

# **UNIVERSITA' DEGLI STUDI DI PADOVA**

**CORSO DI LAUREA IN STATISTICA E GESTIONE DELLE IMPRESE**



**PREVISIONE DEI COSTI DI UN'AZIENDA SANITARIA**

*Relatore: Prof. Omar Paccagnella*

*Dipartimento di Scienze Statistiche*

*Laureando: Nicola Facchin*

*Matricola n° 573577*

**Anno Accademico 2012/2013**



# INDICE

<b>CAPITOLO 1 – PRESENTAZIONE .....</b>	<b>6</b>
1.1 Introduzione .....	6
1.2 L'azienda Sanitaria .....	7
<b>CAPITOLO 2 - METODOLOGIE UTILIZZATE .....</b>	<b>9</b>
2.1 Serie Storiche .....	9
2.2 Il Metodo Holt-Winters .....	11
2.3 La Procedura Box-Jenkins .....	13
2.4 Indici di Valutazione .....	18
<b>CAPITOLO 3 – APPLICAZIONE DEI MODELLI .....</b>	<b>19</b>
3.1 I Dati.....	19
3.2 Analisi Serie Storiche.....	22
3.3 Previsioni con metodo Holt-Winters.....	26
3.3.1 Previsioni con Serie mensile totale Pazienti .....	27
3.3.2 Previsioni con Serie mensile dei Pazienti divisi per Azienda .....	29
3.4 Applicazione modelli SARIMA .....	35
3.4.1 Serie Totale Pazienti .....	35
3.4.2 Serie divise per Azienda .....	43
<b>CAPITOLO 4 - CONSIDERAZIONI FINALI .....</b>	<b>57</b>
<b>BIBLIOGRAFIA.....</b>	<b>60</b>
<b>PACCHETTI STATISTICI.....</b>	<b>60</b>

## INDICE DELLE TABELLE

<i>Tabella 1 – Divisione Spese Totali</i> .....	20
<i>Tabella 2 - Divisione Medici</i> .....	20
<i>Tabella 3 - Divisione Pazienti</i> .....	21
<i>Tabella 4 - H-W Serie mensile totale Pazienti</i> .....	27
<i>Tabella 5 - H-W Serie mensile Azienda 1</i> .....	29
<i>Tabella 6 - H-W Serie mensile Azienda 2</i> .....	30
<i>Tabella 7 - H-W Serie mensile Azienda 3</i> .....	31
<i>Tabella 8 - H-W Serie mensile Azienda 4</i> .....	32
<i>Tabella 9 - H-W Serie mensile Azienda 5</i> .....	33
<i>Tabella 10 - H-W Serie mensile Azienda 6</i> .....	34
<i>Tabella 11 - AIC e SC Serie totale Pazienti</i> .....	41
<i>Tabella 12 - MQE e indice di Theil Serie totale Pazienti</i> .....	43
<i>Tabella 13 - Indici Previsioni SARIMA(0,1,1)(1,0,0)</i> .....	45
<i>Tabella 14 - Indici previsioni SARIMA(0,1,1)(0,0,1)</i> .....	47
<i>Tabella 15 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 3</i> .....	49
<i>Tabella 16 - Indici previsioni SARIMA(1,1,1)(1,0,0) Azienda 4</i> .....	51
<i>Tabella 17 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 5</i> .....	53
<i>Tabella 18 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 6</i> .....	55
<i>Tabella 19 – Riepilogo previsioni serie totale Pazienti</i> .....	57
<i>Tabella 20 - Riepilogo previsioni Azienda 1</i> .....	57
<i>Tabella 21 - Riepilogo previsioni Azienda 2</i> .....	58
<i>Tabella 22 - Riepilogo previsioni Azienda 3</i> .....	58
<i>Tabella 23 - Riepilogo previsioni Azienda 4</i> .....	58
<i>Tabella 24 - Riepilogo previsioni Azienda 5</i> .....	58
<i>Tabella 25 - Riepilogo previsioni Azienda 6</i> .....	59

## INDICE DEI GRAFICI

<i>Grafico 1 - Divisione Spese Totali</i> .....	20
<i>Grafico 2- Divisione Medici</i> .....	20
<i>Grafico 3 - Divisione Pazienti</i> .....	21
<i>Grafico 4 - Serie Mensile Totale Pazienti</i> .....	22
<i>Grafico 5 - Scomposizione Serie Totale Pazienti</i> .....	22
<i>Grafico 6 - Scomposizione Serie Mensile Azienda 1</i> .....	23
<i>Grafico 7 - Scomposizione Serie Mensile Azienda 2</i> .....	24
<i>Grafico 8 - Scomposizione Serie Mensile Azienda 3</i> .....	24
<i>Grafico 9 - Scomposizione Serie Mensile Azienda 4</i> .....	25
<i>Grafico 10 - Scomposizione Serie Mensile Azienda 5</i> .....	25
<i>Grafico 11 - Scomposizione Serie Mensile Azienda 6</i> .....	26
<i>Grafico 12 - H-W Additivo Serie mensile totale Pazienti</i> .....	27

