TEORÍA

Pregunta 1 (1 p.)

Decide, en cada uno de los siguientes modelos ***ARIMA(p,d,q)***, los que son **causales** i/o **invertibles** y determina los valores **p**, **d** y **q** en cada caso

a) $X_t = 0.8X_{t-1} + Z_t - 0.9Z_{t-1}$
$ARIMA(1{,}0{,}1)$, $|\text{roots}(1-0.8\text{B})|>1$ ➔ causal, $|\text{roots}(1-9\text{B})|>1$ ➔ invertible

b) $X_t = 1.5X_{t-1} - 0.5X_{t-2} + Z_t$
$ARIMA(2{,}0{,}0)$, $|\text{roots}(1-1.5\text{B}+0.5\text{B}^2)|\geq 1$ ➔ not causal , AR(2) ➔ invertible

c) $X_t = Z_t - Z_{t-1} + \frac{2}{9}Z_{t-2}$
$ARIMA(0{,}0{,}2)$, MA(2) ➔ causal, $|\text{roots}\left(1-\text{B}+\frac{2}{9}\text{B}^2\right)|>1$ ➔ invertible

d) $(X_t - 1) - 2(X_{t-1} - 1) + (X_{t-2} - 1) = Z_t + 2Z_{t-1}$
$ARIMA(0{,}2{,}1)$, $|\text{roots}(1-2\text{B}+\text{B}^2)|=1$ ➔ not causal, $|\text{roots}(1-2\text{B})|<1$ ➔ not invertible

Pregunta 2 (1 p.)

a) ¿Existe algún modelo *ARMA(p,q)* con un número finito de retardos no nulos tanto en el ACF como en el PACF? Razona la respuesta.
Todo AR(p) presenta infinitos retardos no nulos en el ACF y todo MA(q), q>0 tiene infinitos retardos no nulos en el PACF. Por consiguiente, todo modelo ARMA(p,q) tiene infinitos retardos no nulos en ambas gráficas. El único caso en que no hay infinitos retardos no nulos en el ACF y el PACF es el caso del ruido blanco o ARMA(0,0)

b) Obtén la función de auto-correlación del modelo estacionario ***AR(2)*** de media nula: $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$

$$\gamma(1) = E(X_t X_{t+1}) = \left[X_t\left(\phi_1 X_t + \phi_2 X_{t-1} + Z_{t+1}\right)\right] = \phi_1 E[X_t X_t] + \phi_2 E[X_t X_{t+1}] + E[X_t Z_{t+1}] = \phi_1\gamma(0) + \phi_2\gamma(1) + 0$$

$$\gamma(1) = \frac{\phi_1}{1-\phi_2}\,\gamma(0) \Rightarrow \rho(1) = \frac{\phi_1}{1-\phi_2}$$

$$\gamma(2) = E(X_t X_{t+2}) = \left[X_t\left(\phi_1 X_{t+1} + \phi_2 X_t + Z_{t+2}\right)\right] = \phi_1 E[X_t X_{t+1}] + \phi_2 E[X_t X_t] + E[X_t Z_{t+2}] = \phi_1\gamma(1) + \phi_2\gamma(0) + 0$$
$$= \left(\frac{{\phi_1}^2}{1-\phi_2} + \phi_2\right)\gamma(0) \Rightarrow \rho(2) = \frac{{\phi_1}^2}{1-\phi_2} + \phi_2$$

$$\gamma(k) = E(X_t X_{t+k}) = [X_t(\phi_1 X_{t+k-1} + \phi_2 X_{t+k-2} + Z_{t+k})] = \phi_1 E[X_t X_{t+k-1}] + \phi_2 E[X_t X_{t+k-2}] + E[X_t Z_{t+k}]$$
$$= \phi_1\gamma(k-1) + \phi_2\gamma(k-2) \Rightarrow \rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2), k > 2$$

Pregunta 3 (1 p.)

