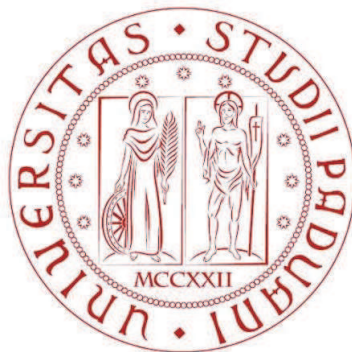


**UNIVERSITA' DEGLI STUDI DI PADOVA**  
Laurea Magistrale in Scienze Statistiche



**L'EFFETTO DELLA TERAPIA  
ANTIBIOTICA SULLA RESISTENZA  
BATTERICA: UNO STUDIO EMPIRICO**

Relatore: Prof.ssa Alessandra R. Brazzale  
Dipartimento di Scienze Statistiche

Correlatore: Dott. Andrea Bortolato  
Dipartimento di Medicina  
U.O.C Istituto di Anestesia e Rianimazione

Laureanda: Dott.ssa Silvia Masiero

**ANNO ACCADEMICO 2011/2012**

# Ringraziamenti

Quando parlo della mia vita universitaria e racconto alle persone di essere arrivata alla tesi, nella maggior parte dei casi mi sento rispondere “Beh, ormai hai finito!”. In realtà, questi ultimi mesi sono stati, per me, i più faticosi e i più impegnativi di tutto il biennio magistrale ma, grazie alle persone che mi sono state vicine, si sono rivelati anche i più entusiasmanti.

Il primo ringraziamento va ai miei genitori, che, come sempre hanno dimostrato di avere un’infinita pazienza, e a tutta la mia famiglia.

Un profondo grazie a Davide, che non mi ha mai lasciata sola, supportando ogni mia scelta e trovando sempre la parola giusta al momento giusto.

Grazie a Katuscia che per me non è una semplice cugina, ma un esempio da seguire.

Un grazie speciale ad Ilaria che, seppur distante, mi è sempre stata vicina con i suoi preziosi consigli.

Un caloroso ringraziamento alle amiche e agli amici che non hanno smesso un attimo di credere in me e non mi hanno mai fatto mancare il loro sostegno.

Un affettuosissimo grazie a tutti i miei compagni di viaggio “statistici” che hanno arricchito ogni mia giornata in Facoltà.

Infine, un grazie a tutte le persone che ho incontrato durante questo cammino, che custodirò gelosamente dentro di me.







# Indice

<b>1. Introduzione</b>	<b>9</b>
1.1. La resistenza batterica . . . . .	9
1.1.1. Cos'è . . . . .	9
1.1.2. Pseudomonas aeruginosa . . . . .	9
1.2. Studi recenti . . . . .	10
1.3. Motivazione del lavoro . . . . .	12
1.4. Organizzazione del testo . . . . .	14
<b>2. Analisi dei dati reali</b>	<b>15</b>
2.1. Concetti teorici . . . . .	15
2.1.1. ACF e PACF . . . . .	17
2.1.2. Processo rumore bianco . . . . .	18
2.1.3. Processi MA, AR e ARMA . . . . .	18
2.1.4. CCF e modelli a funzione di trasferimento . . . . .	20
2.2. Risultati . . . . .	23
2.3. Considerazioni finali . . . . .	27
<b>3. Lo studio di simulazione</b>	<b>29</b>
3.1. Gli scenari . . . . .	29
3.2. Generazione dei dati simulati . . . . .	30
3.2.1. Cenni teorici . . . . .	30
3.2.2. La simulazione della DDD . . . . .	33
3.2.3. La simulazione della resistenza batterica . . . . .	37
3.3. Costruzione di un modello TF . . . . .	40
<b>4. Analisi dei dati simulati</b>	<b>43</b>
4.1. Studio di un caso: lo scenario A . . . . .	44
4.2. Confronti tra scenari . . . . .	48

*Indice*

4.3. Considerazioni finali . . . . .	61
5. Conclusioni e raccomandazioni	65
5.1. Conclusioni . . . . .	65
5.2. Raccomandazioni . . . . .	66
A. Appendice 1	67
A.1. Figure . . . . .	68
A.2. Listato R . . . . .	70
B. Appendice 2	73
C. Appendice 3	75
D. Appendice 4	79

# Elenco delle figure

1.1.	Serie storiche mensili dei dati a nostra disposizione, distinte per reparto.	13
2.1.	Possibile relazione temporale tra la serie di output $y_t$ e quella di input $x_t$ . (Pankratz, 1991)	22
2.2.	Serie storiche DDD (linea continua) e resistenza batterica (linea tratteggiata), distinte per reparto.	25
2.3.	Cross-correlazioni tra DDD e resistenza batterica, distinte per reparto.	26
3.1.	Schema concettuale della stima bootstrap dello scarto quadratico medio. (Efron, 2003)	31
3.2.	Serie DDD simulate con lunghezza pari a 58 mensilità.	34
3.3.	Serie DDD simulate con lunghezza pari a 72 mensilità.	35
3.4.	Serie DDD simulate con lunghezza pari a 96 mensilità.	36
4.1.	Scenario A – Istar 1.	45
4.2.	Scenario A – Istar 2.	47
4.3.	Confronto tra scenari mantenendo fissa la lunghezza temporale – Istar 1. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)	51
4.4.	Confronto tra scenari mantenendo fissa la lunghezza temporale – Istar 2. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)	52
4.5.	Confronto tra scenari mantenendo fissa la numerosità – Istar 1. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)	53
4.6.	Confronto tra scenari mantenendo fissa la numerosità – Istar 2. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)	54
A.1.	Analisi esplorative: ACF e PACF per le resistenze batteriche.	68
A.2.	Analisi esplorative: ACF e PACF per la dose di antibiotico.	69
D.1.	Scenario 1 – Istar 1.	80
D.2.	Scenario 2 – Istar 1.	81



*Elenco delle figure*

<b>D.3. Scenario 3 – Istar 1.</b>	82
<b>D.4. Scenario 1 – Istar 2.</b>	83
<b>D.5. Scenario 2 – Istar 2.</b>	84
<b>D.6. Scenario 3 – Istar 2.</b>	85

# Elenco delle tabelle

1.1. Dati relativi ai reparti Istar 1 e Istar 2 (medie). . . . .	14
3.1. Coefficienti del modello (3.1), applicati per la determinazione della probabilità di manifestare PSAE resistente al MEM. . . . .	39
4.1. Proporzioni di p-value < 0.05 (Scenario A – Istar 1). . . . .	45
4.2. Proporzioni di p-value < 0.05 (Scenario A – Istar 2). . . . .	47
4.3. Potenza (blu) e significatività (rosso) dei diversi ritardi temporali, riferiti alle cross-correlazioni, distinti per scenario – Istar 1. . . . .	49
4.4. Potenza (blu) e significatività (rosso) dei diversi ritardi temporali, riferiti alle cross-correlazioni, distinti per scenario – Istar 2. . . . .	49
4.5. Cross-correlazioni (scenari simulati) – Istar 1. . . . .	56
4.6. Cross-correlazioni (scenari simulati) – Istar 2. . . . .	57
4.7. Stime bootstrap e relativi intervalli di confidenza per i parametri del modello (4.1)- Istar 1. . . . .	59
4.8. Stime bootstrap e relativi intervalli di confidenza per i parametri del modello (4.1) - Istar 2. . . . .	60
4.9. Cross-correlazioni con numerosità pari 20, 40, 60 ed 80 casi – Istar 1 . . .	62
4.10. Cross-correlazioni con numerosità pari a 20, 40, 60 e 80 casi – Istar 2. . .	63



# 1. Introduzione

## 1.1. La resistenza batterica

### 1.1.1. Cos'è

La resistenza batterica è la capacità di un microorganismo di non subire l'attività inibente da parte di un farmaco antibatterico.

In letteratura si possono trovare due tipi di resistenza: quella naturale e quella acquisita. La prima si sviluppa quando la specie batterica non ha mai risentito dell'attività dell'antibiotico, ovvero, lo incontra per la prima volta; la seconda, invece, si presenta, spesso in corso di trattamento, nelle specie già precedentemente esposte a farmaci. Ai fini di una terapia antibiotica l'interesse è rivolto esclusivamente alla resistenza acquisita poiché, in pratica clinica, non si prescrive mai una cura verso la quale c'è resistenza naturale.

Alcuni batteri, sviluppando una resistenza acquisita, riescono ad opporsi ai medicinali tramite diversi meccanismi: il primo è il caso in cui il farmaco somministrato non riesce a penetrare nella cellula che deve debellare poiché non riconosce i canali attraverso i quali entrarvi; il secondo caso è quello in cui il farmaco raggiunge il bersaglio ma subito viene inattivato dagli enzimi del batterio; l'ultimo caso è quello in cui il farmaco raggiunge integro la cellula nemica ma lo trova alterata, non la riconosce e, di conseguenza, non la neutralizza.

### 1.1.2. *Pseudomonas aeruginosa*

Il caso che noi andremo a considerare è quello del batterio *Pseudomonas aeruginosa* (abbreviabile con l'acronimo PSAE). Questo agente patogeno è ampiamente diffuso poiché ha fabbisogni nutritivi minimi, tollera un'ampia gamma di temperature (da 4° a 42° C) ed è resistente a molti antibiotici e disinfettanti. In realtà, il riscontro di questo batterio in una fonte ambientale ospedaliera (alimenti, fiori recisi, lavandini, servizi igienici, attrezzature usate per la terapia, ecc.) significa molto poco, a meno che non vi è evidenza epidemiologica che il sito sia anche una fonte di infezione: secondo molti esperti, quindi, è necessaria

## 1. Introduzione

l'interazione di fattori multipli affinché esso causi malattia. Inoltre, le patologie causate da PSAE sono principalmente opportunistiche, cioè limitate a quei pazienti trattati con antibiotici ad ampio spettro che sopprimono la flora batterica intestinale o a degenti con difese immunitarie compromesse.

*Pseudomonas aeruginosa* può causare ai ricoverati diverse sindromi cliniche. Esso può provocare infezioni del tratto respiratorio che possono variare in gravità da una tracheo-bronchite benigna ad una grave broncopolmonite necrocizzante. Un attacco di questo tipo può essere riscontrato in pazienti con fibrosi cistica o con altre malattie polmonari croniche. Più specificatamente, le condizioni che predispongono i malati alle infezioni causate da questo batterio comprendono precedenti terapie con antibiotici ad ampio spettro, i quali eliminano la normale e protettiva flora batterica, e l'uso di apparecchiature per la ventilazione meccanica, che può fare penetrare il microorganismo nelle basse vie respiratorie. PSAE può causare anche una varietà di infezioni primarie della pelle: quelle più riscontrate sono quelle delle ferite da ustione. La follicolite, derivata dall'immersione in acqua contaminata (ad es. piscine e saune), è un'altra comune infezione causata dal batterio, come lo è anche l'infezione delle ossa e della cartilagine del piede (osteocondite) in seguito ad una lesione penetrante. Sono frequenti anche infezioni del tratto urinario, principalmente in pazienti portatori di catetere urinario, e le infezioni dell'orecchio e dell'occhio (Murray, 2010).

*Pseudomonas aeruginosa* è intrinsecamente resistente a molti antibiotici e sviluppa una resistenza di primo tipo poiché può mutare in ceppi ancora più resistenti durante la terapia. Il suo principale meccanismo di resistenza è la mutazione delle proteine porine impedendo, così, all'antibiotico di riconoscere i canali attraverso i quali entrare nella cellula. La modifica delle pareti delle porine può avvenire più volte e, se ciò accade, il germe può sviluppare simultaneamente una resistenza a molte classi di antibiotici.

La terapia antimicrobica per le infezioni da PSAE, per questo, è spesso deludente in quanto, da un lato, i batteri sono tipicamente resistenti alla maggior parte degli antibiotici e, dall'altro, lo stato di immunocompromissione del paziente non facilita l'eliminazione del patogeno. Affinché la terapia abbia successo in degenti con infezioni particolarmente gravi è necessaria una combinazione di antibiotici attivi.

## 1.2. Studi recenti

La letteratura riguardante la resistenza batterica ed, in particolare, sulla la relazione tra il batterio *Pseudomonas aeruginosa* e i farmaci carbapenemi è decisamente ampia.

Il nostro lavoro prende spunto da uno studio svolto nel 1999 da alcuni ricercatori del Microbiology Laboratory and Pharmacy di Alicante (Spagna) in collaborazione con il National Centre for Hospital Hygiene di Copenhagen (Danimarca). Essi, avendo a disposizione dati rilevati dal 1991 al 1998 nell'ospedale danese, riuscirono a dimostrare la presenza di una marcata connessione tra dosi di antibiotici e sviluppo delle resistenze batteriche. Vennero utilizzate due diverse metodologie per svolgere le analisi: in un primo momento, per l'esplorazione e la modellazione dei dati, vennero adottati i modelli ARIMA e, successivamente, per far emergere la relazione tra le serie, si introdussero i modelli a funzione di trasferimento. I risultati ottenuti furono molto soddisfacenti poiché emerse una stretta relazione tra l'uso dei carbapenemici e la resistenza del germe in questione: la regolamentazione dell'uso dei medicinali, quindi, risultò fondamentale per sorvegliare lo sviluppo di eventuali resistenze batteriche.

Fujimura et al. studiarono circa 200 casi di isolamenti di PSAE nell'area di Tohoku in Giappone: nel loro articolo dimostrarono che tra tutti i casi isolati resistenti ai carbapenemici alcuni non sviluppavano la resistenza specifica verso il Meropenem. Confutarono quindi la tesi di Okamoto et al. che sostenevano il fatto che *Pseudomonas aeruginosa* manifestasse resistenze diverse in base ai differenti antibiotici che si trovava ad affrontare.

Rogues et al. isolarono più di 6000 casi di PSAE da 47 ospedali francesi. Essi non focalizzarono la loro analisi esclusivamente sulla relazione tra i carbapenemici e il batterio ma considerarono un'ampia gamma di antibiotici. Conclusero il loro articolo sostenendo una stretta e significativa relazione tra la somministrazione di farmaci con fluoroquinolone e la resistenza sviluppata nei pazienti dal batterio. In altre parole, anche per questi Autori la possibilità di controllare le dosi di farmaco da somministrare consente la regolamentazione della crescita delle resistenze.

Nel 2005 Kollef, della Washington University School of Medicine, cercò di spiegare quali potessero essere le strategie per limitare la resistenza batterica. Secondo lui la prima cosa da fare era sorvegliare le prescrizioni degli antibiotici in modo da regolarne il dosaggio ed ottimizzare le loro proprietà. Kollef sosteneva che avere delle linee guida che regolassero l'uso dei farmaci fosse indispensabile in primo luogo per limitare l'uso inappropriato di antibiotici ed, in secondo luogo, per fare in modo che i batteri da debellare non creassero resistenze a dosi elevate di farmaco che avrebbero comportato un aumento della dose nella terapia successiva. Un'altra strategia da lui proposta era quella di accorciare i tempi terapeutici nei quali somministrare il farmaco: così facendo, il batterio non avrebbe avuto il tempo necessario per adattarsi all'antibiotico.

Dei ricercatori croati dell'University Hospital e della Medical School di Zagabria inve-

## 1. Introduzione

stigarono il problema delle resistenze batteriche adottando due metodologie: da un lato l'analisi delle serie temporali e dall'altra la correlazione lineare semplice (coefficiente di Pearson). Le due tecniche portarono a risultati completamente diversi: se le serie temporali fecero emergere un significativo legame tra le assunzioni di medicinale e lo sviluppo di resistenze, l'applicazione del coefficiente di correlazione di Pearson non evidenziò nulla. Quindi, l'obiettivo secondario, ma non meno importante, dello studio fu quello di dimostrare come i risultati dipendano anche dalla tecnica che si decide di adottare per svolgere le analisi. In questo caso, infatti, se ci si fosse fermati alla sola correlazione, non si sarebbe riscontrato nessun legame tra farmaco e resistenze che, invece, era ben presente nei dati.

### 1.3. Motivazione del lavoro

Lo studio che andremo ad intraprendere è basato su dati rilevati nei due reparti di rianimazione polivalenti (Istar 1 e Istar 2) dell'Ospedale Civile di Padova e ci sono stati forniti dal dott. Andrea Bortolato che lavora presso essi. Egli ha riscontrato numerosi casi di pazienti affetti da *Pseudomonas aeruginosa* resistente agli antibiotici e, poiché il batterio, come si è già detto nel paragrafo precedente, sviluppa una resistenza molto forte, ha chiesto il nostro intervento per capire se esiste una relazione tra questa e le dosi di antibiotico, tale da giustificare eventuali politiche di regolamentazione dei farmaci e, di conseguenza, dei batteri immuni ai medicinali.

L'ipotesi che sta alla base del quesito sottoposto alla nostra attenzione è che lo sviluppo della resistenza batterica sia una diretta conseguenza della quantità di farmaco distribuita ai degenti: sembra, infatti, che ad una quantità alta di medicinale corrisponda una resistenza più elevata e, viceversa, ad una quantità bassa di antibiotico sia collegata una resistenza minore.

L'obiettivo dello studio è, dunque, quello di identificare se esiste e di che tipo è la relazione tra la resistenza sviluppata da *Pseudomonas aeruginosa* e le dosi di antibiotico somministrate ai pazienti ricoverati nei due reparti di rianimazione. Tuttavia, l'analisi dei dati a nostra disposizione, non è stata sufficiente per adempiere a questo scopo, probabilmente a causa della bassa numerosità dei pazienti e alla contenuta lunghezza delle serie storiche. Per questo motivo, si è presa la decisione di utilizzare i dati reali per simulare dei nuovi scenari nei quali si lavora sulla lunghezza delle serie e sull'aumento della numerosità dei casi: così facendo, si avranno a disposizione maggiori informazioni che permetteranno di riscontrare il vero legame tra batterio e farmaco. Inoltre, l'analisi di scenari molto diversi tra loro, ci darà la possibilità di comprendere quale quantità, tra lunghezza del periodo di

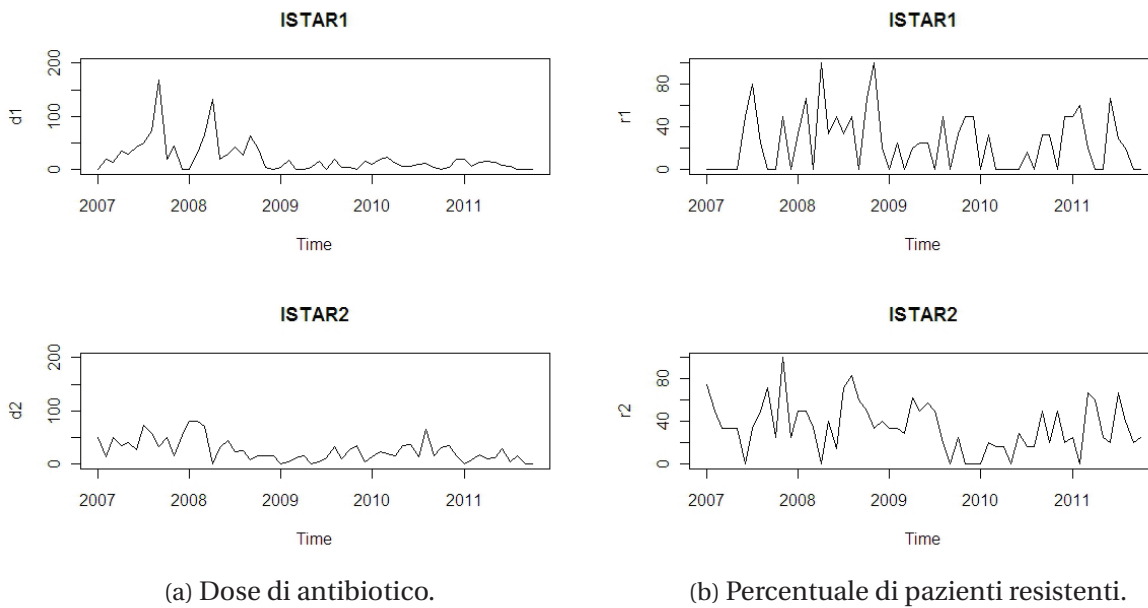


Figura 1.1.: Serie storiche mensili dei dati a nostra disposizione, distinte per reparto.

osservazione e numero di casi, sia quella che incide maggiormente sul conseguimento dei risultati sperati.

Ci è stato chiesto di soffermare l'attenzione solo su un tipo di antibiotico: il Meropenem (abbreviato MEM). Questa scelta è dovuta al fatto che il MEM, appartenente alla famiglia dei farmaci carbapenemici, è, tra questi, quello più incisivo nelle terapie verso gli agenti patogeni multiresistenti. Tuttavia, essendo il Meropenem il farmaco più utilizzato contro i batteri resistenti ed, in particolare, rappresenta l'ultima linea di terapia contro lo *Pseudomonas aeruginosa*, c'è il rischio che il suo frequente e massiccio impiego favorisca lo sviluppo di una sempre più forte resistenza da parte del germe stesso.

I dati ci sono stati forniti dalla farmacia dell'ospedale e riguardano sia la proporzione di pazienti che hanno manifestato la resistenza batterica che le dosi di antibiotico. Si tratta di dati raccolti mensilmente da gennaio 2007 a ottobre 2011, dei quali si possono osservare le rappresentazioni nella Figura 1.1. Inoltre, in Tabella 1.1, si mostrano dei dati relativi al numero dei degenti, alla percentuale di pazienti resistenti e alle dosi di farmaco nei due reparti di rianimazione.

Dati di questo tipo derivano da quella che viene definita sorveglianza microbiologica, ovvero uno strumento tecnico che permette alla farmacia ospedaliera di monitorare i diversi reparti. Il programma di sorveglianza comprende metodiche di campionamento di materiali biologici (e non) con lo scopo di definire l'eventuale agente patogeno responsabile di



## 1. Introduzione

	<b>Istar 1</b>	<b>Istar 2</b>
<b>Numero pazienti</b>	3	5
<b>Pazienti resistenti</b>	25 %	35 %
<b>DDD</b>	21.6	25.8

Tabella 1.1.: Dati relativi ai reparti Istar 1 e Istar 2 (medie).

uno stato infettivo e l'epidemiologia dei germi di un determinato reparto. Da questo tipo di rilevazioni derivano due diversi indici: la pressione di colonizzazione e la pressione di selezione. La prima è la percentuale di isolamenti di una determinata specie batterica sul totale degli isolamenti che ci permette di conoscere la situazione dell'endemia di reparto e anche eventuali epidemie all'interno dell'unità, mentre la seconda riguarda il consumo di antibiotico. Per standardizzare l'utilizzo dei medicinali si usufruisce della dose definita giornaliera<sup>1</sup>, ovvero, un indice di quanto farmaco viene utilizzato sui degenti che transitano in un reparto in un determinato periodo di tempo. E' importante sottolineare che la DDD non è né la dose di antibiotico assunto da un singolo paziente né la dose farmacologica consigliata, bensì esclusivamente uno strumento tecnico adottato internazionalmente.

### 1.4. Organizzazione del testo

Nel Capitolo 2 verranno illustrate le analisi esplorative svolte con i dati consegnateci dal dott. Bortolato. Mostriamo quali sono i risultati ottenuti soffermando l'attenzione soprattutto sui problemi riscontrati.

A causa della mancanza di evidenza empirica nei nostri dati, nel Capitolo 3 verranno introdotte le tecniche di simulazione che si adotteranno per costruire nove diversi scenari, ampliando la lunghezza delle serie temporali e la numerosità dei pazienti reali.

Nel Capitolo 4, con i dati simulati, prima si costruiranno le funzioni di cross-correlazione e i modelli a funzione di trasferimento e, poi, si commenteranno le relazioni riscontrate tra le nostre variabili di interesse.

Nel Capitolo 5 si illustreranno le conclusioni del lavoro e si daranno delle raccomandazioni ai Medici sia in merito alla raccolta dei dati, sia su come si dovranno muovere per riuscire ad evidenziare le relazioni tra antibiotico e resistenza batterica.

---

<sup>1</sup>La dose definita giornaliera o DDD, dall'inglese *defined daily dose*, è l'unità di misura standard della prescrizione farmaceutica, specificata dall'Organizzazione Mondiale della Sanità come la "dose di mantenimento giornaliera media di un farmaco utilizzato per la sua indicazione principale nell'adulto".

## 2. Analisi dei dati reali

In questo capitolo si illustreranno i metodi utilizzati ed i risultati delle analisi descrittive ed inferenziali effettuate sui dati. Si partirà da alcune nozioni teoriche riguardanti i processi stocastici, le funzioni di autocorrelazione (globali e parziali) e quelle di cross-correlazione, per arrivare a spiegare cosa sono i modelli a media mobile, quelli autoregressivi e le funzioni di trasferimento. Nell'ultima parte del capitolo, si esporranno gli esiti dell'esplorazione e le difficoltà riscontrate nello svolgimento.

### 2.1. Concetti teorici

Dati uno spazio di probabilità  $(\Omega, \mathcal{F}, P)$  e uno spazio parametrico  $T$ , un processo stocastico è una funzione  $Z(\omega, t)$  finita a valori reali che, per ogni  $t \in T$ , è una variabile casuale misurabile in funzione di  $\omega \in \Omega$ . Tenendo fisso  $t$ ,  $Z(\omega, t)$  è una unica variabile casuale mentre, fissando  $\omega$ ,  $Z(\omega, t)$  è una realizzazione, in funzione di  $t$ . In altre parole, un processo stocastico rappresenta l'insieme di tutte le possibili realizzazioni di una variabile casuale e, se si considera questa indicizzata dal tempo, una singola realizzazione viene definita serie temporale.

Si consideri ora un insieme finito di variabili casuali  $\{Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}\}$  che hanno origine da un processo stocastico  $\{Z(\omega, t) : t = 0, \pm 1, \pm 2, \dots\}$ . La funzione di distribuzione  $n$ -dimensionale è definita da

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = P\{\omega : Z_{t_1} \leq x_1, \dots, Z_{t_n} \leq x_n\},$$

dove  $x_i, i = 1, \dots, n$  è un qualsiasi numero reale.

Un processo stocastico viene definito stazionario in distribuzione se

$$F_{Z_{t_1}, \dots, Z_{t_n}}(x_1, \dots, x_n) = F_{Z_{t_1+k}, \dots, Z_{t_n+k}}(x_1, \dots, x_n) \quad (2.1)$$

qualsiasi sia  $t$  e qualunque sia  $k$ , ovvero se la famiglia delle ripartizioni finite è invariante per le traslazioni rispetto a  $t$ . Un processo risulta essere stazionario in senso forte se e solo

## 2. Analisi dei dati reali

se la (2.1) vale per qualsiasi  $n$ , i.e.,  $n = 1, 2, \dots$

Avendo compreso che  $Z(\omega, t)$  rappresenta un insieme di variabili casuali indicizzate dal tempo e definite in uno spazio campionario, da questo momento in poi, per richiamare un processo stocastico si userà la notazione  $Z(t)$  o  $Z_t$ .