<i>Grafico 13 - H-W Moltiplicativo Serie mensile totale Pazienti .....</i>	28
<i>Grafico 14 - H-W Additivo Serie mensile Azienda 1.....</i>	29
<i>Grafico 15 - H-W Moltiplicativo Serie mensile Azienda 1.....</i>	29
<i>Grafico 16 - H-W Additivo Serie mensile Azienda 2.....</i>	30
<i>Grafico 17 - H-W Moltiplicativo Serie mensile Azienda 2.....</i>	30
<i>Grafico 18 - H-W Additivo Serie mensile Azienda 3.....</i>	31
<i>Grafico 19 - H-W Moltiplicativo Serie mensile Azienda 3.....</i>	31
<i>Grafico 20 - H-W Additivo Serie mensile Azienda 4.....</i>	32
<i>Grafico 21 - H-W Moltiplicativo Serie mensile Azienda 4.....</i>	32
<i>Grafico 22 - H-W Additivo Serie mensile Azienda 5.....</i>	33
<i>Grafico 23 - H-W Moltiplicativo Serie mensile Azienda 5.....</i>	33
<i>Grafico 24 - H-W Additivo Serie mensile Azienda 6.....</i>	34
<i>Grafico 25 - H-W Moltiplicativo Serie mensile Azienda 6.....</i>	34
<i>Grafico 26 - Correlogramma Serie totale Pazienti .....</i>	36
<i>Grafico 27 - Serie totale Pazienti differenziata .....</i>	36
<i>Grafico 28 - Correlogramma Serie totale Pazienti differenziata.....</i>	37
<i>Grafico 29 - Serie totale Pazienti reale e stimata.....</i>	38
<i>Grafico 30 - Correlogramma residui ARIMA(1,1,1) .....</i>	39
<i>Grafico 31 - Serie totale Pazienti reale e stimata con SARIMA(1,1,1)(1,0,0) .....</i>	40
<i>Grafico 32 - Correlogramma residui SARIMA(1,1,1)(1,0,0).....</i>	41
<i>Grafico 33 - Previsioni Serie totale Pazienti ARIMA(1,1,1).....</i>	42
<i>Grafico 34 - Previsioni Serie totale Pazienti SARIMA(1,1,1)(1,0,0) .....</i>	42
<i>Grafico 35 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 1 .....</i>	44
<i>Grafico 36 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 1 .....</i>	45
<i>Grafico 37 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 1.....</i>	45
<i>Grafico 38 - Correlogramma residui SARIMA(0,1,1)(0,0,1) Azienda 2 .....</i>	46
<i>Grafico 39 - Serie reale e stimata SARIMA(0,1,1)(0,0,1) Azienda 2 .....</i>	47
<i>Grafico 40 - Previsioni SARIMA(0,1,1)(0,0,1) .....</i>	47
<i>Grafico 41 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 3 .....</i>	48
<i>Grafico 42 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 3 .....</i>	49
<i>Grafico 43 - Previsioni SARIMA(0,1,1)(1,0,0) .....</i>	49
<i>Grafico 44 - Correlogramma residui SARIMA(1,1,1)(1,0,0) Azienda 4 .....</i>	50
<i>Grafico 45 - Serie reale e stimata SARIMA(1,1,1)(1,0,0) Azienda 4 .....</i>	51
<i>Grafico 46 - Previsioni SARIMA(1,1,1)(1,0,0) Azienda 4.....</i>	51
<i>Grafico 47 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 5 .....</i>	52
<i>Grafico 48 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda5 .....</i>	53
<i>Grafico 49 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 5.....</i>	53
<i>Grafico 50 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 6 .....</i>	54
<i>Grafico 51 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 6 .....</i>	55
<i>Grafico 52 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 6.....</i>	55



# Capitolo 1

## PRESENTAZIONE

### 1.1 INTRODUZIONE

In questo elaborato effettueremo un'analisi di dati relativi a costi dei medicinali prescritti da un campione di medici di base dell'Azienda Sanitaria del Friuli mediante l'utilizzo di serie storiche. Il campione utilizzato non è rappresentativo dell'intera popolazione, in quanto i dati provengono da un gruppo composto da 107 medici i quali hanno aderito al progetto, non sarà quindi effettuata nessuna analisi inferenziale.

