACIONARIO (CON TENDENCIA)	AUTOREGRESIVOS	ARI(p,d)	SARI(pd)(P,D)
	MEDIAS MOVILES	IMA(d,q)	SIMA(dq)(SQ)
	MIXTOS	ARIMA(p,d,q)	SARIMA(pdq)(PDQ)

SERIES HETEROCEDÁSTICAS

Introducción

Una extensión de los modelos Box y Jenkins, los constituyen los modelos heterocedásticos ARCH y GARCH, los cuales permiten analizar los efectos no lineales en la varianza de los residuos, dicho modelos fue estudiado en un instante por Engle (1982). Estos modelos pueden ser combinados con la estructura de los modelos ARIMA y de Regresión.

Estos modelos en la literatura estadística son llamados modelos con heterocedasticidad condicional y dependiendo de su estructura puede ser autoregresivos (ARCH) o de una estructura generalizada de tipo ARMA (GARCH). En ambos casos el interés es modelar la varianza como función de sus valores pasados.

MODELOS ARCH

Modelo ARCH(1)

$$Z_t = \delta_t \varepsilon_t$$

$$\delta_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 \quad \alpha_0 > 0 \quad \alpha_1 > 0$$

$$\varepsilon_t \sim N(0,1)$$

$$\delta_t^2 \ : \ \ Volatilidad \quad \varepsilon_t \ : \ \ Ruido\ Blando$$

$$\varepsilon_t \ \ y \ \ \delta_t^2 \ \ son\ independientes$$

Esperanzas Marginales y Condicionales

$$E(Z_t) = E(\delta_t)E(\varepsilon_t) = 0$$

$$E(Z_t|Z_{t-1}) = E(\delta_t\varepsilon_t|Z_{t-1})$$

$$E(Z_t|Z_{t-1}) = \delta_t E(\varepsilon_t|Z_{t-1}) = 0$$

Varianza Condicional

$$V(Z_t|Z_{t-1}) = E(Z_t^2|Z_{t-1}) - (E(\varepsilon_t|Z_{t-1})^2$$

$$V(Z_t|Z_{t-1}) = E(\delta_t^2\varepsilon_t^2|Z_{t-1})$$

$$\boxed{V(Z_t|Z_{t-1}) = \delta_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2}$$

Varianza Marginal

$$V(Z_t) = E(Z_t^2) - E(Z_t)^2$$

$$V(Z_t) = E(\alpha_0 + \alpha_1 Z_{t-1}^2)$$

$$V(Z_t) = \alpha_0 + \alpha_1 E(Z_{t-1}^2)$$

$$\boxed{V(Z_t) = \frac{\alpha_0}{1-\alpha_1}}$$

Kurtosis

$$\boxed{K = 3\,\frac{1-\alpha_1^2}{1-3\alpha_1^2}}$$

Conclusiones

- ✓Las esperanzas marginales y condicionales son ceros.
- ✓La varianza marginal es constante.
- ✓La varianza condicional depende de los rezagos pasados.
- ✓La distribución del modelo ARCH(1) es desconocida pero es leptocurtica dado que K > 3 .

Modelo ARCH(q)

$$Z_t = \delta_t \varepsilon_t$$

$$\boxed{\delta_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Z_{t-i}^2}$$

$$\varepsilon_t \sim N(0,1)$$

$$\sum_{i=1}^{q} \alpha_i < 1$$

$$\boxed{V(Z_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i}}$$

MODELOS GARCH

Es una generalización de los modelos ARCH, nace debido a que los modelos ARCH necesita un numero elevado de retardos para describir la volatilidad. Los modelos ARCH consiguen describir el comportamiento de la varianza condicionada, pero no explican las causas de dicho comportamiento.

Modelo GARCH(p,q)

$$Z_t = \delta_t \varepsilon_t \qquad \boxed{\delta_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{p} \beta_j \delta_{t-j}^2}$$

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \qquad \alpha_i > 0 \ y \ \beta_j > 0$$

$$E(Z_t) = 0 \qquad V(Z_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

$$COV(Z_t, Z_s) = 0$$

PRONÓSTICOS

¿Qué son los pronósticos?

Son métodos estadísticos que se utiliza para predecir comportamientos futuros de datos a partir del aprendizaje de una data histórica. Es decir los pronósticos se basan en patrones de datos ya existentes.

Proceso estocástico $(Z_1, Z_2, Z_3, \ldots \ldots \ldots, Z_t)$

Pronóstico $E(Z_{t+h} \mid Z_1, Z_2, Z_3, \ldots \ldots \ldots, Z_t) = Z^*_{t+h}$

Serie

$$Z_t = \phi_0 + \phi_1 Z_{t-1s} + \phi_2 Z_{t-2s} + \ldots \ldots . + \phi_p Z_{t-ps} + \theta_1 \varepsilon_{t-1s} + \theta_2 \varepsilon_{t-2s} + \cdots + \varepsilon_t$$

