Tarea 01

Modelos de Supervivencia y Series de Tiempo

Aaron Mauricio Gómez   Carlos Brito Mejía   Diego Luna
Francisca Vilca Sánchez

Usted invierte dinero en la compra de una acción. Para saber qué tan viable es su inversión, ha decidido hacer un análisis de serie de tiempo. Este tipo de análisis se utiliza para examinar los datos históricos y predecir futuros cambios en el valor de una inversión. Antes de invertir, es importante informarse y evaluar su disposición al riesgo y objetivos, así como hacer su perfil de inversionista. También es fundamental conocer los fundamentos de las inversiones y la rentabilidad esperada. Además, existen diversas metodologías que pueden ayudar a desarrollar estrategias efectivas para la inversión.

Para su tarea busquen en Yahoo finanzas una acción que les llame su atención para su análisis. Yahoo finanzas es una de las principales páginas en la que usted podrá consultar información financiera, en ella además le permitirá ver gráficamente los datos y lo más importante descargar los datos en formato csv. Descargue los datos históricos del **1 de enero de 2001** al **30 Junio de 2020**, y en frecuencia indique periodicidad mensual, guarde estos datos en su carpeta de trabajo e impórtelos en R.

Con base a dichos valores realice:

1. Descargue e importe los datos en R y con ellos convierta el vector de precios a un objeto de series de tiempo y vea el comportamiento de sus datos, es decir calcule la media, moda, cuartiles, máximos, mínimos y varianza. Haga una gráfica de caja para ver visualmente estos resultados.

Nuestro interés está en ver la posibilidad de invertir en las acciones de *CEMEX* una empresa ligada al ámbito de la construcción que se autodenomina ser líder en la venta y producción de Cemento y otros materiales. Para saber si es una buena opción hacer está inversión revisaremos los datos históricos desde el 1 de enero de 2001 al 30 de junio de 2020, usando algunos comandos y librerías de R extraeremos los datos desde Yahoo! Finanzas, tal como se muestra en el siguiente código:

```r
getSymbols("CX", src = "yahoo", from = "2001-01-01", to = "2020-06-30",
           periodicity= "monthly")
base <- CX[,6] # para solo usar la columna de los precios de cierre ajustados
ts_base <- ts(base, start=2001, frequency = 12) # creamos el objeto de series de tiempo
```

Una vez que tenemos la base y el objeto series de tiempo comenzamos a hacer un análisis del comportamiento de los datos obteniendo algunos datos interesantes, que se pueden ver resumidos en la Tabla 1:

	Valores de la acción
Media	9.703
Moda*	7.463
Mínimo	2.120
Cuartil 0.25	6.322
Mediana	7.961
Cuartil 0.75	10.284
Máximo	27.697
Varianza	31.298

Note:

* = Es el valor de la locación estimada de la moda

Tabla 1: Resumen del comportamiento de los datos

Como ya analizamos en valor promedio de la acción es de *9.703*, al graficar el mínimo observamos que se alcanza en marzo de 2020 con valor al cierre ajustado de *2.120*, que como se puede inferir se debió a la pandemia, ya que, al entrar en confinamiento la producción y venta de cemento se vio afectada, después de tocar su mínimo desde 2001, la acción se revalua y empieza a subir de precio, observamos que su mejor precio fue en mayo de 2007 con un valor de 27.697, despues tuvo fluctuaciones y pendientes negativas que coinciden con los años de la crisis de 2008, la gripe porcina 2009, la recesión de 2012, una caída en 2016, y después una caída alcanzando su mínimo hasta la pandemia de COVID-19.

Además a través de un gráfico de caja se puede ver el resumen de estas medidas, tal como se muestra en la Figura 1:

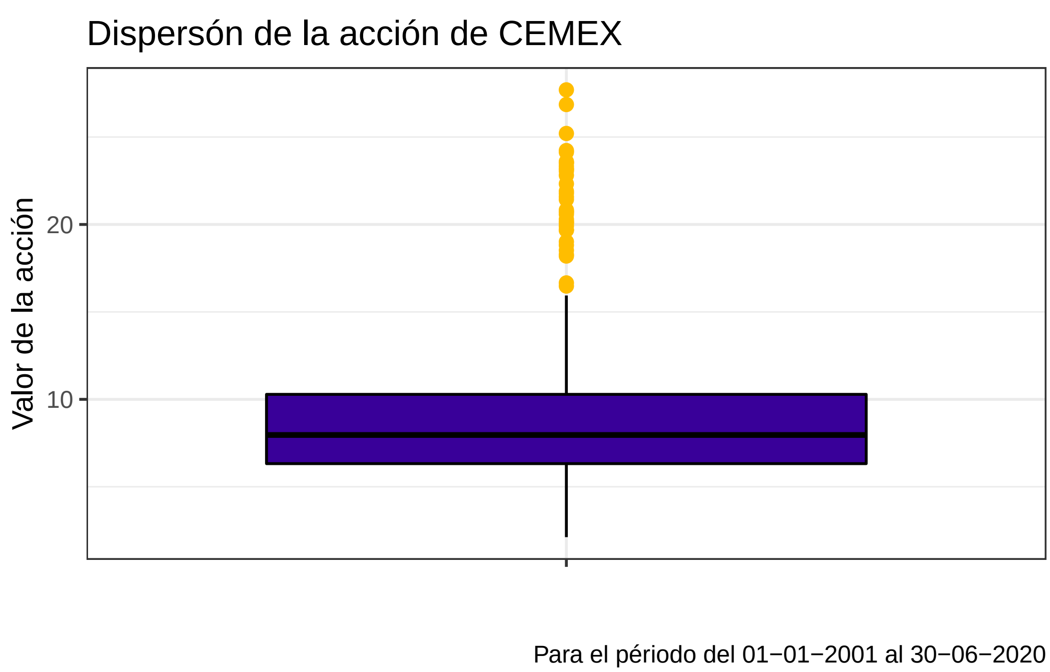


Figura 1: Boxplot de la acción de CEMEX

Graficamente, podemos observar que existen muchos outliers, y hay 20 puntos de varianza con la media y el maximo, eso nos indica que esta muy lejos de el valor que llego a tener la acción.