L'azienda Sanitaria è a tutti gli effetti un'azienda, perciò, a fini amministrativi e direzionali, può essere utile prevedere quali saranno i costi futuri. In una prima analisi considereremo la serie storica complessiva di tutti i pazienti, in seguito utilizzeremo i dati divisi per azienda sanitaria. Per la modellazione della serie e le previsioni ci avvarremo dell'approccio Box-Jenkins, utilizzando i processi ARIMA stagionali, e del metodo Holt-Winters stagionale, additivo e moltiplicativo. Lo scopo di questa analisi è quindi mostrare come sia possibile ottenere delle previsioni coerenti attraverso un metodo di previsione semplice, sia a livello logico, che computazionale, come quello sopracitato, e mediante una procedura più complessa, la procedura Box-Jenkins, facendo riferimento ad un campione non rappresentativo dell'intera popolazione.

## 1.2 L'AZIENDA SANITARIA

Dal punto di vista ordinamentale, la legislazione del Friuli Venezia Giulia ha caratterizzato il complesso dell'organizzazione operativa del proprio servizio sanitario quale gruppo di Aziende, dotate di personalità giuridica e di autonomia imprenditoriale, aventi la comune missione di perseguire il fine pubblico di tutela della salute. Ogni ASS riconosce come propria fondamentale missione il raggiungimento da parte della popolazione del più elevato livello possibile di salute, intesa come condizione di benessere fisico, mentale e sociale e persegue, a sostegno di ciò gli obiettivi sanitari, di salute, di prestazioni, di struttura previsti dalla programmazione nazionale e regionale, assicurando i livelli di assistenza previsti e garantendo gli interventi di prevenzione, diagnostici, terapeutici, educativi più indicati, con il minimo rischio possibile per la persona e con il miglior grado di soddisfazione possibile per gli interventi ricevuti, per i contatti umani con gli operatori e per gli esiti.

Ogni Azienda si articola in Direzione strategica e Strutture operative. La Direzione strategica definisce, sulla base delle indicazioni della pianificazione regionale, le strategie di programmazione aziendale e ne controlla la loro realizzazione. Le Strutture operative gestiscono, in piena autonomia, l'attività sanitaria o amministrativa, garantendo il raggiungimento degli obiettivi assegnati, nell'ambito delle risorse umane, tecniche e finanziarie attribuite; ad esse è assegnato un budget, ovvero obiettivi e risorse predefiniti e negoziati.

La struttura operativa di maggior rilevanza è il distretto. Dotato di autonomia economico-finanziaria e gestionale, con l'attribuzione di specifico budget e con contabilità separata all'interno del bilancio dell'Azienda si configura come la



struttura portante dell'Azienda per i servizi sanitari territoriali. i Distretti sono nati con lo scopo di avvicinare ai cittadini le risposte sanitarie, integrando l'offerta di tipo ospedaliero. Essi garantiscono risposte unitarie e prolungate nel tempo ai problemi sanitari e sociosanitari della comunità, collaborando con i Comuni della Provincia. Ne deriva un percorso che può essere presso il Distretto o a domicilio e le competenze professionali si estendono anche ad interventi socio-sanitari. Ogni Distretto è coordinato da un Direttore nominato dal Direttore Generale dell'Azienda.

Il sistema sanitario della regione Friuli Venezia Giulia risulta così suddiviso:

- Azienda per i servizi sanitari n. 1 Triestina (4 Distretti)
- Azienda per i servizi sanitari n. 2 Isontina (2 Distretti)
- Azienda per i servizi sanitari n. 3 Alto Friuli (2 Distretti)
- Azienda per i servizi sanitari n. 4 Medio Friuli (5 Distretti)
- Azienda per i servizi sanitari n. 5 Bassa Friulana (2 Distretti)
- Azienda per i servizi sanitari n. 6 Friuli Occidentale (5 Distretti)

## Capitolo 2

### METODOLOGIE UTILIZZATE

#### 2.1 SERIE STORICHE

Per serie storica s'intende un insieme di dati ordinati secondo un criterio qualitativo, il tempo. Possiamo pertanto definire serie storica una successione di dati numerici nella quale ogni dato è associato ad un particolare istante od intervallo del tempo. In una serie storica l'ordine dei dati non è casuale e genera dipendenza tra osservazioni successive. Lo scopo dell'analisi delle serie storiche è proprio quello di studiare e modellare questa dipendenza al fine di trovare il meccanismo che l'ha generata.