Sea $\{X_t\}$ una serie a la que ajustamos el modelo MA(2): $X_t = (1 + 0.5B - 0.2B^2)Z_t$ con $Z_t \sim N(0,1)$

a) Sabiendo que: $\tilde{X}_{99|98} = 1$, $X_{99} = 3$, $\tilde{X}_{100|99} = 0$ y $X_{100} = 1$, calcula las previsiones puntuales a largo plazo: $\tilde{X}_{101|100}$, $\tilde{X}_{102|100}$ y $\tilde{X}_{103|100}$

$Z_{99} = X_{99} - \tilde{X}_{99|98} = 3 - 1 = 2$, $Z_{100} = X_{100} - \tilde{X}_{100|99} = 1 - 0 = 1$
$\tilde{X}_{101|100} = E[Z_{101} + 0.5Z_{100} - 0.2Z_{99}|X_{100}, \dots, X_1] = 0.5Z_{100} - 0.2Z_{99} = 0.5 - 0.2 * 2 = 0.1$
$\tilde{X}_{102|100} = E[Z_{102} + 0.5Z_{101} - 0.2Z_{100}|X_{100}, \dots, X_1] = -0.2Z_{100} = -0.2$
$\tilde{X}_{103|100} = E[Z_{103} + 0.5Z_{102} - 0.2Z_{101}|X_{100}, \dots, X_1] = 0$

b) Calcula las varianzas de los errores de las predicciones anteriores
$V(Z_{101}) = V\left[X_{101} - \tilde{X}_{101|100}\right] = V[Z_{101}|X_{100}, \dots, X_1] = V[Z_{101}] = 1$
$V(Z_{102}) = V\left[X_{102} - \tilde{X}_{102|100}\right] = V[Z_{102} + 0.5Z_{101}|X_{100}, \dots, X_1] = 1.25$
$V(Z_{103}) = V\left[X_{103} - \tilde{X}_{103|100}\right] = V[Z_{103} + 0.5Z_{102} - 0.2Z_{101}|X_{100}, \dots, X_1] = 1.29$

Pregunta 4 (1 p.)

a) Escribe la ecuación para X_t de un modelo ***ARIMA(2,1,0)***. Indica cuales son los polinomios característicos y los parámetros del modelo.

$(1-\phi_1 \mathrm{B}-\phi_2 \mathrm{B}^2)\,(1-\mathrm{B})X_t = Z_t$

Polinomio característico de la parte AR: $\phi_p(B) = (1-\phi_1 \mathrm{B}-\phi_2 \mathrm{B}^2)(1-B)$

Polinomio característico de la parte MA: $\theta_q(B) = 0$

Parámetros: (ϕ_1, ϕ_2)

b) El modelo Box-Jenkins permite obtener las previsiones a h-pasos a partir de la expresión: $\tilde{X}_{t+h|t} = E[X_{t+h}|X_t,\ X_{t-1,\dots,}\ X_1]$ Obtén la previsión de X_t a un paso y a dos pasos hacia delante para el modelo ***ARIMA(2,1,0)***, asumiendo que los valores de los parámetros y los valores iniciales de la serie son conocidos.

$(1-\phi_1 \mathrm{B}-\phi_2 \mathrm{B}^2 - (\mathrm{B}-\phi_1 \mathrm{B}^2-\phi_2 \mathrm{B}^3)X_t = Z_t$

$X_t = (\phi_1+1)X_{t-1} + (\phi_2-\phi_1)\mathrm{X}_{t-2} - \phi_2 \mathrm{X}_{t-3} + Z_t$

$\tilde{X}_{t+1|t} = \mathrm{E}(X_{t+1}|X_t, X_{t-1}, \dots, X_1) = (\phi_1+1)X_t + (\phi_2-\phi_1)\mathrm{X}_{t-1} - \phi_2 \mathrm{X}_{t-2}$

$\tilde{X}_{t+2|t} = \mathrm{E}(X_{t+2}|X_t, X_{t-1}, \dots, X_1) = (\phi_1+1)\tilde{X}_{t+1|t} + (\phi_2-\phi_1)\mathrm{X}_t - \phi_2 \mathrm{X}_{t-1}$

Pregunta 5 (1 p.)

a) El proceso ***ARIMA(1,1,0)(1,1,0)₄*** ¿Es un modelo estacionario? ¿Es invertible? Justifica las respuestas.