2. Haga un análisis de sus datos, determinar si hay tendencias crecientes, decrecientes y periodicidad.

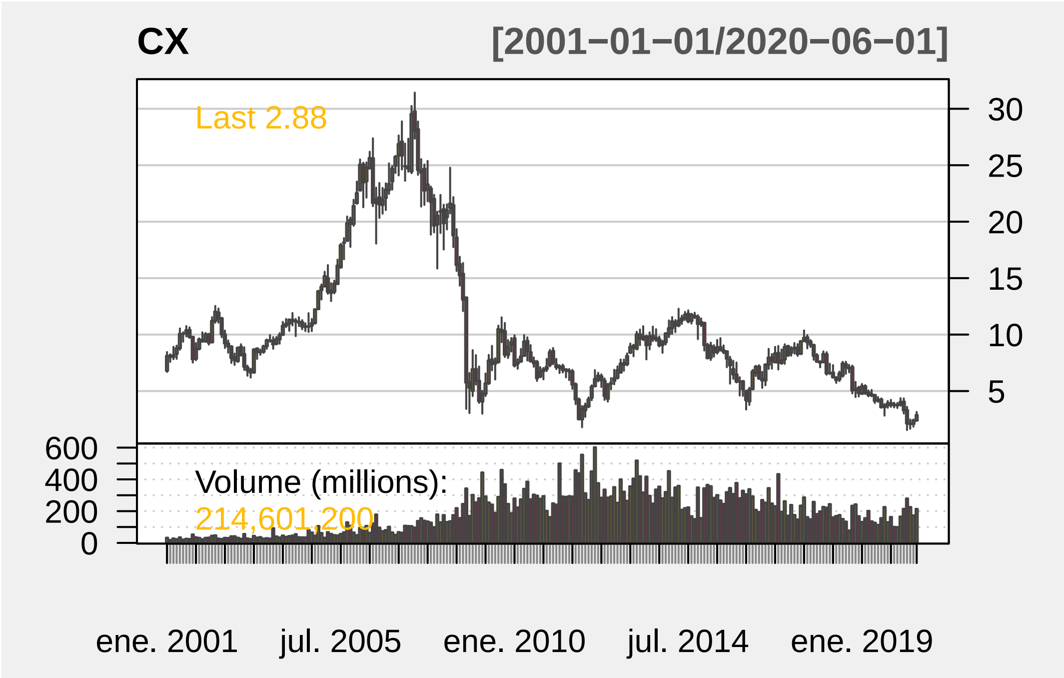


Figura 2: Gráfico Financiero para la acción CX

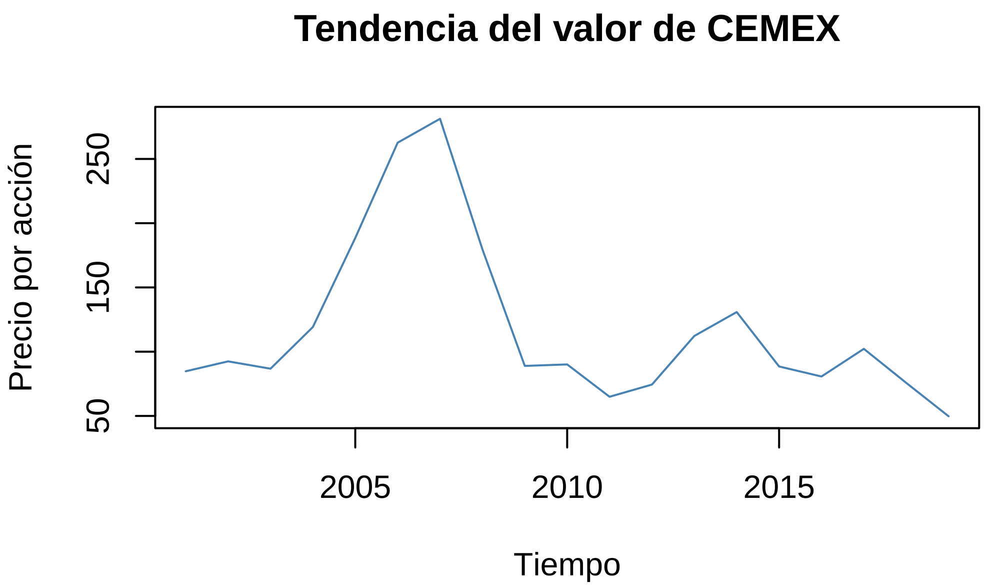


Figura 3: Gráfico de Tendencia CEMEX

Como podemos observar no se ven tendencias a lo largo de los ultimos años, mas bien se ve que no existe tendencias mas allá de las crisis ya descritas, tampoco de observa periodicidad.

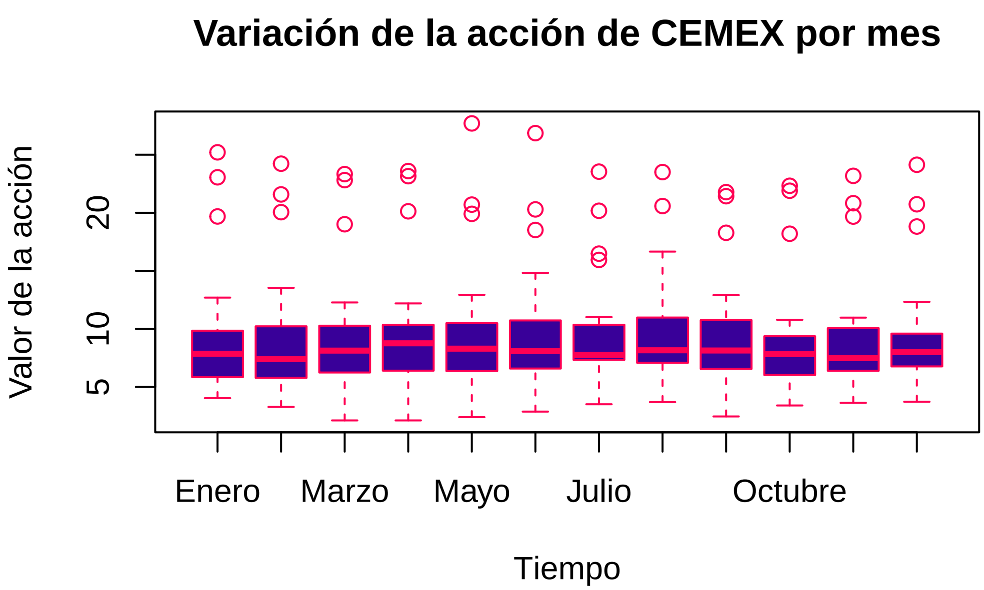


Figura 4: Boxplot para cada mes para la acción CEMEX

En los boxplots mensuales podemos ver casi una recta constante cerca de la media del precio aunque con muchos outliers.