Per un processo stocastico che assume esclusivamente valori reali e, quindi, definito a valori reali  $\{Z_t : t = 0, \pm 1, \pm 2, \dots\}$ , si definiscono nel seguente modo la media

$$\mu_t = E(Z_t),$$

la varianza

$$\sigma_t^2 = E(Z_t - \mu_t)^2,$$

la funzione di covarianza tra  $Z_{t_1}$  e  $Z_{t_2}$

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}),$$

e la funzione di correlazione tra  $Z_{t_1}$  e  $Z_{t_2}$

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}}.$$

Per un processo stazionario in senso forte, poiché la funzione di distribuzione rimane invariata per ogni  $t$ , la media  $\mu_t = \mu$  è una costante, dato  $E(|Z_t|) < \infty$ . Per lo stesso motivo, se  $E(Z_t^2) < \infty$ , allora anche  $\sigma_t^2 = \sigma^2$ , ed è anche essa una costante. Inoltre, poiché  $F_{Z_{t_1}, Z_{t_1}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$  per ogni valore di  $t_1$ ,  $t_2$  e  $k$ , si ottiene che

$$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k)$$

e

$$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k).$$

Infine, ipotizzando  $t_1 = t - k$  e  $t_2 = t$ , ne consegue che

$$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k$$

e

$$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k.$$

In definitiva, per un processo stazionario in senso forte, che ha i primi due momenti finiti,

si ha che le funzioni di covarianza e correlazione tra  $Z_t$  e  $Z_{t+k}$  dipendono soltanto dalla differenza temporale  $k$ .

### 2.1.1. ACF e PACF

Trattandosi di analisi di serie temporali, è utile introdurre alcuni metodi che trattano le serie storiche sotto il profilo della loro evoluzione dinamica. In questo paragrafo si introdurranno la funzione di autocorrelazione (ACF) e la funzione di autocorrelazione parziale (PACF), le quali servono da un lato per investigare la struttura interna dei fenomeni temporali tramite la correlazione dei valori consecutivi, dall'altro per specificare un modello statistico a partire da una serie storica.

Per un processo stazionario  $\{Z_t\}$  si ha che la media  $E(Z_t) = \mu$  e la varianza  $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$  sono costanti mentre la covarianza  $Cov(Z_t, Z_s)$  è funzione solo della differenza temporale  $|t - s|$ . In questo caso, si possono semplificare le notazioni presentate al paragrafo precedente, e, dunque, si scriveranno la covarianza tra  $Z_t$  e  $Z_{t+k}$  come

$$\gamma_k = Cov(Z, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu), \quad (2.2)$$

e la correlazione tra  $Z_t$  e  $Z_{t+k}$  nel seguente modo

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t) Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}. \quad (2.3)$$

Come funzioni di  $k$ , nell'analisi delle serie temporali,  $\gamma_k$  viene definita funzione di autocovarianza, mentre  $\rho_k$  è chiamata funzione di autocorrelazione: esse rappresentano, rispettivamente la covarianza e la correlazione tra  $Z_t$  e  $Z_{t+k}$  derivanti da uno stesso processo stocastico, separate solo da  $k$  ritardi temporali.

E' facile dimostrare che, per un processo stazionario, la funzione di autocovarianza (2.2) e la funzione di autocorrelazione (2.3) godono delle seguenti proprietà:

1.  $\gamma_0 = Var(Z_t)$ ;  $\rho_0 = 1$ ;
2.  $|\gamma_k| \leq \gamma_0$ ;  $|\rho_k| \leq 1$ ;
3.  $\gamma_k = \gamma_{-k}$  e  $\rho_k = \rho_{-k}$  per ogni  $k$ . Sono, quindi, simmetriche rispetto allo 0;
4. sono semidefinite positive; valgono quindi

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0$$

## 2. Analisi dei dati reali

e

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0$$

per qualsiasi momento temporale  $t_i$  e qualsiasi numero reale  $\alpha_i$ .

Per analizzare la correlazione tra le variabili  $Z_t$  e  $Z_{t+k}$ , la funzione ACF tiene conto delle correlazioni seriali di tutti i ritardi intermedi  $t+1, t+2, \dots, t+k-1$ . Per eliminare ciò, si definisce la funzione di autocorrelazione parziale per i ritardi  $k = 1, 2, \dots$ , intesa come la correlazione lineare fra  $Z_t$  e  $Z_{t+k}$  al netto delle correlazioni lineari intermedie. Analiticamente, la PACF per il processo stazionario  $\{Z_t\}$ , è definita mediante la formula

$$\text{Corr}(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1}).$$

### 2.1.2. Processo rumore bianco

Un processo  $\{a_t\}$  viene definito rumore bianco se rappresenta una sequenza di variabili casuali derivante da una distribuzione fissa con media  $E(a_t) = \mu_a$  costante, varianza  $\text{Var}(a_t) = \sigma_a^2$  e  $\gamma_k = \text{Cov}(a_t, a_{t+k}) = 0$  per ogni  $k \neq 0$ . Da quanto appena definito, intuitivamente, deriva che un processo rumore bianco è stazionario, con funzione di autocovarianza pari a

$$\gamma_k = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0; \end{cases}$$

funzione di autocorrelazione pari a

$$\rho_k = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0; \end{cases}$$

e funzione di autocorrelazione parziale pari a

$$\phi_{kk} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

### 2.1.3. Processi MA, AR e ARMA

Nell'analisi delle serie temporali, vi sono due modi per rappresentare ed esprimere un processo temporale.

Il primo è quello di considerare un processo  $\{Z_t\}$  come una combinazione lineare di una sequenza di variabili casuali incorrelate tra loro, ovvero,

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad (2.4)$$

dove  $\psi_0 = 1$ ,  $\{a_t\}$  è un processo rumore bianco con media zero e  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ . E' possibile scrivere la (2.4) anche in forma più compatta, cioè,

$$\dot{Z} = \psi(B) a_t,$$

dove  $\dot{Z} = Z_t - \mu$  e  $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$  con  $B^j x_t = x_{t-j}$ .

La rappresentazione del processo definita con la (2.4) è chiamata a media mobile (MA, dall'inglese *moving average*). Wold (1938) provò che un processo stazionario che è puramente non deterministico può sempre essere espresso nella forma (2.4). Dunque, qualsiasi processo che può essere rappresentato in questo modo, è definito non deterministico.

Il secondo metodo è quello di scrivere un processo  $\{Z_t\}$  tramite un modello autoregressivo (AR), nel quale il valore di  $Z$  al tempo  $t$  viene predetto dai suoi valori passati aggiungendo un termine di disturbo  $a_t$

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \dots + a_t \quad (2.5)$$

o, in modo più sintetico,

$$\pi(B) \dot{Z}_t = a_t,$$

dove  $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$  e  $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$ . Una rappresentazione di questo tipo è utile per capire il meccanismo di predizione. Box e Jenkins (1976) definiscono invertibili i processi che possono essere scritti tramite la rappresentazione autoregressiva.

Sebbene la rappresentazione autoregressiva e quella a media mobile siano degli utili strumenti per esprimere i processi stocastici, esse non sono propriamente i mezzi utilizzati per modellare le realizzazioni di tali processi poiché contengono un infinito numero di parametri, impossibili da stimare da un numero finito di dati osservati. Per questo motivo, quando si decide di studiare un determinato fenomeno, vengono costruiti dei modelli con un numero finito di parametri.

Nella rappresentazione autoregressiva di un processo, se e solo se un numero finito di pesi  $\pi$  sono non nulli, cioè,  $\pi_1 = \phi_1, \pi_2 = \phi_2, \dots, \pi_p = \phi_p$  e  $\pi_k = 0$  per  $k > p$ , si ottiene un modello autoregressivo (AR( $p$ )) di ordine  $p$ , avente la seguente scrittura

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \dots - \phi_p \dot{Z}_{t-p} = a_t.$$

## 2. Analisi dei dati reali

Allo stesso modo, nella rappresentazione a media mobile, se e solo se un numero finito di pesi  $\psi$  sono non nulli, cioè,  $\psi_1 = -\theta_1, \psi_2 = -\theta_2, \dots, \psi_p = -\theta_p$  e  $\psi_k = 0$  per  $k > q$ , si ottiene un modello a media mobile  $q$  (MA( $q$ )), esprimibile con

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Una naturale estensione di un semplice modello autoregressivo e di un puro modello a media mobile è la combinazione dei due che dà origine al modello autoregressivo a media mobile (ARMA). Come si è già detto, processi stazionari ed invertibili possono essere rappresentati rispettivamente da modelli a media mobile e da modelli autoregressivi. Questi ultimi però, possono contenere un numero infinito di parametri e, quindi, per investigare su un determinato fenomeno, vi è la necessità di sviluppare una adeguata approssimazione. In generale, un ampio numero di parametri riduce l'efficienza delle stime così, risulta necessario combinare assieme, in un unico modello, la parte a media mobile e la parte autoregressiva esprimibile nel modo seguente

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t, \quad (2.6)$$

dove

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

e

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q.$$

In un processo ARMA, le condizioni di invertibilità si ottengono dalle radici della sua equazione caratteristica, cioè ponendo  $\theta_p(B) = 0$ . L'invertibilità, quindi, è rispettata quando le radici della equazione sono in modulo  $> 1$ , o, in altre parole, sono esterne al cerchio di raggio unitario. Stessa cosa per la stazionarietà: le sue condizioni si ottengono ponendo  $\phi_p(B) = 0$  ed essa è garantita quando le radici della equazione sono in modulo  $> 1$ .

### 2.1.4. CCF e modelli a funzione di trasferimento

Fino ad ora sono stati considerati metodi utilizzabili nell'analisi di serie storiche univariate. In molte situazioni e studi su dati reali, tuttavia, l'interesse è rivolto all'identificazione di una relazione tra due o più serie, perciò risulta necessario introdurre la funzione di cross-correlazione, strumento utile per l'individuazione dell'eventuale relazione presente tra due variabili, ed i modelli dinamici.

La funzione di cross-correlazione rappresenta la misura di quanto forte sia la relazione tra due variabili casuali. Considerati due processi stocastici  $\{X_t\}$  e  $\{Y_t\}$  per  $t = 0, \pm 1, \pm 2, \dots$ , si può affermare che  $X_t$  e  $Y_t$  sono congiuntamente stazionari se sono entrambi due processi univariati stazionari e la funzione di cross-covarianza tra essi,  $Cov(x_t, y_t)$ , è soltanto una funzione della differenza temporale. In questo caso, la funzione di cross-covarianza assumerà la seguente forma

$$\gamma_{xy}(k) = E[(X_t - \mu_x)(Y_{t+k} - \mu_y)] \quad (2.7)$$

per  $k = 0, \pm 1, \pm 2, \dots$ , dove  $\mu_x = E(X_t)$  e  $\mu_y = E(Y_t)$ . Standardizzando la (2.7), si ha la seguente funzione di cross-correlazione (CCF)

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (2.8)$$

per  $k = 0, \pm 1, \pm 2, \dots$ , dove  $\sigma_x$  e  $\sigma_y$  sono le deviazioni standard di  $X_t$  e  $Y_t$ . E' importante notare che la funzione di cross-covarianza  $\gamma_{xy}(k)$  e la funzione di cross-correlazione  $\rho_{xy}(k)$  sono generalizzazioni delle funzioni di autocovarianza e autocorrelazione in quanto  $\gamma_{xx}(k) = \gamma_x(k)$  e  $\rho_{xx}(k) = \rho_x(k)$ . Tuttavia, a differenza della funzione di autocorrelazione  $\rho_x(k)$  che risulta essere simmetrica rispetto all'origine, la funzione di cross-correlazione non lo è, cioè  $\rho_{xy}(k) \neq \rho_{xy}(-k)$ . Invece, poiché  $\gamma_{xy}(k) = E[(X_t - \mu_x)(Y_{t+k} - \mu_y)] = E[(Y_{t+k} - \mu_y)(X_t - \mu_x)] = \gamma_{xy}(-k)$ , si ha che  $\gamma_{xy}(k) = \gamma_{xy}(-k)$ . Di conseguenza, la funzione CCF non solo misura quanto significativa sia l'associazione tra due variabili casuali ma anche in che direzione questa si muove.

Vengono definiti modelli di regressione dinamici quei sistemi nei quali una serie di output  $y_t$  è linearmente collegata ad una o più serie di input  $x_t$ . Solitamente si assume che le osservazioni delle diverse serie siano state effettuate nei medesimi istanti temporali. Inoltre, si presume che la relazione lineare sia asimmetrica, ovvero, che sia la sola serie di output a dipendere dalle serie di input e non che queste siano causate a loro volta dalla serie di output. L'idea che sta alla base di questi modelli è rappresentata in Figura 2.1.

In generale, si considerino due serie  $x_t$  e  $y_t$ , entrambe stazionarie, dove la prima che rappresenta la serie di input e la seconda quella di output. In un sistema lineare, queste sono legate tra loro dalla seguente relazione

$$y_t = v(B)x_t + n_t, \quad (2.9)$$

dove  $v(B) = \sum_{j=-\infty}^{\infty} v_j B^j$  è l'insieme dei pesi della risposta dell'input della serie e  $n_t$  è la



## 2. Analisi dei dati reali

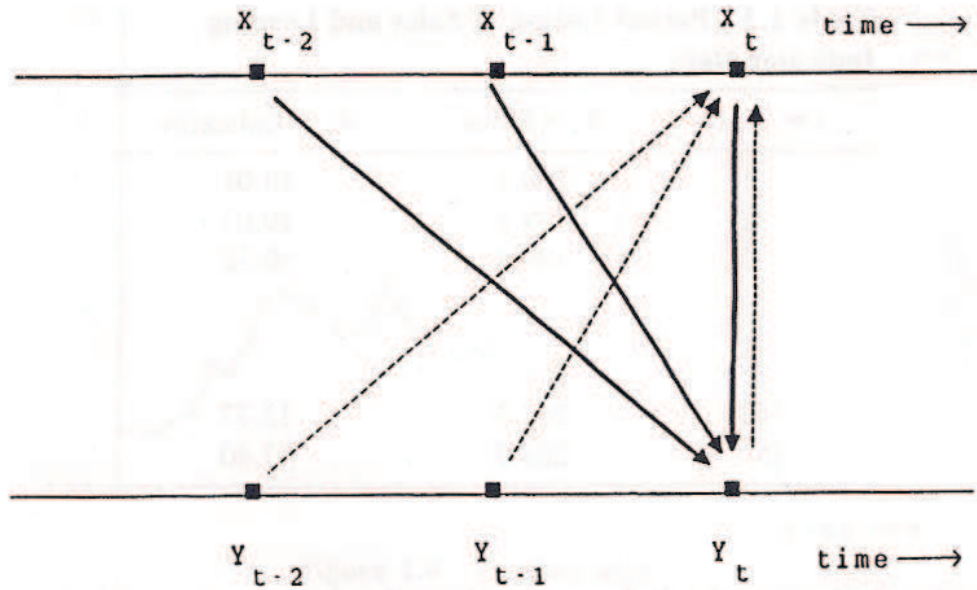


Figura 2.1.: Possibile relazione temporale tra la serie di output  $y_t$  e quella di input  $x_t$ . (Pankratz, 1991)

parte casuale di  $y_t$ , indipendente dall'input della serie. Da sottolineare come l'output del tempo  $t$  dipenda esclusivamente dai valori del tempo  $t$  e da quelli passati. La (2.9) viene definita anche modello a funzione di trasferimento (TF, dall'inglese *transfer function*).

L'obiettivo di una modellazione attraverso la funzione di trasferimento è quello di identificare e stimare la funzione di trasferimento  $v(B)$  e la parte casuale del modello  $n_t$ , basandosi sulle informazioni disponibili circa la funzione di input e quella di output. La difficoltà principale, riscontrata anche nei modelli autoregressivi e a media mobile, è quella che una funzione di trasferimento può contenere infiniti coefficienti mentre le informazioni riguardanti  $x_t$  e  $y_t$  sono finite. Per ovviare a questo problema, si rappresenta la funzione di trasferimento attraverso la forma

$$v(B) = \frac{\omega(B)B^b}{\delta(B)}, \quad (2.10)$$

dove  $\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$ ,  $\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$  e  $b$  è un parametro di ritardo che rappresenta l'attuale spazio temporale che trascorre prima che l'impulso dell'input produca un effetto sulla variabile di output. Per un sistema stabile, ovvero, un modello a funzione di trasferimento nel quale la sequenza dei pesi della risposta dell'input hanno somma finita ( $\sum |v_j| < \infty$ ), si assume che le radici dell'equazione  $\delta(B) = 0$  siano al di fuori del cerchio di raggio unitario.

## 2.2. Risultati

Come descritto nel Paragrafo 1.3, l'obiettivo di questo studio è quello di evidenziare se esiste e di che tipo è la relazione tra le dosi di antibiotico Meropenem somministrate ai pazienti e la resistenza sviluppata dai batteri di *Pseudomonas aeruginosa* verso questo medicinale.

La prima fase lavorativa ha riguardato la creazione di due serie storiche mensili contenenti i dati da gennaio 2007 a ottobre 2011: la prima relativa alla percentuale di degenti che hanno manifestato la presenza del batterio resistente (RES) e la seconda riguardante la dose di antibiotico somministrata (DDD), entrambe distinte per reparto. È stata presa la decisione, in accordo con il dott. Bortolato, di lavorare separando l'Istar 1 dall'Istar 2 poiché questi reparti, pur occupandosi entrambi di rianimazione, non hanno le stesse caratteristiche e, più specificatamente, il secondo ha pazienti mediamente più gravi.

Successivamente, con gli strumenti di cui si è parlato nei paragrafi precedenti, si sono svolte le prime analisi esplorative. La Figura 2.2 mostra quali sono stati, nel periodo di osservazione, gli andamenti delle dosi di antibiotico somministrate e delle resistenze batteriche manifestatesi sui pazienti. Questa non è altro che la visione sovrapposta dei grafici illustrati in Figura 1.1: così facendo, si può osservare come la resistenza batterica si muove rispetto alla DDD. È evidente la differenza tra i due reparti: in Istar 1 gli andamenti di entrambe le serie presentano dei picchi e sembra che i punti più alti della resistenza seguano quelli delle dosi di medicinale. In questo reparto, inoltre, da fine 2008, si riscontra una modificazione di regime, dovuta, da quanto ci ha spiegato il dott. Bortolato, ad un cambiamento della politica di somministrazione che ha portato a dosaggi molto più bassi di antibiotico. In Istar 2 le dosi di farmaco si sono mantenute pressoché costanti in tutti gli anni, eppure, a fine 2008, anche qui si nota una variazione nell'andamento delle dosi di farmaco.

Tuttavia, in entrambe le figure, si riscontra che spesso la linea tratteggiata segue l'andamento della linea continua con qualche mese di ritardo ed, in particolare, in Istar 1 questo avviene soprattutto dopo il cambio di regime, mentre in Istar 2 la situazione è presente in quasi tutto il periodo di osservazione. L'analisi grafica, quindi, non aiuta a capire quale sia effettivamente la relazione presente tra le dosi di antibiotico e le resistenze batteriche: in alcuni casi, infatti, sembra che le resistenze abbiano un andamento piuttosto indipendente

## 2. Analisi dei dati reali

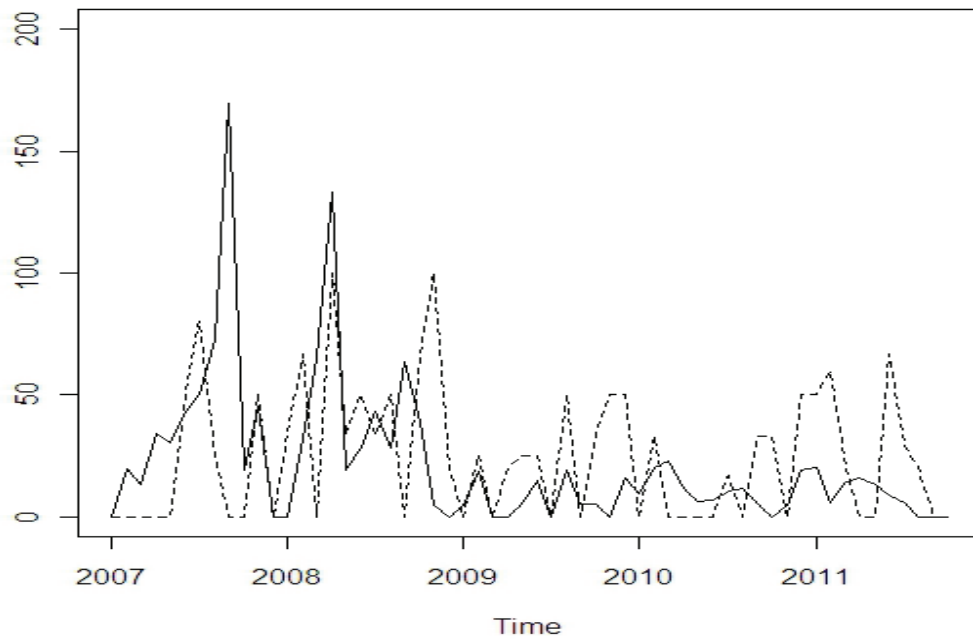
rispetto alle dosi di farmaco, in altri sembra vi sia la possibilità che i batteri diminuiscano la loro resistenza con qualche mese di ritardo rispetto all'abbassamento della DDD.

Ulteriori analisi hanno portato ad osservare i grafici delle autocorrelazioni globali e parziali delle serie ma non è emerso nulla di rilevante in merito alle resistenze batteriche (Figura A.1, Appendice 1). Nella Figura A.2, che si riferisce alla dose di antibiotico, invece, è risultato significativo, in entrambi i reparti, il ritardo 1 ed, intuitivamente, questo significa che la dose del mese  $t+1$  è strettamente connessa con la dose somministrata nel mese  $t$ . E' un risultato del tutto plausibile in quanto le cure dei mesi che si susseguono non possono essere indipendenti tra di loro poiché i degenti che si trovano nei reparti di rianimazione subiscono particolari cicli di cure. Inoltre, i pazienti che entrano in quei reparti nel mese  $t$ , non necessariamente ne escono entro il mese stesso e, di conseguenza, essi continueranno la terapia anche nel mese successivo.

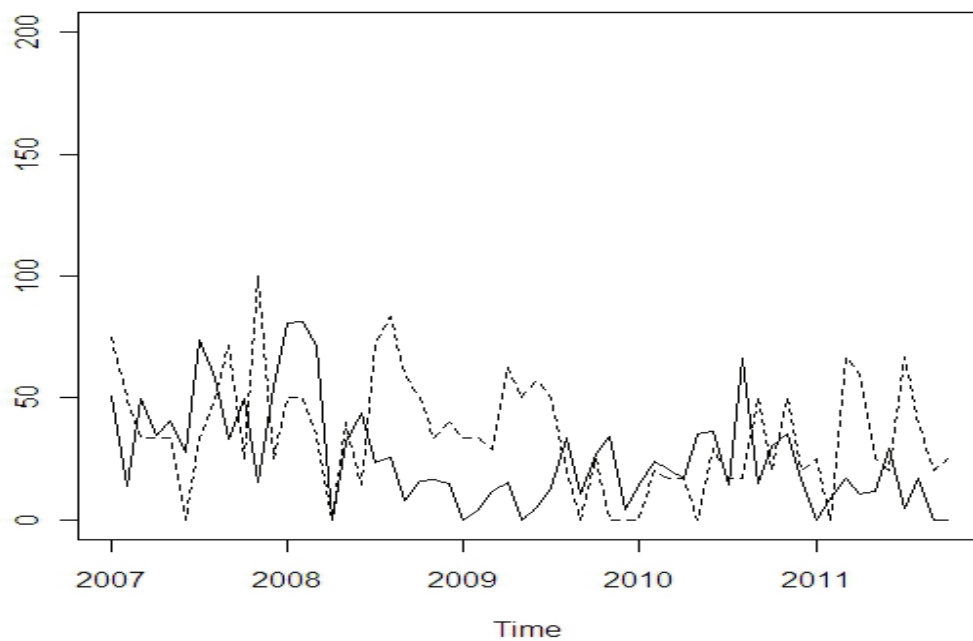
Per quanto riguarda la stazionarietà, si può affermare che tutte le serie di cui ci stiamo occupando sono stazionarie: in nessuna delle quattro, infatti, sembra esserci presenza di una qualche stagionalità o trend particolare.

Per studiare al meglio la relazione tra dose di farmaco somministrato e resistenze batteriche, è utile l'osservazione delle cross-correlazioni in Figura 2.3. Esse sono state ottenute mettendo a confronto i residui della serie delle DDD pre-sbiancata (per i dettagli, si veda il Paragrafo 3.3) e i residui della serie delle resistenze trasformata con lo stesso modello utilizzato per la modellazione della DDD. Intuitivamente, si pensava di ottenere una situazione nella quale fosse solo la quantità di antibiotico ad influenzare l'emergere delle resistenze batteriche, e, per questo, la serie storica del farmaco viene considerata come la serie di input, mentre quella della manifestazione delle resistenze come quella di output. In realtà, le analisi esplorative effettuate smentiscono in parte la nostra supposizione: in Istar 1, infatti, abbiamo una situazione nella quale da una parte, la DDD ai tempi  $t-2$ ,  $t-6$  e  $t-7$  è significativamente correlata con le resistenze e, dall'altra, la resistenza è correlata con le dosi di medicinale. In Istar 2, invece, la situazione sembra essere più chiara: soltanto la DDD al tempo  $t-8$  hanno effetti sulla resistenza al tempo  $t$ . Queste evidenze empiriche, tuttavia, non sono semplici da spiegare poiché, trattandosi di reparti di rianimazione, sono dei risultati anomali. Il tempo massimo per cui un malato rimane in uno di questi due reparti, infatti, è di 3 mesi: il fatto di riscontrare una correlazione significativa tra la DDD di mesi molto lontani ( $t-6$ ,  $t-7$ ,  $t-8$ ) e le resistenze al tempo  $t$ , quindi, fa presumere che i nostri dati facciano emergere dei legami spuri anziché evidenziare le vere relazioni presenti.