Un modello stocastico generale per descrivere il processo generatore dei dati di una serie storica relativa ad una variabile  $Y$  può essere espresso dalla somma di una parte completamente deterministica  $f(t)$  e di una casuale  $\varepsilon(t)$ , che rappresenta la parte stocastica della serie ed obbedisce ad una determinata legge di probabilità.

$$Y_t = f(t) + \varepsilon_t$$

Si assume che la parte sistematica  $f(t)$  sia la risultante della combinazione di tre componenti: trend, ciclo e stagionalità.

Il trend è definito come la tendenza di fondo a crescere o a decrescere. Il ciclo è costituito dalle fluttuazioni attribuibili al succedersi nel fenomeno considerato di fasi ascendenti e di fasi discendenti. La stagionalità, invece, è composta dai movimenti del fenomeno nel corso dell'anno che tendono a ripetersi nel medesimo periodo di anni successivi.

I principali modelli di combinazione di queste tre componenti sono:

il modello additivo:

$$Y_t = T_t + C_t + S_t + \varepsilon_t$$

nel modello additivo le quattro componenti sono tutte espresse nella stessa unità di misura di  $Y_t$

il modello moltiplicativo :

$$Y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t$$

questo modello può essere ricondotto a quello moltiplicativo mediante trasformazione logaritmica. In questo caso solo  $T_t$  è espressa nella stessa unità di misura di  $Y_t$ , le altre sono espresse sotto forma di numeri puri.

il modello misto:

$$Y_t = T_t \cdot C_t \cdot S_t + \varepsilon_t$$

in questo modello  $T_t$  e  $\varepsilon_t$  sono espresse nella stessa metrica di  $Y_t$  mentre  $C_t$  e  $S_t$  assumono la forma di numeri indici.

Lo studio della parte deterministica rappresenta l'approccio classico, si suppone che i residui di  $Y_t$  siano imputati al caso ed assimilati come errori accidentali. La componente stocastica è quindi generata da un processo white noise, cioè una successione di variabili casuali iid di media nulla e varianza costante.

Nell'approccio moderno, invece, l'attenzione viene posta sulla componente stocastica  $\varepsilon_t$  che si ipotizza essere un processo a componenti correlate che va trattato con opportune tecniche statistiche.

## 2.2 IL METODO HOLT-WINTERS

Il metodo Holt-Winters è la naturale generalizzazione del metodo del liscio esponenziale semplice e permette di essere applicata a serie che presentano un trend ed eventualmente una componente stagionale di periodo  $s$ . Holt-Winters ammette sia un modello non stagionale sia uno stagionale. Prenderemo in considerazione solo quest'ultimo, che è presente nella forma additiva e moltiplicativa.

### Stagionale additivo

Il modello stagionale additivo suppone che nelle vicinanze di  $n$ , la serie storica  $\{y_t\}_{t=1}^n$  sia localmente approssimabile dalla retta di equazione:

$$L_n + T_n(t - n) + S_n$$

dove con  $L_n$  e  $T_n$  rappresentano il livello e la pendenza della serie nel periodo  $n$ , mentre  $S_n$  è un fattore stagionale di periodo pari a  $s$

Indichiamo con  $\hat{L}_n$ ,  $\hat{T}_n$  e  $\hat{S}_n$  le stime di questi tre fattori le cui formule di aggiornamento sono:

$$\begin{cases} \hat{L}_n = \alpha[\hat{L}_{n-1} + \hat{T}_{n-1}] + (1 - \alpha)(y_n - \hat{S}_{n-s}) \\ \hat{T}_n = \beta\hat{T}_{n-1} + (1 - \beta)[\hat{L}_n - \hat{L}_{n-1}] \\ \hat{S}_n = \gamma\hat{S}_{n-s} + (1 - \gamma)[y_n - \hat{L}_n] \end{cases}$$

La previsione a  $k$  passi è data da:

$$\hat{y}_{n+k} = \hat{L}_n + \hat{T}_n k + \hat{S}_{n+k-s} \quad \text{se } 1 \leq k \leq s$$

$$\hat{y}_{n+k} = \hat{L}_n + \hat{T}_n k + \hat{S}_{n+k-2s} \quad \text{se } 1 \leq k \leq 2s$$

