

práctica 1

Econometría Ampliación



Eugenio Martín Gallego

Alejandro Olmos Montero

Alberto Ribera Segura

**Ejercicio 1. Caracterización de propiedades de un estadístico mediante simulación.**

Se va a calcular probabilidad de cometer error de tipo I, es decir, rechazar la hipótesis nula siendo cierta.

En nuestro caso la hipótesis nula va a ser que la muestra se distribuye como una normal.

**H0: Normalidad**

Para ello se ha programado en Matlab un programa que realiza 10000 simulaciones sobre una muestra de 20, 50 y 100 datos que se distribuyen como una normal de media cero y varianza uno.

El resultado obtenido para un nivel de significación del 1%, 5% y 10% ha sido:

Tabla 1 Probabilidad de aceptar normalidad para una muestra que se distribuye como una N(0,1)

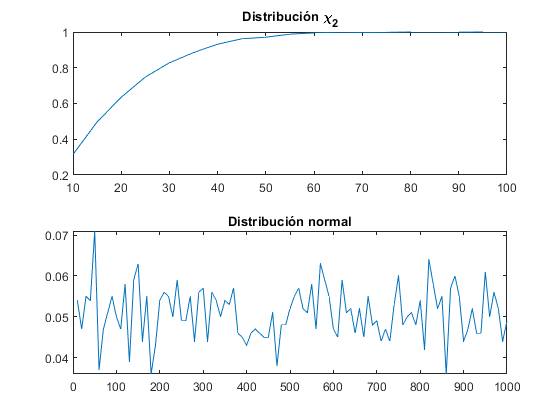
|  |
| --- |
| Nivel de significación veinte cincuenta cien  \_\_\_\_\_\_ \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ \_\_\_\_  0.01 1.09 1.03 1.05  0.05 5.18 5.14 5.16  0.1 10.02 10.02 10.03 |
|
|
|
|

Se puede observar que a mayor nivel de significación la desviación en tanto por cien sobre el nivel de significación objetivo es cada vez menor. Por ejemplo, para una muestra de 20 datos la desviación sobre el nivel de significación es del 9% para un alfa del 0.01, un 3,6% para un alfa 0.05 y 0.2% para un alfa 0.1.

Lo lógico sería que a mayor tamaño de muestra, menor error tipo 1. Pero observando los resultados obtenidos de la simulación en la Tabla 1, este suceso no ocurre ya que con 50 simulaciones el error es menor que con 100.

Dado esta anomalía se ha hecho el Gráfico 1, donde se aprecia que a pesar de aumentar el tamaño de la muestra de una normal con media cero y varianza uno, nunca converge al valor teórico. Mientras que con una muestra de Chi-cuadrados a partir de 60 datos el test Jarque Bera siempre rechaza la hipótesis nula de normalidad para un nivel de significación del 5%.

Gráfico 1 La probabilidad de rechazar normalidad para un nivel de significación del 5% dependiendo del tamaño de la muestra



Ahora se quiere estimar la potencia de contraste la probabilidad de rechazar la Hipótesis Nula cuando es falsa siendo la hipótesis la normalidad de los datos. Para ellos vamos a repetir lo anterior para una muestra que se distribuya como una Exponencial con media 1 y una Chi-cuadrado con dos grados de libertad.

Tabla 2 Potencia de contraste para una muestra con una distribución exponencial

|  |
| --- |
| Nivel de significación veinte cincuenta cien  \_\_\_\_\_\_\_ \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ \_\_\_\_\_  0.01 35.21 78.81 99.41  0.05 62.73 97.36 100  0.1 79.96 99.75 100 |
|
|
|
|

Tabla 3 Potencia de contraste para una muestra con distribución Chi-cuadrado con 2 grados de libertad

|  |
| --- |
| Nivel de significación veinte cincuenta cien  \_\_\_\_\_\_\_ \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ \_\_\_\_\_  0.01 34.64 78.83 99.33  0.05 62.9 97.67 100  0.1 80.21 99.71 100 |
|
|
|
|

El resultado es muy parecido para los dos tipos de muestras. Con pocos datos rechaza la hipótesis falsa 35 de cada 100 veces a un nivel de significación del 1%. A medida que aumenta la muestra rechaza con mayor frecuencia la hipótesis falsa. Como se muestra en el Gráfico 1 para un nivel de significación del 5% con 60 o más datos se rechaza siempre la hipótesis falsa.