Nelle ultime analisi esplorative sono stati creati dei modelli lineari con i quali si è cercato di determinare in che modo la dose farmacologica, somministrata ai pazienti in diversi



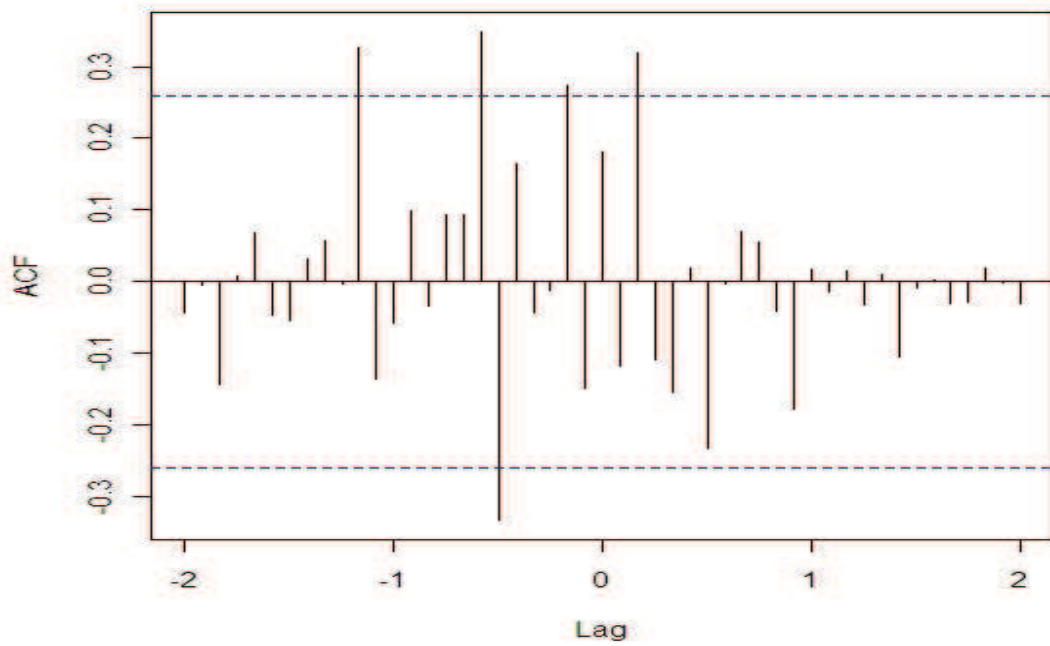
(a) Istar 1



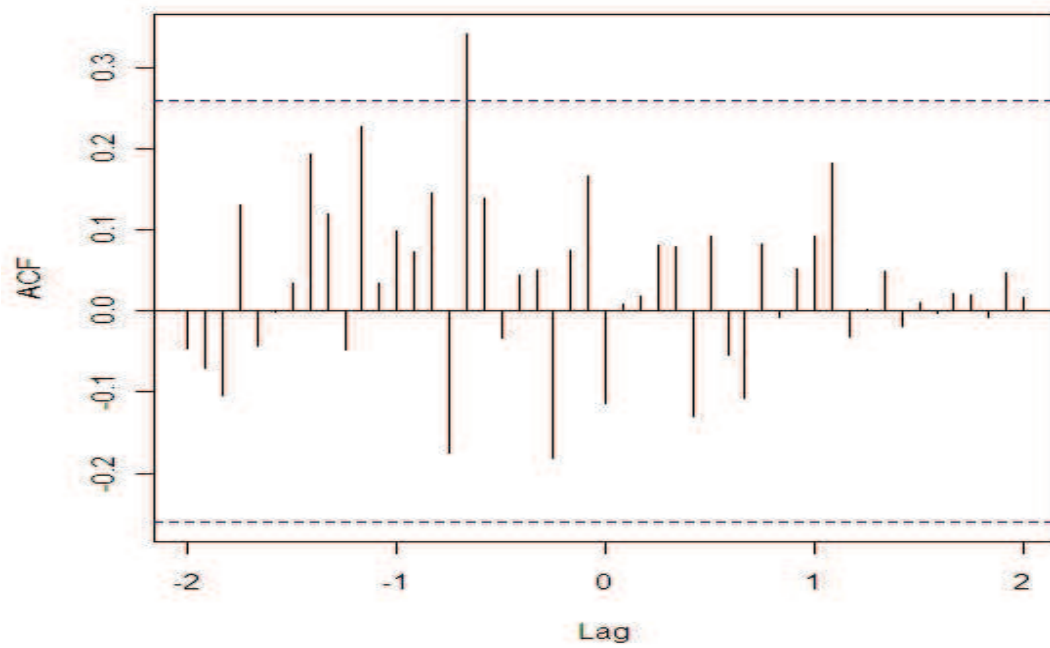
(b) Istar 2

Figura 2.2.: Serie storiche DDD (linea continua) e resistenza batterica (linea tratteggiata), distinte per reparto.

2. Analisi dei dati reali



(a) Istar 1



(b) Istar 2

Figura 2.3.: Cross-correlazioni tra DDD e resistenza batterica, distinte per reparto.

mesi, abbia effetti sulla resistenza sviluppata dai germi (si veda Appendice 1). La variabile dipendente è quindi  $RES_t$ , cioè la resistenza manifestata nei pazienti al tempo  $t$ , mentre le variabili esplicative sono  $DDD_t$ ,  $DDD_{t-1}$ ,  $DDD_{t-2}$  e  $DDD_{t-3}$ , ovvero, la dose antibiotica ai tempi  $t$ ,  $t-1$ ,  $t-2$  e  $t-3$ . Purtroppo anche in questo caso non si è registrato nulla di significativo: solo in Istar 1 viene confermato quanto emerso dalla CCF, ovvero la significatività della dose al tempo  $t-2$  (p-value < 0.05), mentre le altre non lo sono affatto; in Istar 2 non emerge nessuna dipendenza tra farmaco e resistenza batterica.

## 2.3. Considerazioni finali

Dopo aver condotto le analisi esplorative ed aver constatato che non vi è nessuna evidenza empirica per determinare una relazione tra dose di antibiotico e il manifestarsi della resistenza batterica, ci siamo chiesti il perché di tale risultato.

Il fatto che i nostri dati non denuncino nessuna dipendenza tra farmaco e vita del batterio, non significa che, in realtà, questa non esista: vi è, infatti, una copiosa letteratura in merito a questo argomento.

Le condizioni che ci hanno portato a non avere i risultati sperati sono, probabilmente, due: da una parte la lunghezza del periodo di osservazione, composto soltanto da 58 mensilità, dall'altra il basso numero di casi nei due reparti. Mentre sul primo problema non c'è possibilità di intervento, per ovviare al secondo, si era pensato di unire i due reparti in un unico campione. Questa soluzione avrebbe certamente facilitato le analisi e, probabilmente, condotto a risultati più soddisfacenti, ma per i fini pratici dello studio non sarebbe stata utilizzabile: si è già detto, infatti, che i due reparti, pur occupandosi entrambi di rianimazione, non ospitano la stessa tipologia di degenti, i quali necessitano di cure differenti e, per questo motivo, lo studio non può non considerare tale diversità. Un'altra soluzione proposta era quella di estendere lo studio anche ad unità di rianimazione di altri policlinici, in modo da avere numerosità più elevate. Tuttavia, si sarebbe presentata nuovamente la problematica appena descritta: i reparti di rianimazione non hanno tutti la stessa tipologia di paziente e, in più, in questo caso, avremmo avuto anche una serie di variabili legate al contesto e all'ambiente circostante completamente diverse.

Era emersa anche la proposta di ampliare il campione includendo altri tipi di batteri e di antibiotici. Il problema è sempre il medesimo: comprendere germi diversi, e, quindi, pazienti con patologie molto differenti, significherebbe avere una gamma troppo ampia di cure e, di conseguenza, caratteristiche a priori troppo diverse. La scelta di restringere il campo soltanto al *Pseudomonas aeruginosa* e al Meropenem era stata fatta, appunto, per

## *2. Analisi dei dati reali*

ovviare al problema delle diversità: in questo modo si sarebbe potuto lavorare su campioni omogenei sia per patologia che per tipo di cura.

Per porre rimedio a queste difficoltà, si è deciso, pertanto, di eseguire uno studio tramite simulazioni per valutare quanto i fattori “lunghezza delle serie” e “numerosità di casi” incidano sulla scelta e sulla validazione del modello finale. Si potranno, perciò, costruire degli scenari diversi, modificando le quantità menzionate, per capire le motivazioni che, nella realtà, ci hanno portato a non avere nessuna evidenza empirica. Avremmo, quindi, la possibilità di dare una risposta ai Medici del Policlinico, suggerendo loro quali sono i cambiamenti nella raccolta dei dati da apportare operativamente sul campo.

## 3. Lo studio di simulazione

Come già esposto nel capitolo precedente, le analisi esplorative effettuate sulle serie temporali a nostra disposizione non hanno portato ai risultati sperati, pertanto non si è riusciti ad evidenziare nessuna relazione tra lo sviluppo delle resistenze batteriche e le dosi di antibiotico somministrate ai pazienti.

In questo capitolo, inizialmente si andranno ad illustrare quali sono gli scenari che si vogliono ricreare tramite lo studio di simulazione; successivamente, si descriveranno le tecniche utilizzate per riprodurre le serie relative alle dosi di antibiotico e alle resistenze batteriche; infine, si mostrerà il procedimento per l'applicazione dei modelli a funzione di trasferimento.

### 3.1. Gli scenari

L'obiettivo delle nostre simulazioni è quello di poter dare, tramite determinati scenari, delle risposte concrete ai Medici del Policlinico patavino in merito alla relazione tra l'utilizzo degli antibiotici e la vita di PSAE, così da offrire loro dei suggerimenti su come ottimizzare la raccolta delle informazioni e le relative elaborazioni.

L'idea è quella di costruire tre scenari:

1. Un primo (Scenario 1) nel quale si simuleranno le serie mantenendo come lunghezza quella reale, ossia, 58 mensilità. Si lavorerà, poi, creando tre sotto-scenari, i quali avranno rispettivamente numerosità dei casi pari a 10, 100 e 500 unità.
2. Un secondo (Scenario 2) nel quale si lavorerà con una lunghezza pari a 72 mensilità, ovvero, simulando i dati fino a dicembre 2012. Anche in questo caso si avranno tre sotto-scenari con numerosità diverse.
3. Infine, un terzo (Scenario 3) che sarà quello con lunghezza maggiore, cioè 96 mensilità, che ci permetterà di ottenere i dati fino a dicembre 2014. Ovviamente, come nei due precedenti casi, si lavorerà con tre sotto-scenari di numerosità dei casi differenti.



### 3. Lo studio di simulazione

Costruendo tre diverse situazioni ci sarà la possibilità di analizzare quanto le quantità “lunghezza della serie” e “numerosità campionaria” incidano sulle risposte ottenute dalle analisi. In particolare, nel primo scenario l’obiettivo è quello di capire se il problema della nostra mancata evidenza empirica dipenda esclusivamente dai pochi dati a disposizione; il secondo ci permetterà di capire se l’aggiunta di 14 mensilità potrebbe risultare significativa per ottenere i risultati sperati in breve tempo ed, infine, tramite il terzo ci avvicineremo alle lunghezze presenti in letteratura e quindi si verificherà in che misura queste incidano sulle conclusioni del lavoro.

Una volta preparati i tre diversi contesti, si lavorerà su due fronti: da un lato si confronteranno i risultati ottenuti dalle diverse numerosità mantenendo fissa le lunghezze delle serie, mentre, dall’altro, si paragoneranno gli esiti avuti dalle stesse numerosità date, però, lunghezze temporali diverse.

Ricordiamo al lettore che, per facilitare la lettura e la comprensione, i sotto-scenari con numerosità 10, 100 e 500 sono stati denominati rispettivamente con le lettere A, B e C se appartenenti allo scenario 1; D, E ed F se appartenenti allo scenario 2; G, H ed I se appartenenti allo scenario 3.

## 3.2. Generazione dei dati simulati

### 3.2.1. Cenni teorici

Per dare origine ai dati simulati abbiamo fatto affidamento sulle tecniche di bootstrap, introdotte da Bradley Efron nel 1979 ed in Figura 3.1 viene illustrato lo schema concettuale della stima bootstrap dello scarto quadratico medio, sviluppato da egli stesso. L’obiettivo di questa metodologia è quello di ottenere dal campione che si ha a disposizione, una stima di un qualche parametro della popolazione intera. In altre parole, si assuma di avere un campione  $x = (x_1, \dots, x_n)$  da una distribuzione ignota  $\mathcal{F}(x, \theta)$  e di avere una stima  $\hat{\theta}(x)$ , di un qualche parametro  $\theta$ , che può essere la media, la varianza, lo scarto quadratico medio o un qualsiasi altro indicatore. Se si potessero ottenere altri campioni di numerosità  $n$  dalla popolazione, si potrebbero calcolare diverse stime di  $\theta$  dalle quali poi trarre la stima di nostro interesse. In realtà, non disponendo dell’intera popolazione ma soltanto dell’unico campione osservato, l’idea del bootstrap è quella di sostituire l’ignota distribuzione  $\mathcal{F}$  con una sua stima ottenuta empiricamente dall’osservazione.

In linea generale, per dati indipendenti ed identicamente distribuiti, il ricampionamento bootstrap può essere descritto come segue: sia  $(x_1, \dots, x_n)$  un campione iniziale estratto da

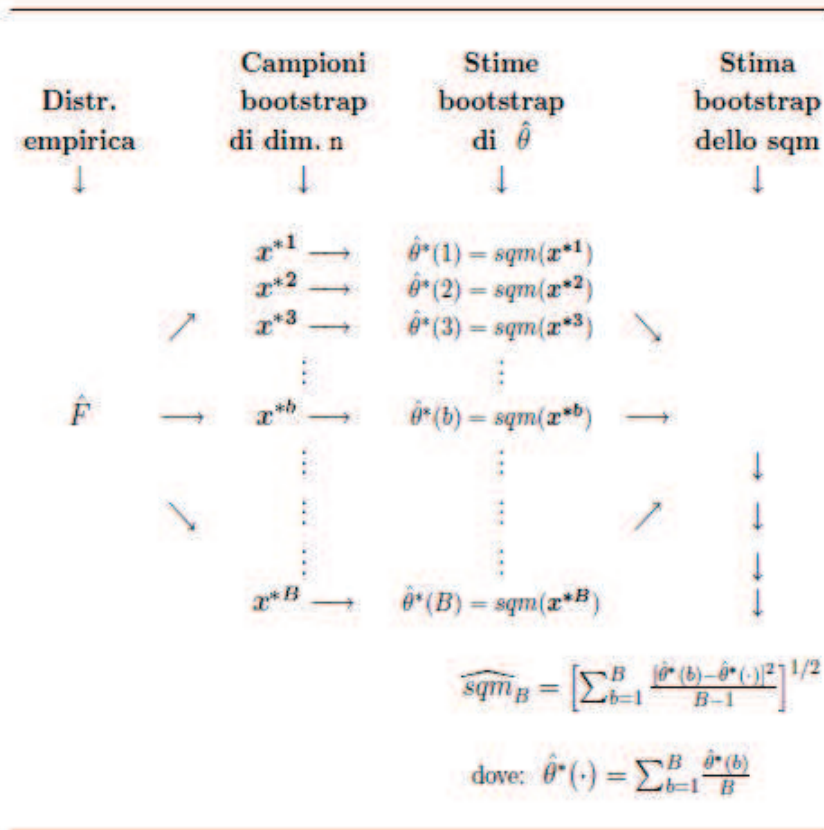


Figura 3.1.: Schema concettuale della stima bootstrap dello scarto quadratico medio. (Efron, 2003)

una variabile casuale  $X$  con distribuzione  $\mathcal{F}(x, \theta)$  e si supponga di essere interessati ad un parametro  $\theta$  tale che  $\hat{\theta} = T(x_1, \dots, x_n)$ , dove  $T(\cdot)$  è una statistica; dal campione iniziale si estraggono con reinserimento altri  $B$  campioni  $(x_{1,b}^*, \dots, x_{n,b}^*)$ , con  $b = 1, \dots, B$ . Su ciascuno di questi campioni bootstrap si stima, con lo stesso stimatore  $\hat{\theta}_b^*$ , il parametro di interesse. L'insieme dei  $\hat{\theta}_b^*$  è chiamato distribuzione bootstrap di  $\hat{\theta}$ .

Risulta evidente, dunque, che la metodologia si divide in due fasi principali: la prima è quella in cui un campione  $S$ , di ampiezza  $n$ , viene estratto senza ripetizione da una popolazione  $X$ , che ha come distribuzione  $\mathcal{F}(x, \theta)$ , e la seconda è quella in cui vengono estratti da  $S$ , con reinserimento,  $B$  sotto-campioni anch'essi di ampiezza  $n$ .

La metodologia bootstrap appena descritta, viene definita "classica" e presenta una significativa limitazione in quanto si può applicare solo nel caso in cui il campione sia i.i.d. e, di conseguenza, nell'ambito dell'analisi delle serie temporali sarebbe improponibile. Nel contesto di nostro interesse risulta, quindi, necessario affidarsi ad una metodologia che, ri-

### 3. Lo studio di simulazione

scontrata la struttura di dipendenza temporale tra le osservazioni, la mantenga nelle serie ricampionate.

Lo strumento da noi scelto è il metodo *model-based resampling*, detto anche bootstrap parametrico. Questo è applicabile se viene assunto che i dati da ricampionare siano stati generati da un processo lineare, stazionario ed invertibile, come ad esempio un  $ARMA(p, q)$ , del quale è possibile stimare i residui. Sotto l'assunzione di corretta specificazione del modello, i residui sono indipendenti ed identicamente distribuiti, e ciò rende possibile applicare a questi il bootstrap classico. Le pseudo-serie bootstrap sono quindi costruite utilizzando i residui ricampionati e i parametri stimati.

Le fasi principali del bootstrap parametrico sono le seguenti:

1. adattare un modello appropriato ai dati, per esempio un  $AR(p)$ , scegliendo il miglior modello in base ad un certo criterio (AIC, BIC, ecc.);
2. stimare i residui in base al modello scelto ed, eventualmente, ricentrarli;
3. applicare ai residui il bootstrap classico;
4. usare i parametri stimati e i residui bootstrap ricampionati per costruire le serie bootstrap che possiedono proprietà simili alla serie originale.

Si consideri ora il caso generale di un processo  $AR(p)$  che ha la forma

$$\phi_p(B)(Z_t - \mu) = a_t,$$

dove  $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,  $\mu = E(Z_t)$ ,  $a_t$  è un processo rumore bianco con  $E(a_t) = 0$  e  $Var(a_t) = \sigma_a^2$ . Assumendo che la varianza sia pari ad 1, è possibile stimare i residui nel seguente modo

$$\hat{a}_t = Z_t - \hat{\phi}_1 Z_{t-1} - \dots - \hat{\phi}_p Z_{t-p}$$

con  $t = p + 1, \dots, n$ , dove  $\hat{\phi}_i$  sono stimatori consistenti, ottenuti con il metodo dei minimi quadrati. Il bootstrap parametrico deve quindi procedere con un campionamento con reinserimento da  $\tilde{a}_{p+1}, \dots, \tilde{a}_n$ , i quali hanno tutti la stessa probabilità  $\frac{1}{n-p}$  di essere estratti, per generare le innovazioni  $a_t^*$ . I residui centrati sono  $\tilde{a}_{p+1}, \dots, \tilde{a}_n$  e si ottengono da

$$\tilde{a}_t = \hat{a}_t - \bar{a}_t,$$

con  $t = p + 1, \dots, n$ , dove  $\bar{a}_t = \sum_{t=p+1}^n \frac{\hat{a}_t}{n-p}$ .

A questo punto, si può identificare una pseudo-serie

$$Z_t^* = \hat{\phi}_1 Z_{t-1}^* + \dots + \hat{\phi}_p Z_{t-p}^* + a_t^*$$

al fine di ottenere la nostra serie bootstrap  $\mathbf{Z}_{t,b}^* = (Z_{1,b}^*, \dots, Z_{n,b}^*)$ .

Infine, si calcoli la statistica di interesse  $\hat{\theta}$  su questa serie

$$\hat{\theta}_b^* = T_n(Z_{1,b}^*, \dots, Z_{n,b}^*)$$

e si ripeta la procedura illustrata (dalla fase di ricampionamento dei residui in poi) B volte, con B numero intero positivo molto grande, calcolando per ogni nuovo campione bootstrap  $\hat{\theta}_b^*$ , in modo da poter ricavare la distribuzione empirica della statistica di interesse.

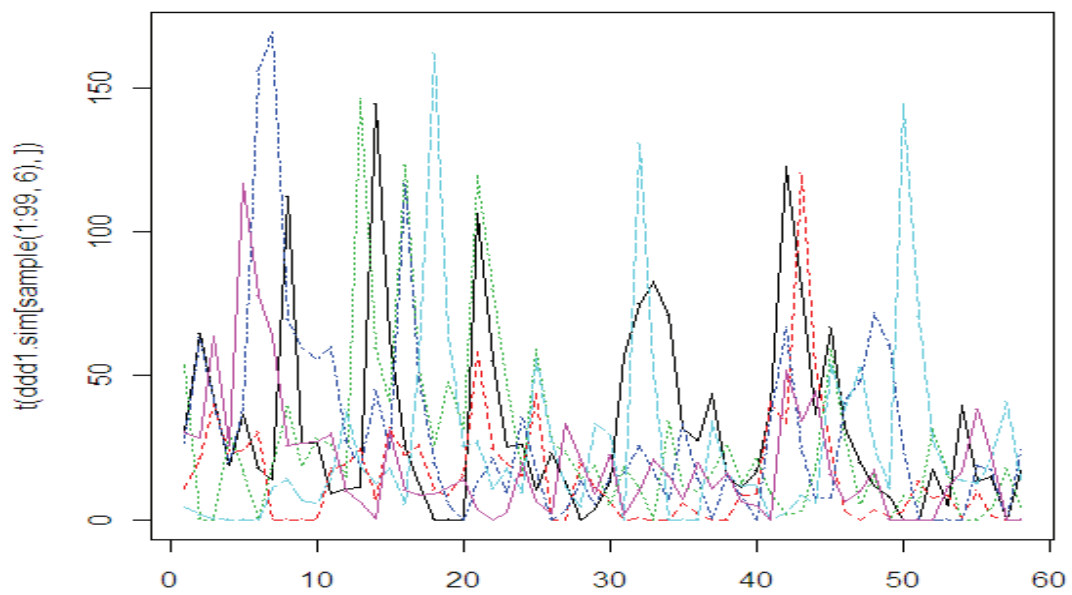
### 3.2.2. La simulazione della DDD

Le analisi esplorative illustrate nel Paragrafo 2.2 avevano evidenziato, per la serie delle dosi di antibiotico, un modello autoregressivo di ordine 1 come possibile processo generatore dei dati. Partendo da questo presupposto, le nostre simulazioni hanno avuto come obiettivo quello di replicare tale processo.

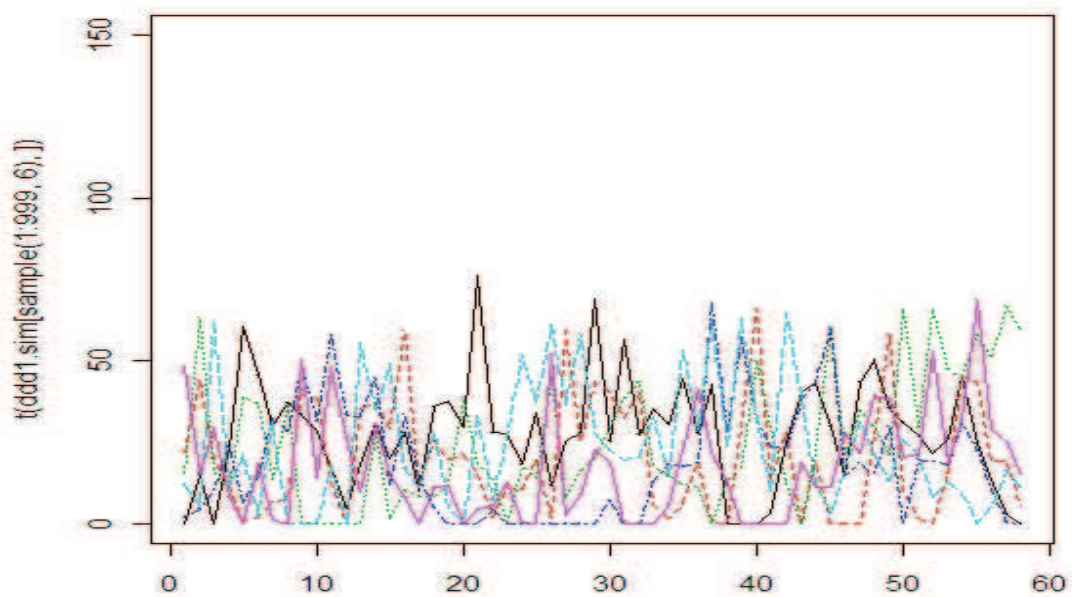
Utilizzando il metodo *model-based resampling*, e, quindi, stimando prima i coefficienti e poi i residui del modello AR(1) emerso dalla serie originale, sono state effettuate 999 repliche. Le Figure 3.2, 3.3 e 3.4 mostrano alcune delle serie simulate nei tre diversi scenari: è evidente che le serie ricreate hanno, nella loro interezza, numerosi picchi e valori nulli, come quelli che si riscontrano nella serie originale.

Tuttavia, un lettore attento, osservando le serie ricreate, noterà la mancanza di valori negativi che, nel processo di simulazione da noi adottato, possono sicuramente verificarsi. In realtà, nelle serie che abbiamo generato vi era la presenza di alcuni valori negativi ma la scelta che abbiamo compiuto è stata quella di sostituire tali elementi con valori nulli poiché la variabile che stiamo considerando non può certo avere quantità inferiori a 0. Tale operazione è stata effettuata in seguito alla generazione delle serie e non durante il processo stesso di generazione perché il nostro obiettivo era quello di simulare un processo autoregressivo di ordine 1: se avessimo apportato delle modifiche al processo di simulazione, modificando la logica sottostante al bootstrap parametrico, avremmo rischiato di non rispettare quello che è il processo generatore dei nostri dati. In ogni caso, le percentuali di valori negativi riscontrate nei nove scenari si aggiravano tra il 12 e il 14%.