$(1-\phi_1 \mathrm{B})(1-\Phi_4 \mathrm{B}^4)(1-\mathrm{B})(1-\mathrm{B}^4)X_t = Z_t$

No es estacionario (causal) puesto que hay raíces unitarias en el polinomio característico de la parte AR.

Es invertible porque es un proceso AR (no hay polinomio en la parte MA)

b) Para este modelo, escribe la ecuación que expresa la X_t en función de las observaciones anteriores.

$(1-\phi_1 \mathrm{B}-\Phi_4 \mathrm{B}^4+\phi_1\Phi_4 \mathrm{B}^5)(1-\mathrm{B}-\mathrm{B}^4+\mathrm{B}^5)X_t = Z_t$

$(1-\phi_1 \mathrm{B}-\Phi_4 \mathrm{B}^4+\phi_1\Phi_4 \mathrm{B}^5-\mathrm{B}+\phi_1 \mathrm{B}^2+\Phi_4 \mathrm{B}^5-\phi_1\Phi_4 \mathrm{B}^6-\mathrm{B}^4+\phi_1 \mathrm{B}^5+\Phi_4 \mathrm{B}^8-\phi_1\Phi_4 \mathrm{B}^9+\mathrm{B}^5-\phi_1 \mathrm{B}^6-\Phi_4 \mathrm{B}^9 + \phi_1\Phi_4 \mathrm{B}^{10})X_t = Z_t$

$(1-(\phi_1+1)\mathrm{B}+\phi_1 \mathrm{B}^2-(\Phi_4+1)\mathrm{B}^4+(\phi_1\Phi_4+\Phi_4+\phi_1+1)\mathrm{B}^5-(\phi_1\Phi_4+\phi_1)\mathrm{B}^6+\Phi_4 \mathrm{B}^8-(\phi_1\Phi_4+\Phi_4)\mathrm{B}^9 + \phi_1\Phi_4 \mathrm{B}^{10})X_t = Z_t$

$X_t = (\phi_1+1)X_{t-1}-\phi_1 X_{t-2}+(\Phi_4+1)X_{t-4}-(\phi_1\Phi_4+\Phi_4+\phi_1+1)X_{t-5}+(\phi_1\Phi_4+\phi_1)X_{t-6}-\Phi_4 X_{t-8}+(\phi_1\Phi_4 + \Phi_4)X_{t-9}-\phi_1\Phi_4 X_{t-10}+Z_t$

c) Este modelo es un caso particular de un modelo AR(p). ¿Cuál es el valor de p? ¿Cuántos coeficientes ϕ_i son diferentes de cero?

AR(10). Hay 2 coeficientes nulos: el tercero y el séptimo. En total hay 8 coeficientes diferentes de cero

LABORATORIO

La serie **NUCLEAR** (Fuente: US Energy Information Administration. http://www.eia.gov/totalenergy/data/monthly/#nuclear) contiene los datos mensuales de los miles de millones de KwH de energía eléctrica generados por las centrales nucleares en USA, desde enero de 1990 hasta diciembre de 2018 (figura 1). El objetivo consiste en obtener las previsiones para el 2019.

En todas las cuestiones planteadas justifica la respuesta.

En las figuras 2 y 3 aparecen el boxplot por años y el gráfico de medias-varianzas para la serie original.

Pregunta 1 (0,5 p.)

¿Para qué se utilizan estos gráficos cuando se trabajan con series temporales? Explica cómo se interpretan y a qué conclusión se llega en este caso.

Una de las premisas para identificar modelos ARIMA es transformar la serie en estacionaria, y en particular, la serie ha de tener la varianza constante. Estos gráficos permiten discutir si la varianza varía en función del nivel de la serie. En el Box-plot por años, si el IQR (amplitud de la caja) no es del mismo orden para los diferentes niveles de la serie y aumenta cuando aumenta el nivel de la serie, hay que aplicar la transformación logaritmo para homogeneizar la varianza. En el gráfico medias varianzas, si la varianza es del mismo orden para las diferentes medias, no hace falta aplicar la transformación. Si crece la varianza a medida que crece la media, hay que tomar logaritmos. En este caso, se puede considerar que la varianza es constante y por lo tanto no hay que cambiar la escala aplicando logaritmos.