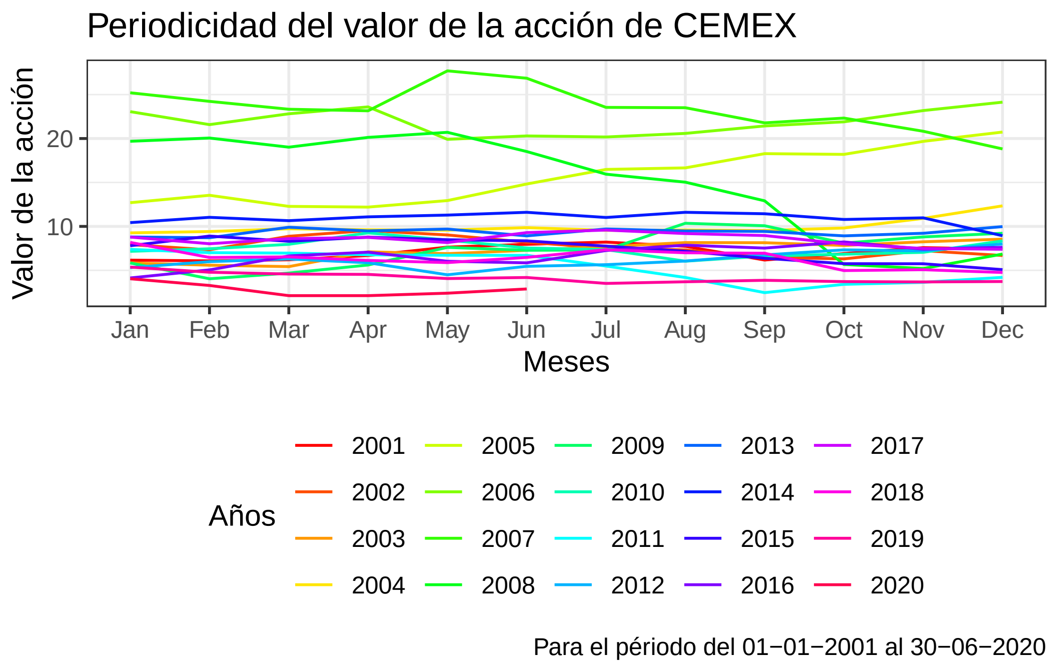


Figura 5: Gráfico de líneas para estudiar periodicidad de la acción de CX

Es más facil observar en esta gráfica que no existe estacionalidad ni periodicidad, ya que cada año la grafica cambia mucho en contraste con los otros años.

3. Realice una gráfica de a serie de tiempo usando las metodologías vistas en clases (ggplot2,plot y xts)

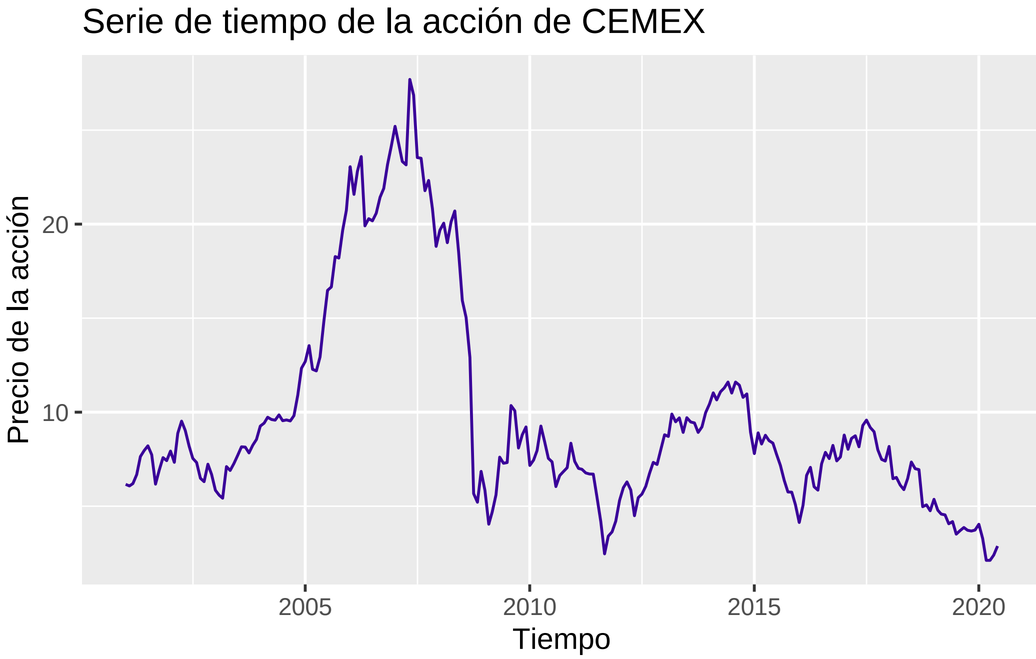


Figura 6: Serie de Tiempo de la acción de CEMEX

Esta es la grafica obtenida al usar la libreria ggplot2, que para lo cual como ya vimos necesitamos hacer una base de datos.

Ahora, usando la serie de tiempo que ya creamos de nuestros datos podemos usar la función plot como se observa a continuación.

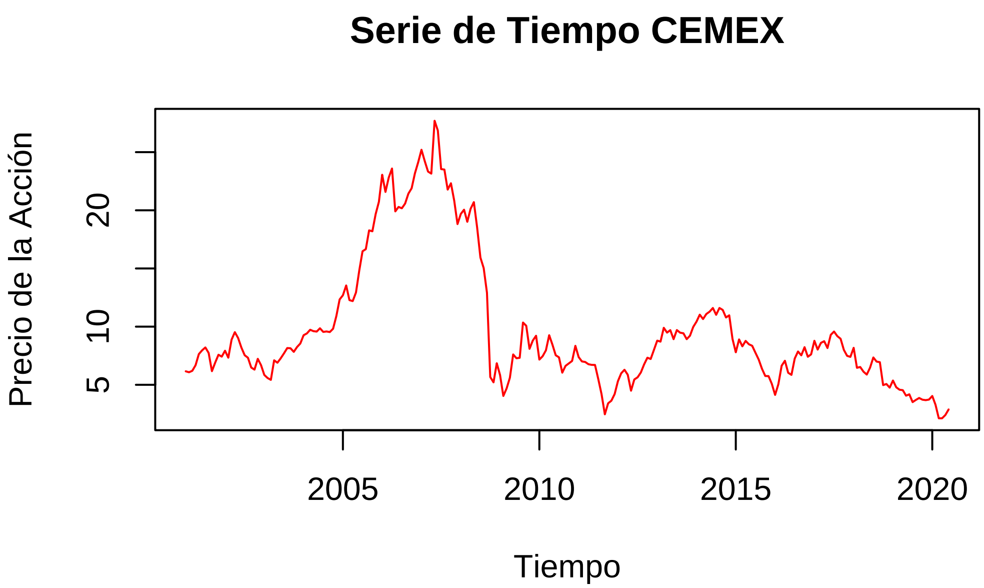


Figura 7: Serie de Tiempo de la acción de CEMEX con plot

Otra forma vista para graficar es usando la libreria xts:

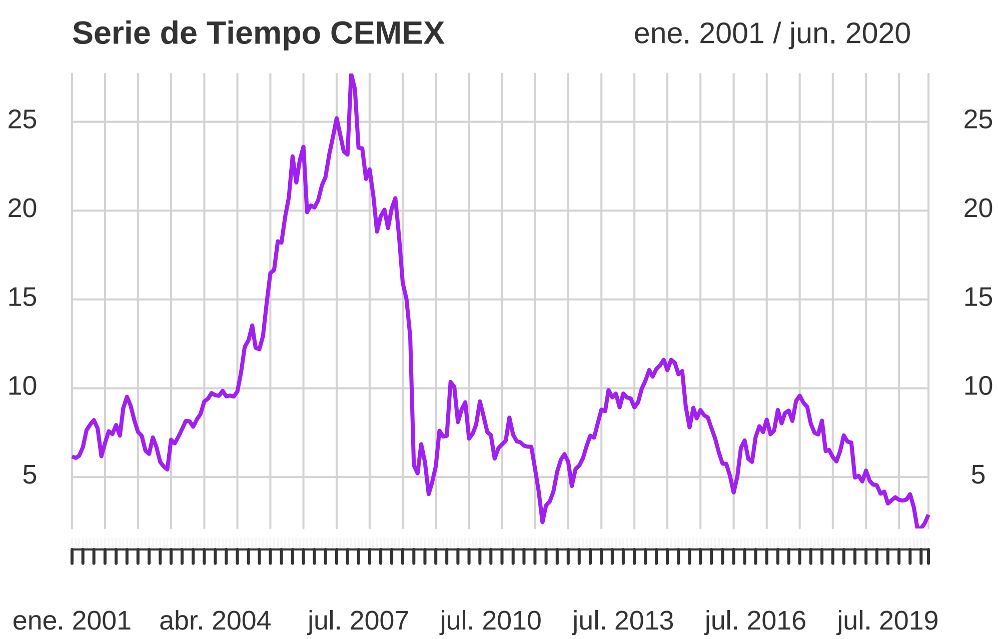


Figura 8: Serie de Tiempo de la acción de CEMEX con xts

Esta opción nos agrega lineas verticales que nos hacen mas fácil la interpretación de los datos, ya que es más fácil ubicar la temporalidad de la serie.

4. Realiza las siguientes suavizaciones:

- Suavización de promedios móviles simple (PS) de orden 2 y 10.

```
p.mov.2 = ma(ts_base, order = 2)
p.mov.10 = ma(ts_base, order = 10)
```

- Suavización de promedios móviles ponderados (PPM) de orden 4 con los pesos $(0.5, 0.2, 0.2, 0.1)$ para $(x_t, x_{t-1}, x_{t-2}, x_{t-3})$ y otra modelo (PPM) de orden 12 seleccionando ustedes mismos los pesos y mencionado el porque de esos pesos.

```
pesos.b1 <- c(0.5,0.2,0.2,0.1)
p.mov.pond.4 <- stats::filter(ts_base, pesos.b1, sides=1)

ts <- c(ts_base, rep(0,6))
ts_matrix <- matrix(ts, ncol = 12, byrow = T)
suma <- sum(colMeans(ts_matrix))
pesos.b2 <- round(colMeans(ts_matrix)/suma,4)

p.mov.pond.12 <- stats::filter(ts_base, pesos.b2, sides=1)
```

Para obtener nuestros pesos, al notar que estamos tomando en cuenta el año anterior, decidimos hacer el promedio historico de cada mes y así obtuvimos el valor de cada alfa para cada mes, de esta forma intentamos minimizar la varianza a corto plazo ya que nuestros alfas son pequeños.

- Suavización de promedios móviles simples centrados (PSC) de orden 2 y 10.

```
p.mov.2.c = ma(ts_base, order = 2, centre = TRUE)
p.mov.10.c = ma(ts_base, order = 10, centre = TRUE)
```

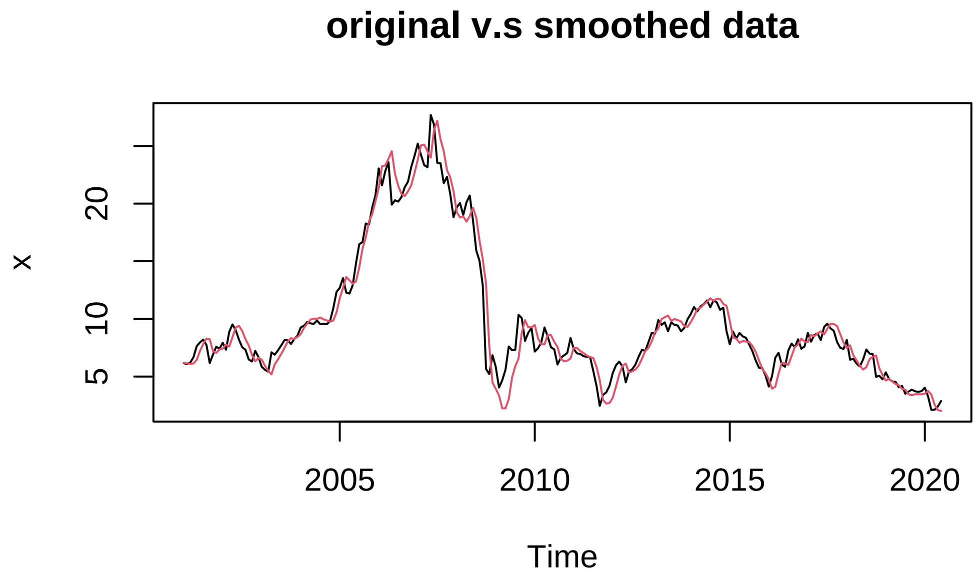
- Suavización exponencial simple con $\alpha = 0.01, \alpha = 0.5, \alpha = 0.99$, ¿Cuál es el papel de α en la estimación de valores pronostico?

```
ts.base.sua_001 = ses(ts_base, alpha=0.01, initial = 'simple')
ts.base.sua_05 = ses(ts_base, alpha=0.5, initial = 'simple')
ts.base.sua_099 = ses(ts_base, alpha=0.99, initial = 'simple')
```