**Ejercicio 2. Predicción de precios y de rentabilidades**

Se va a intentar encontrar el modelo que mejor ajuste a los rendimientos y a los precios de tres activos financieros de diferente naturaleza. El periodo de estimación escogido corresponde al 01/01/2000 hasta 31/12/2005, siendo el tamaño de la muestra de 1565 observaciones de carácter diaria. Los activos elegidos son el azúcar (commodity), FTSE 100 (índice) e Iberdrola (acción).

Los modelos propuestos o estimados estarán formado por autoregresivos de orden 1 ,2 o 3 aproximadamente y medias móviles de orden 5, 10 o 15. Podría tener sentidos que los rendimientos estén relacionados con los del día anterior o próximo porque el mercado no es totalmente eficiente y podría tardar en digerir shocks o noticias. Por lo contrario, si el mercado fuera eficiente en sentido fuerte no tendría sentido que estuvieran relacionados. Por parte de las medias móviles de orden 5,10 y15 son las más probables porque los datos son diarios y cada 5 datos hay dos días sin mercado, donde se puede acumular información y repercutir los lunes en el precio o rendimiento. ***Azúcar:***

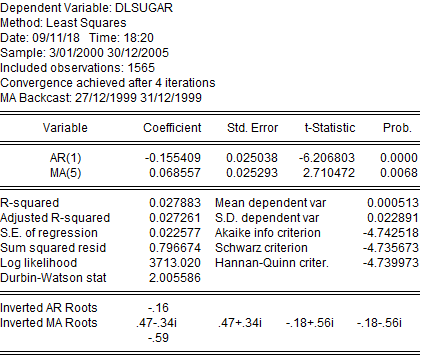
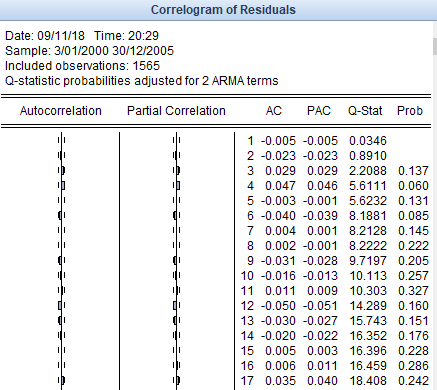
El modelo obtenido para los rendimientos del precio del azúcar ha sido un arima(1,1,lags(5)):

Gráfico 2 Estimación del modelo, ACF y PACF de los rendimientos logarítmicos del precio del azúcar entre el 2000 y 2005

A partir del modelo obtenido se han hecho dos tipos de predicciones una estática y otra dinámica. Con la estática se calcula el valor esperado de los rendimientos **,** solo toma información los valores hasta **el instante t.** Mientras que con las predicciones dinámicas asignamos aleatoriamente valores N(0,1) a épsilon durante toda la predicción del tal forma que se toman como información conocida los valores hasta instante a predecir. Las predicciones se harán aproximadamente de los 100 siguientes datos desde 01/01/2006 el hasta 26/05/2006

Gráfico 3 Predicción del rendimiento logarítmico del azúcar desde el 01/2006 hasta el 06/2006