### 3. Lo studio di simulazione

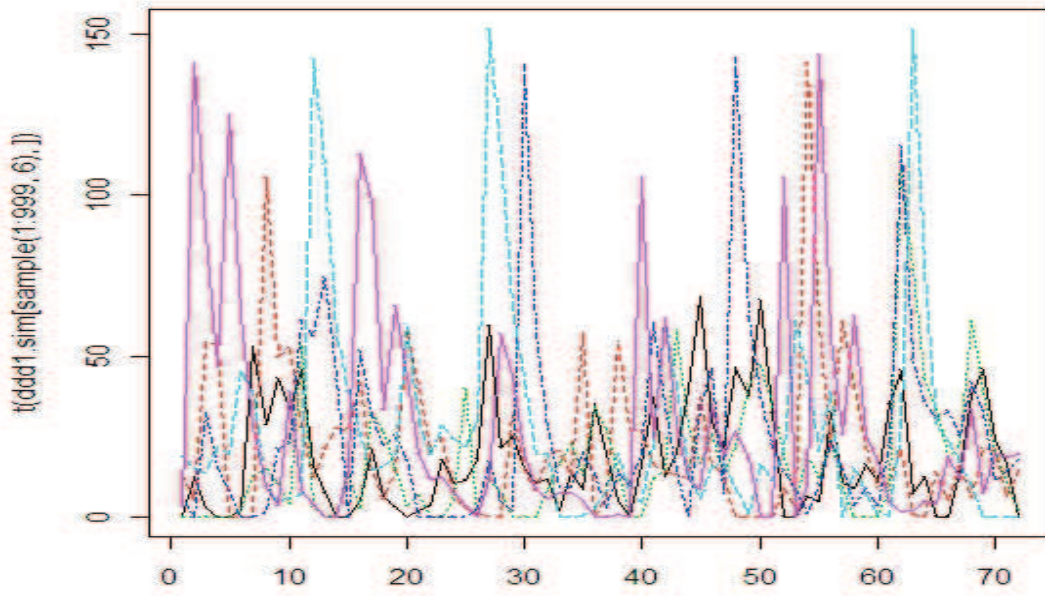


(a) Istar 1

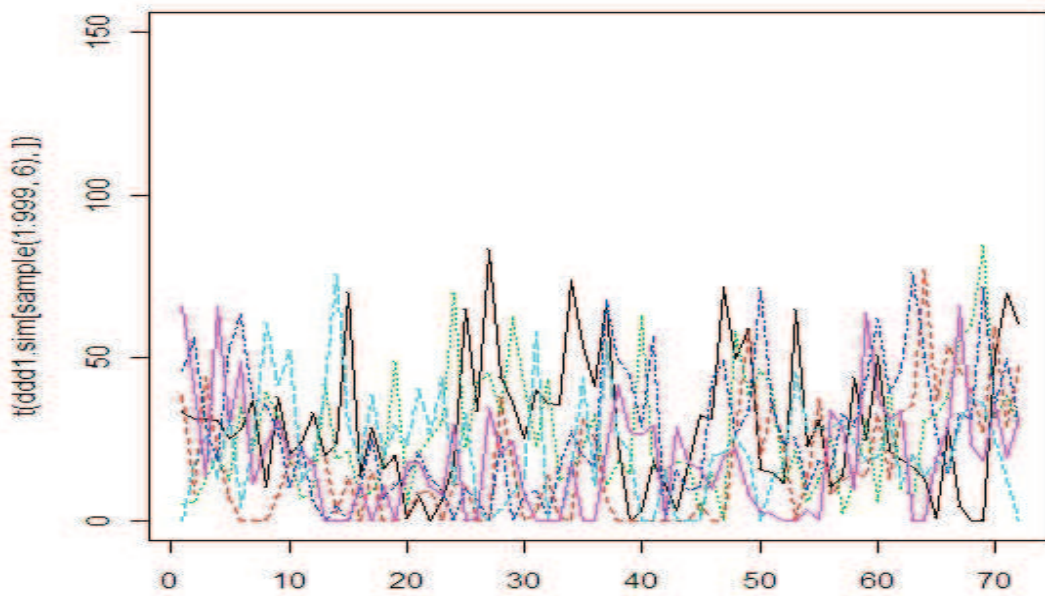


(b) Istar 2

Figura 3.2.: Serie DDD simulate con lunghezza pari a 58 mensilità.



(a) Istar 1

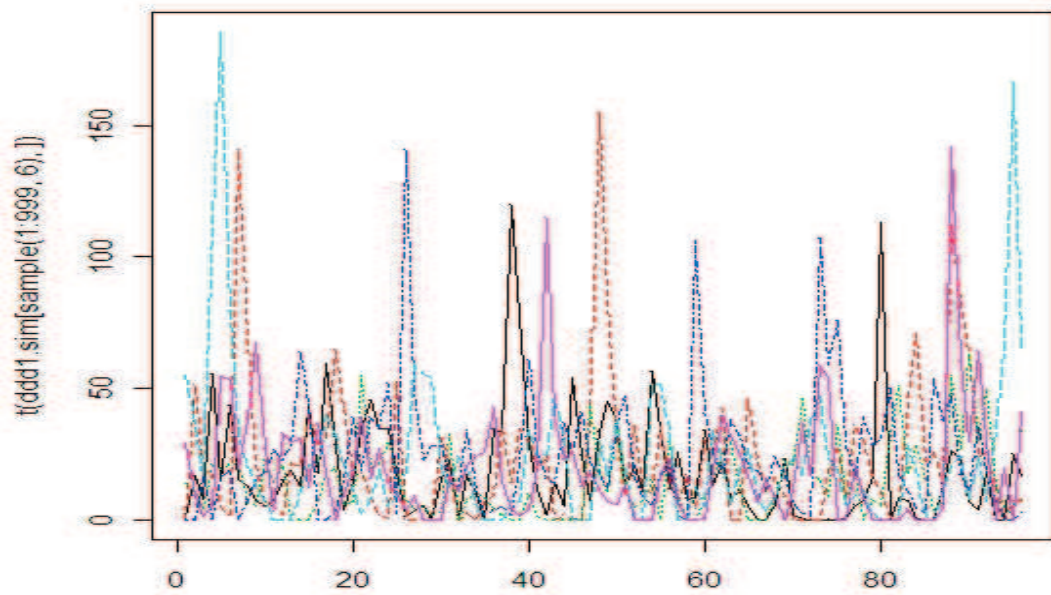


(b) Istar 2

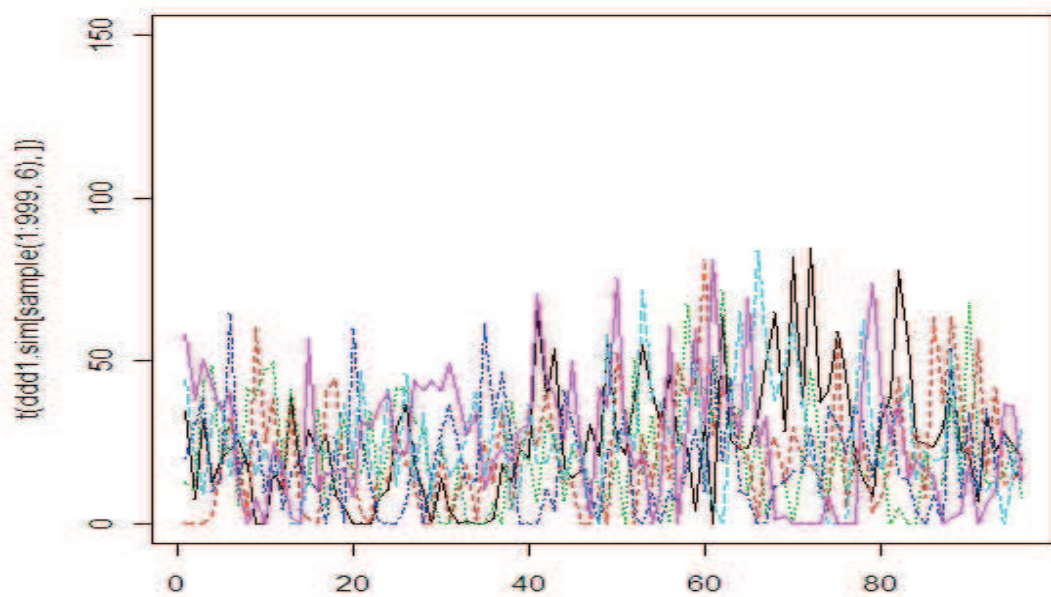
Figura 3.3.: Serie DDD simulate con lunghezza pari a 72 mensilità.



### 3. Lo studio di simulazione



(a) Istar 1



(b) Istar 2

Figura 3.4.: Serie DDD simulate con lunghezza pari a 96 mensilità.

Necessariamente va ricordato che nelle analisi preliminari dei dati ospedalieri si era parlato di un cambio di regime a partire dall'anno 2008, dovuto ad una modificazione della politica di somministrazione. Nelle simulazioni effettuate, però, non si è tenuto conto di questo cambiamento poiché, avendo l'idea di ricreare tre tipi di scenari con lunghezze temporali diverse, sarebbe stato difficile per noi capire quando collocare temporalmente tale variazione. Nonostante ciò, non riteniamo che questa scelta possa avere dei riscontri fuorvianti su quelli che possono essere i risultati dell'analisi poiché l'obiettivo è quello di evidenziare quale relazione c'è tra le dosi di antibiotico e la resistenza batterica, indipendentemente da quelle che sono le politiche di somministrazione.

### 3.2.3. La simulazione della resistenza batterica

La serie originale della resistenza batterica consiste nella raccolta mensile delle percentuali di pazienti che hanno sviluppato la presenza di PSAE resistente al Meropenem. Ciò che noi abbiamo dovuto ricreare è, appunto, il numero dei casi che mensilmente hanno manifestato la resistenza batterica: si è deciso, per questo motivo, di considerare questi come realizzazioni di una variabile casuale binomiale

$$Bi(n_t, \pi_t) = \binom{n_t}{k_t} \pi_t^{k_t} (1 - \pi_t)^{n_t - k_t},$$

dove  $n_t$ ,  $\pi_t$  e  $k_t$  rappresentano rispettivamente la numerosità di pazienti ricoverati con PSAE, la probabilità di manifestare resistenza all'antibiotico MEM e il numero di casi che hanno manifestato la resistenza, in un determinato mese  $t$ .

Per quanto riguarda la numerosità, nelle simulazioni si è deciso di mantenerla costante nel tempo: si lavorerà, quindi, con serie dove la numerosità è fissata ogni mese a 10, 100 o 500 unità. Essendo uno degli obiettivi scoprire quanto la numerosità incida nell'emergere dei risultati, lavorare con numerosità così diverse ci permetterà di perseguire al meglio il nostro scopo.

Fissato  $n_t$ , non resta che definire come ottenere le probabilità di sviluppare la resistenza all'antibiotico nel mese  $t$ . La scelta è ricaduta sul modello

$$\logit(\pi_t) = \alpha + \beta_1 DDD_t + \beta_2 DDD_{t-1} + \beta_3 DDD_{t-2} + \beta_4 DDD_{t-3} + \beta_5 DDD_{t-4} + \beta_6 DDD_{t-5}, \quad (3.1)$$



### 3. Lo studio di simulazione

dove  $\pi_t$  può essere esplicitata nel modo seguente

$$\pi_t = \frac{e^{\alpha + \sum_{j=0}^5 \beta_j DDD_{t-j}}}{1 + e^{\alpha + \sum_{j=0}^5 \beta_j DDD_{t-j}}}. \quad (3.2)$$

Nel nostro modello le variabili esplicative sono le DDD e, più specificatamente, le DDD al tempo  $t$  e quelle ritardate fino al tempo  $t-5$ : introdurre, infatti, le dosi di antibiotico come covariate, permette di implementare la struttura di dipendenza temporale tra resistenza batterica e medicinale, ipotizzata nei dati reali.

La scelta di inserirvi solo cinque ritardi è dovuta a due fattori: uno statistico ed uno medico. Dal punto di vista statistico, si considera il fatto che in uno dei tre scenari si lavora con serie piuttosto brevi ed, in uno dei suoi sotto-scenari si hanno anche basse numerosità. In merito al fattore medico, che risulta essere ai fini pratici, il fattore più importante, ci si ferma a cinque ritardi poiché, dal punto di vista clinico andare oltre non avrebbe senso: il tempo massimo per cui un paziente rimane in rianimazione, infatti, è 2.5 mesi; dopodiché o viene introdotto in altri reparti o viene a mancare. Inserire molti ritardi, o, in altre parole, decidere di stimare la resistenza del mese  $t$  inserendo nel modello le DDD somministrate molti mesi prima (ad es. nei mesi  $t-6$ ,  $t-7$ ,  $t-8$ , ecc.), quindi, non avrebbe alcun senso.

Per la definizione della probabilità di nostro interesse, è stato necessario imputare i valori dei coefficienti  $\beta$ . In accordo con il dott. Bortolato, questi sono stati determinati sulla base della situazione endemica ospedaliera ( $\pi_0$ ). In tutto il Policlinico vi è un'endemia di PSAE pari a circa il 20% e questa è riprodotta dal valore assegnato all'intercetta

$$\alpha = \log\left(\frac{\pi_0}{1 - \pi_0}\right) = -1.4$$

che, appunto, rappresenta l'endemia "di base", cioè al netto dell'utilizzo di antibiotico.

L'interpretazione che viene attribuita agli altri coefficienti è legata al concetto di *odds ratio* (OR):

$$OR_t^\Delta = \frac{e^{\alpha + \sum_{j \neq i} \beta_j DDD_{t-j} + \beta_i (DDD_{t-i} + \Delta)}}{e^{\alpha + \sum_{j \neq i} \beta_j DDD_{t-j} + \beta_i (DDD_{t-i})}} = \frac{\pi_t^\Delta}{1 - \pi_t^\Delta} \times \frac{1 - \pi_t}{\pi_t} = e^{\beta_i \Delta},$$

dove  $\pi_t$  rappresenta la probabilità di sviluppare resistenza batterica in corrispondenza ad una determinata combinazione di DDD e  $\pi_t^\Delta$  rappresenta la probabilità di sviluppare re-

Coefficiente	Valore	Variazione $\pi_0$ con $\Delta=1$	Variazione $\pi_0$ con $\Delta=10$
$\alpha$	-1.4	-	-
$\beta_1$	0.006	0.45 %	4.5 %
$\beta_2$	0.005	0.37 %	3.7 %
$\beta_3$	0.004	0.31 %	3.1 %
$\beta_4$	0.002	0.16 %	1.6 %
$\beta_5$	0.0008	0.06 %	0.64 %
$\beta_6$	0.0006	0.04 %	0.48 %

Tabella 3.1.: Coefficienti del modello (3.1), applicati per la determinazione della probabilità di manifestare PSAE resistente al MEM.

sistenza in corrispondenza alla stessa combinazione di antibiotico aumentata di  $\Delta$  unità. Ogni singolo coefficiente, quindi, rappresenta la variazione del  $\log(OR_t^\Delta)$  nel caso vi fosse un incremento di 1 unità di  $DDD_{t-i}$ , al netto di tutte le altre covariate. Tali variazioni sono riportate nella seconda colonna della Tabella 3.1.

Nella terza e quarta colonna della stessa tabella, vi sono, invece, le variazioni dell'endemia di base nel caso una dose di antibiotico venga aumentata di 1 e 10 unità, al netto di tutte le altre covariate. Si è deciso di lavorare per via numerica, e quindi illustrando la variazione della resistenza batterica rispetto all'endemia di base per facilitare l'interpretazione e la collaborazione con i Medici. Sappiamo che, così facendo, non si tiene in considerazione il legame temporale presente tra le diverse DDD ma, lo scopo, in questa fase, era quello di determinare dei coefficienti ragionevoli per proseguire con le simulazioni. Inoltre, con questa combinazione di coefficienti, abbiamo garantito un valore atteso di resistenze presoché simile a quello osservato nella serie reale ed abbiamo anche riscontrato che queste nostre ipotesi trovano supporto nel lavoro di Mahamat A. (2005).

Infine, facciamo notare che i coefficienti sono stati scelti decrescenti man mano che si va a ritroso nel tempo poiché, alla base, vi è la consapevolezza medica che, per la determinazione della resistenza al tempo  $t$ , sono le DDD dei mesi vicini a  $t$  quelle più significative, in quanto i batteri sviluppano una resistenza cromosomica in tempi molto brevi. In pratica, si lavora ragionando sul peso che ha la  $DDD_{t-i}$  sulla probabilità di sviluppo della resistenza batterica, nel caso la dose venisse aumentata di  $\Delta$  unità.

### 3.3. Costruzione di un modello TF

Come già presentato nel Paragrafo 2.1.4, il modello a funzione di trasferimento segue la forma

$$y_t = v(B)x_t + n_t$$

dove la funzione di trasferimento può essere scritta come segue

$$v(B) = \frac{\omega(B)B^b}{\delta(B)} \quad (3.3)$$

con  $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r$  e  $\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$ .

Per l'identificazione di un modello a funzione di trasferimento, Wei (2006) suggerisce di procedere come segue:

1. stimare un modello ARMA sulla serie di input pre-sbiancata

$$\phi_x(B)x_t = \theta_x(B)\alpha_t,$$

cioè,

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)} x_t,$$

dove  $\alpha_t$  è un rumore bianco con media zero e varianza  $\sigma_\alpha^2$ .

2. ottenere la serie di output filtrata, ovvero, trasformare la serie di output  $y_t$  usando il modello della serie pre-sbiancata al passo 1 per generare la serie

$$\beta_t = \frac{\phi_y(B)}{\theta_y(B)} y_t,$$

dove  $\beta_t$  è un rumore bianco con media zero e varianza  $\sigma_\beta^2$ .

3. calcolare la funzione di cross-correlazione,  $\hat{\rho}_{\alpha\beta}(k)$  tra  $\alpha_t$  e  $\beta_t$  per stimare  $v_k$

$$\hat{v}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha} \hat{\rho}_{\alpha\beta}(k).$$

La significativà della CCF ed, equivalentemente, la significatività di  $\hat{v}_k$  può essere valutata comparandola al suo errore standard  $\frac{1}{\sqrt{(n-k)}}$ .

4. identificare  $b$  e selezionare adeguatamente anche  $r$  ed  $s$  in modo da stimare  $\hat{\omega}_j$  ed  $\hat{\delta}_j$ .  
Una stima preliminare della funzione di trasferimento quindi risulta essere

$$\hat{v}(B) = \frac{\hat{\omega}(B)}{\hat{\delta}(B)} B^b. \quad (3.4)$$

5. una volta stimata la funzione di trasferimento preliminare, è possibile calcolare la stima della serie del rumore  $n_t$

$$\hat{n}_t = y_t - \hat{v}(B)x_t = y_t - \frac{\hat{\omega}(B)}{\hat{\delta}(B)} B^b x_t.$$

Il modello appropriato per il rumore può essere, per esempio, identificato esaminando le funzioni di autocorrelazione globale e parziale del processo

$$\phi(B)n_t = \theta(B)\alpha_t. \quad (3.5)$$

6. infine, combinando assieme la (3.4) e la (3.5), si ottiene il seguente modello a funzione di trasferimento:

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} \alpha_t. \quad (3.6)$$



## 4. Analisi dei dati simulati

Nelle analisi dei dati simulati si sono considerati tre aspetti:

1. in un primo momento ci si è concentrati sulla verifica della normalità della distribuzione degli stimatori di massima verosimiglianza dei parametri del modello (3.1). Si sono, quindi, stimati dei modelli di regressione logistica e dall'output di questi si sono estrapolate le stime dei coefficienti. Infine, si è verificato, tramite analisi grafica con diagrammi a scatola, che le loro distribuzioni empiriche seguissero, approssimativamente, una distribuzione normale centrata nel valore che si era attribuito loro;
2. nella seconda fase, l'obiettivo è stato quello di ricavare le distribuzioni dei valori delle cross-correlazioni ad ogni determinato istante temporale quali mettono in luce le dosi di antibiotico più significative per la determinazione della resistenza.

Per costruire le CCF si è proceduto come segue:

- a) per le serie di input, ovvero quelle delle dosi di antibiotico, attraverso la funzione `arimaId`<sup>1</sup>, abbiamo selezionato il modello che minimizzava il criterio di Akaike (AIC) e questo è risultato essere, nella maggior parte dei casi, un modello autoregressivo di ordine 1; si è, pertanto, deciso, per la modellazione degli input, di usare sempre degli AR(1) e, una volta stimati questi, ne sono stati estrapolati i coefficienti (da utilizzare al punto (b)) ed i residui (da utilizzare al punto (c));
- b) sono state trasformate le serie di output, cioè le serie delle resistenze batteriche simulate, sulla base dei coefficienti stimati attraverso il modello AR(1) al punto (a) e ne sono stati ricavati i residui;
- c) sono state ricavate le funzioni di cross-correlazioni dai residui ottenuti al punto (a) e al punto (b);

---

<sup>1</sup>La funzione `arimaId` è contenuta nel pacchetto "ast.R" che il prof. Guido Masarotto (Facoltà di Scienze Statistiche - UniPD) ha predisposto per l'insegnamento di "Analisi delle Serie Temporal". Tale funzione determina il miglior modello ARMA che si adatta ai dati che stiamo considerando.

#### 4. Analisi dei dati simulati

3. il terzo aspetto ha riguardato la costruzione dei modelli a funzione di trasferimento. Questi, tuttavia, non sono stati costruiti nel modo classico presentato nel Paragrafo 3.3, poiché, nel modello, non sono stati inseriti solo i ritardi emersi significativi dalle CCF al passo 2, bensì anche i ritardi fino al tempo  $t-10$ . L'idea di inserire così tanti ritardi temporali, nasce dal fatto che, pur essendo consapevoli che i ritardi significativi dovrebbero essere compresi tra il tempo  $t$  e il tempo  $t-5$ , si vuole verificare che non vi siano casi in cui emergono significatività anche in tempi più lontani. Il modello a funzione di trasferimento risulta, quindi, essere il seguente

$$\begin{aligned} resistenza_t = & \omega_1 resistenza_{t-1} + \delta_0 DDD_t + \delta_1 DDD_{t-1} + \delta_2 DDD_{t-2} \\ & + \delta_3 DDD_{t-3} + \delta_4 DDD_{t-4} + \delta_5 DDD_{t-5} + \delta_6 DDD_{t-6} + \\ & + \delta_7 DDD_{t-7} + \delta_8 DDD_{t-8} + \delta_9 DDD_{t-9} + \delta_{10} DDD_{t-10}. \end{aligned} \quad (4.1)$$

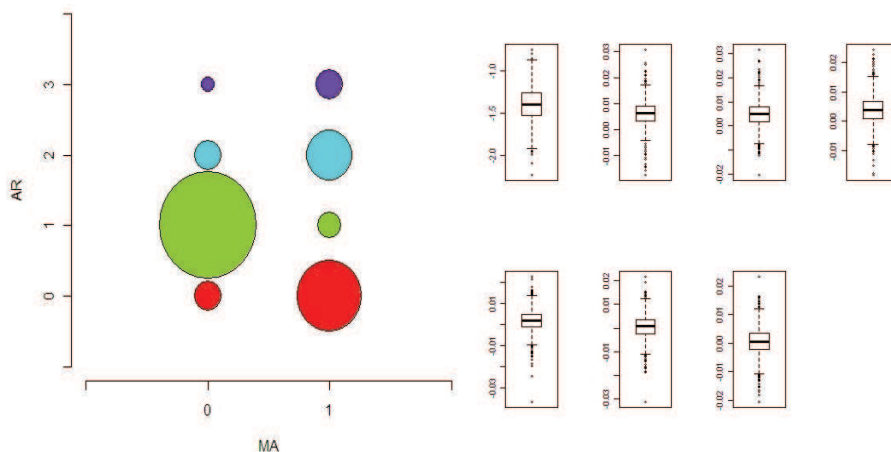
Sui modelli così formulati sono state, infine, estrapolate le stime dei coefficienti relativi alle covariate e si è verificata la loro significatività attraverso il test di Wald.

Nel prossimo paragrafo verranno illustrati, per un particolare scenario, lo svolgimento delle fasi appena descritte: dapprima, quindi, la verifica della normalità degli stimatori dei parametri del modello (3.1), in seguito le distribuzioni delle cross-correlazioni ed, infine, le significatività dei ritardi introdotti nel modello a funzione di trasferimento. Nel paragrafo successivo, invece, verranno messi a confronto i risultati dei nove scenari considerati nelle simulazioni e nell'ultimo paragrafo si illustreranno le considerazioni finali.

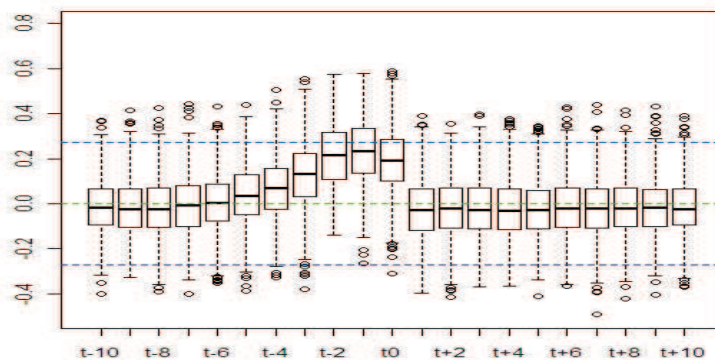
### 4.1. Studio di un caso: lo scenario A

Lo scenario A è quello che vogliamo sviluppare per intero poiché, oltre a presentare una lunghezza della serie uguale a quella dei dati reali, ha come indice binomiale una numerosità di pazienti che si avvicina a quella presente in Istar 1 e Istar 2.

Analizziamo ora i risultati ottenuti simulando i dati del reparto di rianimazione Istar 1. In Figura 4.1(a) si può osservare il diagramma a bolle che illustra graficamente quanto emerso con la funzione `arimaId`, confermando la nostra decisione di adottare per le dosi di antibiotico simulate un modello autoregressivo di ordine 1. In Figura 4.1(b) si può verificare come le distribuzioni empiriche degli stimatori dei coefficienti del modello (3.1) risultino normali nonostante la lunghezza della serie sia pari a 58 mensilità e il numero dei casi sia pari a 10.



(a) Diagramma a bolle per rappresentare la selezione del modello migliore attraverso la funzione arimaId. (b) Diagramma a scatola delle distribuzioni empiriche degli stimatori dei coefficienti nel modello (3.1).



(c) Diagrammi a scatola delle distribuzioni empiriche delle cross-correlazioni.

Figura 4.1.: Scenario A – Istar 1.

$RES_{t-1}$	$DDD_{t-0}$	$DDD_{t-1}$	$DDD_{t-2}$	$DDD_{t-3}$	$DDD_{t-4}$
0.06	0.19	0.17	0.11	0.06	0.06
$DDD_{t-5}$	$DDD_{t-6}$	$DDD_{t-7}$	$DDD_{t-8}$	$DDD_{t-9}$	$DDD_{t-10}$
0.06	0.05	0.05	0.05	0.06	0.06

Tabella 4.1.: Proporzione di p-value < 0.05 (Scenario A – Istar 1).

I risultati più interessanti sono quelli osservabili nella Figura 4.1(c), nella quale sono rappresentate le distribuzioni delle correlazioni tra le dosi di antibiotico e le resistenze batte-



#### 4. Analisi dei dati simulati

riche per vari istanti temporali. E' evidente, anche grazie alle bande di confidenza<sup>2</sup> tratteggiate in blu, che esiste una correlazione positiva tra le DDD al tempo  $t$ ,  $t-1$  e  $t-2$  e la resistenza batterica al tempo  $t$ . In altre parole, aumentando le dosi di antibiotico nei mesi  $t$ ,  $t-1$  e  $t-2$ , aumenterà anche la resistenza batterica.

Il passo successivo è stato quello di stimare il modello a funzione di trasferimento definito con la (4.1) e l'analisi della significatività, attraverso il test di Wald, delle stime dei coefficienti relativi alle covariate di tale modello. Per facilitare la presentazione dei risultati, in Tabella 3.2, vengono illustrate le percentuali delle stime dei coefficienti che hanno presentato un p-value inferiore a 0.05 per ogni covariata.

