

Un altro modello accettabile è il modello ARIMA(1,0,0)(0,1,<sup>1</sup>0), mentre i modelli ARIMA(1,0,0)(1,0,0) e ARIMA(1,0,0)(1,1,1) sono considerati plausibili ma i primi sono migliori

**SERIE STORICHE - prova A**

Padova, 17 Febbraio 2017

Nome \_\_\_\_\_ Cognome \_\_\_\_\_ N. matricola \_\_\_\_\_

1.1 [6]	1.2 [2]	1.3 [2]	1.4 [2]	2.1 [3]	2.2 [2]	3.1 [3]	3.2 [2]	3.3 [1]	3.4 [2]	4.1 [3]	4.2 [2]	Totale

**Esercizio 1** Il file *serie\_A.txt* contiene una serie storica,  $y_t$ , di 240 osservazioni mensili a partire dal gennaio 1997.

1. Identificare gli ordini del modello e riportarli nella seguente tabella:

<i>Serie</i>	$p$	$d$	$q$	$P$	$D$	$Q$	$s$
$y_t$	1	0	0	0	1	0	12

2. Stimare il modello identificato al punto precedente e riportare i risultati nella seguente tabella (significatività al 5%):

<i>Parametro stimato</i>	$\phi_1$						
<i>Stima</i>	0,5006						
<i>p-value</i>	0,0000						

3. **a.** Scrivere il numero di ritardi,  $K$ , utilizzato per stimare le autocorrelazioni del modello:  
**b.** Riportare quante e quali autocorrelazioni risultano esterne alla bande di confidenza ( $\alpha = 0.05$ ) per  $i = 1, \dots, K$ : **60**  
**c.** Riportare il valore della statistica di Ljung-Box,  $Q_K$ , assieme al suo  $p$ -value: **ACF 2: 6, 51, PACF 3: 12, 25, 36**  
**d.** Sulla base dei valori di cui al punto precedente dire se si accetta o rifiuta l'ipotesi nulla di incorrelazione dei residui: **si accetta**
4. **a.** La serie  $y_t$  è correlata? **SI NO**  
**b.** La serie  $y_t$ , presenta un trend? **SI NO**  
**c.** La serie  $y_t$  è stazionaria? **SI NO**  
**d.** I residui del modello sono incorrelati? **SI NO**

**Esercizio 2** Si consideri ancora la serie  $y_t$  dell'esercizio precedente e le due medie mobili

$$M_1 = \left\{ [7]; \frac{1}{16} [1, 2, 3, 4] \right\} \quad \text{e} \quad M_2 = \left\{ [5]; \frac{1}{8} [1, 2, 3, 1, 1] \right\}.$$

1. Si calcolino i valori di  $M_1$  e di  $M_2$  per il mese di gennaio dell'anno 2010:  $M_1 = \mathbf{1,59066}$ ,  $M_2 = \mathbf{1,994324}$
2. Si dica se le due medie mobili mantengono un trend lineare e perché.  $M_1$  **si, perché**  $\sum \theta_i = 1$  e  $\sum i\theta_i = 0$ .  $M_2$  **no, perché**  $\sum i\theta_i \neq 0$ .

**Esercizio 3** Si consideri la serie mensile, *temperature*, relativa alle anomalie nelle temperature Europee (gradi Celsius), per il periodo 1910:1 - 2016:12, contenuta nel file *temperature.gdt*. Si consideri tale serie per il periodo gennaio 1960 - dicembre 1995 e la si denoti con  $x$ .

1. Sulla serie  $x$  si stimino i due modelli ARIMA(1,0,0) e ARIMA(0,1,1) - decidendo se introdurre o meno il termine costante - e si completino le seguenti tabelle, indicando con un asterisco eventuali coefficienti non significativi ( $\alpha = 0.05$ ):

Parametri ARIMA(1,0,0)	$\phi_0$	$\phi_1$	AIC	SC	HQ
Stima		0,3672	1111,91	1119,328	1114,404
s.e.		0,0447			

Parametri ARIMA(0,1,1)	$\phi_0$	$\theta_1$	AIC	SC	HQ
Stima		-0,8671	1149,413	1157,55	1152,626
s.e.		0,0410			

