# Examen Parcial de Sèries Temporals

**Anàlisi de Dades. Àlex Batlle Casellas (DNI: 39413434J).**

## Teoria

**Pregunta 1.**

**a)** Es tracta d'un model $ARIMA(1,0,0)$, és a dir, un $AR(1)$. Per tant, sabem que és causal. Però no és invertible, ja que la única arrel de la part $MA$ és $\frac{1}{3} < 1$.

**b)** Es tracta d'un model $ARIMA(1,0,1)$. És causal, ja que les arrels de la part $AR$ són totes majors que 1. En canvi, no és invertible, ja que la única arrel de la part $MA$ és menor o igual a 1.

**c)** És un $ARIMA(2,0,0)$, és a dir, un $AR(2)$. Per tant, és causal. Però no és invertible, ja que les arrels de la part $MA$ són ambdues menors que 1.

**d)** El model és

$$(1 - 1.2B + 0.2B^2)X_t = (1 - 0.5B)Z_t$$

i per tant, un $ARIMA(2,0,1)$. És invertible, doncs les arrels de la part $MA$ són majors que 1. Però no és invertible, doncs a la part $AR$ hi ha una arrel unitària.

**Pregunta 2.**

**a)** Per trobar la variància d'un procés $AR(1)$, fem

$$\gamma(0) = \mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[(\phi X_{t-1} + Z_t)^2\right] = \mathbb{E}\left[\phi^2 X_{t-1}^2 + Z_t^2 + 2\phi X_{t-1}Z_t\right] =$$
$$= \phi^2 \mathbb{E}\left[X_{t-1}^2\right] + \mathbb{E}\left[Z_t^2\right] + 2\phi\, \cancel{\mathbb{E}\left[X_{t-1}Z_t\right]}^{\,0} = \phi^2\gamma(0) + \sigma_Z^2 \implies$$
$$\implies \gamma(0) = \frac{\sigma_Z^2}{1-\phi^2}.$$

$$\gamma(1) = \mathbb{E}\left[X_t X_{t-1}\right] = \mathbb{E}\left[(\phi X_{t-1} + Z_t)X_{t-1}\right] = \phi\mathbb{E}\left[X_{t-1}^2\right] + \cancel{\mathbb{E}\left[Z_t X_{t-1}\right]}^{\,0} = \phi\gamma(0).$$

$$\gamma(h) = \mathbb{E}\left[X_t X_{t-h}\right] = \mathbb{E}\left[(\phi X_{t-1} + Z_t)X_{t-h}\right] = \phi\gamma(h-1) = \phi^h\gamma(0) = \boxed{\phi^h \frac{\sigma_Z^2}{1-\phi^2}.}$$

La funció d'autocorrelació ACF és l'autocovariància entre la variància, que és

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \boxed{\phi^h.}$$

**b)** Si $\phi$ no fos menor que 1, l'autocorrelació augmentaria amb el retard, cosa que no tindria sentit en un model com aquest, ja que estem predient una observació en funció de just l'anterior. És necessari, per tant, que $\phi < 1$, ja que així ens assegurem que el model té sentit.

**c)** L'expressió general de la predicció és:

$$\tilde{X}_{t+h|t} = \mathbb{E}\left[X_{t+h}|X_t, X_{t-1}, \ldots, X_1\right] = \mathbb{E}\left[\phi X_{t+h-1} + Z_{t+h}\right] = \phi\mathbb{E}\left[X_{t+h-1}\right] =$$
$$\phi\mathbb{E}\left[\phi X_{t+h-2} + Z_{t+h-1}\right] = \cdots = \phi^h\mathbb{E}\left[X_t\right] = \boxed{\phi^h X_t.}$$

La predicció convergeix a zero independentment de $X_t$ quan $h \to \infty$, ja que $|\phi| < 1$.

**Pregunta 3.**

**a)** Si $\beta, \gamma \neq 0$, aleshores és un $ARIMA(4,0,1)$ amb constant. Si $\beta = 0, \gamma \neq 0$, aleshores és un $ARIMA(0,0,1)$, i si $\beta \neq 0, \gamma = 0$, és un $ARIMA(4,0,0)$. Si ambdues fossin 0, seria un model constant igual a $\alpha$ amb soroll blanc $ARIMA(0,0,0)$.

**b)** Per a que sigui estacionari, $\alpha = (1-\beta)\mathbb{E}[X_t]$. Si a més volem que sigui invertible, el polinomi $1 - \beta B^4$ no ha de tenir arrels amb mòdul menor a 1, i per tant, $0 < |\beta| < 1$. $\gamma$ ha de fer que el model sigui causal, i per tant, $1 + \gamma B$ no ha de tenir arrels menors que 1, és a dir, que $0 < |\gamma| < 1$.