### Stagionale moltiplicativo

Con questo metodo supponiamo invece che nelle vicinanze di  $n$  la serie sia localmente approssimabile dalla retta di equazione:

$$[L_n + T_n(t - n)]S_n$$

le formule di aggiornamento delle stime in questo caso saranno:

$$\begin{cases} \hat{L}_n = \alpha[\hat{L}_{n-1} + \hat{T}_{n-1}] + (1 - \alpha) \frac{y_n}{\hat{S}_{n-s}} \\ \hat{T}_n = \beta \hat{T}_{n-1} + (1 - \beta)[\hat{L}_n - \hat{L}_{n-1}] \\ \hat{S}_n = \gamma \hat{S}_{n-s} + (1 - \gamma) \frac{y_n}{\hat{L}_n} \end{cases}$$

e le previsioni a  $k$  passi:

$$\hat{y}_{n+k} = (\hat{L}_n + \hat{T}_n k) \hat{S}_{n+k-s} \quad \text{se } 1 \leq k \leq s$$

$$\hat{y}_{n+k} = (\hat{L}_n + \hat{T}_n k) \hat{S}_{n+k-2s} \quad \text{se } 1 \leq k \leq 2s$$

Le costanti di liscio  $\alpha, \beta$  e  $\gamma$  sono comprese tra 0 e 1; più sono vicine a 1 maggiore è l'influenza delle osservazioni passate, rendendo la previsione rigida, se invece sono vicine a 0 la previsione sarà influenzata dalle previsioni più recenti e sarà, quindi, più flessibile.

Le costanti vanno scelte cercando di minimizzare l'MQE e l'indice di Theil.

Il metodo Holt-Winters necessita di un'inizializzazione, possiamo scegliere di partire al tempo  $t = s$  calcolando le stime iniziali dei 3 fattori utilizzando le seguenti formule:

per il modello additivo

$$\begin{cases} \hat{L}_s = \frac{1}{s} \sum_{t=1}^s y_t \\ \hat{T}_s = \hat{L}_s - \hat{L}_{s-1} \\ \hat{S}_j = y_j - \hat{L}_s \quad \text{con } j = 1, 2, \dots, s \end{cases}$$

per il modello moltiplicativo

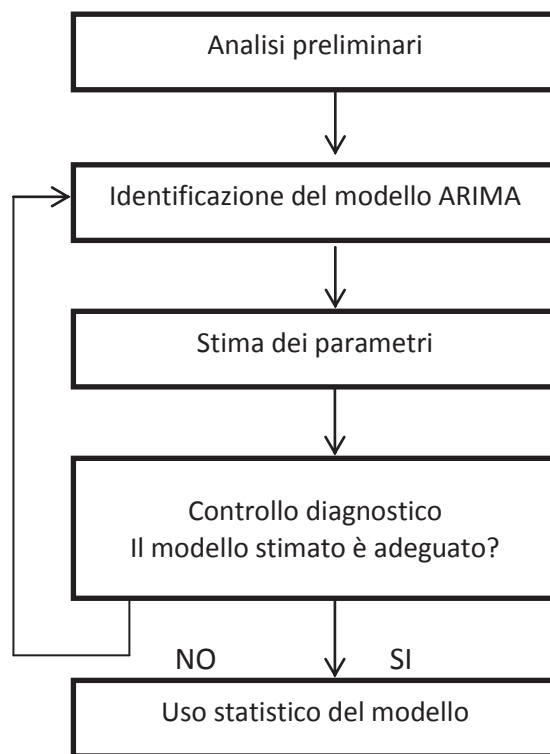
$$\left\{ \begin{array}{l} \hat{L}_s = \frac{1}{s} \sum_{t=1}^s y_t \\ \hat{T}_s = \hat{L}_s - \hat{L}_{s-1} \\ \hat{S}_j = \frac{y_j}{\hat{L}_s} \quad \text{con } j = 1, 2, \dots, s \end{array} \right.$$

## 2.3 LA PROCEDURA DI BOX JENKINS

Il *metodo di Box Jenkins* è una procedura utilizzata per stimare il modello generatore dei dati, che successivamente può essere utilizzato a fini predittivi. In particolare i modelli più importanti sono: il *modello lineare autoregressivo (AR)* ed il *modello a media mobile (MA)*. Dalla loro combinazione risultano il *modello autoregressivo a media mobile (ARMA)* ed il *modello autoregressivo integrato a media mobile (ARIMA)*; nel caso la serie storica presenti stagionalità vengono utilizzati rispettivamente il modello **SARMA** o **SARIMA**.

Il metodo proposto da Box e Jenkins è costituito principalmente da quattro fasi:

- i) Analisi preliminare
- ii) Identificazione
- iii) Stima dei parametri
- iv) Controllo diagnostico



Vediamo in dettaglio le quattro fasi:

### **i) Analisi preliminare**

Lo scopo principale della prima fase è quello di analizzare l'andamento della serie: verificare la sua stazionarietà ed analizzarla graficamente. Per poter stimare il modello, la serie deve essere stazionaria, ovvero sia la sua funzione di autocorrelazione (ACF) che la sua funzione di autocorrelazione parziale (PACF) devono tendere ad annullarsi a velocità esponenziale al crescere di  $k$ . Nel caso in cui la serie non sia stazionaria, quindi quando le due funzioni tendono ad annullarsi troppo lentamente o non si annullano, si procede nella eliminazione di questa caratteristica, differenziando  $d$  volte i dati.