2. Si considerino i residui dei due modelli e si completi la seguente tabella:

Statistica di Ljung-Box		Riportare quante e quali autocorrelazioni risultano esterne alla bande di confidenza ( $\alpha = 0.05$ ) per $i = 1, \dots, 48$	
ARIMA(1,0,0)	ARIMA(0,1,1)	ARIMA(1,0,0)	ARIMA(0,1,1)
$Q_{24} = 29,8094$ $p - value = 0,155$	$Q_{24} = 82,1628$ $p - value = 0,000$	ACF 3: 7, 13, 48	ACF 5: 1,4,15,16,17
$Q_{48} = 47,3495$ $p - value = 0.458$	$Q_{48} = 96,2060$ $p - value = 0,000$	PACF 4: 7, 11, 13, 48	PACF 6: 1,4,9,15,33,48

3. Sulla base dei precedenti risultati dire quale fra i due modelli è preferibile: **ARIMA(1,0,0)**
4. Utilizzando il modello scelto calcolare le previsioni di  $x$  per l'anno 1996 e riportare i valori previsti per i mesi di febbraio e dicembre. **1996.02=-0,144 1996.12=-0,000**

**Esercizio 4** Sulla serie  $x$  si stimi un opportuno trend polinomiale (di ordine massimo 4).

1. Nella tabella sottostante riportare **il grado del polinomio scelto e le stime dei parametri del modello scelto** indicando con un asterisco eventuali coefficienti non significativi.

grado del polinomio	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
2	5,89548	-0,0157883	0,0000		

2. Commentare brevemente l'adattamento ai dati del modello scelto.  $R^2 = 0,0445$  **Adattamento pessimo.**

## SERIE STORICHE - prova B

Padova, 17 Febbraio 2017

Nome \_\_\_\_\_ Cognome \_\_\_\_\_ N. matricola \_\_\_\_\_

1.1 [6]	1.2 [2]	1.3 [2]	1.4 [2]	2.1 [3]	2.2 [2]	3.1 [3]	3.2 [2]	3.3 [1]	3.4 [2]	4.1 [3]	4.2 [2]	Totale

**Esercizio 1** Il file *serie\_B.txt* contiene una serie storica,  $y_t$ , di 240 osservazioni mensili a partire dal gennaio 1997.

1. Identificare gli ordini del modello e riportarli nella seguente tabella:

<i>Serie</i>	$p$	$d$	$q$	$P$	$D$	$Q$	$s$
$y_t$	0	0	1	0	0	1	12

2. Stimare il modello identificato al punto precedente e riportare i risultati nella seguente tabella (significatività al 5%):

<i>Parametro stimato</i>	$\theta_1$	$\Theta_1$					
<i>Stima</i>	0,6922	0,7030					
<i>p-value</i>	0,0000	0,0000					

3. **a.** Scrivere il numero di ritardi,  $K$ , utilizzato per stimare le autocorrelazioni del modello: **60**  
**b.** Riportare quante e quali autocorrelazioni risultano esterne alla bande di confidenza ( $\alpha = 0.05$ ) per  $i = 1, \dots, K$ : **ACF 1 al lag 34, PACF 0**  
**c.** Riportare il valore della statistica di Ljung-Box,  $Q_K$ , assieme al suo  $p$ -value: **46,2156 (0,868)**  
**d.** Sulla base dei valori di cui al punto precedente dire se si accetta o rifiuta l'ipotesi nulla di incorrelazione dei residui: **si accetta**
4. **a.** La serie  $y_t$  è eteroschedastica? **SI NO**  
**b.** La serie  $y_t$ , presenta un trend? **SI NO**  
**c.** La serie  $y_t$  è non stazionaria? **SI NO**  
**d.** I residui del modello sono stazionari? **SI NO**

**Esercizio 2** Si consideri ancora la serie  $y_t$  dell'esercizio precedente e le due medie mobili

$$M_1 = \left\{ [13]; \frac{1}{13} [1, 1, 1, 1, 1, 1, \mathbf{1}] \right\} \quad \text{e} \quad M_2 = \left\{ [5]; \frac{1}{8} [1, 2, \mathbf{2}] \right\}.$$

1. Si calcolino i valori di  $M_1$  e di  $M_2$  per il mese di gennaio dell'anno 2010:  $M_1$  : **0,3825**,  $M_2$  : **0,5249**
2. Si dica se le due medie mobili possono essere usate per destagionalizzare la serie  $y_t$  e perché. Nessuna delle due, perché  $M_1$  annulla stagionalità costanti di periodo 13, mentre  $y_t$  ha periodo 12 e  $M_2$  annulla stagionalità di periodo 4.