**c)** Sí que coincidiran. Vegem-ho:

$$\begin{aligned}\tilde{X}_{102|100} &= \mathbb{E}[X_{102}|X_{100},\ldots] = \mathbb{E}[\alpha + \beta X_{98} + Z_{102} + \gamma Z_{101}] = \alpha + \beta X_{98}.\\ \tilde{X}_{103|100} &= \mathbb{E}[X_{103}|X_{100},\ldots] = \mathbb{E}[\alpha + \beta X_{99} + Z_{103} + \gamma Z_{102}] = \alpha + \beta X_{99}.\end{aligned}$$

Com que $X_{99} = X_{98}$, aquests valors són el mateix.

**Pregunta 4.**

La predicció sabent només fins a $t$ és:

$$\begin{gathered}\tilde{X}_{t+2|t} = \mathbb{E}[X_{t+2}|X_t,\ldots] = \mathbb{E}\left[\sum_{i\geq 1}\pi_i X_{t+2-i}\middle|X_t\right] = \sum_{i\geq 1}\pi_i \mathbb{E}\left[X_{t+2-i}\middle|X_t\right] = \\ \pi_1\mathbb{E}[X_{t+1}|X_t] + \sum_{i\geq 2}\pi_i X_{t+2-i} = \pi_1\tilde{X}_{t+1|t} + \sum_{i\geq 2}\pi_i X_{t+2-i}.\end{gathered}$$

La predicció sabent fins a $t+1$ és:

$$\begin{gathered}\tilde{X}_{t+2|t+1} = \mathbb{E}[X_{t+2}|X_{t+1},\ldots] = \mathbb{E}\left[\sum_{i\geq 1}\pi_i X_{t+2-i}\middle|X_{t+1}\right] = \sum_{i\geq 1}\pi_i \mathbb{E}\left[X_{t+2-i}\middle|X_{t+1}\right] = \\ \sum_{i\geq 1}\pi_i X_{t+2-i}\end{gathered}$$

Per tant, efectivament

$$\tilde{X}_{t+2|t+1} = \tilde{X}_{t+2|t} + \pi_1\left(X_{t+1} - \tilde{X}_{t+1|t}\right) = \pi_1\tilde{X}_{t+1|t} + \sum_{i\geq 2}\pi_i X_{t+2-i} + \pi_1 X_{t+1} - \pi_1\tilde{X}_{t+1|t} = \sum_{i\geq 1}\pi_i X_{t+2-i}.$$

**Pregunta 5.**

**a)** El model és

$$\begin{gathered}(1-B^4)(1-B)X_t = (1+\Theta B^4)(Z_t + \theta Z_{t-1}), \text{ es a dir,}\\ \boxed{X_t = X_{t-4} + X_{t-1} - X_{t-5} + Z_t + \Theta Z_{t-4} + \theta Z_{t-1} + \Theta\theta Z_{t-5}.}\end{gathered}$$

La diferenciació estacional es fa per aconseguir que la mitjana de la sèrie sigui constant, mentre que la diferenciació regular es fa per aconseguir que sigui zero.

**b)** Vegem-ho:

$$\begin{gathered}\tilde{X}_{t+2|t} = \mathbb{E}[X_{t+2}|X_t,\ldots] = \\ \mathbb{E}[X_{t-2} + X_{t+1} - X_{t-3} + Z_{t+2} + \Theta Z_{t-2} + \theta Z_{t+1} + \Theta\theta Z_{t-3}|X_t,\ldots] = \\ \tilde{X}_{t+1|t} + \mathbb{E}[X_{t-2} - X_{t-3} + Z_{t+2} + \Theta Z_{t-2} + \theta Z_{t+1} + \Theta\theta Z_{t-3}|X_t,\ldots] = \\ \tilde{X}_{t+1|t} + X_{t-2} - X_{t-3} + \Theta Z_{t-2} + \Theta\theta Z_{t-3} + \mathbb{E}[Z_{t+2} + \theta Z_{t+1}|X_t,\ldots] = \\ \boxed{\tilde{X}_{t+1|t} + X_{t-2} - X_{t-3} + \Theta Z_{t-2} + \Theta\theta Z_{t-3}.}\end{gathered}$$