En la figura 4 aparece el resultado de la función `monthplot` aplicado a la serie. Las figuras 5, 6 y 7 contienen las transformaciones de la serie mediante sucesivas diferenciaciones. Las varianzas de la serie y de sus transformaciones son:

```
> var(serie)
[1] 61.90153
> var(d12serie)
[1] 9.791821
> var(d1d12serie)
[1] 5.829643
> var(d1d1d12serie)
[1] 13.49588
```

Pregunta 2 (1 p.)

Explica brevemente por qué es necesario transformar la serie en estacionaria, antes de proceder a analizar su ACF y PACF. Describe el patrón estacional de la serie. En base a los resultados obtenidos hasta este momento ¿qué transformación de la serie seleccionarías para analizar el ACF y PACF?

Una serie estacionaria tiene la media y varianza constante y la estructura de autocorrelación no depende del origen. DE esta manera, con una serie temporal podemos estimar los parámetros del modelo. Si la serie no es estacionaria, el número de parámetros del modelo es excesivo y no se puede hacer una estimación eficiente de los mismos.
Mirando el monthplot, el patrón estacional incluye dos máximos en los meses de invierno (diciembre y enero) y verano (julio y agosto) y mínimos en primavera (abril) y otoño (octubre). El patrón estacional parece relacionado con la demanda de energía.
Teniendo en cuenta las diferentes diferenciaciones, la estacional es necesaria para eliminar el patrón estacional y con una diferenciación regular obtenemos una serie con media y varianza claramente constante y una nueva diferenciación aumentaría la varianza. Por ello, la transformación sería: $W_t = (1-B)(1-B^{12})X_t$

Finalmente, se obtiene la serie W_t transformada de la original y que se considera estacionaria. En las figuras 8 y 9 aparecen su ACF y PACF. Las flechas indican los retardos múltiplos de la estacionalidad (s=12)

Pregunta 3 (1 p.)

Identifica al menos dos posibles modelos **para la serie W_t**, explicando claramente en qué te basas para formular estos modelos. Escribe la expresión de estos modelos, indicando los polinomios característicos y los parámetros de cada modelo propuesto.

Parte estacional: El último retardo significativo en el ACF sería el primero y hay pautas de decrecimiento en el PACF para los retardos múltiplos de 12 ➔MA(1)$_{12}$.
Parte regular:

- El último retardo significativo entre los primeros del ACF sería el segundo y hay pautas de decrecimiento en el PACF ➔MA(2)
- El último retardo significativo entre los primeros del PACF sería el sexto y hay pautas de decrecimiento en el ACF ➔AR(6)
- Consideramos pautas de decrecimiento en ambos gráficos ➔ ARMA(1,1)

Posibles modelos:
W_t~ARIMA(0,0,2)(0,0,1)$_{12}$ $W_t = (1+\theta_1 B + \theta_2 B^2)(1+\Theta_{12}B^{12})Z_t$
W_t~ARIMA(6,0,0)(0,0,1)$_{12}$ $(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6)W_t = (1+\Theta_{12}B^{12})Z_t$
W_t~ARIMA(1,0,1)(0,0,1)$_{12}$ $(1-\phi_1 B)W_t = (1+\theta_1 B)(1+\Theta_{12}B^{12})Z_t$

Independientemente de la respuesta anterior, se decide estimar dos modelos: uno con cuatro parámetros y otro con tres:

Modelo 1:
```
      0.6213  0.0270  -0.9141  -0.8490
s.e.  0.0834  0.0678   0.0634   0.0392
sigma^2 estimated as 2.966:  log likelihood = -665.69,  aic = 1341.39
```
Modelo 2:
```
      -0.2845  -0.2296  -0.8620
s.e.   0.0544   0.0606   0.0378
sigma^2 estimated as 3.016:  log likelihood = -668.56,  aic = 1345.11
```

Pregunta 4 (1 p.)