**Esercizio 3** Si consideri la serie mensile, *temperature*, relativa alle anomalie nelle temperature Europee (gradi Celsius), per il periodo 1910:1 - 2016:12, contenuta nel file *temperature.gdt*. Si consideri tale serie per il periodo gennaio 1980 - dicembre 2016 e la si denoti con  $x$ .

1. Sulla serie  $x$  si stimino i due modelli ARIMA(1,0,0) e ARIMA(3,0,0) - decidendo se introdurre o meno il termine costante - e si completino le seguenti tabelle, indicando con un asterisco eventuali coefficienti non significativi ( $\alpha = 0.05$ ):

<i>Parametri</i> <i>ARIMA(1,0,0)</i>	$\phi_0$	$\phi_1$	AIC	SC	HQ
<i>Stima</i>	0,6231	0,4420	1195,363	1207,651	1200,209
<i>s.e.</i>	0,0784	0,0427			

<i>Parametri</i> <i>ARIMA(3,0,0)</i>	$\phi_0$	$\phi_1$	$\phi_2^*$	$\phi_3$	AIC	SC	HQ
<i>Stima</i>	0,6180	0,3565	0,0915	0,1456	1180,32	1200,799	1188,396
<i>s.e.</i>	0,1050	0,0471	0,0499	0,0471			

2. Si considerino i residui dei due modelli e si completi la seguente tabella:

Statistica di Ljung-Box		Riportare quante e quali autocorrelazioni risultano esterne alla bande di confidenza ( $\alpha = 0.05$ ) per $i = 1, \dots, 48$	
<i>ARIMA(1,0,0)</i>	<i>ARIMA(3,0,0)</i>	<i>ARIMA(1,0,0)</i>	<i>ARIMA(3,0,0)</i>
$Q_{24} = 74,9137$	$Q_{24} = 29,0828$	ACF 7: 3, 7, 11, 13, 22, 25, 30	ACF 2: 13 30
$p - value = 0,000$	$p - value = 0,112$		
$Q_{48} = 120,2397$	$Q_{48} = 61,0247$	PACF 5: 3, 7, 11, 13, 32	PACF 4: 11, 13, 30, 48
$p - value = 0,000$	$p - value = 0,056$		

3. Sulla base dei precedenti risultati dire quale fra i due modelli è preferibile: **ARIMA(3,0,0)**
4. Utilizzando il modello scelto calcolare le previsioni di  $x$  per l'anno 2017 e riportare i valori previsti per i mesi di febbraio e dicembre. **2017.02=0,594 2017.12=0.619**

**Esercizio 4** Sulla serie  $x$  si stimi un opportuno trend polinomiale (di ordine massimo 4).

1. Nella tabella sottostante riportare **il grado del polinomio scelto e le stime dei parametri del modello scelto** indicando con un asterisco eventuali coefficienti non significativi.

<i>grado del polinomio</i>	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
1	-3.24578	0,00364			

2. Commentare brevemente l'adattamento ai dati del modello scelto.  $R^2 = 0,206$  **Adattamento non molto soddisfacente. Residui correlati.**

## SERIE STORICHE

Padova, 17 febbraio 2017

Nome

Cognome

N. matricola

Corso di Laurea

1.1 [2]	1.2 [3]	1.3 [2]	2 [6]	3.1 [3]	3.2 [2]	4.1 [2]	4.2 [2]	5 [4]	6 [4]	Totale

**Domanda 1** Si consideri il modello:

$$(1 - B)X_t = (1 - \theta B)\epsilon_t$$

dove  $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$  e  $|\theta| < 1$ .