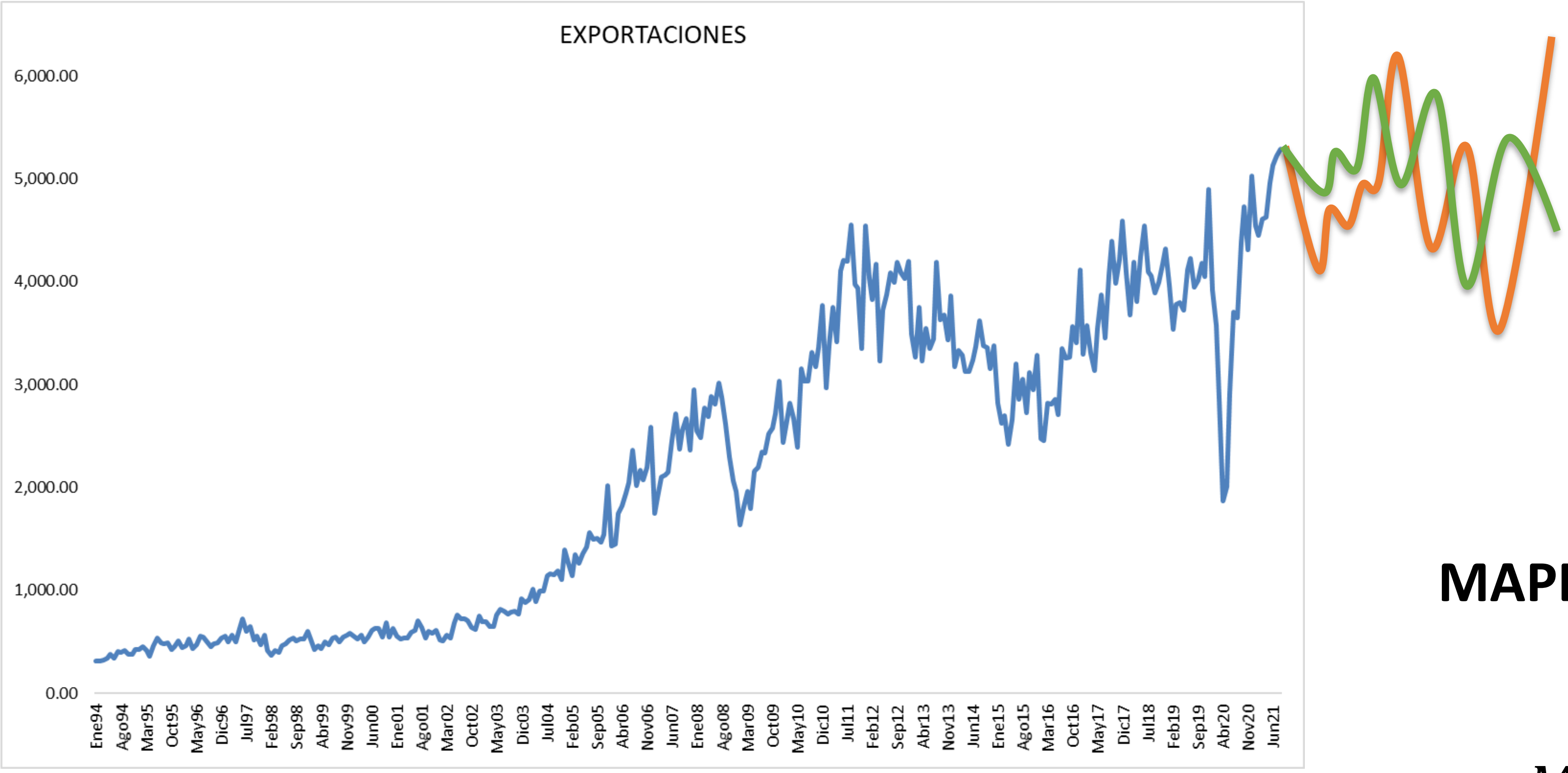
Pronóstico estático

$$\underbrace{Z_1 \quad Z_2 \qquad\qquad Z_t}_{\textbf{Data histórica}} \qquad \underbrace{Z_{t+h} \quad Z_{t+h} \quad Z_{t+h} \qquad Z_{t+h}}_{\textbf{Pronósticos constantes en K periodos}}$$

Pronóstico dinámicos

$$\underbrace{Z_1 \quad Z_2 \qquad\qquad Z_t}_{\textbf{Data histórica}} \qquad \underbrace{Z_{t+1} \quad Z_{t+2} \quad Z_{t+3} \qquad Z_{t+h}}_{\textbf{Pronósticos variables en K periodos}}$$

Pronóstico estático



Pronósticos

Valor real en el futuro

Data histórica

Pronóstico dinámicos

Métricas de performance en la predicción

MAE (Error Absoluto Medio)

$$MAE = \frac{\sum_{k=1}^{k=t}(Z_k - Z_k^*)}{t}$$

MAPE (Error Porcentual Medio)

$$MAPE = \frac{\sum_{k=1}^{k=t}(\frac{MAE_t}{Z_t})}{t}$$

APLICACIONES

CASO IMPORTACIONES Y EXPORTACIONES

UNIVERSIDAD NACIONAL DE INGENIERÍA

ESCUELA CENTRAL DE POSGRADO

Facultad de Ingeniería Económica, Estadística y CCSS

Unidad de Posgrado - FIEECS

Curso: Forecasting for Data Science

Docente: Mg. Omar A. Chincaro Del C.