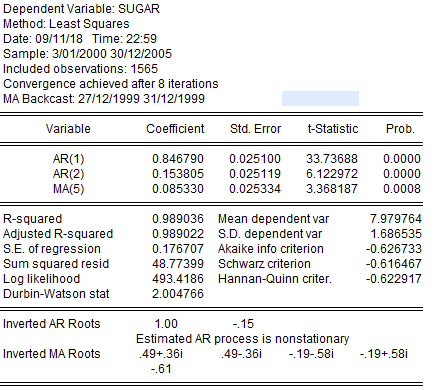
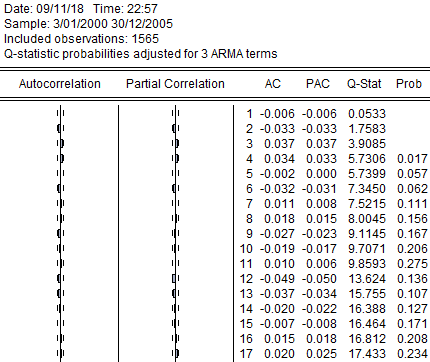
Por otra parte, el modelo obtenido para los precios del azúcar ha sido un ARMA(lags(1,2), 5)):

Gráfico 4 Estimación del modelo, ACF y PACF de los precios del azúcar entre el 2000 y 2005

Si estimamos una predicción de los próximos 100 días el resultado es:

Gráfico Predicción del precio del azúcar desde el 01/2006 hasta el 06/2006



**FTSE-100:**

El modelo obtenido para los rendimientos del índice ha sido un arima(lags(1,3,8),1,lags(10)).

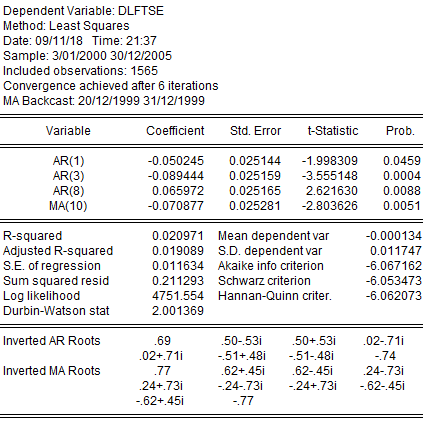
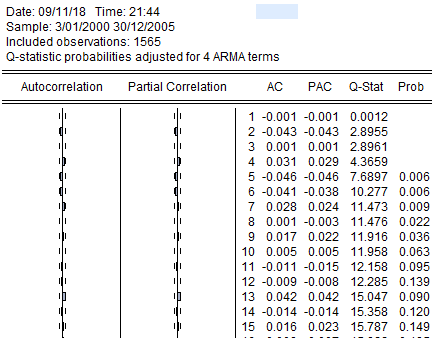


Gráfico 6 Estimación del modelo, ACF y PACF de los rendimientos logarítmicos del índice FTSE-100 entre el 2000 y 2005

La estimación obtenida en este caso es:

Gráfico Predicción del rendimiento logarítmico del FTSE-100 desde el 01/2006 hasta el 06/2006



Por otra parte, el modelo obtenido para el índice ha sido un ARMA(lags(1,3),lags(3,8)):

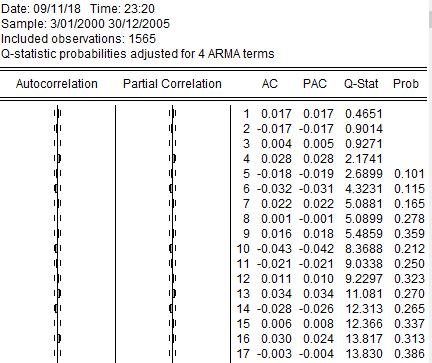
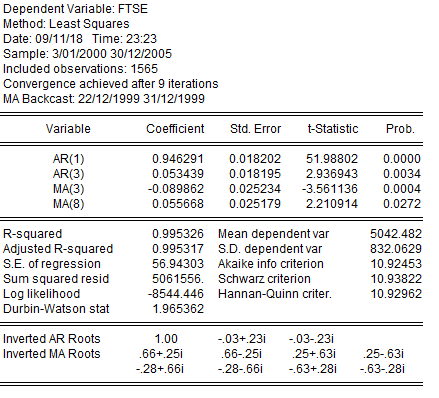


Gráfico 8 Gráfico 3 Estimación del modelo, ACF y PACF de índice FTSE-100 entre el 2000 y 2005