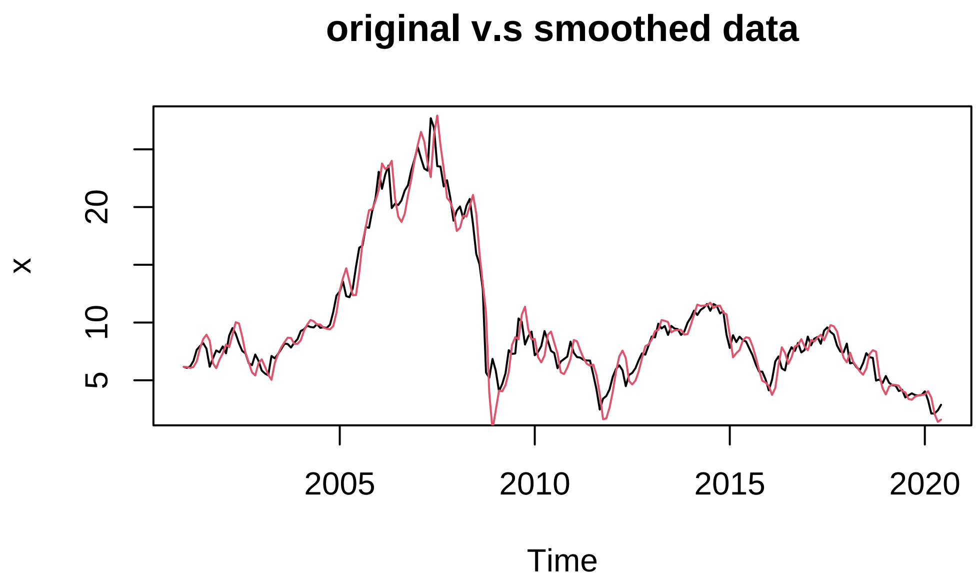
Como vimos en clases el papel de las alphas es suavizar la varianza de nuestros pronosticos, en este caso por el ejemplo para $ =0.01 nos indica que los valores mas recientes de nuestra serie de tiempo un peso de 0.01 es decir muy poco peso a los datos mas reciente en consecuencia se le da mas pesos a los datos mas viejos.

- Suavización Holt, con $\alpha = 0.5$ y $\beta = 0.2, \beta = 0.8$ ¿Cuál es el papel de β en la estimación de valores pronostico?

```
ts.holt_0502 = Holt(ts_base, alpha = 0.5, beta=0.2)
```




```
ts.holt_0508 = Holt(ts_base, alpha = 0.5, beta=0.8)
```



El papel de las betas es suavizar o exponenciar los cambios en las tendencias, una beta pequeña suaviza las tendencias recientes, es decir les da menor peso.

- Suavización Holt-Winters, con $\alpha = 0.5y\beta = 0.2, \gamma = 0.8$ y otro modelo Holt-Winterscon una selección automática de R.

```
hw.ts.1 = hw(ts_base, h=12,optim.start = c(0.5,0.2,0.8))     # :c
hw.ts.2 = hw(ts_base, h=12, seasonal='additive')
```

Notemos que en este modelo el valor de gamma suaviza la estacionalidad.

5. Cada inciso del punto anterior tiene dos o tres modelo del mismo tipo, así que una misma gráfica muestra los resultados junto con la serie original por modelo.

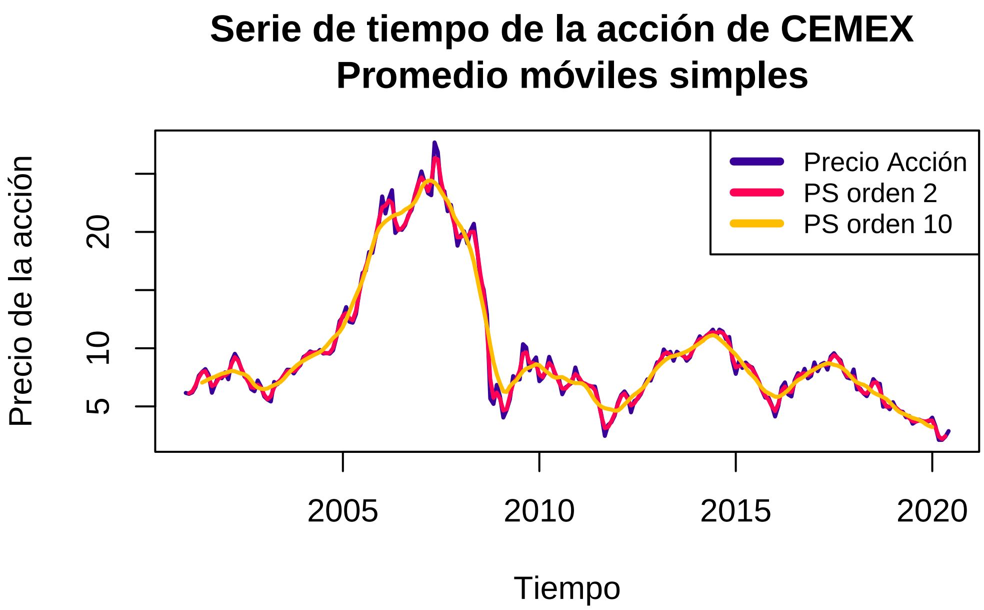


Figura 9: Serie de Tiempo de la acción de CEMEX junto con el ajuste de Promedio móviles simples

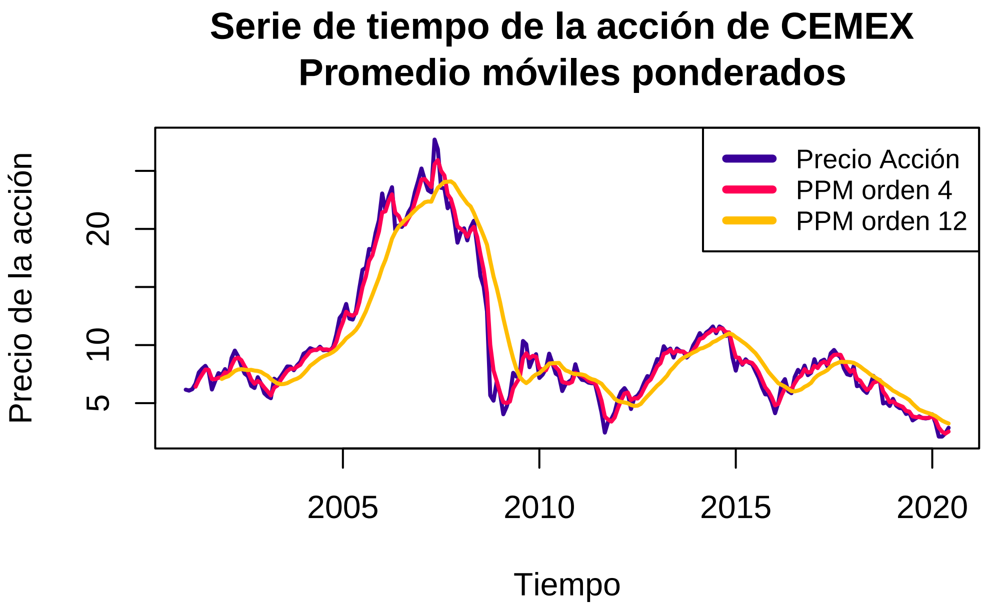


Figura 10: Serie de Tiempo de la acción de CEMEX junto con el ajuste de Promedio móviles ponderados