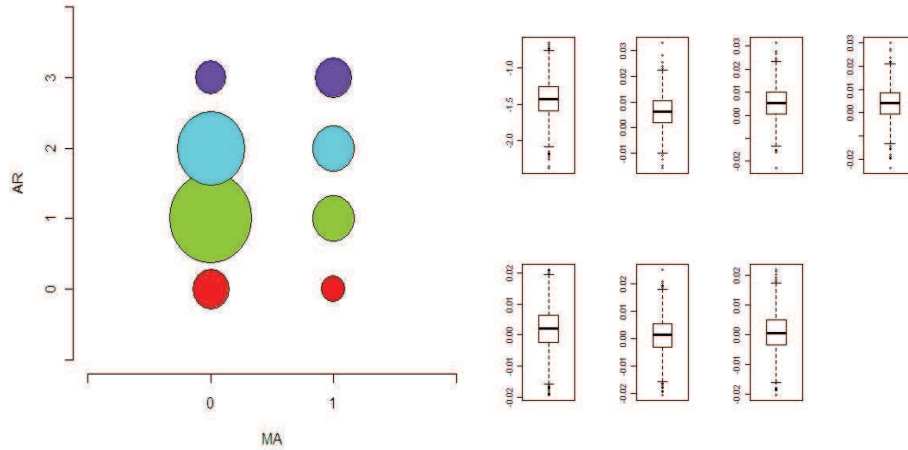
Come si può osservare, man mano che aumentano i ritardi temporali delle dosi di antibiotico, le proporzioni di p-value significativi vanno via via diminuendo. Ovviamente, la tabella conferma quanto già illustrato nei diagrammi a scatola relativi alla cross-correlazione: sono, infatti, la  $DDD_t$ , la  $DDD_{t-1}$  ed, in misura minore, la  $DDD_{t-2}$  le covariate che evidenziano maggior influenza sulla resistenza al tempo  $t$ .

In conclusione, però, si può affermare che, nonostante si riscontri una correlazione positiva tra i primi ritardi delle DDD e la resistenza batterica, questa sia piuttosto debole. In altre parole, dall'analisi grafica è evidente che, in relazione alla resistenza al tempo  $t$ , tendono ad emergere significativamente 5 ritardi, tuttavia, le esigue numerosità, in termini di lunghezza delle serie e di casi, fanno sì che la potenza associata alla  $DDD_t$ , alla  $DDD_{t-1}$  e alla  $DDD_{t-2}$  rimanga a livelli piuttosto bassi: infatti, solo in circa 1 caso su 5 risultano rilevanti le dosi al tempo  $t$  e  $t-1$  mentre la DDD al tempo  $t-2$  lo è solo in un caso su 10. Adirittura, la potenza associata alla  $DDD_{t-3}$ , alla  $DDD_{t-4}$  e alla  $DDD_{t-5}$  si confondono con i livelli di significatività dei ritardi più lontani.

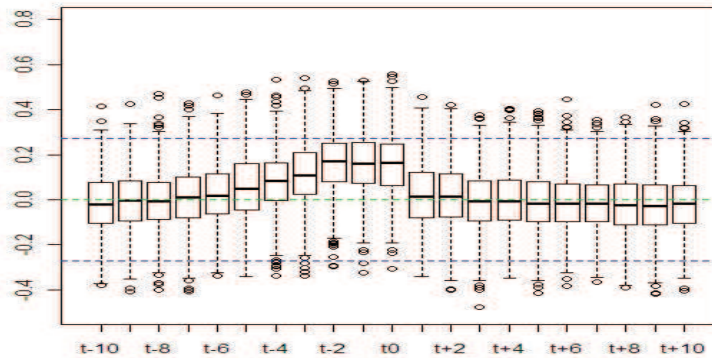
Analizziamo ora quali risultati si sono ottenuti utilizzando come base per le simulazioni le serie appartenenti all'Istar 2. Partendo dall'osservazione del diagramma a bolle in Figura 4.2(a), l'applicazione di un modello autoregressivo di ordine 1 per l'identificazione delle dosi di farmaco simulate è anche in questo caso confermata. E' vero che sembra che anche un modello AR(2) potrebbe essere utilizzato per la stima delle serie di DDD però per, motivi di continuità, si è preferito proseguire l'analisi applicando gli stessi modelli utilizzati fino ad ora. Per quanto riguarda Figura 4.2(b), invece, possiamo confermare quanto già riscontrato per i dati dell'Istar 1, ovvero che le distribuzioni empiriche degli stimatori dei coefficienti introdotti nel modello (3.1) seguono una distribuzione normale centrata nei valori che avevamo attribuito loro.

---

<sup>2</sup>La funzione `ccf()`, in R, restituisce di default delle bande di confidenza al 95%.



(a) Diagramma a bolle per rappresentare la selezione del modello migliore attraverso la funzione arimaId. (b) Diagramma a scatola delle distribuzioni empiriche degli stimatori dei coefficienti nel modello (3.1).



(c) Diagrammi a scatola delle distribuzioni empiriche delle cross-correlazioni.

Figura 4.2.: Scenario A – Istar 2.

$RES_{t-1}$	$DDD_{t-0}$	$DDD_{t-1}$	$DDD_{t-2}$	$DDD_{t-3}$	$DDD_{t-4}$
0.05	0.14	0.12	0.08	0.05	0.07
$DDD_{t-5}$	$DDD_{t-6}$	$DDD_{t-7}$	$DDD_{t-8}$	$DDD_{t-9}$	$DDD_{t-10}$
0.05	0.04	0.06	0.05	0.06	0.05

Tabella 4.2.: Proporzione di p-value < 0.05 (Scenario A – Istar 2).

Ciò che cambia rispetto alle simulazioni dell’Istar 1 sono le cross-correlazioni: se in quel reparto, infatti, avevamo riscontrato una, seppur debole, positiva correlazione tra le dosi di antibiotico e la resistenza batterica, in questo caso non è emersa nessuna relazione si-

#### 4. Analisi dei dati simulati

gnificativa. Le distribuzioni delle cross-correlazioni, infatti, si sviluppano principalmente entro le bande di confidenza e, dunque, non emerge - anche se dovrebbe - nessun legame tra quelle che sono le dosi di medicinale e l'emergere delle resistenze. Lo stesso riscontro si ottiene osservando la Tabella 4.2, nella quale si nota come solo in poco più di 1 caso su 10 risultano rilevanti la  $DDD_t$  e la  $DDD_{t-1}$  mentre le altre dosi di antibiotico che dovrebbero emergere hanno una potenza che si confonde con i livelli di significatività delle dosi di antibiotico con ritardi più lontani.

### 4.2. Confronti tra scenari

Come già spiegato in precedenza, lo Scenario 1, nella sua interezza, ci permette di capire se la causa della mancata evidenza empirica nei nostri dati originali sia esclusivamente da attribuire alla scarsa numerosità dei casi nei reparti di rianimazione; lo Scenario 2 è utile per capire se si riusciranno ad avere i risultati sperati in breve tempo; lo Scenario 3, avendo lunghezza temporale simile a quelle che si riscontrano in letteratura, ci servirà per verificare quanto questa incida sull'ottenimento di risultati soddisfacenti.

Iniziamo il confronto tra gli scenari osservando le Tabelle 4.3 e 4.4, contenenti le proporzioni di p-value significativi dei diversi ritardi temporali, ottenuti con il test di significatività di Wald, dopo lo sviluppo delle cross-correlazioni.

Come si osserva, i valori riportati nelle Tabelle sono in parte blu ed in parte rossi: la nostra intenzione, infatti, è quella di distinguere i valori che identificano la potenza (blu) da quelli che rappresentano la significatività (rosso) del test di Wald. Ricordiamo che con il test alla Wald, di cui si è parlato nel primo paragrafo di questo capitolo, abbiamo voluto verificare

$$\begin{cases} H_0: \delta_i = 0 \\ H_1: \delta_i \neq 0 \end{cases},$$

utilizzando come livello di significatività  $\alpha = 0.05$ .

I valori che identificano la potenza del test sono in corrispondenza delle variabili  $DDD_t$ ,  $DDD_{t-1}$ ,  $DDD_{t-2}$ ,  $DDD_{t-3}$ ,  $DDD_{t-4}$  e  $DDD_{t-5}$ , ossia le covariate introdotte nel modello a funzione di trasferimento (4.1) che figurano anche nel modello (3.1). In corrispondenza alle prime covariate ( $DDD_t$ ,  $DDD_{t-1}$ ,  $DDD_{t-2}$ ) sembra vi sia un effetto combinato delle quantità "lunghezza della serie" e "numerosità dei casi": la potenza del test, infatti, aumenta sia con l'aumentare della lunghezza della serie temporale, che con l'ampliamento dell'indice binomiale. Addirittura, se si osservano le righe in corrispondenza della dose di

	A	B	C	D	E	F	G	H	I
$DDD_t$	.19	.99	.99	.27	.99	.99	.45	.99	.99
$DDD_{t-1}$	.17	.81	.97	.22	.89	.99	.33	.97	.99
$DDD_{t-2}$	.11	.56	.86	.14	.76	.97	.21	.90	.99
$DDD_{t-3}$	.06	.16	.44	.07	.25	.60	.08	.35	.77
$DDD_{t-4}$	.06	.06	.12	.06	.09	.23	.06	.13	.37
$DDD_{t-5}$	.06	.06	.12	.06	.07	.18	.05	.10	.30
$DDD_{t-6}$	.05	.06	.06	.06	.06	.05	.06	.06	.06
$DDD_{t-7}$	.05	.05	.06	.06	.05	.05	.05	.05	.05
$DDD_{t-8}$	.05	.05	.06	.05	.05	.06	.06	.05	.05
$DDD_{t-9}$	.06	.06	.06	.06	.06	.05	.05	.06	.05
$DDD_{t-10}$	.06	.06	.04	.06	.06	.06	.05	.06	.05

Tabella 4.3.: Potenza (blu) e significatività (rosso) dei diversi ritardi temporali, riferiti alle cross-correlazioni, distinti per scenario – Istar 1.

	A	B	C	D	E	F	G	H	I
$DDD_t$	.14	.94	.99	.16	.87	.99	.23	.95	.99
$DDD_{t-1}$	.12	.68	.95	.12	.68	.98	.18	.84	.99
$DDD_{t-2}$	.08	.37	.80	.11	.48	.92	.11	.64	.98
$DDD_{t-3}$	.05	.13	.33	.06	.16	.45	.07	.22	.63
$DDD_{t-4}$	.07	.06	.11	.05	.08	.17	.05	.09	.22
$DDD_{t-5}$	.05	.06	.09	.06	.06	.14	.06	.08	.17
$DDD_{t-6}$	.04	.05	.05	.05	.05	.05	.04	.05	.06
$DDD_{t-7}$	.06	.05	.05	.05	.06	.05	.05	.05	.06
$DDD_{t-8}$	.05	.05	.05	.05	.05	.04	.05	.05	.06
$DDD_{t-9}$	.06	.05	.05	.05	.05	.05	.06	.05	.05
$DDD_{t-10}$	.05	.06	.06	.04	.05	.05	.05	.05	.05

Tabella 4.4.: Potenza (blu) e significatività (rosso) dei diversi ritardi temporali, riferiti alle cross-correlazioni, distinti per scenario – Istar 2.

antibiotico al tempo  $t$  e al tempo  $t-1$ , nel caso di numerosità pari a 500 pazienti, in entrambi i reparti, vediamo che la probabilità di accettare  $H_0$  in quanto falsa è inferiore al 5%.

Per le covariate  $DDD_{t-3}$ ,  $DDD_{t-4}$  e  $DDD_{t-5}$ , invece, non si ottiene lo stesso risultato: con numerosità pari a 10, infatti, in entrambi i reparti, abbiamo che la potenza del test si aggira attorno allo 0.06 anche variando la lunghezza della serie. I cambiamenti si hanno lavorando con numerosità di casi più elevati, tuttavia, più in Istar 2 che in Istar 1, questi ritardi fanno fatica ad emergere come strettamente relazionati alla resistenza batterica.

Infine, un buon risultato è osservare i valori in rosso delle due tabelle: relativamente alle variabili  $DDD_{t-6}$ ,  $DDD_{t-7}$ ,  $DDD_{t-8}$ ,  $DDD_{t-9}$  e  $DDD_{t-10}$ , in entrambi i reparti, si ha

#### 4. Analisi dei dati simulati

che di significatività empirico è sempre vicino al valore 0.05, il che corrisponde al livello di di significatività adottato. Questo esito ci permette di affermare che, in nessun caso, emergono dosi di antibiotico somministrate in tempi molto lontani da  $t$  relazionate significativamente con la resistenza batterica presente al tempo  $t$ .

In conclusione, analizzando sia la potenza, che diminuisce via via che aumentano i ritardi temporali delle dosi di antibiotico, e la significatività, che si mantiene pressoché costante per tutte le dosi di antibiotico molto lontane nel tempo, possiamo affermare che sono i farmaci somministrati nei mesi più vicini al tempo  $t$  ad emergere per primi.

In questa seconda parte confrontiamo gli scenari, dapprima mantenendo fissa la lunghezza temporale, poi mantenendo fissa il valore dell'indice binomiale osservando i grafici rappresentati nelle Figure 4.3, 4.4, 4.5 e 4.6 ottenuti mediante i dati presentati nelle Tabelle 4.3 e 4.4.

In Figura 4.3, si osservano, nei diversi scenari, a sinistra della linea nera verticale le potenze e a destra della stessa le significatività dei diversi ritardi temporali riferiti alle cross-correlazioni per quanto riguarda l'Istar 1, mentre in Figura 4.4, si analizza la stessa situazione per l'Istar 2. I risultati nei due reparti sono piuttosto simili: in tutti i casi, infatti, gli andamenti delle spezzate si assomigliano ed, inoltre, si osserva come le numerosità incidano fortemente su quello che è l'emergere delle dosi di antibiotico significative. Tuttavia, la quantità dei ritardi significativi, cambia in base alla lunghezza della serie temporale: se nello Scenario 1 simulato in entrambi i reparti, ad emergere sono le dosi di farmaco assunte nei tempo  $t$  e  $t-1$ , negli Scenari 2 e 3 anche la DDD del mese  $t-2$  ha una potenza molto elevata.

Le Figure 4.5 e 4.6 sono, dal nostro punto di vista, molto utili poiché confrontano gli scenari con le stesse numerosità. Quello che salta subito agli occhi è come, in entrambi i reparti, le spezzate che corrispondono alle numerosità elevate siano pressoché uguali, il che ci fa presupporre che il parametro "numerosità" sia fondamentale per ottenere dei risultati soddisfacenti. Tuttavia, una diversità tra gli scenari con numerosità 100 e quelli con numerosità 500, c'è: nel primo caso, infatti, le spezzate evidenziano una potenza molto elevata dei ritardi  $t$ ,  $t-1$  e  $t-2$ , mentre nel secondo caso aumenta anche la potenza dei ritardi  $t-3$  e  $t-4$ . In questo ultimo caso, inoltre, le spezzate, in corrispondenza dei ritardi  $t$ ,  $t-1$  e  $t-2$ , formano un plateau che rappresenta il fatto che questi istanti temporali sono ugualmente importanti nell'identificazione della resistenza al tempo  $t$ .

Infine, evidenziamo che, in tutti i grafici presentati, alle dosi di antibiotico più lontane corrispondono dei livelli di significatività empirici che si aggirano attorno al livello da noi stabilito ( $\alpha = 0.05$ ), rappresentato orizzontalmente con la linea nera.

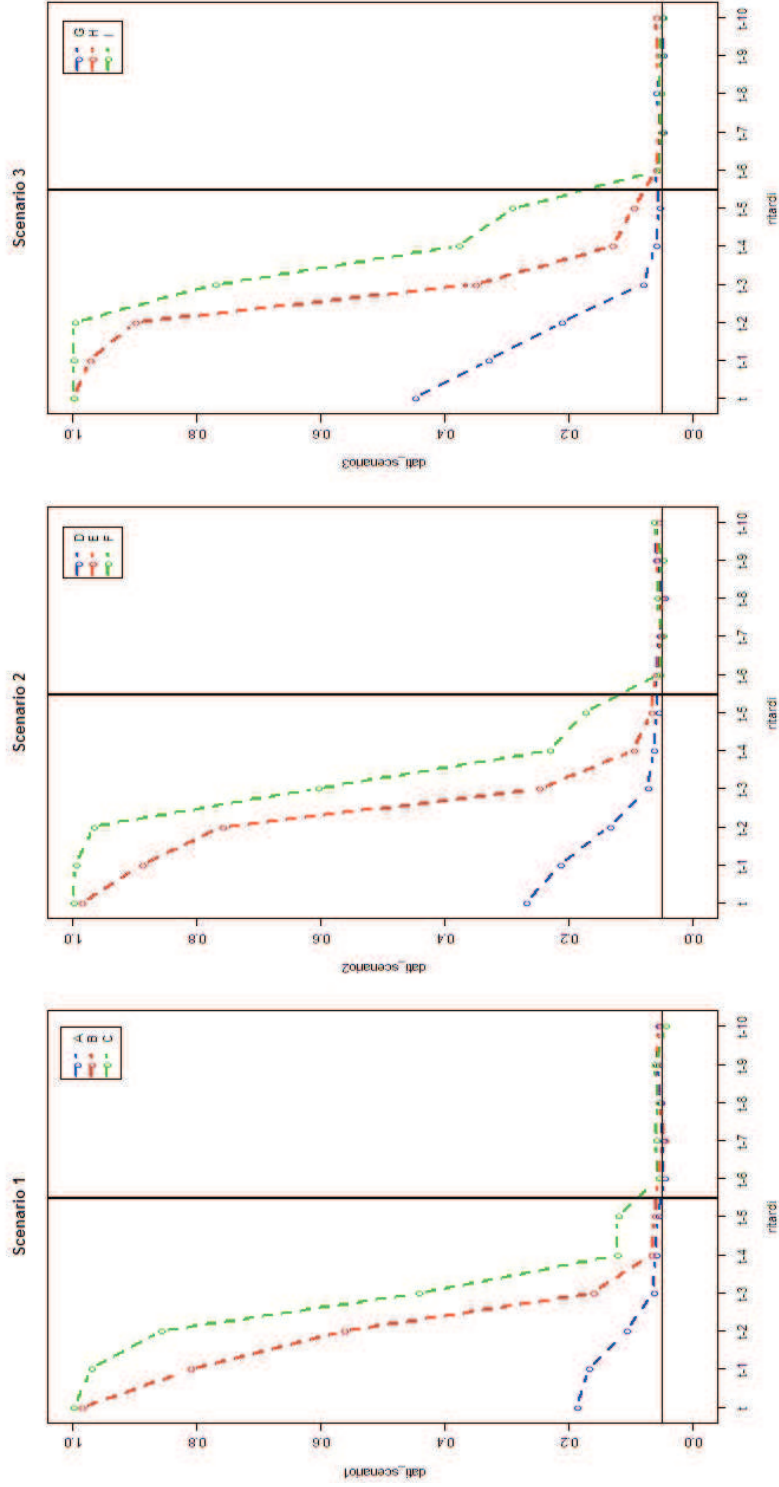
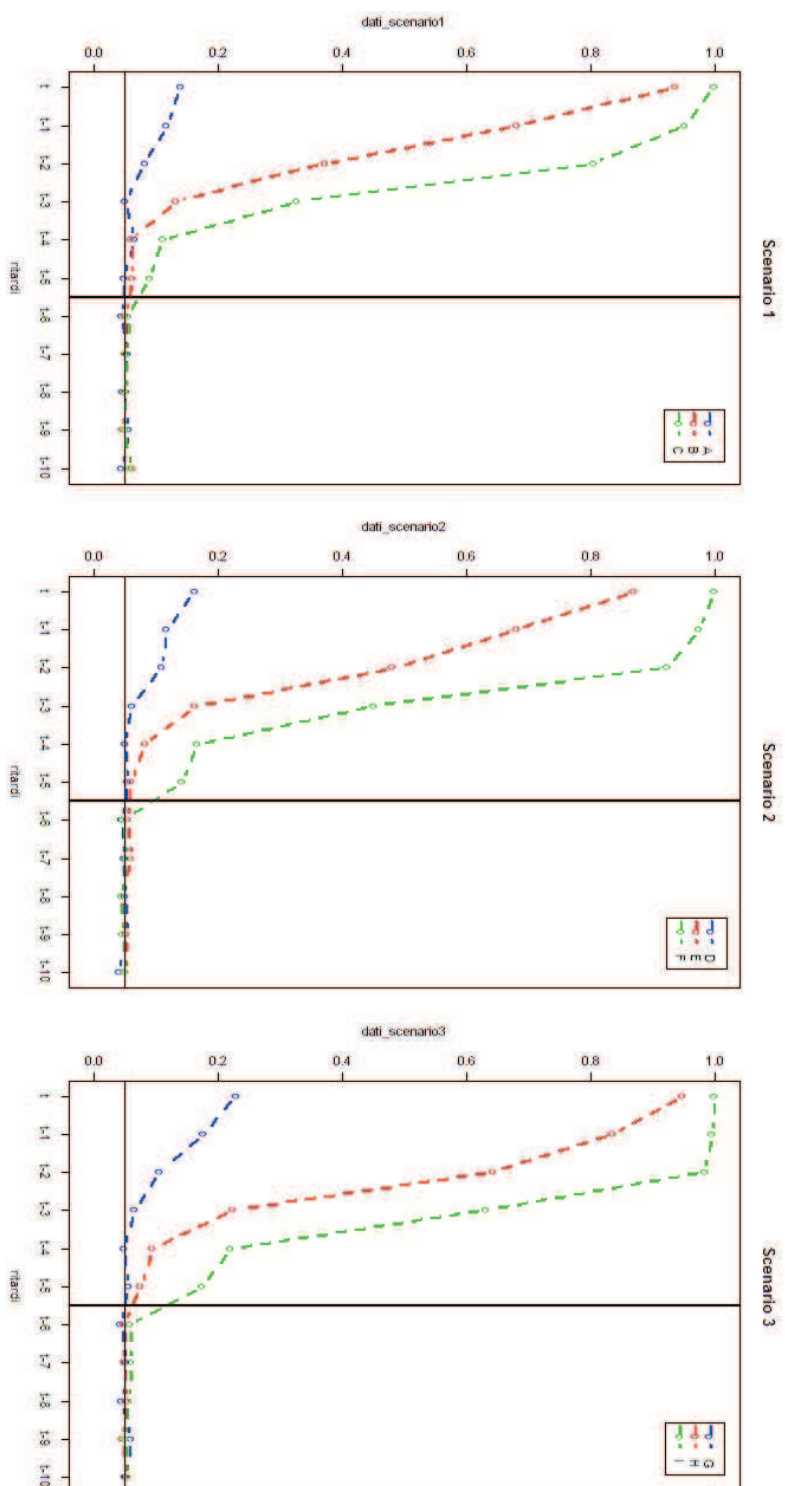


Figura 4.3.: Confronto tra scenari mantenendo fissa la lunghezza temporale – Istar 1. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)

#### 4. Analisi dei dati simulati

Figura 4.4.: Confronto tra scenari mantenendo fissa la lunghezza temporale – Istar 2. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)





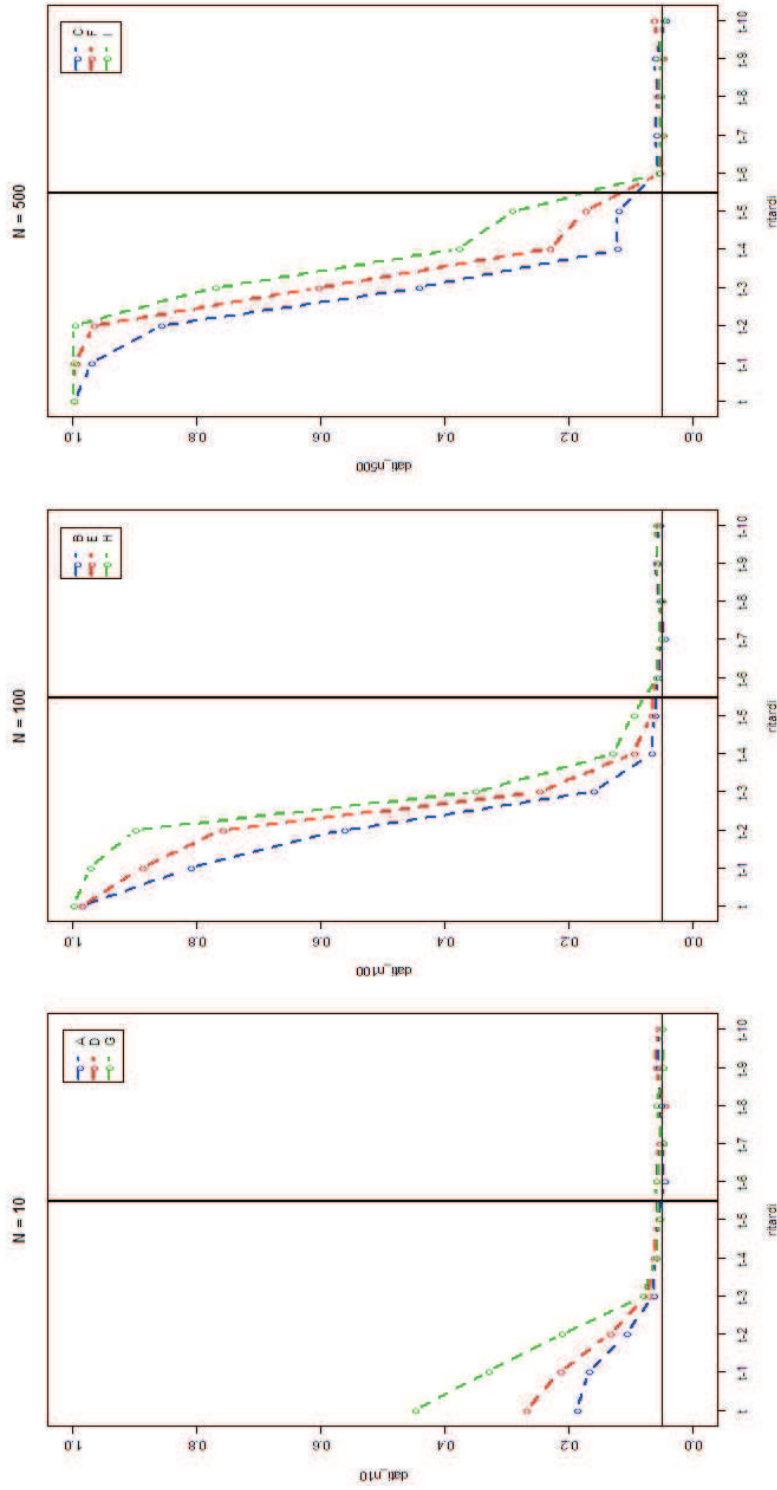


Figura 4.5.: Confronto tra scenari mantenendo fissa la numerosità – Istar 1. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)



#### 4. Analisi dei dati simulati

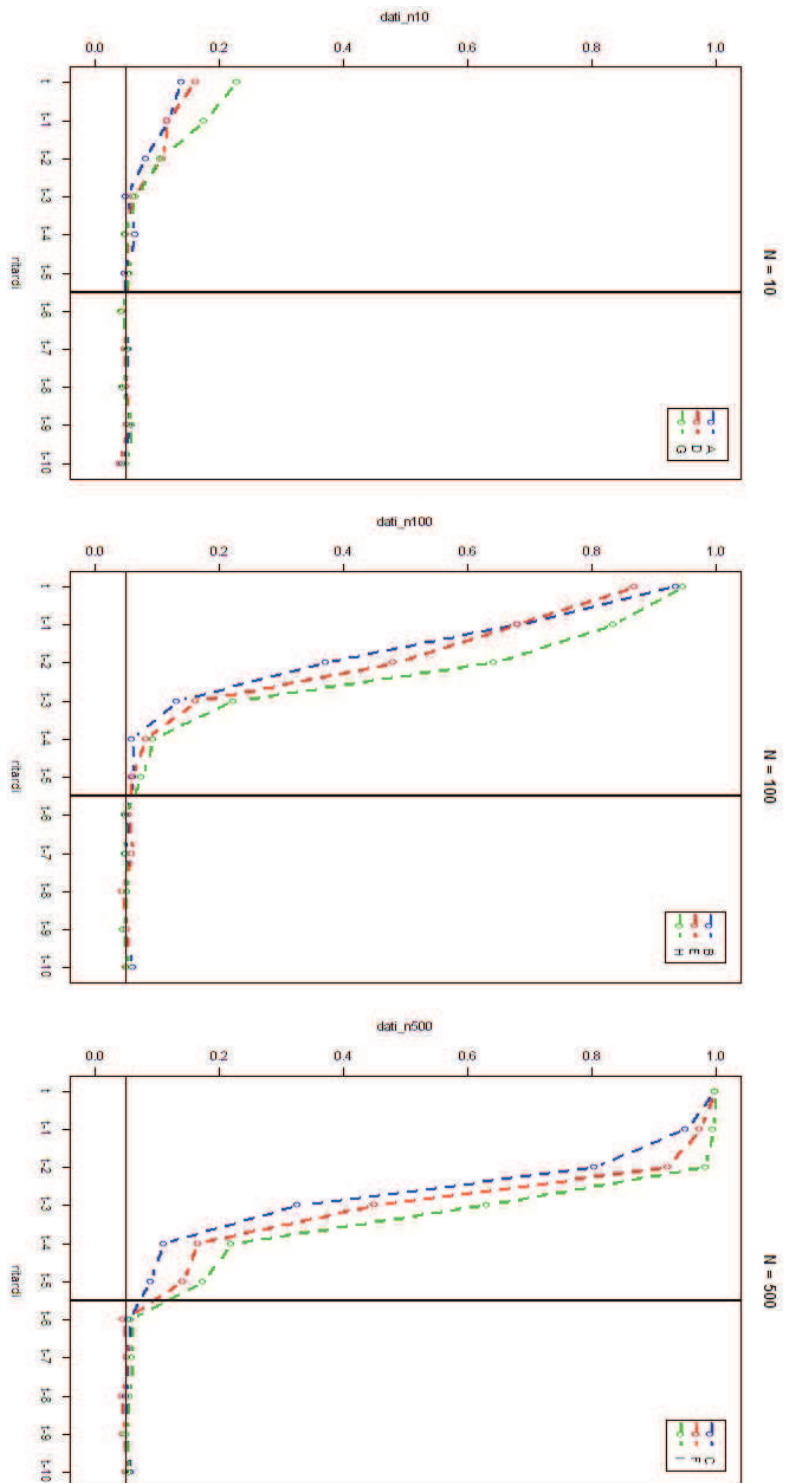


Figura 4.6.: Confronto tra scenari mantenendo fissa la numerosità – Istar 2. (In ascissa i ritardi temporali; in ordinata le proporzioni di p-value < 0.05.)