Y la predicción obtenida:

Gráfico Predicción del FTSE-100 desde el 01/2006 hasta el 06/2006



**Acción Iberdrola.**

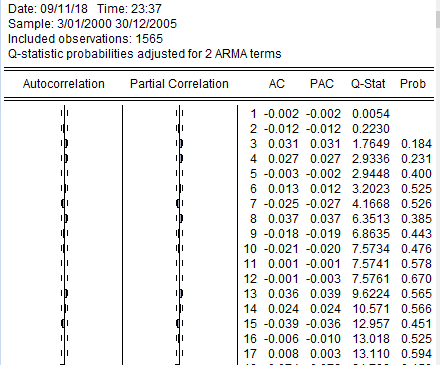
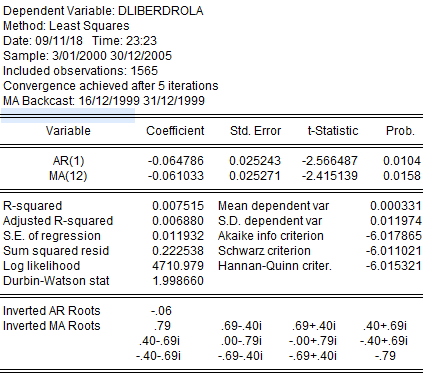
El modelo obtenido para los rendimientos del precio de la acción de Iberdrola ha sido un ARIMA (1, 1, lags(12)):

Gráfico 10 Estimación del modelo, ACF y PACF de los rendimientos logarítmicos de la acción de Iberdrola entre el 2000 y 2005



Siguiendo el modelo estimado, la predicción es:

Gráfico Predicción del rendimiento logarítmico del precio de Iberdrola desde el 01/2006 hasta el 06/2006



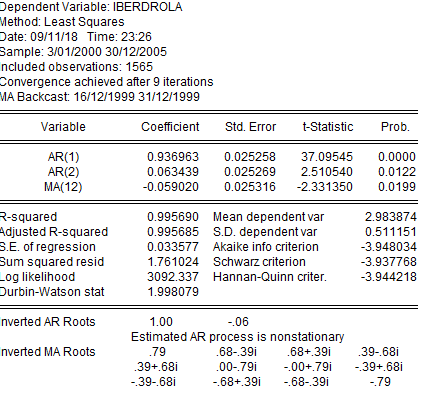
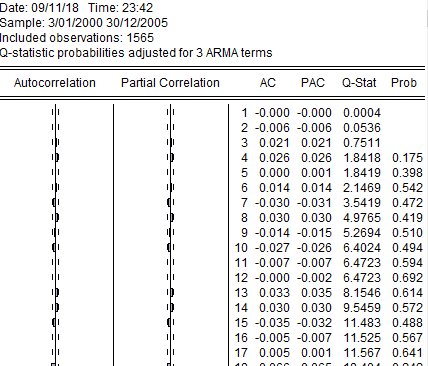
Por otra parte, el modelo obtenido para el í ha sido un ARMA(2,lags(12))):

Gráfico 12 Estimación del modelo, ACF y PACF de los precios de Iberdrola entre el 2000 y 2005

Y su predicción con el modelo obtenido:

Gráfico Predicción del precio de Iberdrola desde el 01/2006 hasta el 06/2006



**Comentarios sobre el resultado:**

1. Se puede observar que los modelos estimados para los rendimientos logarítmicos de los activos (Gráficos 3, 7, 11) sobrepasa las banda de confianza con mayor frecuencia que los datos reales. Esto es debido a que estos modelos no permiten captar correctamente la curtosis.
2. Los precios de los activos dependen en gran parte de los precios de los dos días anteriores. Observando los gráficos 4,8 y 12 se obtiene que los coeficientes de AR(1)>0.85 y los AR(2)>0.05 .
3. Estadísticos de bondad de predicción:

Tabla Estadísticos de bondad de predicción para precios y rendimientos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Rendimiento | | | Precios | | |
|  | Sugar | FTSE | Iberdrola | Sugar | FTSE | Iberdrola |
| RMSE | 0.0248 | 0.0075 | 0.0248 | 3.7893 | 3.6464 | 3.5174 |
| MAE | 0.019338 | 0.005719 | 0.019402 | 3.588444 | 3.449365 | 3.323824 |
| RMSEp | - | - | - | 0.2057 | 0.0613 | 0.0728 |
| MAEp | - | - | - | 0.2057 | 0.0613 | 0.0783 |

Dos estadísticos de bondad de predicción altamente utilizados son la raíz del error cuadrático medio (RMSE) y el error absoluto medio (MAE). Ambos se pueden obtener en su forma porcentual o estándar, dependiendo del tipo de datos.

En cuanto a la preferencia de un método sobre el otro (RMSE vs MAE), el estimador RMSE es más sensible a predicciones extremadamente erróneas. Esto a su vez implica que el estimador MAE está sesgado hacia métodos cuyas predicciones sean más bajas.

Si se trabaja sobre rentabilidades será correcto aplicar la forma estándar de los estimadores, pues la porcentual no puede aplicarse ya que el cociente diverge si el rendimiento es nulo en un periodo, mientras que si se trabaja sobre precios será más correcto aplicar la porcentual pues se obtiene un resultado más fácilmente interpretable (porcentaje de desviación en los errores de la predicción). Además, si se aplica la forma estándar de los estadísticos a los precios se obtienen resultados que dependerán de la escala de la serie, por lo tanto no se puede deducir qué modelo ajustará mejor solamente a partir del valor del estadístico. Es decir, para el caso de los rendimientos ambos estadísticos son un medida correcta del error en la predicción, mientras que no lo será para la estimación de los precios.

En el caso de los precios, sabemos que el FTSE está alrededor de los 5000 puntos, mientras que el azúcar sobre los 15. Esto implicará que los estimadores estándar darán, en general, menor error en la predicción al modelo del azúcar, puesto que es varios órdenes de magnitud menor. Observando el gráfico se observa que ambos tienen un RMSE y un MAE similares, lo cual podría llevar erróneamente a deducir que ambas estimaciones son igual de (in)correctas, sin embargo si se observan los estimadores porcentuales es menor el correspondiente al FTSE que el del azúcar, lo cual indica que el modelo que mejor predice los precios futuros es del FTSE, demostrando que los estimadores estándar están sesgados por la escala de los valores de la serie

**Ejercicio 3. Utilización de las predicciones**

**a) Suponga un fondo de inversión cuyo rendimiento mensual anualizado sigue el siguienteproceso estocástico:**

**Si tiene invertido 10 000 euros, dentro de tres meses, al 99% de confianza, ¿cuál será la pérdida esperada. Suponga que el rendimiento en el mes actual ha sido un 2% por encima de su valor esperado incondicional.**

El rendimiento en el mes actual ha sido del 12%, un 2% por encima de su valor esperado incondicional, que viene dado por la expresión:

Para calcular la pérdida esperada al 99% de confianza producida dentro de tres meses en una inversión de 10.000 euros necesitamos conocer el error de predicción. El valor obtenido gracias a la siguiente expresión arroja un valor dentro de 3 periodos de 0.033%.

También necesitamos el rendimiento esperado condicionado en el tercer periodo, cuyo valor calculado es del 0.115%. Ahora obtener el VaR al 0.01 gracias a la expresión:

La pérdida máxima esperada con un 99% de confianza que hemos obtenido es -417.14€ sobre la ganancia media esperada. Obteniendo un beneficio neto de 1150-417.14= 732.86 .Esto es debido a que el rendimiento esperado para el tercer periodo es un valor positivo muy alto en comparación con la volatilidad.