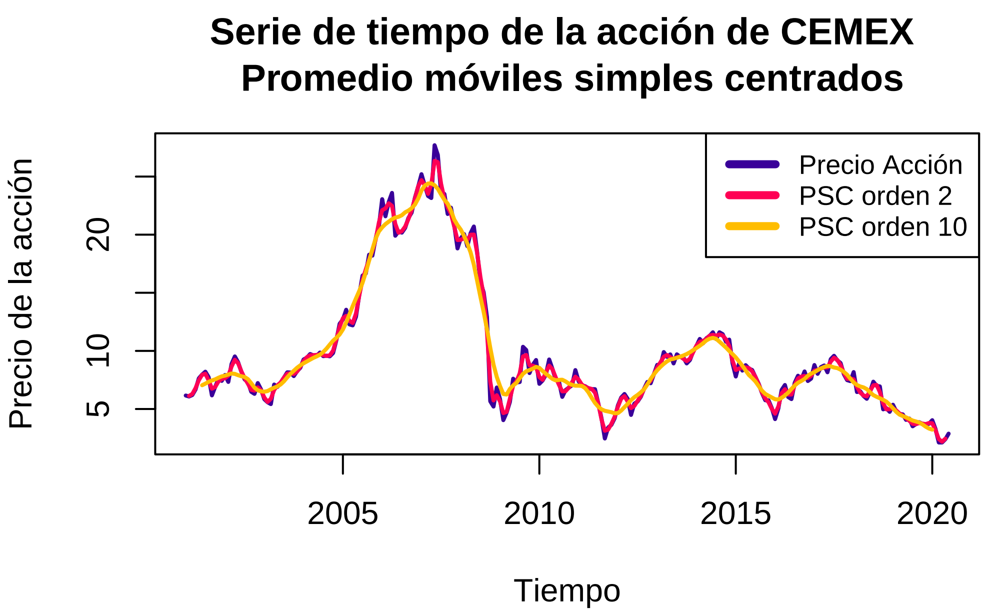


Figura 11: Serie de Tiempo de la acción de CEMEX junto con el ajuste de Promedio móviles simples centrados

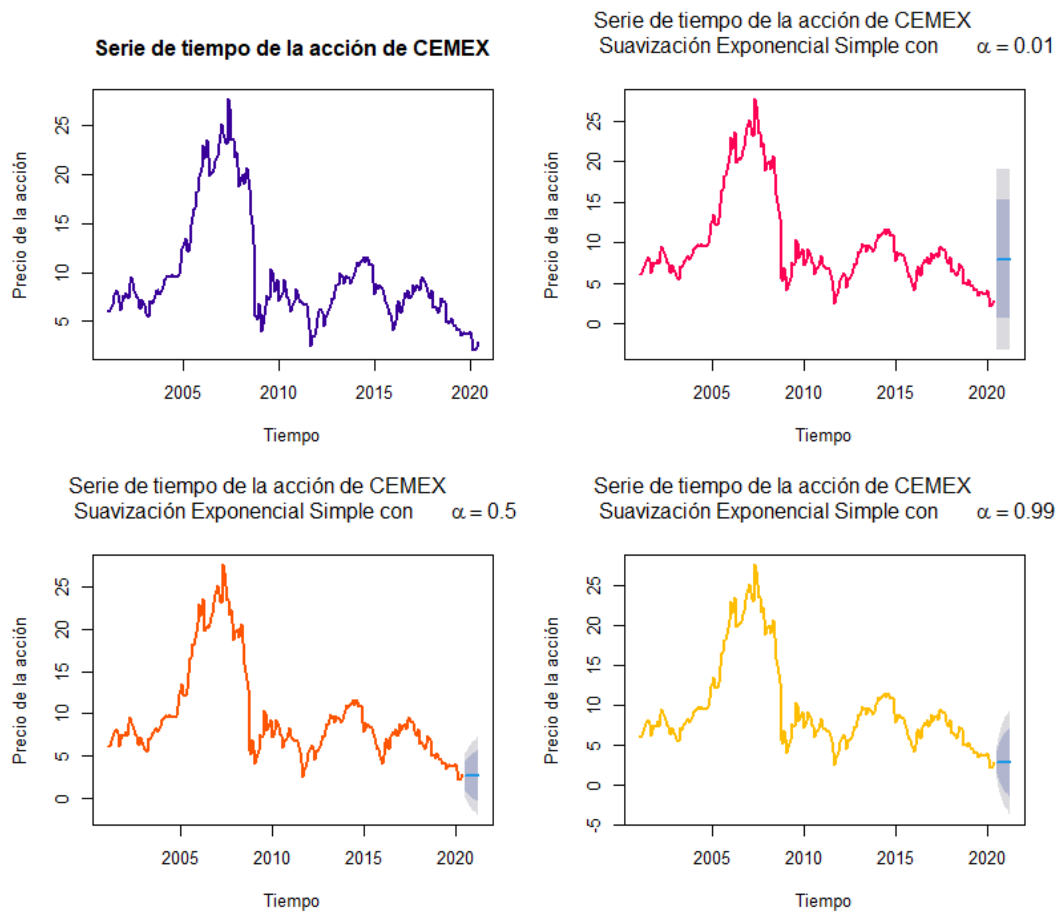


Figura 12: Serie de Tiempo de la acción de CEMEX junto con la suavización exponencial