1. Si dica, giustificando la risposta, se il modello considerato è stazionario e/o invertibile e a che classe di processi ARIMA appartiene.
2. Si calcolino le previsioni  $\hat{X}_{n+k}$ .
3. Si trovino gli intervalli di previsione al 95% per  $k = 1, 2$ .

**Soluzione:**

1. Il modello è un  $ARIMA(0, 1, 1)$ , quindi non stazionario in quanto possiede una radice unitaria, ma invertibile essendo  $|\theta| < 1$ .
2. Il modello può essere riscritto come:

$$X_t = X_{t-1} - \theta\epsilon_{t-1} + \epsilon_t$$

pertanto le previsioni richieste, usando le regole del valore atteso condizionato, si ottengono nel modo seguente:

$$\hat{X}_{n+k} = E[X_{n+k} | I_n] = E[X_{n+k-1} | I_n] - \theta E[\epsilon_{n+k-1} | I_n] + E[\epsilon_{n+k} | I_n]$$

pertanto, per  $k = 1$ , il previsore è dato da:

$$\begin{aligned}\hat{X}_{n+1} &= E[X_n | I_n] - \theta E[\epsilon_n | I_n] + E[\epsilon_{n+1} | I_n] = \\ &= X_n - \theta\epsilon_n\end{aligned}$$

e la previsione si ottiene usando il valore osservato  $x_n$  e il residuo stimato  $e_n$ .  
Per  $k = 2$  si ha:

$$\begin{aligned}\hat{X}_{n+2} &= E[X_{n+1} | I_n] - \theta E[\epsilon_{n+1} | I_n] + E[\epsilon_{n+2} | I_n] = \\ &= \hat{X}_{n+1} = X_n - \theta\epsilon_n\end{aligned}$$

quindi, in generale:

$$\hat{X}_{n+k} = \hat{X}_{n+k-1} = \hat{X}_{n+1} = X_n - \theta\epsilon_n$$

3. Per calcolare gli intervalli di previsione abbiamo bisogno di calcolare la varianza degli errori di previsione 1 e 2 passi in avanti.

L'errore di previsione un passo in avanti è:

$$e_{n+1} = X_{n+1} - \hat{X}_{n+1} = X_n - \theta\epsilon_n + \epsilon_{n+1} - X_n + \theta\epsilon_n = \epsilon_{n+1}$$

(come deve essere), e quindi:

$$Var(\epsilon_n) = \sigma_\epsilon^2$$

L'errore di previsione per  $k = 2$  è invece:

$$\begin{aligned} e_{n+2} &= X_{n+2} - \hat{X}_{n+2} = \\ &= X_{n+1} - \theta\epsilon_{n+1} + \epsilon_{n+2} - X_n + \theta\epsilon_n = \\ &= (X_{n+1} - X_n) - \theta\epsilon_{n+1} + \epsilon_{n+2} + \theta\epsilon_n = \\ &= -\theta\epsilon_n + \epsilon_{n+1} - \theta\epsilon_{n+1} + \epsilon_{n+2} + \theta\epsilon_n \\ &= \epsilon_{n+1}(1 - \theta) + \epsilon_{n+2} \end{aligned}$$

(osservando che  $X_{n+1} = X_n - \theta\epsilon_n + \epsilon_{n+1}$  e quindi  $X_{n+1} - X_n = -\theta\epsilon_n + \epsilon_{n+1}$ ) pertanto:

$$Var(e_{n+2}) = Var(\epsilon_{n+1}(1 - \theta) + \epsilon_{n+2}) = \sigma_\epsilon^2[(1 - \theta)^2 + 1]$$

Gli intervalli di confidenza richiesti sono, pertanto:

$$IC(1) = \hat{X}_{n+1} \pm 1.96\sigma_\epsilon$$

$$IC(2) = \hat{X}_{n+2} \pm 1.96\sigma_\epsilon\sqrt{(1 - \theta)^2 + 1}$$

**Domanda 2** Si trovi un processo, appartenente alla classe dei processi ARIMA, che abbia la seguente funzione di autocovarianza:

$$\gamma_0 = 10, \quad \gamma_1 = 0, \quad \gamma_2 = 4, \quad \gamma_k = 0 \quad \forall k \geq 2$$