Para comprobar el resultado obtenido analíticamente hemos realizado un simulación en Matlab que nos ha dado un valor del VaR de -415€. (Código incluido en el anexo)

**b) A partir de los datos del IBEX que se encuentran en la hoja de cálculo ibex.xlsx, especificar y estimar un modelo ARIMA. Después estime a partir de la distribución de las predicciones y mediante simulación, la probabilidad de que cierre el año con una ganancia respecto de la situación actual.**

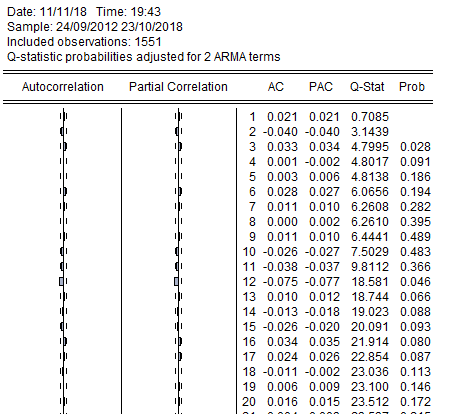
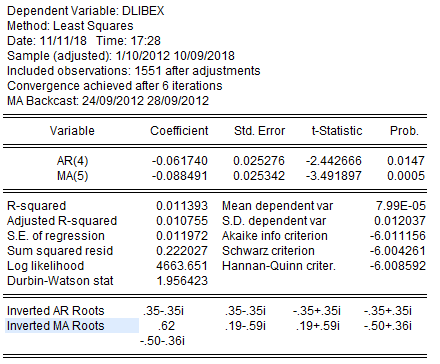
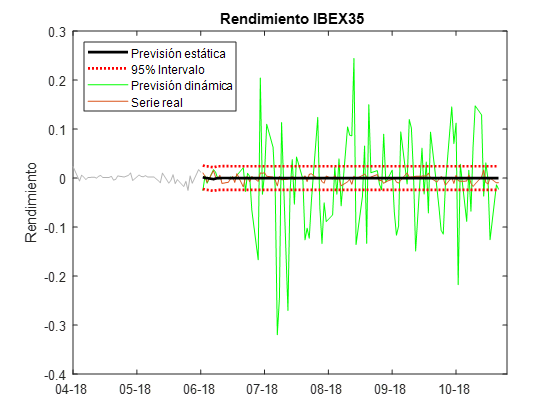
A partir de los datos del IBEX hemos estimado el modelo ARIMA(lags(4),1,lags(5)) que siguen los rendimientos, para ello hemos introducido la serie en Eviews y hemos analizado la función de autocorrelación total y parcial.

Gráfico Estimación del modelo, ACF y PACF de los rendimientos logarítmicos del IBEX·35 entre el 2000 y 2005

Se ha testeado el modelo haciendo una predicción desde junio del 2018 hasta el último dato de la muestra y sigue sucediendo lo mismo que en el ejercicio 2, nuestro modelo tiene rendimientos frecuentemente por encima de las bandas de confianza.

Gráfico Predicción del ibex desde el 06/2018 hasta el 23/10/2018



Por último, Introduciendo los coeficientes del modelo en Matlab hemos realizado 10000 simulaciones para estimar la probabilidad de que cierre el año con una ganancia respecto de la situación actual. El resultado obtenido indica que con una probabilidad aproximada del 50% el IBEX35 acabará por encima de 8726.1 puntos a fecha 23/10/2018 (código incluido en el anexo).

**c) Descargue los datos del PIB de USA de la página web de la Reserva Federal de St Louis (https://fred.stlouisfed.org/series/GDPC1). Ésta es una serie trimestral desestacionalizada en términos reales. Identifique y estime un modelo ARIMA para el PIB de Estados Unidos. Estime, dado su modelo, con qué probabilidad el crecimiento del PIB será menor del 1% en el último trimestre del año.**

Para analizar el modelo que sigue PIB de Estados Unidos, hemos introducido la serie s en Eviews y hemos realizado el mismo análisis que en el apartado anterior.