## ii) Identificazione

Identificare il modello significa specificare il suo ordine, ovvero identificare i parametri  $p, d, q$  (in aggiunta anche  $P, D, Q$ ). Per giungere a questo obiettivo vengono utilizzate la funzione di autocorrelazione e la funzione di autocorrelazione parziale (stimate sul campione), in maniera tale da poter arrivare ad una similitudine con una funzione di autocorrelazione teorica.

Da questo confronto vengono stimati degli andamenti che portano all'identificazione della classe del modello di appartenenza:

- il modello risulta essere un MA ( $q$ ) se ACF  $\rho_k$  è diversa da zero per  $k \leq q$  e se si annulla per  $k \geq q + 1$ ;
- il modello risulta essere un AR ( $p$ ) se la PACF  $P_k$  è diversa da zero per  $k \leq p$  e si annulla per  $k \geq p + 1$ ;
- il modello risulta essere un ARMA ( $p, d, q$ ) se vengono combinate le caratteristiche di un MA ( $q$ ) e di un AR ( $p$ ), il suo corrispettivo stagionale è SARMA;
- il modello risulta essere un ARIMA in presenza di un ARMA con  $d$  differenziazioni, il suo corrispettivo stagionale è SARIMA;

L'utilizzo di un ARMA o ARIMA, ovvero di modelli misti, è dato dal fatto che è più immediato trovarsi in presenza contemporanea di componenti di media mobile e di componenti autoregressive. Un altro vantaggio di questi strumenti è quello di essere più parsimoniosi in termini di numero di parametri.



Una volta preso atto del tipo di modello bisogna procedere alla definizione del numero di parametri. Per ovviare al problema della sovrapparametrizzazione sono stati introdotti due criteri: il criterio di AIC e il criterio di SC (essi assegnano un "costo" all'introduzione di ogni nuovo parametro addizionale).

*Il criterio di AIC o AKAIKE (Asymptotic Information Criterion) è dato da:*

$$AIC(k) = -\frac{2}{n} (\log L(\hat{\delta}) - k)$$

dove  $k$  è il numero di parametri nel modello,  $\hat{\delta}$  è il vettore ( $k \times 1$ ) contenente i parametri stimati e  $L(\delta)$  è la funzione di verosimiglianza calcolata in  $\hat{\delta}$ .

*Il criterio di SC o SCHWARZ (Schwarz Criterion) viene utilizzato per ovviare alla tendenza del criterio AIC alla sovrapparametrizzazione, ed è dato da:*

$$SC(k) = -\frac{2 \log L(\hat{\delta})}{n} + k \frac{\log n}{n}$$

I valori di  $p$  e  $q$  scelti sono quindi quelli che minimizzano i due criteri visti.

### **iii) Stima dei parametri**

Definito il numero di parametri, si procede con la loro stima. Le metodologie di stima utilizzate maggiormente sono *il metodo dei minimi quadrati non lineari* ed *il metodo della massima verosimiglianza*. Il pregio del metodo dei minimi quadrati non lineari è quello di non dipendere da specifiche ipotesi distributive, dall'altro lato il metodo della massima verosimiglianza impone di conoscere la distribuzione ma fornisce degli stimatori con migliori proprietà statistiche.

#### iv) Controllo diagnostico

L'ultima fase serve a verificare l'adeguatezza del modello stimato attraverso opportune analisi e test diagnostici. Esso è correttamente identificato se sui valori dei residui  $e_t = (y_t - \hat{y}_t)$  si riscontrano le ipotesi fatte a priori sui disturbi  $\varepsilon_t$ .

Esistono diversi tipi di analisi diagnostiche per arrivare a questo risultato. La prima avviene attraverso *un'analisi grafica*. Essa consiste nel consultare il grafico dei residui o il diagramma di dispersione dei punti, i quali non dovrebbero mostrare nessuna regolarità e dovrebbero dar luogo ad una nuvola di punti senza alcuna struttura.

Un altro metodo è quello dell'*autocorrelazione dei residui*: la serie dei residui viene trattata come una serie storica e si può quindi calcolare la funzione di autocorrelazione empirica. Se la funzione stimata è significativamente diversa da quella di un processo *white noise*, il modello stimato è buono. Se infatti tutti i valori si trovano all'interno di un intervallo di confidenza  $[-1,96 / \sqrt{n}, 1,96 / \sqrt{n}]$  con livello di significatività 5% allora non esiste correlazione fra i residui.