Le Tabelle 4.5 e 4.6 mostrano gli andamenti delle cross-correlazioni di tutti gli scenari, in Istar 1 e Istar 2, rispettivamente.

Come già osservato nei grafici precedenti man mano che aumentano la numerosità e la lunghezza temporale della serie, più facile è l'individuazione dei ritardi temporali significativi. In particolare, vi è la sempre più marcata correlazione positiva tra le dosi di antibiotico ai tempi  $t$ ,  $t-1$  e  $t-2$  e la resistenza al tempo  $t$ .

Tuttavia, tra i due reparti vi è una netta differenza: se nel caso di Istar 1, l'evidenza appena descritta appare già visibile quando il numero dei casi è pari a 10, in Istar 2 è necessario avere numerosità più elevate per confermare la relazione tra dose di farmaco e resistenza batterica.

In ogni caso, interessante è osservare come le dosi di antibiotico lontane nel tempo non siano mai ritenute relazionate con resistenza al tempo  $t$ . In entrambi i reparti, inoltre, non vi è nessun caso in cui la resistenza batterica sembra influire sulle dosi di antibiotico.

Queste considerazioni sono importanti, alla luce di quanto era emerso nelle analisi di esplorazione iniziali. Ricordiamo, infatti, che nelle cross-correlazioni (si veda Figura 2.3) erano emerse delle relazioni che, dal punto di vista medico, erano difficili da spiegare: in Istar 1, infatti, da un lato, la DDD ai tempi  $t-2$ ,  $t-6$  e  $t-7$  era significativamente correlata con le resistenze, dall'altro questa era correlata con le dosi di medicinale, mentre, in Istar 2, era la DDD al tempo  $t-8$  ad avere effetti sulla resistenza al tempo  $t$ .

Il fatto che negli scenari da noi simulati si sia evidenziata la vera struttura del legame tra dosi di medicinale e resistenza batterica, ci fa ipotizzare che i dati iniziali non siano sufficienti per far emergere le relazioni presenti tra le variabili di interesse. Inoltre, ci sentiamo di affermare che la causa della mancata evidenza empirica non sia da attribuire alla lunghezza della serie, bensì all'esigua numerosità dei casi reali poiché, come si è visto, siamo riusciti ad ottenere dei discreti risultati anche con numerosità pari a 10: anche con 10 pazienti simulati, infatti, in 1 caso su 5 emerge la relazione tra resistenza batterica al tempo  $t$  con la  $DDD_t$  e la  $DDD_{t-1}$ .

4. Analisi dei dati simulati

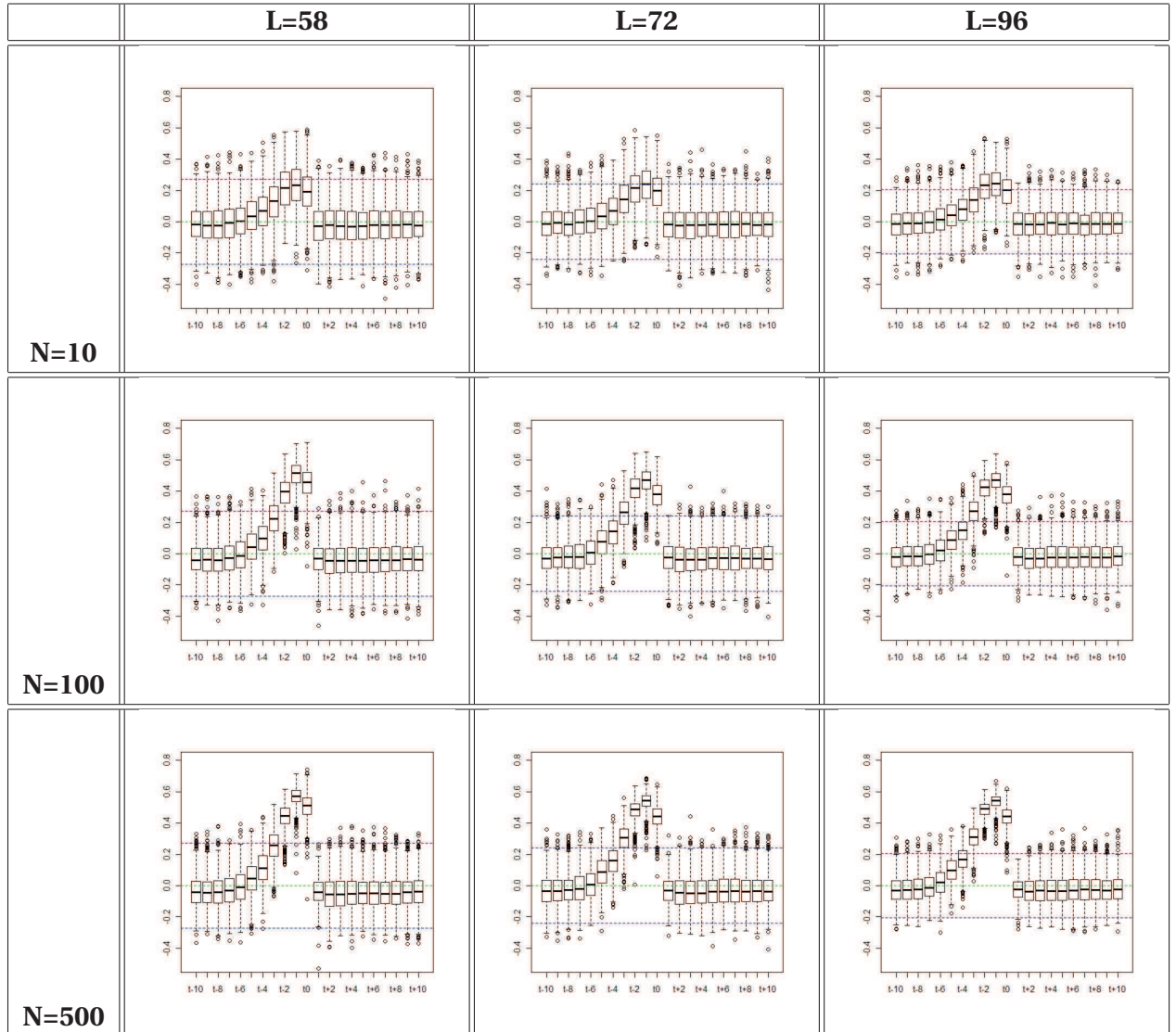


Tabella 4.5.: Cross-correlazioni (scenari simulati) – Istar 1.

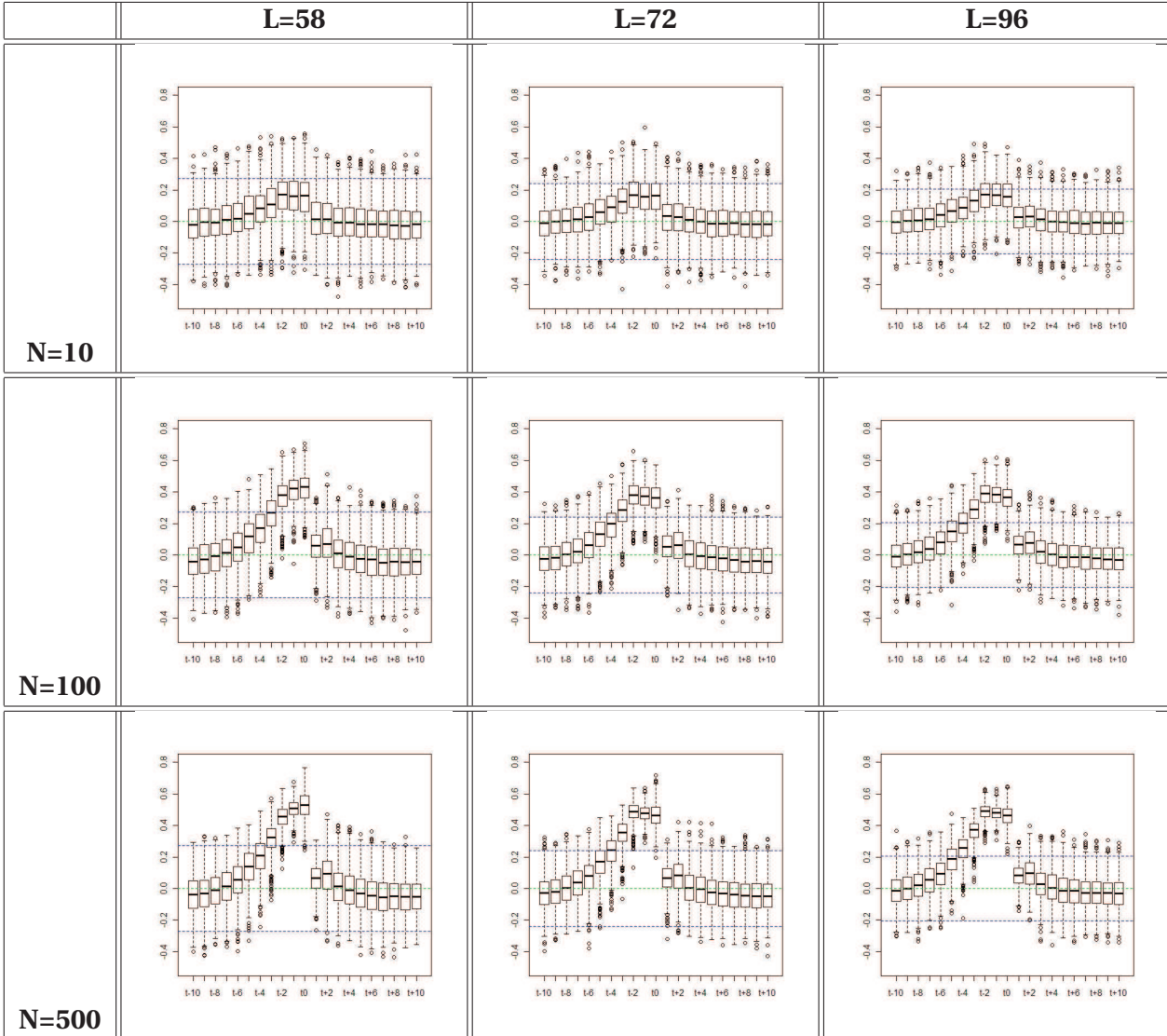


Tabella 4.6.: Cross-correlazioni (scenari simulati) – Istar 2.

#### 4. Analisi dei dati simulati

Concludiamo le analisi presentando le Tabelle 4.7 e 4.8 le quali contengono le stime bootstrap e i relativi intervalli di confidenza per i parametri del modello a funzione di trasferimento (4.1). Osservando i valori riportati in entrambe le tabelle, si può giungere alle stesse considerazioni emerse dall'analisi delle cross-correlazioni.

Sia in Istar 1 che in Istar 2, infatti, è evidente che aumentando la lunghezza della serie temporale e la numerosità dei casi considerati, le stime si discostano sempre più dallo 0.

Analizziamo in dettaglio ciò che avviene nel caso dell'Istar 1. Se consideriamo lo Scenario A (lunghezza pari a 58 mensilità e numerosità pari a 10), osserviamo che le stime dei coefficienti relativi alle covariate  $DDD_t$ ,  $DDD_{t-1}$  e  $DDD_{t-2}$  sono tutti pari a 0.01 e i relativi intervalli di confidenza contengono lo 0. Questa situazione si ripete anche nei casi degli Scenari D e G che mantengono come indice binomiale numerosità pari a 10. Aumentando la lunghezza della serie, tuttavia, si arriva ad avere che le stime relative ai coefficienti delle variabili  $DDD_{t-3}$ ,  $DDD_{t-4}$  e  $DDD_{t-5}$  si discostano da 0.

Situazione analoga nel caso dell'Istar 2, sebbene i valori delle stime si discostino in modo leggermente meno marcato rispetto a quanto avviene in Istar 1.

Come i valori delle stime sono sempre più diverse da 0, si osservano, ovviamente, anche i loro relativi intervalli di confidenza spostarsi verso valori distanti da 0: questo evidenzia la significatività di tali parametri poiché, effettuando un test di Wald per verificare l'ipotesi di diversità dallo 0, ad essi corrisponderebbe un p-value inferiore a 0.05.

La volontà di presentare tali risultati è data dal fatto che quanto appena descritto supporta appieno le analisi sviluppate precedentemente. In pratica, anche l'osservazione delle stime dei parametri nel modello a funzione di trasferimento ci permette di concludere che sono maggiormente le dosi di farmaco somministrate nei tempi  $t$ ,  $t-1$  e  $t-2$  ad emergere come significative per la determinazione della resistenza batterica al tempo  $t$ . Inoltre, come già ribadito più volte, non vi è nessuna evidenza che lega le dosi di farmaco somministrate prima del tempo  $t-5$  alla resistenza batterica del tempo  $t$ .

	A	B	C	D	E	F	G	H	I
$DDD_t$	0.01 [-0.01; 0.03]	0.17 [0.09; 0.24]	0.86 [0.66; 1.02]	0.01 [-0.01; 0.03]	0.13 [0.06; 0.18]	0.63 [0.48; 0.75]	0.01 [0.00; 0.03]	0.13 [0.08; 0.17]	0.63 [0.51; 0.74]
$DDD_{t-1}$	0.01 [-0.01; 0.03]	0.13 [0.03; 0.22]	0.66 [0.34; 0.97]	0.01 [-0.01; 0.03]	0.11 [0.04; 0.18]	0.53 [0.31; 0.75]	0.01 [0.00; 0.03]	0.11 [0.05; 0.16]	0.53 [0.34; 0.72]
$DDD_{t-2}$	0.01 [-0.02; 0.03]	0.09 [0.00; 0.17]	0.42 [0.15; 0.68]	0.01 [-0.01; 0.03]	0.08 [0.01; 0.15]	0.42 [0.21; 0.61]	0.01 [-0.01; 0.02]	0.09 [0.03; 0.14]	0.58 [0.26; 0.48]
$DDD_{t-3}$	0.00 [-0.02; 0.03]	0.04 [-0.04; 0.12]	0.20 [-0.01; 0.41]	0.00 [-0.01; 0.02]	0.04 [-0.02; 0.10]	0.20 [0.01; 0.39]	0.00 [-0.01; 0.02]	0.04 [-0.01; 0.09]	0.33 [0.05; 0.29]
$DDD_{t-4}$	0.00 [-0.02; 0.02]	0.02 [-0.06; 0.09]	0.08 [-0.10; 0.24]	0.00 [0.02; 0.02]	0.02 [-0.04; 0.08]	0.09 [-0.05; 0.24]	0.00 [-0.01; 0.02]	0.02 [-0.03; 0.07]	0.21 [-0.02; 0.19]
$DDD_{t-5}$	0.00 [-0.02; 0.02]	0.01 [-0.06; 0.09]	0.06 [-0.11; 0.23]	0.00 [0.02; 0.02]	0.01 [-0.05; 0.07]	0.08 [-0.06; 0.21]	0.00 [-0.01; 0.02]	0.02 [-0.03; 0.06]	0.19 [-0.04; 0.15]
$DDD_{t-6}$	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.07]	0.00 [-0.17; 0.17]	0.00 [0.02; 0.02]	0.00 [-0.06; 0.06]	0.00 [-0.13; 0.13]	0.00 [-0.01; 0.02]	0.00 [-0.05; 0.05]	0.11 [-0.11; 0.08]
$DDD_{t-7}$	0.00 [-0.02; 0.02]	0.00 [-0.08; 0.07]	0.00 [-0.16; 0.18]	0.00 [0.02; 0.02]	0.00 [-0.06; 0.06]	0.00 [-0.13; 0.13]	0.00 [-0.01; 0.02]	0.00 [-0.05; 0.05]	0.11 [-0.10; 0.25]
$DDD_{t-8}$	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.07]	0.00 [-0.16; 0.16]	0.00 [0.02; 0.02]	0.00 [-0.06; 0.06]	0.00 [-0.14; 0.13]	0.00 [-0.01; 0.02]	0.00 [-0.05; 0.05]	0.10 [-0.11; 0.33]
$DDD_{t-9}$	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.07]	0.00 [-0.15; 0.17]	0.00 [0.02; 0.02]	0.00 [-0.07; 0.06]	0.00 [-0.12; 0.13]	0.00 [-0.01; 0.02]	0.00 [-0.04; 0.04]	0.11 [-0.11; 0.08]
$DDD_{t-10}$	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.07]	0.00 [-0.16; 0.18]	0.00 [0.02; 0.02]	0.00 [-0.05; 0.06]	0.00 [-0.12; 0.13]	0.00 [-0.01; 0.01]	0.00 [-0.04; 0.04]	0.11 [-0.09; 0.34]

Tabella 4.7.: Stime bootstrap e relativi intervalli di confidenza per i parametri del modello (4.1)-Istar 1.

4. Analisi dei dati simulati

	A	B	C	D	E	F	G	H	I
$DDD_t$	0.01 [-0.02; 0.04]	0.17 [0.07; 0.25]	0.83 [0.61; 1.05]	0.01 [-0.01; 0.04]	0.12 [0.04; 0.21]	0.61 [0.42; 0.78]	0.01 [-0.01; 0.03]	0.12 [0.06; 0.18]	0.61 [0.46; 0.74]
$DDD_{t-1}$	0.01 [-0.02; 0.04]	0.13 [0.02; 0.24]	0.63 [0.27; 0.96]	0.01 [-0.02; 0.04]	0.10 [0.02; 0.18]	0.52 [0.26; 0.76]	0.01 [-0.01; 0.03]	0.10 [0.03; 0.17]	0.51 [0.31; 0.71]
$DDD_{t-2}$	0.01 [-0.02; 0.04]	0.09 [-0.02; 0.19]	0.42 [0.12; 0.71]	0.01 [-0.02; 0.03]	0.08 [-0.01; 0.17]	0.41 [0.16; 0.63]	0.01 [-0.01; 0.03]	0.08 [0.01; 0.15]	0.41 [0.22; 0.59]
$DDD_{t-3}$	0.01 [-0.02; 0.04]	0.04 [-0.06; 0.15]	0.21 [-0.05; 0.48]	0.00 [-0.02; 0.03]	0.04 [-0.04; 0.13]	0.21 [-0.02; 0.41]	0.00 [-0.02; 0.03]	0.04 [-0.02; 0.11]	0.20 [0.01; 0.38]
$DDD_{t-4}$	0.00 [-0.03; 0.03]	0.01 [-0.09; 0.10]	0.08 [-0.17; 0.29]	0.00 [-0.02; 0.03]	0.02 [-0.06; 0.10]	0.10 [-0.10; 0.29]	0.00 [-0.02; 0.02]	0.02 [-0.05; 0.09]	0.10 [-0.07; 0.25]
$DDD_{t-5}$	0.00 [-0.03; 0.03]	0.01 [-0.09; 0.12]	0.06 [-0.17; 0.28]	0.00 [-0.02; 0.03]	0.01 [-0.07; 0.10]	0.08 [-0.12; 0.26]	0.00 [-0.02; 0.02]	0.02 [-0.05; 0.08]	0.07 [-0.08; 0.22]
$DDD_{t-6}$	0.00 [-0.03; 0.03]	0.00 [-0.10; 0.10]	0.00 [-0.22; 0.22]	0.00 [-0.02; 0.03]	0.00 [-0.08; 0.08]	0.00 [-0.18; 0.17]	0.00 [-0.02; 0.02]	0.00 [-0.06; 0.07]	0.00 [-0.15; 0.15]
$DDD_{t-7}$	0.00 [-0.03; 0.03]	0.00 [-0.10; 0.10]	0.00 [-0.20; 0.22]	0.00 [-0.03; 0.02]	0.00 [-0.08; 0.08]	0.00 [-0.19; 0.18]	0.00 [-0.02; 0.02]	0.00 [-0.06; 0.06]	0.00 [-0.15; 0.15]
$DDD_{t-8}$	0.00 [-0.03; 0.03]	0.00 [-0.10; 0.10]	0.00 [-0.21; 0.21]	0.00 [-0.03; 0.02]	0.00 [-0.08; 0.09]	0.00 [-0.17; 0.19]	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.06]	0.00 [-0.15; 0.14]
$DDD_{t-9}$	0.00 [-0.03; 0.03]	0.00 [-0.10; 0.09]	0.00 [-0.22; 0.22]	0.00 [-0.02; 0.03]	0.00 [-0.08; 0.08]	0.00 [-0.18; 0.18]	0.00 [-0.02; 0.02]	0.00 [-0.06; 0.06]	0.00 [-0.15; 0.15]
$DDD_{t-10}$	0.00 [-0.03; 0.03]	0.00 [-0.09; 0.09]	0.00 [-0.21; 0.21]	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.08]	0.00 [-0.017; 0.16]	0.00 [-0.02; 0.02]	0.00 [-0.07; 0.06]	0.00 [-0.14; 0.15]

Tabella 4.8.: Stime bootstrap e relativi intervalli di confidenza per i parametri del modello (4.1) - Istar 2.

### 4.3. Considerazioni finali

Alla luce delle considerazioni sviluppate al paragrafo precedente, nel quale abbiamo affermato che la probabile causa della mancata evidenza empirica nei nostri dati originali è la numerosità esigua dei degenti all'interno dei reparti di rianimazione, vogliamo mostrare i risultati delle nostre simulazioni in base ad altri indici binomiali. Tale scelta è dettata da quelle che sono le esigenze reali dei Medici che lavorano in Istar 1 ed Istar 2, ovvero, essi hanno bisogno di avere dei riscontri su numerosità che ritengono plausibili. Ricordiamo, infatti, che le numerosità da noi proposte si riferiscono a pazienti che hanno contratto il germe PSAE e che, quindi, hanno probabilità di sviluppare la resistenza batterica al MEM. Sulla base di queste ipotesi, 500 degenti mensili è, per l'equipe medica, una quantità troppo elevata per essere considerata plausibile mentre la possibilità di averne 100 potrebbe essere plausibile nel caso dell'accorpamento delle due rianimazioni.

Tuttavia, dal punto di vista statistico, era utile simulare quantità molto diverse da quelle reali proprio per evidenziare l'incidenza delle quantità che abbiamo considerato fondamentali in questo lavoro, cioè la lunghezza della serie e il numero di pazienti ricoverati. Riteniamo, infatti, che, se si fosse lavorato con valori molto vicini alle serie originali, non si sarebbe avuta la possibilità di capire a cosa fossero dovute le difficoltà delle prime analisi esplorative.

In questa ultima parte, quindi, mostriamo quelle che sono le distribuzioni delle cross-correlazioni tra le dosi di antibiotico e le resistenze batteriche (ottenute nello stesso modo con il quale si sono sviluppate le analisi nel Paragrafo 4.2), utilizzando, come indici binomiali, quantità pari a 20, 40, 60 ed 80.

Possiamo già immaginare i risultati che stiamo per descrivere: man mano che la numerosità aumenta, più facile è l'individuazione dei ritardi significativi (in particolare emergono la  $DDD_t$ , la  $DDD_{t-1}$  e la  $DDD_{t-2}$ ) ed, ovviamente, anche l'allungamento della serie temporale apporta notevoli migliorie in termini di potenza relativa alle dosi di antibiotico somministrate al tempo  $t$ ,  $t-1$ ,  $t-2$ ,  $t-3$ ,  $t-4$  e  $t-5$ .