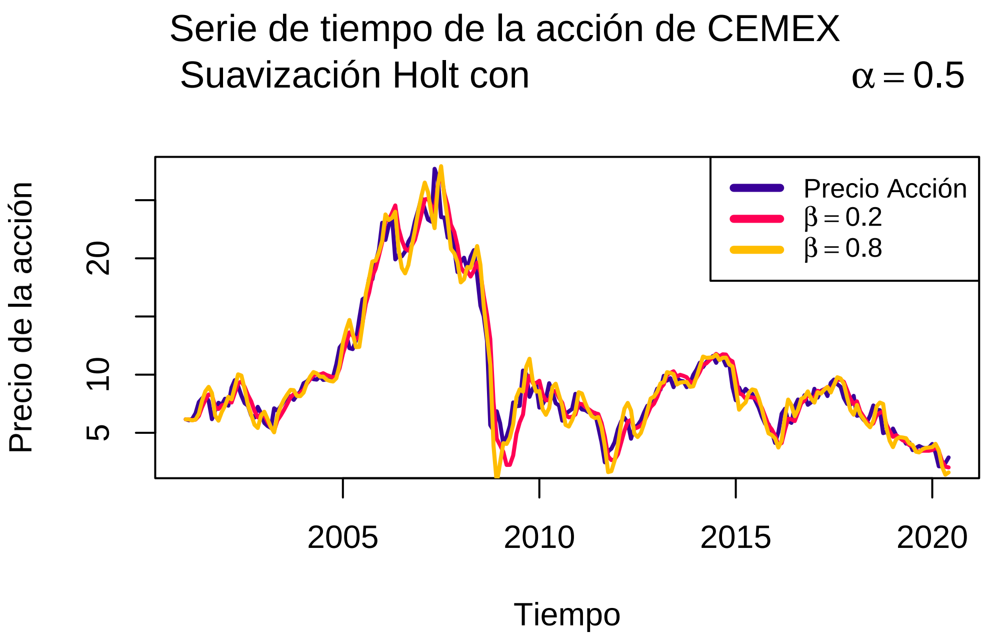


Figura 13: Serie de Tiempo de la acción de CEMEX junto con la suavización Holt

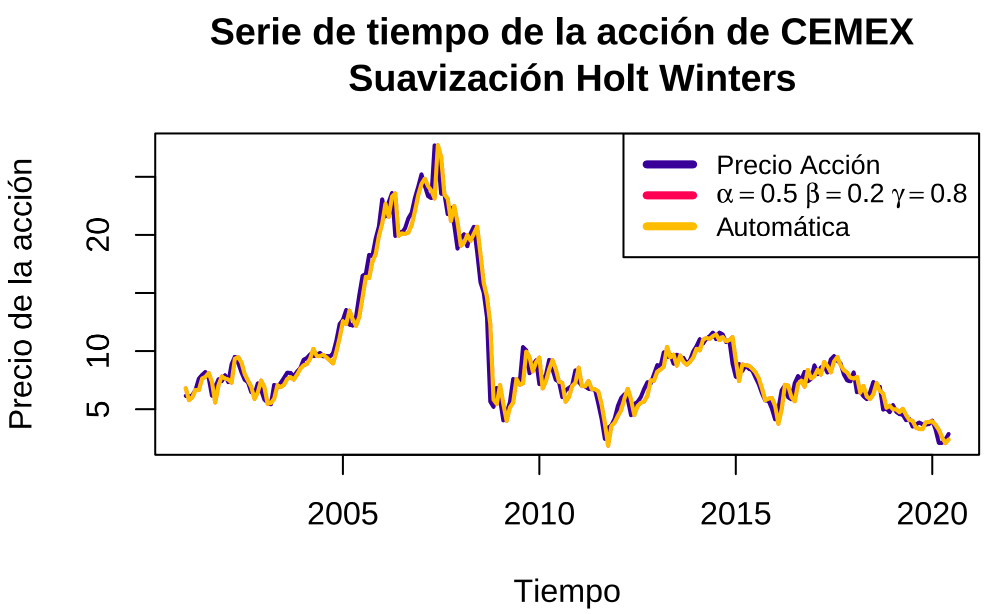


Figura 14: Serie de Tiempo de la acción de CEMEX junto con la suavización Holt-Winters

6. Seleccione el mejor modelo y mencione el por qué de su elección.

El mejor modelo es: el de la Suavización Holt con alpha = 0.5 y Beta = 0.2; veamos porque:

¿Por qué elegir un modelo Holt? Tanto teóricamente como dadas las visualizaciones de los modelos de series de tiempo considerados en la sección anterior, es inmediato concluir que necesitamos o un modelo Holt o un modelo Holt-Winters porque necesitamos un modelo que considere la tendencia.

Ahora ¿Por qué consideramos que es mas apropiada la suavización Holt que la suavización Holt-Winters? Porque en esta acción, en el inciso 2 detectamos que no hay una estacionalidad o periodicidad significativa. Esto es importante porque ambos modelos consideran la tendencia en la suavización, sin embargo el Holt-Winters toma en cuenta la estacionalidad, los patrones ciclicos, por lo que es más apropiado un modelo Holt ya que:

- La suavización Holt es mas útil cuando la serie no tiene un patrón estacional o cuando es muy débil.
- Es mas simple y esto lo hace mas sencillo de interpretar, describir, presentar y trabajar.
- Ya que el Holt-Winters estaría haciendo más complejo (quizás más de lo que sería suficiente) corremos el riesgo de estar "sobre-ajustando" la serie,

Ya que decidimos optar por el modelo de Suavización Holt; ¿Por qué esos parámetros? Simplemente porque se ajustan mejor a la serie de tiempo. Observese que escoger una Beta tan "pequeña" significa que le estamos dando un menor peso a las tendencias recientes (Por el contrario el modelo con Beta = 0.8 les da mucho peso).

7. Utilizar la autocorrelación para determinar si hay patrones en los residuos de una serie de tiempo.

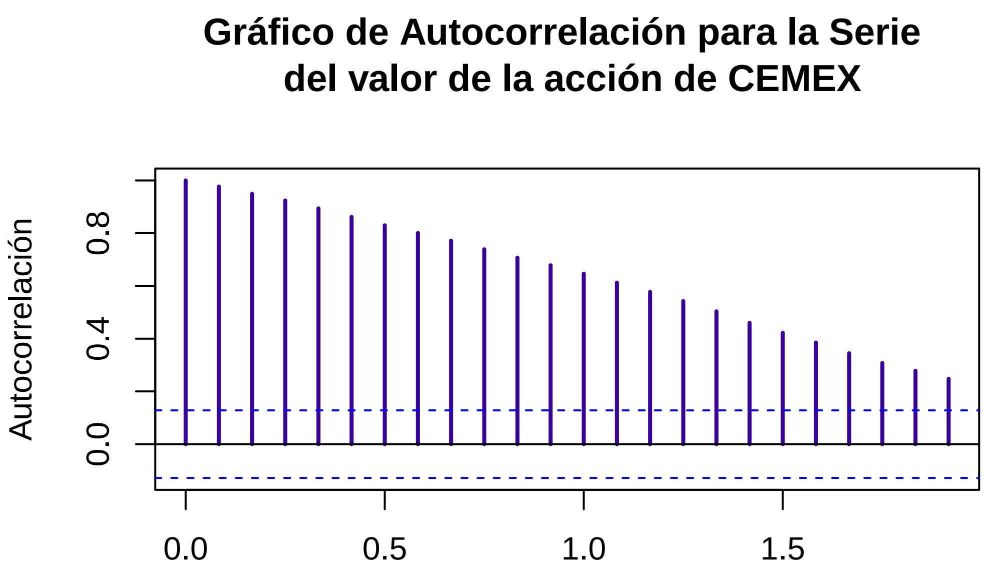


Figura 15: Gráfico de Autocorrelación

Con el gráfico de autocorrelación se vuelve bastante sencillo ver que a medida que avanza el tiempo cada nuevo análisis depende menos del primero y más de su valor inmediatamente anterior.