¿Sobra algún parámetro en alguno de los modelos? De los tres criterios que aparecen (`sigma^2, log likelihood y aic`) ¿cuáles se pueden utilizar para comparar modelos y cuales no son adecuados para la comparación? Suponiendo que ambos modelos pasan la validación ¿cuál consideras que es mejor modelo?

Al hacer el test de significación de los parámetros, el segundo parámetro del primer modelo da un t-ratio mucho menor que 2 en valor absoluto, por lo que decidiríamos el test por no rechazar el hecho de que este parámetro del modelo teórico sea nulo. Para comparar modelos sobre los mismos datos, cuanto mayor número de parámetros disminuye la varianza residual y aumenta el log-likelihood, por lo que estas medidas no son adecuadas para seleccionar el mejor modelo. El AIC penaliza la adecuación del modelo a los datos si el modelo se complica en exceso (principio de parsimonia). Por ello, la medida para compara modelos sería el AIC (cuanto menor valor, mejor modelo).

En este caso, el AIC menor es el primero de los dos, por lo que sería preferible. El hecho de suprimir el parámetro no significativo mejoraría aún más el AIC.

En las figuras 10 a 15 aparecen los gráficos que permiten hacer la validación de los residuos del modelo 1.

Pregunta 5 (1 p.)

Indica brevemente que representa cada gráfico, que premisa se comprueba en cada figura y realiza la validación del modelo 1.

En el plot de los residuos vemos que hay una varianza bastante constante y no parece haber datos atípicos.
El qqplot indica que los residuos se pueden considerar normales y tampoco se observan atípicos ni colas pesadas (exceso de curtosis).
El histograma de los residuos nos indica también la distribución gaussiana de los residuos
En la ACF y PACF de los residuos, vemos que hay varios retrasos (el 18 y el 20) que son ligeramente significativos pero por estar alejados del origen y no relacionados con el periodo estacional podemos considerar que son residuos independientes, y que el modelo explica toda la estructura de autocorrelación observada en los datos originales.
En el gráfico de los diagnósticos, todos los p-valores de los estadísticos de Ljung-Box están por encima del 0.05 confirmando que se puede asumir que la distribución de los residuos es un ruido blanco.
En conclusión, el modelo parece ser válido porque cumple las hipótesis de homocedasticidad (varianza constante), normalidad e independencia.

A continuación, se estima el modelo 1 para la serie sin las últimas 12 observaciones:

Modelo 1 sin las últimas 12 observaciones:

```
Coefficients:
      0.6182  0.0199  -0.9060  -0.8496
s.e.  0.0915  0.0720   0.0734   0.0403

sigma^2 estimated as 2.986:  log likelihood = -643.2,  aic = 1296.4
```

En la figura 20 aparece el plot de la serie original y las previsiones para 2018 y sus intervalos a partir del modelo obtenido habiendo reservado las últimas 12 observaciones.

Pregunta 6 (0.5 p.)

¿Se puede considerar **estable** el modelo 1? ¿Por qué es importante verificar que el modelo es estable? En base a las previsiones obtenidas ¿consideras que el modelo hace buenas previsiones? ¿En qué basas tu respuesta?

Se puede observar que los coeficientes del modelo con todos los datos y los del que se ajusta sin las últimas observaciones son de la misma magnitud, signo y significación. Por ello, se puede decir que corresponden aproximadamente al mismo modelo y que por lo tanto en el último año el modelo se ha mantenido vigente, lo cual implica que el modelo es estable. Es importante verificarlo para garantizar que las métricas para evaluar la capacidad de previsión (RMSPE y MAPE) de acuerdo al segundo modelo son extrapolables al primero, que es con el que calcularemos las predicciones futuras.
La previsiones obtenidas parecen reproducir los valores que se observan posteriormente (sobre todo en los primeros ocho meses), con ligeras desviaciones en los meses de otoño. En cualquier caso, las observaciones obtenidas se hallan dentro de los límites de los intervalos de confianza, por lo que podemos afirmar que el modelo hace buenas predicciones.