In ogni caso, ai fini pratici del lavoro, è utile osservare come in Istar 1, con numerosità di 20, si possano già ottenere dei risultati soddisfacenti a fine 2012. Purtroppo, in Istar 2, con numerosità pari a 20 non si possono trarre le stesse conclusioni entro l'anno e nemmeno attendendo la raccolta di altre informazioni: per questo reparto, sembra sia necessaria una numerosità di degenti più elevata.



4. Analisi dei dati simulati

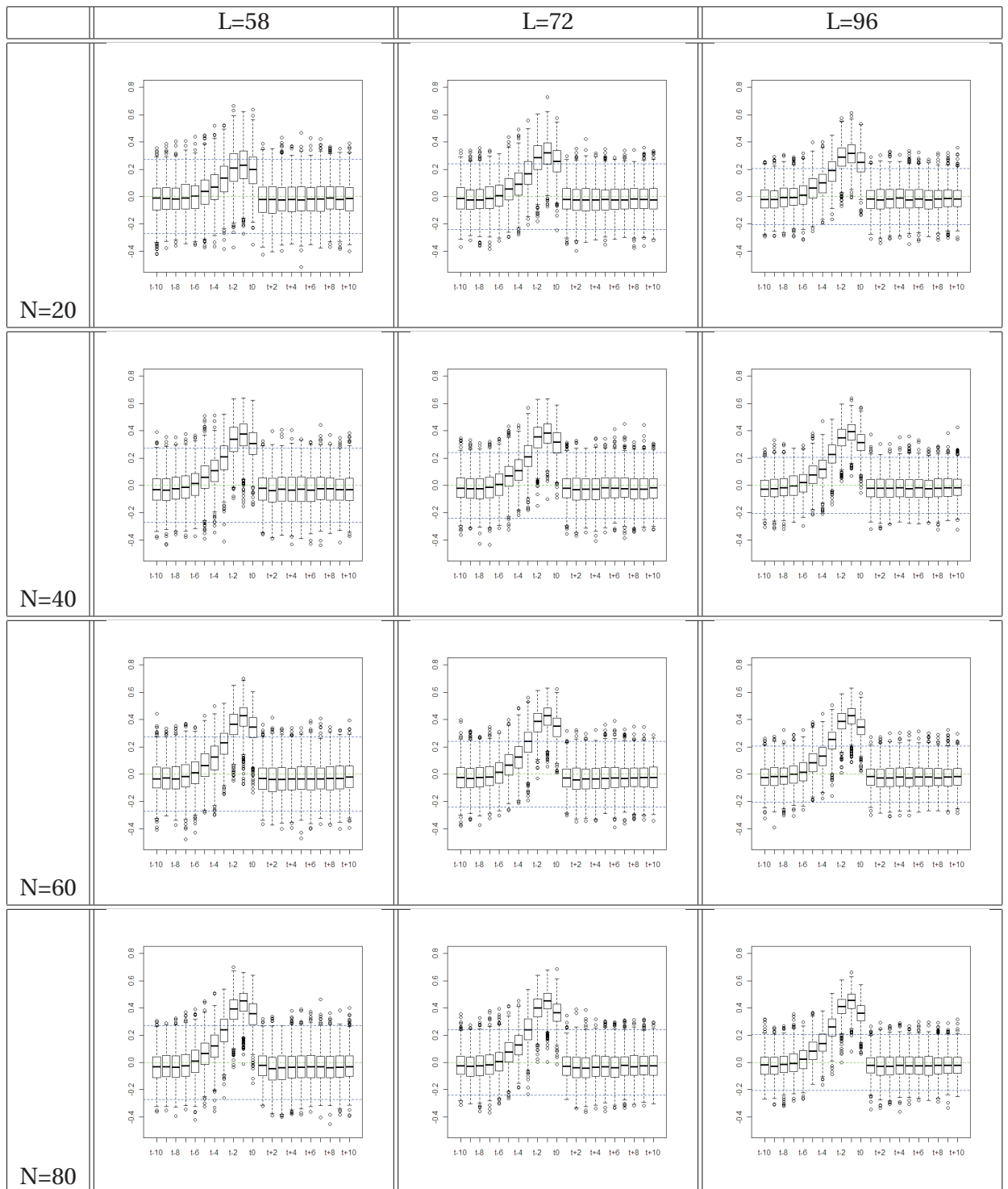


Tabella 4.9.: Cross-correlazioni con numerosità pari 20, 40, 60 ed 80 casi – Istar 1

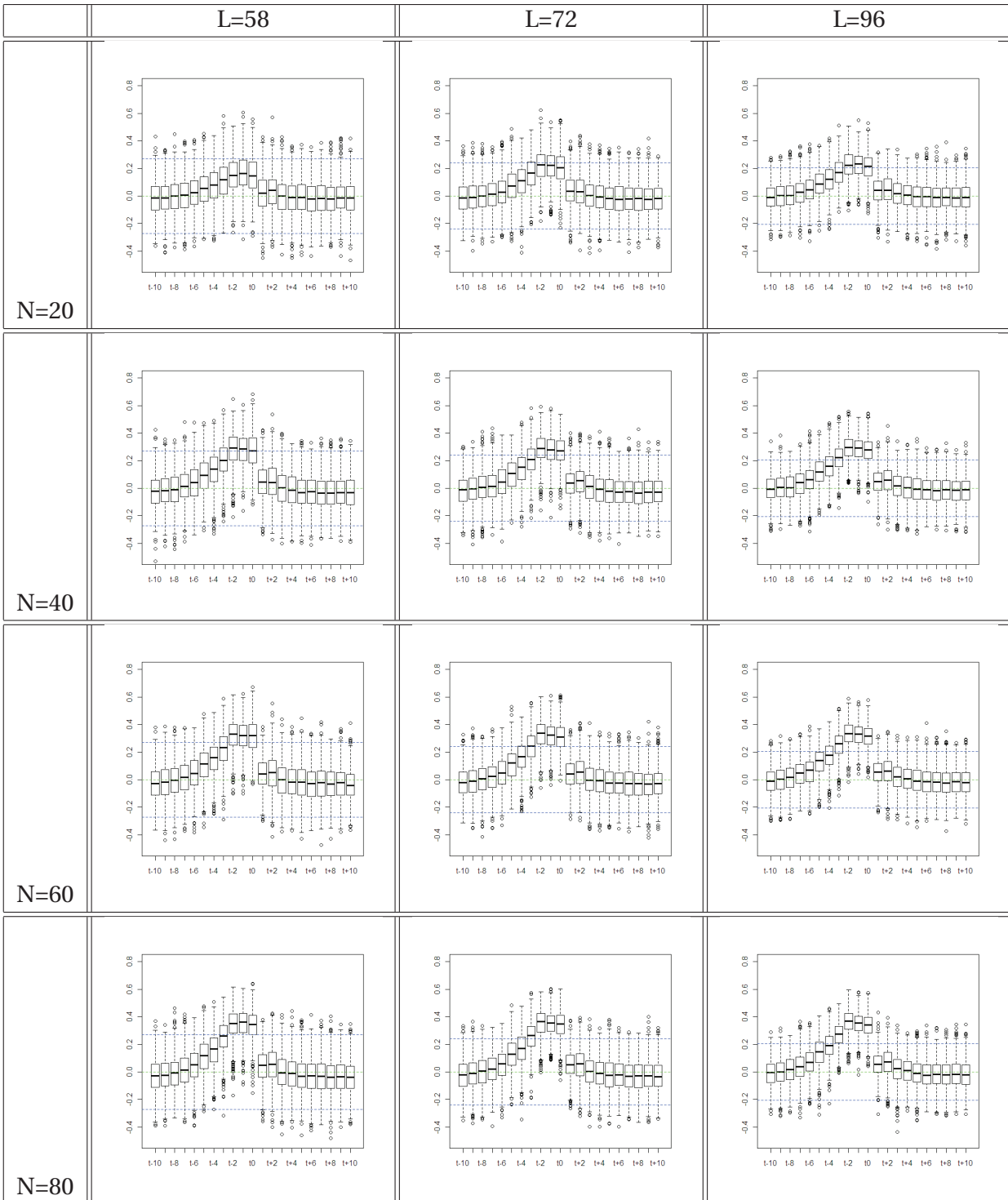


Tabella 4.10.: Cross-correlazioni con numerosità pari a 20, 40, 60 e 80 casi – Istar 2.



## 5. Conclusioni e raccomandazioni

### 5.1. Conclusioni

Riprendendo quello che è l'obiettivo iniziale dello studio, ovvero quello di identificare se esiste e di che tipo è la relazione tra la resistenza sviluppata da *Pseudomonas aeruginosa* e le dosi di antibiotico assegnate ai pazienti ricoverati nei due reparti di rianimazione, possiamo affermare che solo grazie alle simulazioni, siamo riusciti a portarlo a termine.

Costruire diversi scenari, con lunghezze temporali e numero di pazienti diversi, ci ha permesso di capire quanto queste due quantità incidano sul conseguimento di risultati più o meno soddisfacenti.

Nel nostro lavoro è risultato essere fondamentale il “numero di degenti”, poiché, solo con numerosità più alte rispetto a quelle reali si è manifestata la relazione tra dosi di medicinale e resistenza batterica: in particolare, nella maggior parte degli scenari, sono emerse le dosi al tempo  $t$ ,  $t-1$  e  $t-2$  come quelle maggiormente collegate alla resistenza al tempo  $t$ . Tuttavia, per avere i risultati sperati basta avere numerosità sull'ordine di qualche decina e, non necessariamente arrivare ad un centinaio come nelle nostre simulazioni.

La “lunghezza delle serie” incide meno rispetto all'altra grandezza: infatti, nonostante l'ampliamento delle serie temporali rappresenti una facilitazione nell'individuazione della relazione presente tra le nostre variabili di interesse, questo passa in secondo piano se confrontato con l'incidenza della quantità “numero di degenti”.

Infine, ci teniamo a ricordare che il lavorare sulla lunghezza delle serie e sul numero dei casi, non obbligatoriamente corrisponde all'evidenziare la struttura dinamica che si ipotizza: certo, l'aumento delle grandezze rappresenta una garanzia in quanto vi è un corrispettivo aumento della potenza dei test e quindi un minor rischio di commettere errori, tuttavia, la struttura dinamica si riscontra se questa è realmente presente nei dati, nel caso non ci fosse, non si troverebbe riscontro nemmeno nei dati simulati.

## 5.2. Raccomandazioni

Alla luce delle conclusioni sviluppate al paragrafo precedente, il suggerimento che ci sentiamo di dare ai Medici è quello di continuare a lavorare su due fronti: da un lato proseguire nella raccolta delle informazioni per estendere la serie temporale a loro disposizione, dall'altra verificare la compatibilità di alcuni reparti per poterli accorpate in eventuali analisi future.

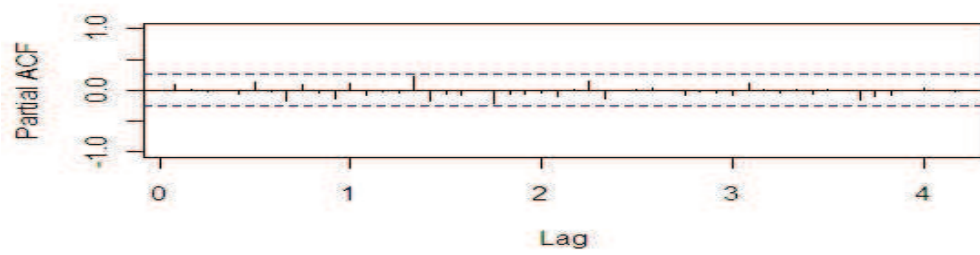
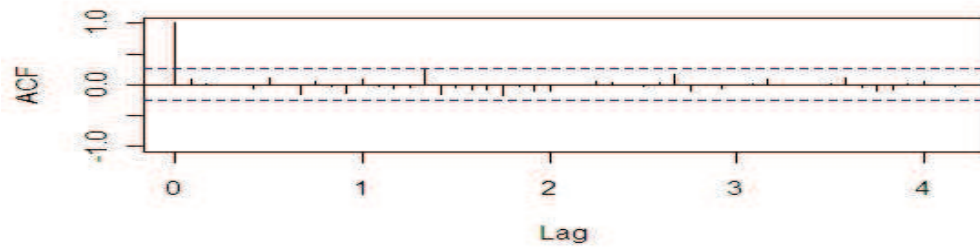
In ogni caso, lavorare sulla numerosità e, quindi, ampliare il campione da analizzare, permetterà di ottenere risultati in più breve tempo e offrirà a chi di competenza la possibilità di attuare delle politiche di regolamentazione della resistenza batterica più velocemente.

Tuttavia, se l'accorpamento di alcuni reparti non fosse attuabile, consigliamo di continuare a raccogliere informazioni per più di un anno ancora, soprattutto se si vogliono ottenere dei buoni risultati anche per il reparto Istar 2.

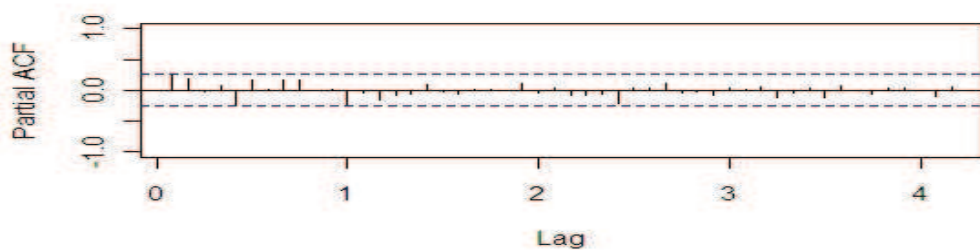
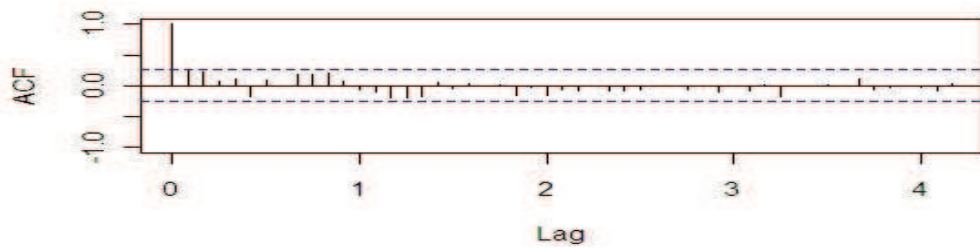
# A. Appendice 1

Nelle pagine seguenti si possono trovare alcuni dei grafici che non sono stati inseriti nel Capitolo 2 ed il listato dei programmi eseguiti per sviluppare le analisi esplorative.

## A.1. Figure

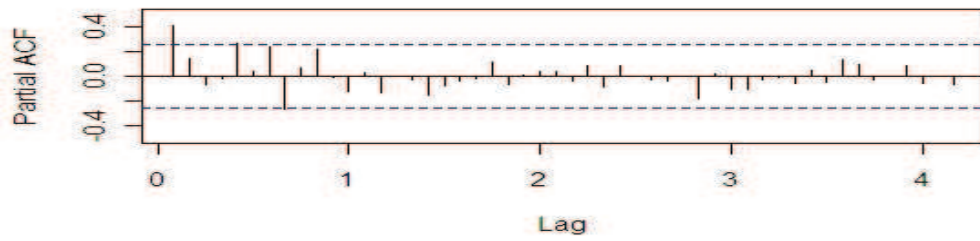
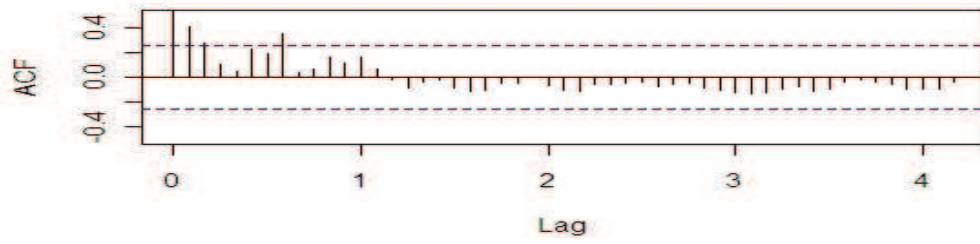


(a) Istar1

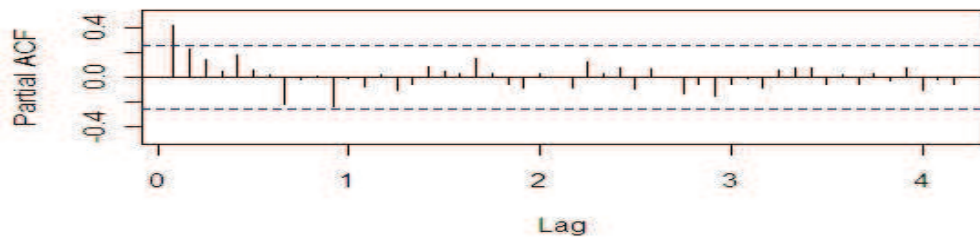
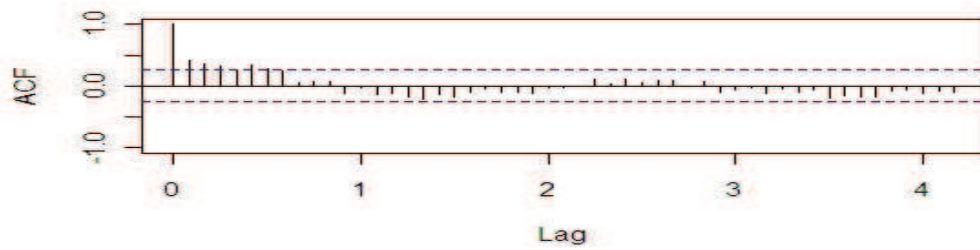


(b) Istar2

Figura A.1.: Analisi esplorative: ACF e PACF per le resistenze batteriche.



(a) Istar1



(b) Istar2

Figura A.2.: Analisi esplorative: ACF e PACF per la dose di antibiotico.



## A.2. Listato R

Costruzione delle serie temporali:

```
ts_RES_R1MEM <- ts(R1MEM$RES, start=2007, frequency=12)
ts_RES_R2MEM <- ts(R2MEM$RES, start=2007, frequency=12)
r1 = ts_RES_R1MEM
r2 = ts_RES_R2MEM
ts_DDD_R1MEM <- ts(R1MEM$DDD, start=2007, frequency=12)
ts_DDD_R2MEM <- ts(R2MEM$DDD, start=2007, frequency=12)
d1 = ts_DDD_R1MEM
d2 = ts_DDD_R2MEM
```

Analisi descrittive:

```
par(mfrow=c(2,1))
plot(d1, ylim=c(0,200), main="ISTAR1")
plot(d2, ylim=c(0,200), main="ISTAR2")
plot(r1, ylim=c(0,100), main="ISTAR1")
plot(r2, ylim=c(0,100), main="ISTAR2")
acf(d1,50, ylim=c(-0.5,0.5), main="ACF DDD ISTAR1")
pacf(d1,50, ylim=c(-0.5,0.5), main="PACF DDD ISTAR1")
acf(d2,50, ylim=c(-1,1), main="ACF DDD ISTAR2")
pacf(d2,50, ylim=c(-1,1), main="PACF DDD ISTAR2")
acf(r1,50, ylim=c(-1,1), main="ACF %RESISTENTI ISTAR1")
pacf(r1,50, ylim=c(-1,1), main="PACF %RESISTENTI ISTAR1")
acf(r2,50, ylim=c(-1,1), main="ACF %RESISTENTI ISTAR2")
pacf(r2,50, ylim=c(-1,1), main="PACF %RESISTENTI ISTAR2")
ts.plot(d1,r1, ylim=c(0,200))
ts.plot(d2,r2, ylim=c(0,200))
d1.ar <- arima(d1, order=c(1,0,0)) d1.ar
res.input1 <- residuals(d1.ar) - mean(residuals(d1.ar))
par(mfrow=c(2,1))
acf(res.input1)
pacf(res.input1)
r1.centrata <- r1-mean(r1)
res.output1 <- r1.centrata-0.4044*(lag(r1.centrata))
par(mfrow=c(2,1))
acf(res.output1)
```

```

pacf(res.output1)
par(mfrow=c(1,1))
ccf(res.input1, res.output1, lag.max=24)
d2.ar <- arima(d2, order=c(1,0,0))
d2.ar res.input2 <- residuals(d2.ar) - mean(residuals(d2.ar))
par(mfrow=c(2,1))
acf(res.input2)
pacf(res.input2)
r2.centrata <- r2-mean(r2)
res.output2 <- r2.centrata - 0.4360*(lag(r2.centrata))
par(mfrow=c(2,1))
acf(res.output2)
pacf(res.output2)
par(mfrow=c(1,1))
ccf(res.input2, res.output2, lag.max=24)

n = length(d1)
my.R1MEM <- data.frame(RES_t=R1MEM$RES[-c(1:3)],
  DDD_t=R1MEM$DDD[-c(1:3)], DDD_t1=R1MEM$DDD[-c(1,2,n)],
  DDD_t2=R1MEM$DDD[-c(1,n-1,n)], DDD_t3=R1MEM$DDD[-c(n-2,n-1,n)],
  PAZIENTI_RES_t=R1MEM$PAZIENTI_RES[-c(1:3)],
  TOTALI_t=R1MEM$TOTALI[-c(1:3)])
lm.1 = lm(RES_t~DDD_t+DDD_t1+DDD_t2+DDD_t3, data=my.R1MEM)
lm.2 = lm(RES_t~DDD_t+DDD_t1+DDD_t2, data=my.R1MEM)
lm.3 = lm(RES_t~DDD_t+DDD_t2, data=my.R1MEM)
lm.4 = lm(RES_t~DDD_t2, data=my.R1MEM)
m = length(d2)
my.R2MEM <- data.frame(RES2_t=R2MEM$RES[-c(1:3)],
  DDD2_t=R2MEM$DDD[-c(1:3)], DDD2_t1=R2MEM$DDD[-c(1,2,n)],
  DDD2_t2=R2MEM$DDD[-c(1,n-1,n)], DDD2_t3=R2MEM$DDD[-c(n-2,n-1,n)],
  PAZIENTI2_RES_t=R2MEM$PAZIENTI_RES[-c(1:3)],
  TOTALI2_t=R2MEM$TOTALI[-c(1:3)])
lm.5 = lm(RES2_t~DDD2_t+DDD2_t1+DDD2_t2+DDD2_t3, data=my.R2MEM)
lm.6 = lm(RES2_t~DDD2_t+DDD2_t1+DDD2_t2, data=my.R2MEM)
lm.7 = lm(RES2_t~DDD2_t+DDD2_t1, data=my.R2MEM)
lm.8 = lm(RES2_t~DDD2_t2, data=my.R2MEM)

```



## B. Appendice 2

Di seguito si riporta il listato dei programmi utilizzati nelle simulazioni di cui si è parlato nel Capitolo 3.

SIMULAZIONE DDD (La lunghezza della serie simulata è 58 mensilità.)

```
d1.ar <- ar(d1)
d1.ar$order
d1.fun <- function(d1)
{
ar.fit <- ar(d1, order.max=10)
c(ar.fit$order, mean(d1), d1)
}
d1.model <- list(order = c(d1.ar$order, 0, 0), ar = d1.ar$ar)
d1.res <- d1.ar$resid[!is.na(d1.ar$resid)]
d2.res <- d1.res - mean(d1.res)
d1.sim <- function(res, n.sim, ran.args)
{
rg1 <- function(n, res) sample(res, n, replace = TRUE)
ts.orig <- ran.args$ts
ts.mod <- ran.args$model
mean(ts.orig) + ts(arima.sim(model = ts.mod, n = n.sim,
rand.gen = rg1, res = as.vector(res)))
}
ddd1.sim <- matrix(0, nrow=999, ncol=58)
resist1.sim <- matrix(0, nrow=999, ncol=53)
round(apply(ddd1.sim,1,function(x) sum(x<0))/58, 2)*100
for(R in 1:99)
{
ddd1.sim[R,] = d1.sim(d2.res, 58,
ran.args=list(ts = (d1), model = d1.model))
```

## B. Appendice 2

```
}  
ddd1.sim[ddd1.sim < 0] = 0  
matrice1.sim <- cbind(ddd1.sim, resist1.sim)
```

SIMULAZIONE RESISTENZE (La lunghezza della serie simulata è 58 mensilità e 10 è il valore dell'indice binomiale.)

```
for(R in 1:999)  
{  
my.data<-data.frame(d1.I=rep(1,53),  
d1.t=ddd1.sim[R, -c(1:5)],  
d1.t1 = ddd1.sim[R,-c(1:4,58)],  
d1.t2 =ddd1.sim[R,-c(1:3,57:58)],  
d1.t3 =ddd1.sim[R,-c(1:2,56:58)],  
d1.t4 =ddd1.sim[R,-c(1:1,55:58)],  
d1.t5 =ddd1.sim[R,-c(54:58)])  
eta.sim <- data.matrix(my.data) %*% matrix(c(-1.4, 0.006,  
0.005, 0.004, 0.002, 0.0008, 0.0006), ncol=1)  
p.sim <- exp(eta.sim)/(1+exp(eta.sim))  
for(i in 1:53)  
resist1.sim[R, i] <- rbinom(1, size=10, p=p.sim[i])/10*100  
}
```

## C. Appendice 3

Di seguito si riporta il listato dei programmi utilizzati nelle analisi dei dati simulati di cui si è parlato nel Capitolo 4.