**c)**

$$\tilde{X}_{t+h|t} = \tilde{X}_{t+1} + \tilde{X}_{t+4} - X_t.$$

# Laboratori

**Pregunta 6.**

- **Figura 1:** es mostra la sèrie. No es comprova necessàriament cap premisa.
- **Figura 2:** són els residus del model. Es comprova que l'amplada de la gràfica dels residus sigui globalment constant (**homoscedasticitat**), és a dir, que els residus tinguin tots la mateixa variància.
- **Figura 3:** aquí hi veiem l'ajust suau de les arrels dels valors absoluts dels residus del model. Es torna a comprovar que els residus tinguin tots una variància/desviació típica semblants (**homoscedasticitat**). En aquest cas, no es validaria, ja que la corba no és plana.
- **Figura 4:** és un plot Quantil-Quantil. Es comprova que els quantils empírics dels residus s'ajustin als quantils teòrics d'una distribució normal amb els seus valors de mitjana i variància (**normalitat**). En aquest cas, veiem un excés de curtosi a la cua esquerra.
- **Figura 5:** es fa el test de Ljung-Box per 84 retards. Aquest test serveix per comprovar que els residus no tinguin correlacions significatives entre ells. En aquest cas, no validaríem la premisa d'**independència** dels residus.
- **Figura 6&7:** són les ACF i PACF dels residus del model. Es comprova altre cop la correlació entre els residus (**independència**). Veiem com en aquest cas no validaríem la hipòtesi, ja que hi ha diversos valors de l'autocorrelació que són significatius.

Per tot el que hem vist (els residus no s'ajusten a normalitat, presenten retards significatius a l'autocorrelació, no sembla que tinguin variància constant) no validaríem els supòsits dels models ARMA.

**Pregunta 7.**

S'han detectat 5 atípics: 2 atípics additius, 2 atípics de canvi de nivell i 1 atípic de canvi transitori. Aquests atípics tenen els següents efectes en la sèrie:

- Atípic de l'observació 14: és un canvi de nivell on la sèrie decreix.
- Atípic de l'observació 64: es tracta d'un atípic additiu, un canvi puntual en la sèrie. En aquest cas, la fa augmentar puntualment.
- Atípic de l'observació 100: es tracta d'un atípic additiu. També fa augmentar puntualment la sèrie.
- Atípic de l'observació 107: es tracta d'un canvi de nivell bastant pronunciat, que fa disminuïr el valor de la sèrie.
- Atípic de l'observació 109: es tracta d'un canvi transitori, que fa disminuïr el valor de la sèrie i l'efecte del qual es nota fins ben entrat l'any posterior a la seva observació.

**Pregunta 8.**

Com a primer model, podríem intentar ajustar un $ARMA(3,6)(2,1)_{12}$, ja que observem que hi ha sis retards significatius a l'ACF (per la part $MA$) i tres retards significatius a la PACF (per la part $AR$). Aquí la discussió podria estar en per què no agafem nou retards de l'ACF: potser els dos retards 7 i 8 sortirien no significatius en aquest cas, i seria una complicació innecessària del model. Per la part estacional, observem que el segon retard estacional de l'ACF és significatiu, el que ens indica que podríem utilizar-lo per estimar el model.

També podria ser que els paràmetres pels retards 7 i 8 sortissin significatius per la part $AR$, i aleshores també podríem mirar d'ajustar un $ARMA(3,9)(2,1)_{12}$.

**Pregunta 9.**

En aquest segon cas, els residus del model s'ajusten molt millor que abans als quantils d'una normal. A més, l'ACF i la PACF dels residus són més semblants al que anomenaríem soroll blanc. Podem veure, en les prediccions, com les primeres del model 2 cauen més a prop de la realitat que les del model 1, però després se n'allunyen més del que ho fan les del model 1. L'amplada dels intervals de confiança sembla ser molt semblant per tots dos models.

**Pregunta 10.**

L'AIC i el BIC són mesures que representen la bondat d'ajust del model a les dades. Ambdues estan penalitzades pel nombre de paràmetres, i el BIC és més estricte que l'AIC. La Sigma2Z és la variància dels residus. Com més petita, més ens interessa el model. El RMSPE i el MAPE són mesures d'exactitud de les prediccions dels models, i representen l'error que comet el model en predir dades que ja tenim. La meanLength és una representació de la precisió del model, és a dir, la longitud mitjana dels intervals de confiança de les prediccions.

En aquest cas, el millor model sembla ser el segon: els seus AIC i BIC són menors que els del primer model, i és més precís, tot i que amb lleugerament més biaix. La variància dels errors també és menor.