Un terzo tipo di analisi è dato dal test di tipo portmanteau, ad esempio si può utilizzare la statistica elaborata da *Ljung e Box*. Essa è basata sulle ipotesi:

$$\begin{cases} H_0 : p_1 = p_2 = \dots = p_h = 0 \\ H_1 : p_i \neq 0 \text{ per almeno un } i \in (1, \dots, h) \end{cases} \quad \begin{cases} \text{residui incorrelati} \\ \text{residui correlati} \end{cases}$$

e sulla statistica test:

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{p}_k^2$$

se il modello è corretto allora  $Q(m)$  è approssimato da una variabile casuale  $X_{m-p-q}^2$ .

La *verifica di casualità dei residui* consiste in una serie di test di casualità o di indipendenza, come per esempio attraverso il diagramma di dispersione.

## 2.4 INDICI DI VALUTAZIONE

Definiamo come errore della previsione al tempo  $t$  la differenza tra valore realizzato,  $r_t$ , e valore previsto  $p_t$  :

$$e_t = r_t - p_t$$

Gli indici più utilizzati sono l'*MQE* (Radice quadrata dell'errore quadratico medio di previsione)

$$MQE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

e l'indice di Theil:

$$U = \frac{MQE}{\sqrt{\frac{1}{n} \sum_{t=1}^n r_t^2}}$$

Quest'ultimo rappresenta una normalizzazione dell'*MQE* consentendo il confronto tra previsioni relative a più serie storiche. Assume valori tra 0 e infinito. Se l'indice ha valore 0 significa che le previsioni coincidono esattamente con i valori realizzati, se invece assume valori prossimi o superiori ad 1 significa che il modello non si adatta bene alla serie.

## Capitolo 3

### APPLICAZIONE DEI MODELLI

#### 3.1 DATI

I dati utilizzati ci sono stati forniti su tre dataset. Il primo contenente tutte le prescrizioni dei farmaci da gennaio 2003 a dicembre 2007, il secondo e il terzo contenenti le informazioni dei pazienti e dei medici rispettivamente. Ad ogni paziente e ad ogni medico è stato assegnato un codice identificativo univoco (chiamate `key_anagrafe` e `Userid_medico` rispettivamente), i quali hanno permesso di associare i dati dei 3 dataset. Per ogni dataset sono stati raccolti dati riguardanti una molteplicità di attributi, ma ai fini della nostra analisi sono risultati pertinenti solo la data di spedizione dei farmaci, gli importi totali delle singole prescrizioni, l'azienda di appartenenza del medico e i codici univoci di medici e pazienti.

Per creare la serie storica comprendente la totalità dei pazienti, è stato utilizzato solo il primo dataset, dividendo mensilmente le prescrizioni sulla base della data di spedizione e sommando gli importi per ogni mese. E' stata creata una serie storica mensile in quanto ha fornito maggiori dettagli sulla fluttuazione della serie rispetto alle serie trimestrale e annuale.

La divisione per azienda sanitaria è stata possibile incrociando i dati dei tre dataset. Per ogni ASS sono stati isolati i medici e i pazienti appartenenti ai medici selezionati, ottenendo così i pazienti divisi per Azienda. Le serie storiche suddivise per azienda sono state create sfruttando la corrispondenza tra la key anagrafe presente nel

dataset delle prescrizioni e quella presente nel dataset dei pazienti. Non sono stati considerati i cambi di medico perciò la somma delle serie di ogni azienda non corrisponde esattamente alla somma della serie iniziale, tuttavia questa omissione genera un errore trascurabile.

Azienda	Spese Totali (in €)
1	41656026
2	20887943
3	5599573
4	45904843
5	6371504
6	42080392