Costruzione del modello (3.1):

```
coef.sim = matrix(0, nrow=999, ncol=7)
for(R in 1:999)
{
dati_sim = round(matrice1.sim[R,59:111]/100*10,0)
dati_sim = cbind(dati_sim, matrice1.sim[R,6:58],
matrice1.sim[R,5:57], matrice1.sim[R,4:56],
matrice1.sim[R,3:55], matrice1.sim[R,2:54],
matrice1.sim[R,1:53])
dimnames(dati_sim)[[2]] = list("res","ddd0","ddd1","ddd2",
"ddd3","ddd4","ddd5")
mA <- glm(cbind(res,10-res)~ddd0+ddd1+ddd2+ddd3+ddd4+ddd5,
data=as.data.frame(dati_sim), family=binomial("logit"))
coef.sim[R,] <- coef(mA)
}
par(mfrow=c(2,4))
apply(coef.sim,2,boxplot)
```

Rappresentazione tramite un diagramma a bolle dell'output della funzione arimaId:

```
library(labstatR)
order.DDD.1.sim = matrix(NA, nrow=99, ncol=3)
for(R in 1:99)
{
tmp = print(arimaId(ddd1.sim[R,], c(3,0,1)))
order.DDD.1.sim[R,] <- unlist(tmp[tmp$rankAIC==1,1:3])
}
```

### C. Appendice 3

```
par(mfrow=c(1,3))
apply(order.DDD.1.sim, 2, function(x) plot(table(x)))
plot(order.DDD.1.sim[,1] ~ order.DDD.1.sim[,3])
AR <- factor(order.DDD.1.sim[,1])
MA <- factor(order.DDD.1.sim[,3])
table(AR, MA)
bubbleplot(table(AR, MA), main="Bubble plot scenario A")
```

#### Costruzione delle CCF:

```
ddd1.sim.for <- ddd1.sim[,6:58]
ddd.centrata <- matrix(NA, nrow=999, ncol=53)
coef.ar1 <- matrix(NA, nrow=999, ncol=1)
res.input <- matrix(NA, nrow=999, ncol=53)
res.centrata <- matrix(NA, nrow=999, ncol=53)
res.output <- matrix(NA, nrow=999, ncol=53)
coef.ccf.A <- matrix(NA, nrow=999, ncol=21)
for(R in 1:999)
{
ddd.centrata[R,] <- ddd1.sim.for[R,] - mean(ddd1.sim.for[R,])
tmp = arima(ddd.centrata[R,], order=c(1,0,0), include.mean=F)
coef.ar1[R,] <- unlist(tmp[1])
res.input[R,] <- unlist(tmp[8])
res.centrata[R,] <- resist1.sim[R,]-mean(resist1.sim[R,])
res.output[R,]<-res.centrata[R,]-coef.ar1[R,]*lag(res.centrata[R,])
tmp1 = ccf(res.input[R,], res.output[R,], lag.max=10)
coef.ccf.A[R,] <- unlist(tmp1$acf)
}
coef.ccf dim(coef.ccf)
dimnames(coef.ccf.A)[[2]] <- list("t-10", "t-9", "t-8", "t-7", "t-6",
"t-5", "t-4", "t-3", "t-2", "t-1", "t0", "t+1", "t+2", "t+3", "t+4",
"t+5", "t+6", "t+7", "t+8", "t+9", "t+10")
l58_n10 <- boxplot(coef.ccf.A, ylim=c(-0.5,+0.8))
coef.mxA = matrix(0, nrow=999, ncol=13)
pval.mxA = matrix(0, nrow=999, ncol=13)
resid.mxA = matrix(0, nrow=999, ncol=48)
for(R in 1:999)
{
```

```

res_7 = round(matrice1.sim[R,64:111]/100*10,0)
res_6 = round(matrice1.sim[R,63:110]/100*10,0)
dati_mx = cbind(res_7, res_6, matrice1.sim[R,11:58],
matrice1.sim[R,10:57], matrice1.sim[R,9:56],
matrice1.sim[R,8:55], matrice1.sim[R,7:54],
matrice1.sim[R,6:53], matrice1.sim[R,5:52],
matrice1.sim[R,4:51], matrice1.sim[R,3:50],
matrice1.sim[R,2:49], matrice1.sim[R,1:48])
dimnames(dati_mx)[[2]] = list("res", "res_m1", "ddd0",
"ddd_m1","ddd_m2", "ddd_m3", "ddd_m4", "ddd_m5",
"ddd_m6", "ddd_m7", "ddd_m8", "ddd_m9", "ddd_m10")
mxA <- lm(res ~ res_m1 + ddd0 + ddd_m1 + ddd_m2
+ ddd_m3 + ddd_m4 + ddd_m5 + ddd_m6 + ddd_m7 + ddd_m8
+ ddd_m9 + ddd_m10, data=as.data.frame(dati_mx))
coef.mxA[R,] <- coef(mxA)
pval.mxA[R,] <- summary(mxA)$coef[,4]
resid.mxA[R,] <- residuals(mxA)
}

```

Costruzione del modello a funzione di trasferimento (4.1):

```

for(R in 1:999)
{
m_etaA <- lm(res ~ ddd0 + ddd_m1 + ddd_m2 + ddd_m3 + ddd_m4 +
ddd_m5 + ddd_m6 + ddd_m7 + ddd_m8 + ddd_m9 + ddd_m10,
data=as.data.frame(dati_mx))
residui.etaA[R,] <- residuals(m_etaA)
tmp2 = ccf(resid.mxA[R,], residui.etaA[R,], lag.max=10)
coef.ccf.tf[R,] <- unlist(tmp2$acf)
}

```

Test di Wald:

```

apply(2*pmin(pnorm(coef.mxA/matrix(apply(coef.mxA,2,sd),nrow=999,
ncol=13, byrow=T)), 1-pnorm(coef.mxA/matrix(apply(coef.mxA,2,sd),
nrow=999, ncol=13, byrow=T))),2, function(x) sum(x<0.05)/1000 )

```

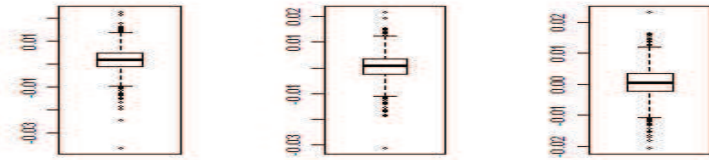




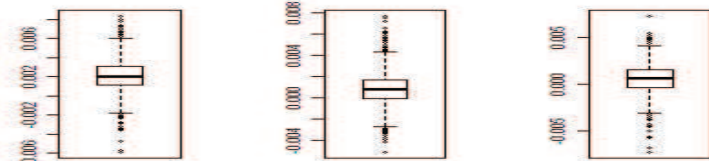
## D. Appendice 4

Nelle pagine seguenti si riportano i grafici contenenti i boxplot per la verifica della consistenza dei coefficienti introdotti nel modello logit, distinti per scenario e reparto, di cui si è parlato nel Capitolo 4.

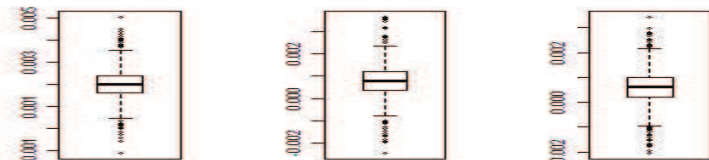
D. Appendice 4



(a) A

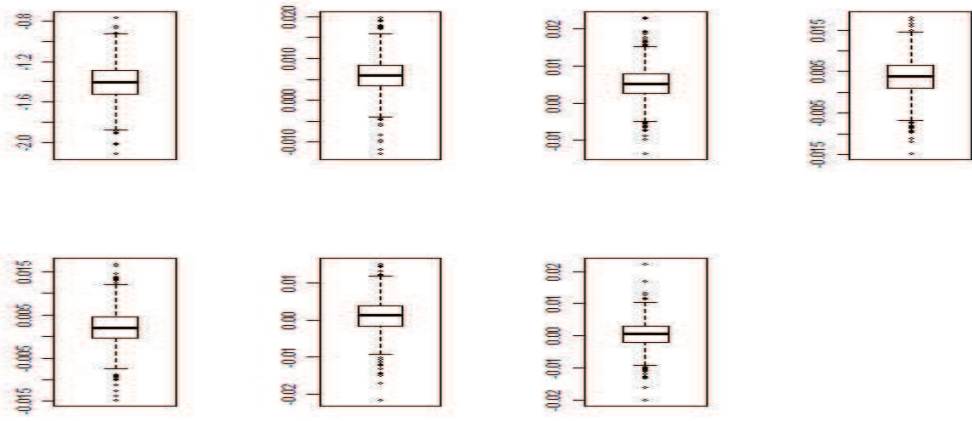


(b) B

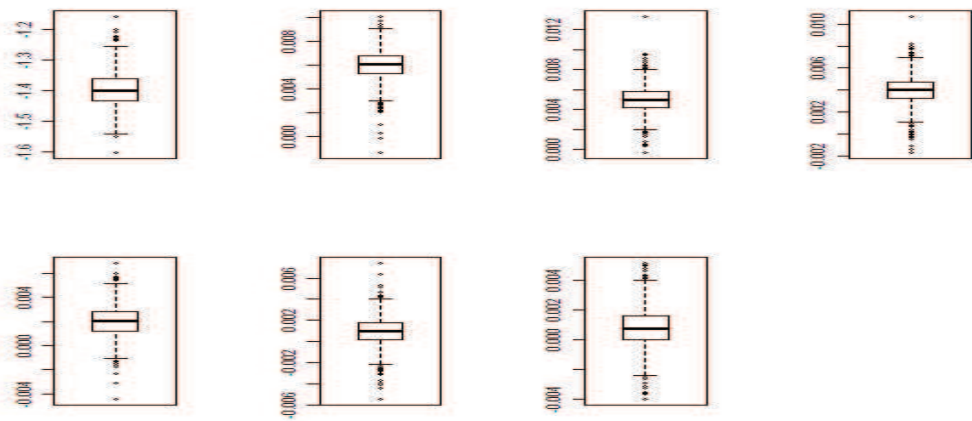


(c) C

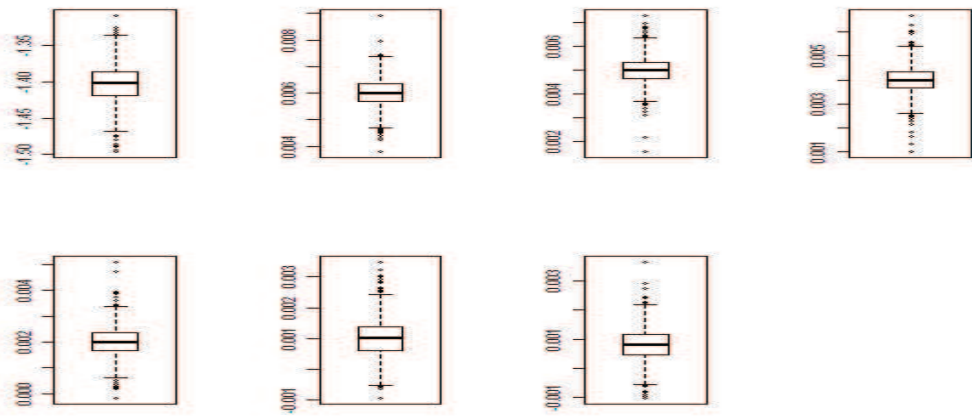
Figura D.1.: Scenario 1 – Istar 1.



(a) D



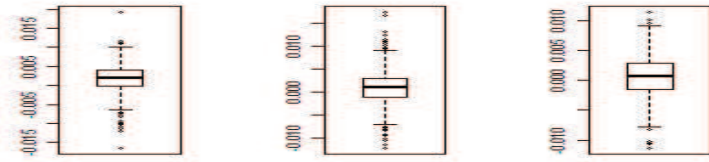
(b) E



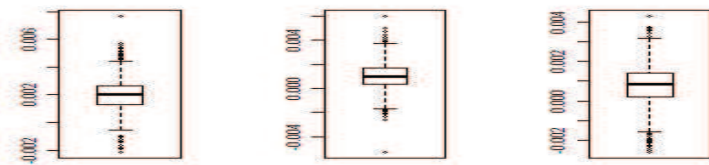
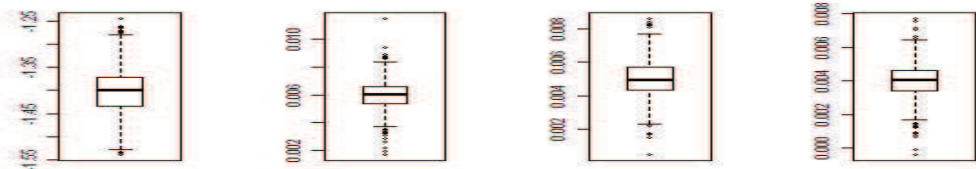
(c) F

Figura D.2.: Scenario 2 – Istar 1.

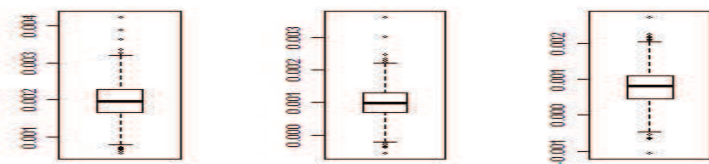
D. Appendice 4



(a) G

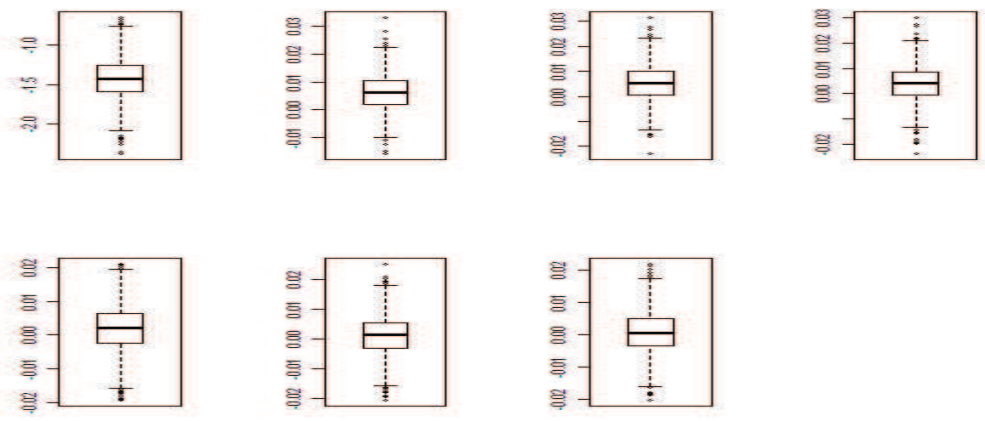


(b) H

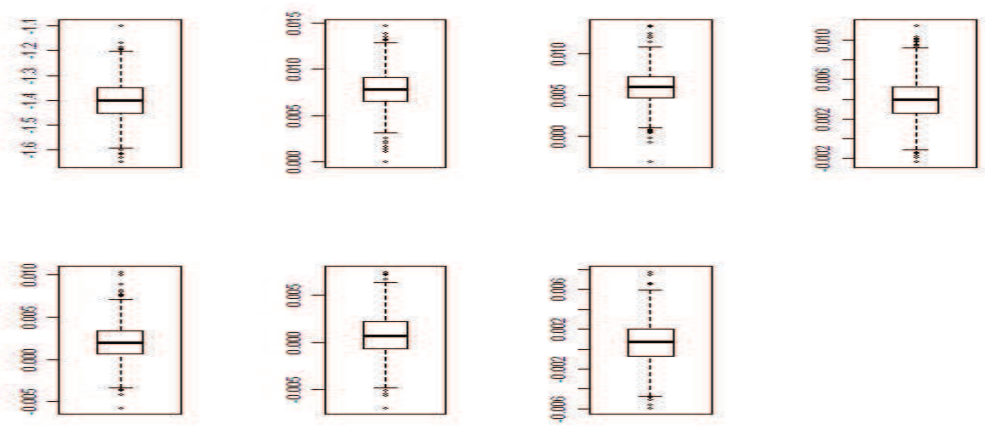


(c) I

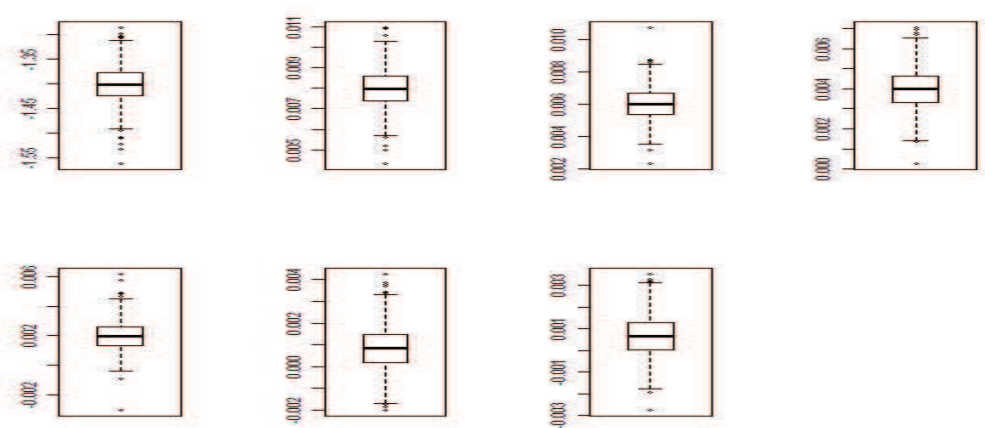
Figura D.3.: Scenario 3 – Istar 1.



(a) A



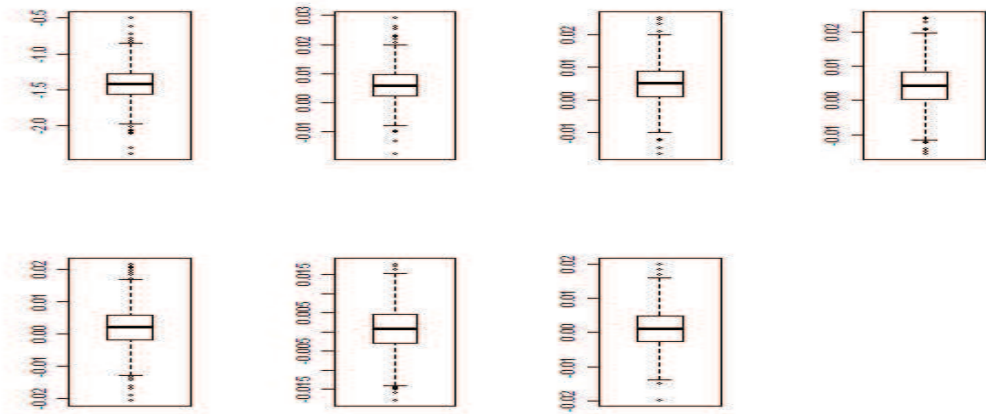
(b) B



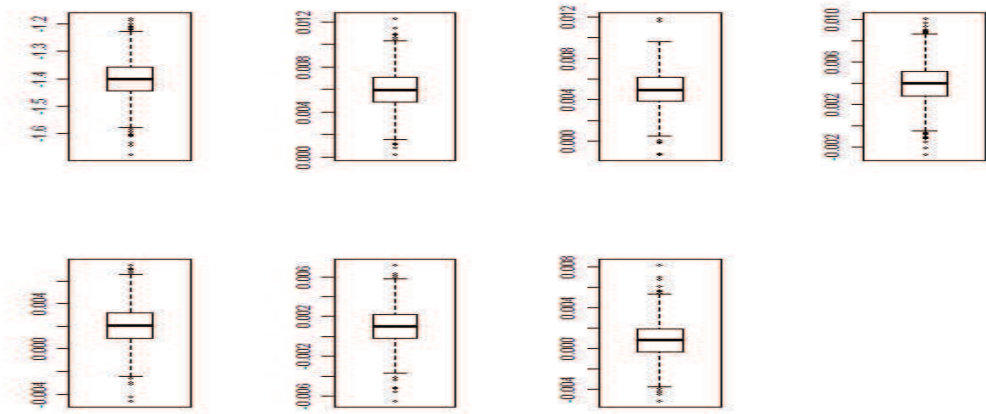
(c) C

Figura D.4.: Scenario 1 – Istar 2.

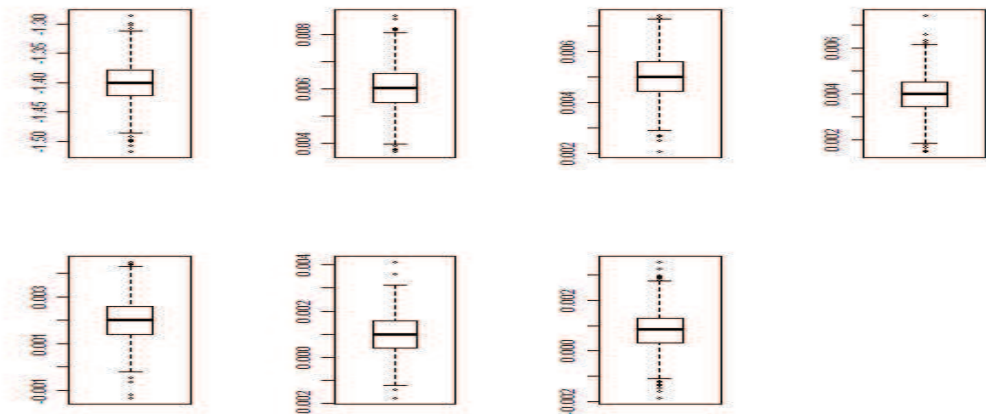
D. Appendice 4



(a) D

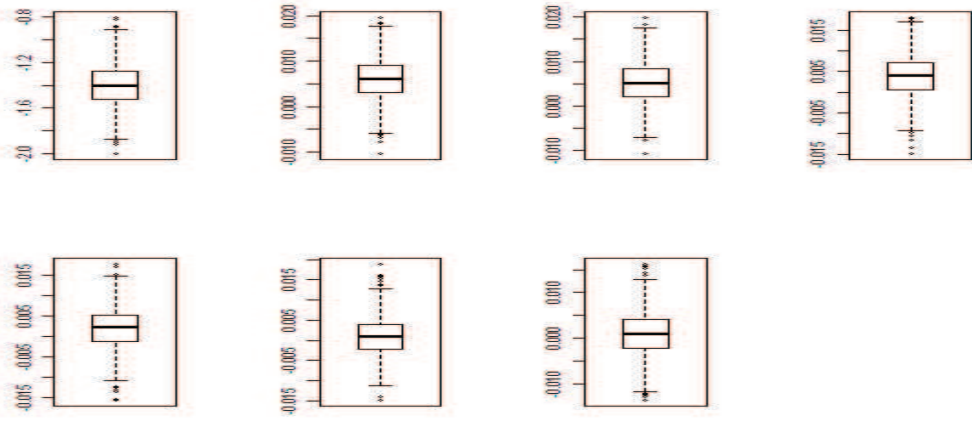


(b) E

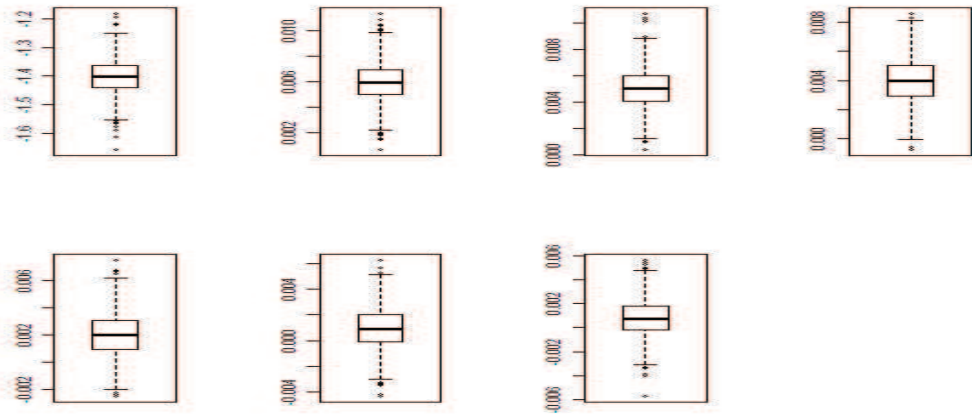


(c) F

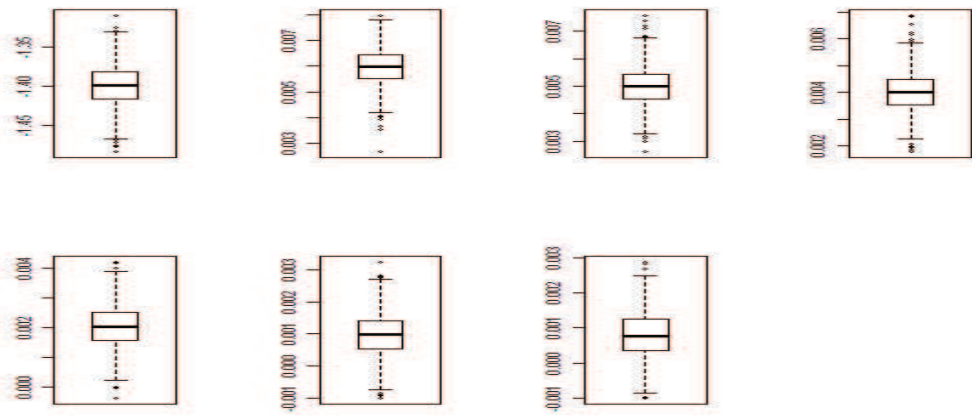
Figura D.5.: Scenario 2 – Istar 2.



(a) G



(b) H



(c) I

Figura D.6.: Scenario 3 – Istar 2.





# Bibliografia

- [1] DAVISON A. C., HINKLEY D.V. (1997), *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge
- [2] ERDELJIC V. ET AL. (2011), “Distributed lags time series analysis versus linear correlation analysis (Pearson’s r) in identifying the relationship between antipseudomonal antibiotic consumption and the susceptibility of *Pseudomonas aeruginosa* isolates in a single Intensive Care Unit of a tertiary hospital.”, *International Journal of Antimicrobial Agents* 37 (2011) 467–471.
- [3] FUJIMURA S. ET AL. (2007), “Relationship between the usage of carbapenem antibiotics and the incidence of imipenem-resistant *Pseudomonas aeruginosa*.”, *Journal of Infection and Chemotherapy* 2007; 13:147–150.
- [4] KOLLEF M. H. (2005), “Bench-to-bedside review: Antimicrobial utilization strategies aimed at preventing the emergence of bacterial resistance in the intensive care unit.” Disponibile al sito <http://ccforum.com/content/9/5/459>.
- [5] LOPEZ-LOZANO J. ET AL. (2000), “Modelling and forecasting antimicrobial resistance and its dynamic relationship to antimicrobial use: a time series analysis.”, *International Journal of Antimicrobial Agents* 14 2000; 21–31.
- [6] MAHAMAT A. ET AL. (2005), “Evolution of fluoroquinolone resistance among *Escherichia coli* urinary tract isolates from a French university hospital: application of the dynamic regression model”, *Clinical Microbiology and Infection - Journal Information* 2005; 11: 301–306
- [7] MURRAY P. R. ET AL. (2010), *Microbiologia Medica*, Elsevier, Milano
- [8] OKAMOTO K, GOTOH N, NISHINO T. (2002), “Alterations of susceptibility of *Pseudomonas aeruginosa* by overproduction of multidrug efflux systems, MexAB-OprM, MexCD-OprJ, and MexXY/OprM to carbapenems: substrate specificities of the efflux systems.”, *Journal of Infection and Chemotherapy* 2002; 8:371–3.

## *Bibliografia*

- [9] PANKRATZ A. (1991), *Forecasting with Dynamic Regression Models*, Wiley-Interscience Publication, New York
- [10] ROGUES A. M. ET AL. (2007), "Relationship Between Rates of Antimicrobial Consumption and the Incidence of Antimicrobial Resistance in *Staphylococcus aureus* and *Pseudomonas aeruginosa* Isolates From 47 French Hospitals.", *Infection Control and Hospital Epidemiology*, december 2007, vol. 28, no. 12.
- [11] SHUMWAY R. H., STOFFER D. S. (2006), *Time Serie Analysis and its Applications. With R Examples*, Springer, New York
- [12] WEI W. (2006), *Time Serie Analysis. Univariate and Multivariate Methods*, Pearson Education, USA