**Soluzione:** La funzione di autocovarianza data è coerente con quella di un processo  $MA(2)$  con il primo coefficiente a media mobile pari a zero:

$$X_t = \epsilon_t - \theta\epsilon_{t-2}$$

dove  $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ . Per tale processo avremo:

$$\gamma_0 = Var(X_t) = \sigma_\epsilon^2(1 + \theta^2) = 10$$

$$\gamma_1 = Cov(X_t, X_{t-1}) = 0$$

$$\gamma_2 = Cov(X_t, X_{t-2}) = E[(\epsilon_t - \theta\epsilon_{t-2})(\epsilon_{t-2} - \theta\epsilon_{t-4})] = -\theta\sigma_\epsilon^2 = 4$$

$$\gamma_k = Cov(X_t, X_{t-k}) = 0, \quad \forall k \geq 2$$

Dobbiamo quindi risolvere il sistema di due equazioni in due incognite:

$$\begin{cases} \sigma_\epsilon^2(1 + \theta^2) = 10 \\ -\theta\sigma_\epsilon^2 = 4 \end{cases}$$

Professori che porta ad ottenere le due coppie di parametri:  $(\theta = -2, \sigma_\epsilon^2 = 2)$  e  $(\theta = -0.5, \sigma_\epsilon^2 = 8)$ . Osserviamo che abbiamo quindi due processi compatibili con la funzione di autocovarianza data:

1.  $X_t = \epsilon_t + 2\epsilon_{t-2}$  dove  $\epsilon_t \sim WN(0, 2)$
2.  $X_t = \epsilon_t + 0.5\epsilon_{t-2}$  dove  $\epsilon_t \sim WN(0, 8)$

Osserviamo che il primo processo è non invertibile, in quanto

$$1 + 2B^2 = 0$$

porta a due radici (complesse) in modulo minori di uno.

### Domanda 3

Si consideri la seguente serie storica:

$$x = (-6, -5, -3, -4, -3, -2, -2, -1, 0, 2, 3, 4, 4, 4, 5, 4).$$

1. Si completi la seguente tabella (**approssimando alla terza cifra decimale e scrivendo a parte i calcoli necessari per completarla**):

$k$	$\hat{\rho}(k)$	$\hat{P}(k)$
1	???	???
2	0.646	-0.140
3	0.510	0.049
4	0.306	-0.330
5	???	-0.071

2. Sulla base dei risultati di cui al punto 1, dire quale modello potrebbe essere adattato alla serie storica in questione.

### Soluzione:

1. Dobbiamo calcolare le autocorrelazioni globali empiriche ai ritardi  $k = 1$  e  $5$  e il coefficiente di autocorrelazione parziale al ritardo  $k = 1$ . Per fare questo abbiamo bisogno della media e della varianza campionaria:

$$\bar{x} = \frac{1}{16} \sum_{t=1}^{16} x_t = \frac{1}{16}(-6 - 5 + \dots + 5 + 4) = 0,$$

$$\hat{\gamma}_0 = Var(x) = \frac{1}{16} \sum_{t=1}^{16} x_t^2 = 12.875$$

Essendo  $\bar{x} = 0$ , l'autocovarianza empirica al ritardo 1 si calcola come:

$$\hat{\gamma}(1) = \frac{1}{16} \sum_{t=1}^{15} x_t x_{t+1} = \frac{1}{16}[(-6) \cdot (-5) + (-5)(-3) + (-3)(-4) + \dots + 4 \cdot 5 + 5 \cdot 4] = 10.688$$

da cui si ricava:

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = 0.830.$$



Per calcolare il coefficiente di autocorrelazione globale al ritardo 5, dobbiamo prima calcolare l'autocovarianza al ritardo 5:

$$\hat{\gamma}(5) = \frac{1}{16} \sum_{t=1}^{11} x_t x_{t+5} = \frac{1}{16} [(-6) \cdot (-2) + (-5)(-2) + (-3)(-1) + \dots + 2 \cdot 5 + 3 \cdot 4] = 1.438$$

da cui l'autocorrelazione empirica al ritardo 5:

$$\hat{\rho}(5) = \frac{\hat{\gamma}(5)}{\hat{\gamma}(0)} = 0.112.$$

Per quanto riguarda l'autocorrelazione parziale al ritardo 1, non abbiamo bisogno di calcoli. Professorilare nulla essendo pari a

$$\hat{P}(1) = \hat{\rho}(1) = 0.830.$$

2. La lunghezza della serie è  $n = 16$  per cui le bande di fiducia al 5% per l'ipotesi di non significatività dei coefficienti di autocorrelazione sono:

$$\pm 1.96/\sqrt{16} = \pm 0.49$$

Questo significa che, per quanto riguarda la funzione di autocorrelazione empirica,  $\hat{\rho}(1)$ ,  $\hat{\rho}(2)$  e  $\hat{\rho}(3)$  sono significativamente diversi da zero, mentre per quanto riguarda la funzione di autocorrelazione parziale, solo  $\hat{P}(1)$  è significativo. Ciò è coerente con un processo  $AR(1)$ .

#### Domanda 4

Si considerino le due medie mobili

$$M_1 = -y_{t-2} + \frac{1}{2}y_{t-1} + \frac{5}{4}y_t + \frac{3}{4}y_{t+1} - \frac{1}{2}y_{t+2}$$
$$M_2 = \frac{1}{4}y_{t-1} + \frac{1}{2}y_t + \frac{1}{4}y_{t+1}$$

e si supponga di volerle applicare (i) ad una serie storica incorrelata al fine di ridurne la variabilità e (ii) ad una serie storica generata da un modello con trend evolutivo e componente erratica, al fine di stimarne il trend.

1. Quale delle due medie mobili sarebbe preferibile e perché?
2. Ci sono motivi per ritenere che una delle due medie mobili non debba essere utilizzata?

**Soluzione:** Esercizio n. 3.4 dell'Eserciziario.

1. La riduzione di variabilità di una serie a seguito dell'applicazione di una media mobile si misura con il rapporto di riduzione della varianza residua,  $\sum_{i=-m}^m \vartheta_i^2$ . Per  $M_1$  tale quantità risulta pari a

$$1 + 1/4 + 25/16 + 9/16 + 1/4 = (16 + 4 + 25 + 9 + 4)/16 = 58/16 = 3,625.$$

Pertanto, questa media mobile produce una serie storica la cui componente accidentale ha una variabilità che è oltre tre volte e mezzo rispetto a quella della serie originale! È chiaro quindi che l'obiettivo di riduzione della componente erratica nel caso (i) può essere perseguito solo mediante l'utilizzo di  $M_2$ , con un rapporto di riduzione pari a

$$1/16 + 1/4 + 1/16 = (1 + 4 + 1)/16 = 6/16 = 0,375.$$

Anche per il caso (ii) la media  $M_2$  risulta più appropriata. Per essa infatti valgono le condizioni  $\sum_{i=-m}^m \vartheta_i = 1$  e  $\sum_{i=-m}^m i\vartheta_i = 0$ , mentre per  $M_1$  si ha

$$\sum_{i=-2}^2 i\vartheta_i = \frac{8 - 2 + 3 - 4}{4} = \frac{5}{4} \neq 0.$$

Ne consegue che l'applicazione di  $M_1$  non lascerebbe inalterato un trend di tipo lineare.

2. Entrambe le medie mobili sono centrate, mentre solo  $M_2$  è simmetrica. La proprietà di simmetria dei coefficienti è fondamentale per ottenere i risultati desiderabili dall'applicazione di una media mobile, ovvero la riduzione della componente d'errore, la conservazione del trend o la depurazione della componente stagionale. L'utilizzo della media mobile  $M_1$  è quindi sconsigliabile.

**Domanda 5** Dare la definizione di processo trend-stazionario e a trend stocastico e fornire un esempio in entrambi i casi.

**Domanda 6** Descrivere brevemente come stimare la componente stagionale secondo l'approccio classico all'analisi delle serie storiche.