8. Desea proyectar la información un año, ¿Cuál sería el mejor modelo para la proyección? realice la proyección y concluya si la proyección tiene sentido o no, detallando sus observaciones

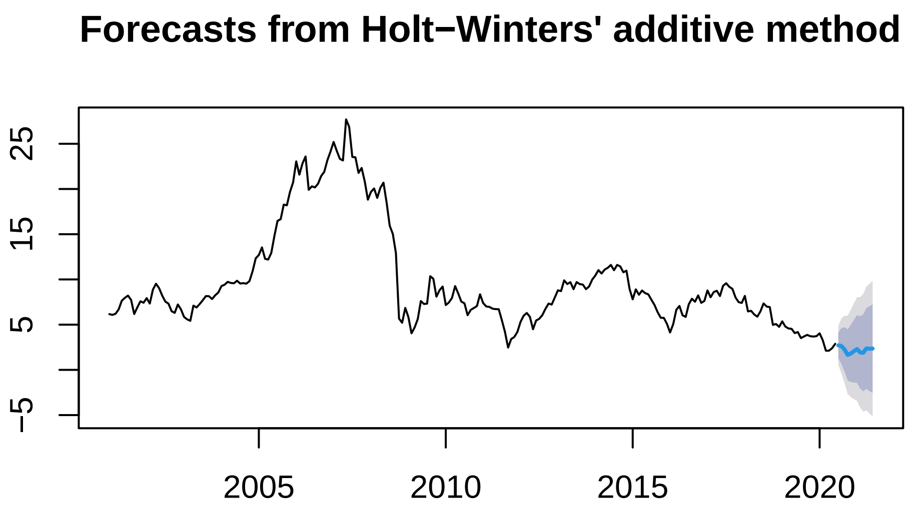


Figura 16: Gráfico de Proyección

Al observar la proyección, nos damos cuenta que tiene mucho sentido. Se aprecia natural y hasta intuitivo que lo que proyectamos sea una proyección muy verosimil.

9. Demuestra que si una serie de tiempo es estacionaria en media y varianza, entonces su autocorrelación es

$$\rho(h) = Corr(x_t, x_{t+h}) = \frac{E[x_t, x_{t+h}] - \mu}{\sigma^2}$$

9. Demuestra que si una serie de tiempo es estacionaria en media y varianza, entonces su autocorrelación es

$$\rho(h) = Corr(x_t, x_{t+h}) = \frac{E[x_t x_{t+h}] - \mu^2}{\sigma^2}.$$

Demostración.

Por definición, se tiene que

$$Corr(x_t, x_{t+h}) = \frac{cov(x_t, x_{t+h})}{\sqrt{var(x_t)}\sqrt{var(x_{t+h})}}.$$

Por propiedad de la covarianza, se cumple

$$cov(x_t, x_{t+h}) = E(x_t x_{t+h}) - E(x_t)E(x_{t+h}).$$

Ahora bien, por hipótesis tenemos que la serie de tiempo es estacionaria en media y varianza, es decir, la media y la varianza se mantienen constantes en el tiempo, por lo tanto,

$$E(x_t) = E(x_{t+h}) = \mu \quad y \quad var(x_t) = var(x_{t+h}) = \sigma^2$$

y entonces se tiene que

$$corr(x_t, x_{t+h}) = \frac{E(x_t x_{t+h}) - E(x_t)E(x_{t+h})}{\sqrt{var(x_t)}\sqrt{var(x_{t+h})}}$$

$$= \frac{E[x_t x_{t+h}] - \mu^2}{\sigma^2}$$

Figura 17: Solución Pregunta 9

10. Demuestra mediante inducción que en el suavizado exponencial esta expresión es válida

$$\hat{x}_{t+1} = ax_t + (1-\alpha)\hat{x}_t$$

$$\bar{X}_{t+1} = \alpha X_t + (1-\alpha)\bar{X}_t$$

10 Demuestre mediante inducción que en el suavizado exponencial esta expresión es valida.

$$\hat{x}_{t+1} = \alpha x_t + (1-\alpha)\hat{x}_t$$

Demostración.

- Veamos que se cumple para $\hat{x}_2$

Por definición

$$\underbrace{\hat{x}_2} = \alpha x_1 + \alpha(1-\alpha)\cancel{x_0}^{\,0} = \underbrace{\alpha x_1} \checkmark$$

Pues $\hat{x}_1$ no está definido.

Veamos igual que se cumple para $\hat{x}_3$

Por definición.

$$\hat{x}_3 = \alpha x_2 + \alpha(1-\alpha)x_1$$

$$= \alpha x_2 + (1-\alpha)\underbrace{\alpha x_1} = \alpha x_2 + (1-\alpha)\hat{x}_2 \checkmark$$

- Hipótesis de inducción (fuerte)

Supongamos que se cumple para todo $n \le t+1$ es decir,

$$\hat{x}_n = \alpha x_{n-1} + (1-\alpha)\hat{x}_{n-1} \quad \forall n \le t+1.$$

- Paso inductivo.

Pd. $\hat{x}_{t+2} = \alpha x_{t+1} + (1-\alpha)\hat{x}_{t+1}$.

Figura 18: Solución Pregunta 10

Por definicion

$$\hat{X}_{t+2} = \alpha X_{t+1} + \alpha(1-\alpha)X_t + \alpha(1-\alpha)^2 X_{t-1} + \dots + \alpha(1-\alpha)^{t+1} X_1$$

$$= \alpha X_{t+1} + \dots + \alpha(1-\alpha)^t X_2 + \alpha(1-\alpha)^{t+1} X_1$$

factorizando $(1-\alpha)^t \rightsquigarrow$

$$= \alpha X_{t+1} + \dots + (1-\alpha)^t \left[\alpha X_2 + (1-\alpha)\underbrace{\alpha X_1}\right]$$

Hip. induc. fuerte $\rightsquigarrow$

$$= \alpha X_{t+1} + \dots + (1-\alpha)^t \underbrace{\left[\alpha X_2 + (1-\alpha)\underbrace{\hat{X}_2}\right]}_{\hat{X}_3}$$

← Por hip. induc.

$$= \alpha X_{t+1} + \alpha(1-\alpha)X_t + \dots + \alpha(1-\alpha)^{t-1} X_3 + (1-\alpha)^t \underbrace{\hat{X}_3}$$

$$= \alpha X_{t+1} + \alpha(1-\alpha)X_t + \dots + (1-\alpha)^{t-1} \underbrace{\left[\alpha X_3 + (1-\alpha)\underbrace{\hat{X}_3}\right]}_{\hat{X}_4}$$

← Por hip. induc.

$$= \alpha X_{t+1} + \dots + \alpha(1-\alpha)^{t-2} X_4 + (1-\alpha)^{t-1} \underbrace{\hat{X}_4}$$

Continuando de esta forma obtendríamos que

$$= \alpha X_{t+1} + (1-\alpha)\hat{X}_{t+1}$$

justo lo que se quería probar.

Figura 19: Solución Pregunta 10