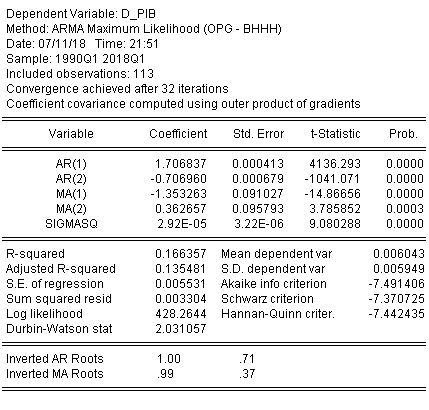
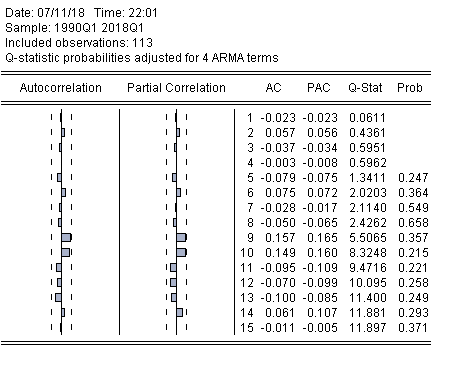


Gráfico Gráfico 14 Estimación del modelo, ACF y PACF el crecimiento del PIB entre el 2000 y 2005

Todos los coeficientes son significativos y además todos los retardos del correlograma de los residuos quedan dentro de las bandas de confianza, por tanto el modelo ARMA(2,2) el que mejor se ajusta a la serie de rendimientos del PIB.

Introduciendo los coeficientes del modelo en Matlab hemos realizado 10000 simulaciones para concluir que con una probabilidad del 69% el crecimiento del PIB será inferior al 1% en el último trimestre del año. El código está incluido en el anexo

**ANEXO**

**Código 1**

muestras = 10000;

h = zeros(muestras,9); p\_value = zeros(muestras,9);

jbstat = zeros(muestras,9) ; critval = zeros(muestras,9);

alpha = [0.01, 0.05, 0.1];

simsize = [20, 50, 100];

for k = 1:length(alpha)

sim = exprnd(1,muestras,simsize(k));

%sim = lognrnd(0,1,muestras, simsize(k));

%sim = chi2rnd(2,muestras, simsize(k));

%sim = randn(muestras,simsize(k));

for j = 1:length(simsize)

for i = 1:size(sim,1)

[h(i,3\*(k-1)+j), p\_value(i,3\*(k-1)+j), jbstat(i,3\*(k-1)+j), critval(i,3\*(k-1)+j)] = jbtest(sim(i,:),alpha(j));

end

end

end

tamano = sum(h,1)/muestras\*100;

%% tabla con los datos

veinte = tamano(1:3)';

cincuenta = tamano(4:6)';

cien = tamano(7:9)';

p\_value = alpha';

T\_exp = table(p\_value,veinte,cincuenta,cien)

**Código 2**

%% Importo los datos

y1 = xlsread('datos\_financieros','S6612:S8176');

y2 = xlsread('datos\_financieros','S8177:S8277');

[date1,date2,date3] = xlsread('datos\_financieros','A6612:A8277');

% yr1=log(y1(2:end)./y1(1:end-1)); % creo los rendimientos logarítmicos

% yr2=log(y2(2:end)./y2(1:end-1)); % creo los rendimientos logarítmicos

yr1 = y1(2:end); % precios

yr2 = y2(1:end-1); % precios

%% Definición de los parámetros

n = 100; % observaciones a estimar

size\_data = length(y1); % longitud del vector de datos

size\_forecast = length(y2); % longitud del numero de datos a predecir

%% Estimación estática del modelo

ARLAG = [1,2];

MALAG = [12];

Model = arima('ARlags',ARLAG,'MAlags',MALAG); % no consigo quitar la cte de la estimación

[ModelFit,a,b,coef] = estimate(Model,yr1);

[Y,YMSE] = forecast(ModelFit,n,'Y0',yr1);

lower = Y - 1.96\*sqrt(YMSE);

upper = Y + 1.96\*sqrt(YMSE);

%% Estimación dinámica del modelo

max\_lag = max(max(ARLAG),max(MALAG)); % cantidad máxima de lags entre AR y MA

% Estimación del ruido a partir de los datos anteriores

for i = 1:max\_lag

ruido(i) = yr1(end+1-i)-yr1(end-i:-1:end-i-length(ARLAG)+1)'\*coef.X(2:1+length(ARLAG));

end

% Se incluye la opción de simular varias trayectorias

sim = 1;

ruido = [repmat(ruido,sim,1) , randn(sim,100)];

% Posteriormente la 2a parte del vector es escalado con la varianza del modelo

r\_din = repmat(yr1(end-max\_lag+1:end)',sim,1);

coeficiente = [coef.X(1:end-1) ; sqrt(coef.X(end))]

% Generación de la previsión dinámica

for i=max\_lag+1:length(yr2)

r\_din(:,i) = [zeros(sim,1),r\_din(:,i-1), r\_din(:,i-2), ruido(:,i-12), ruido(:,i)] \* coeficiente;

end

% prob = sum(r\_din(:,end)>0)/sim

%% Gráfica de las predicciones

figure(1)

% Fecha y formato de la gráfica

date2 = cell2mat(date2);

a1=datetime(date2,'InputFormat','dd/MM/yyyy');

plot(a1(end-200:end-100),yr1(end-100:end),'Color',[.7,.7,.7]);

xlim([a1(end)-200 a1(end)+4]);

xticks(a1(end)-200:30:a1(end)+4)

xtickformat('MM-yy')

% Introduzco las series en la gráfica

hold on

r\_estimated = plot(a1(end-99:end),r\_din,'g');

serie\_real = plot(a1(end-99:end),yr2);

h1 = plot(a1(end-99:end),lower,'r:','LineWidth',2);

plot(a1(end-99:end),upper,'r:','LineWidth',2)

h2 = plot(a1(end-99:end),Y,'k','LineWidth',2);

legend([h2 h1 r\_estimated, serie\_real],'Previsión estática','95% Intervalo','Previsión dinámica','Serie real',...

'Location','NorthWest')

title(' Rendimiento')

ylabel('Rendimiento');

hold off

%% ACF, PACF e histograma de modelo

% figure(2)

% subplot(2,2,1)

% autocorr(yr1)

% subplot(2,2,3)

% parcorr(yr1)

% subplot(2,2,[2 4])

% hist(yr1,20)

% title('Histogram')

%% Estadísticos de bondad de predicción

% Error cuadrático medio

RMSE = sqrt(mean((Y-yr2).^2))

% Error absoluto medio

MAE = sum(abs(Y-yr2))

% Raíz del error cuadrático porcentual

RMSEp = sqrt(mean(((Y-yr2)./yr2)).^2)

% Error del error medio porcentual

MAEp = mean(abs((Y-yr2)./yr2))

**Código 3.1**

%% Estimacion del VaR por Montecarlo

e = 0.02 \* randn(100000,3); %% N(0,1)

c = 0.12 \* ones(100000,1); %% Rendimiento del mes actual

%% Simulacion de trayectorias

for i = 1:3

c(:,i+1) = 0.01 + 0.9 \* c(:,i) + e(:,i);

end

rend = c(:,end);

rend\_2 = sort(rend,1); %% ordenamos los datos

a = rend\_2(1000); %% nos quedamos con el cuantil 0.01

Var\_001 = -a\*10000

**Código 3.3**

%% Datos

z= xlsread('GDPC1','C12:C297');

zr=log(z(2:end)) - log(z(1:end-1));

%% coeficientes del modelo ARMA

AR1 = 1.706837;

AR2 = -0.706960;

MA1 = -1.353263;

MA2 = 0.362657;

e = 0.0054 \* randn(10000,3); %% N(0,1)

%% SimulaciÛn de trayectorias

r = AR1 \* zr(:,end-1) + AR2 \* zr(:,end-2) + MA1 \* e(:,1) + MA2 \* e(:,2) + e(:,3);

%% C·culo de la probabilidad

prob = sum(r<0.01,1)/10000