Tabella 1 – Divisione Spese Totali

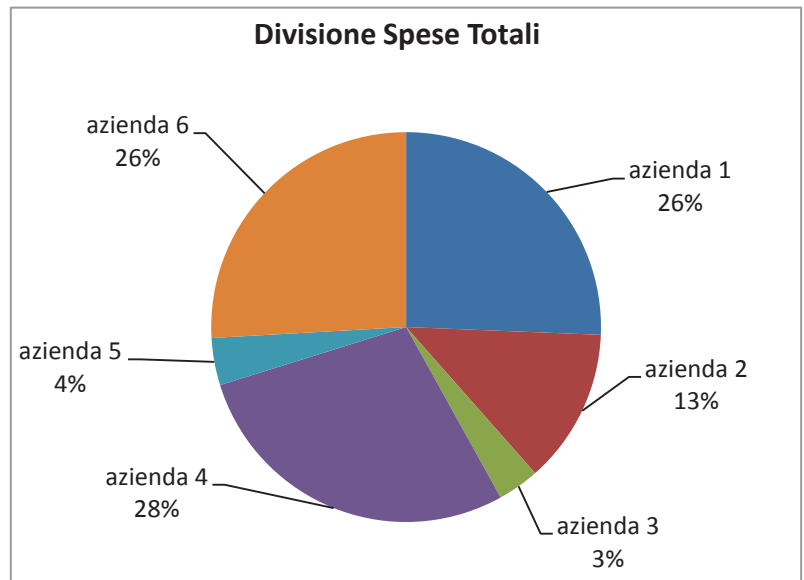


Grafico 1 - Divisione Spese Totali

Azienda	N° Medici
1	24
2	13
3	5
4	33
5	4
6	28

Tabella 2 - Divisione Medici

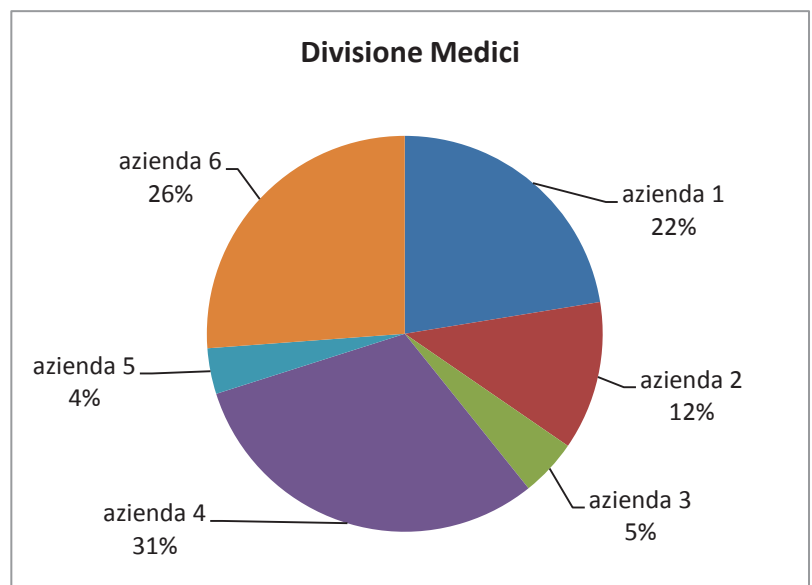


Grafico 2 - Divisione Medici

Azienda	N° Pazienti
1	37039
2	20593
3	6336
4	50561
5	6169
6	48053

Tabella 3 - Divisione Pazienti

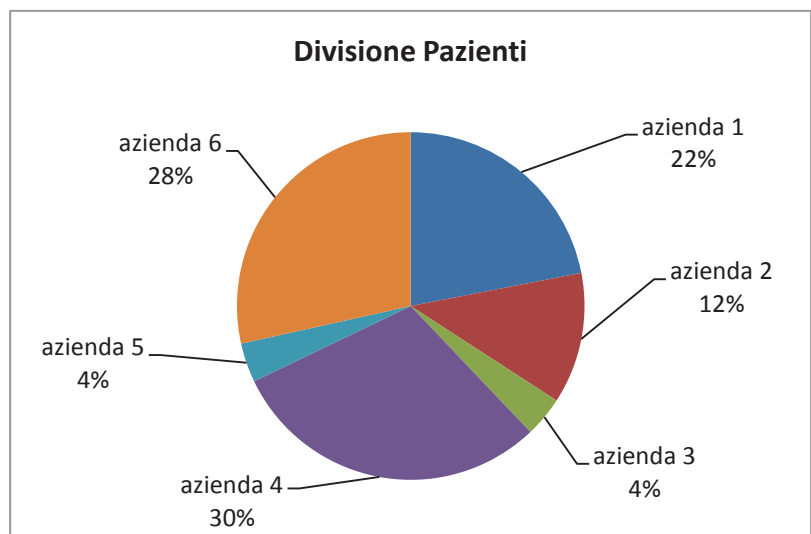


Grafico 3 - Divisione Pazienti

I grafici a torta ci mostrano le percentuali del campione esaminato di pazienti, spese totali e medici. Questi grafici ci danno una panoramica della composizione del nostro campione non casuale. Notiamo che maggiore è il numero dei distretti, maggiore è il numero di medici. Di conseguenza maggiore saranno il numero dei pazienti e le spese totali.

## 3.2 ANALISI SERIE STORICHE

### Serie Mensile Totale Pazienti

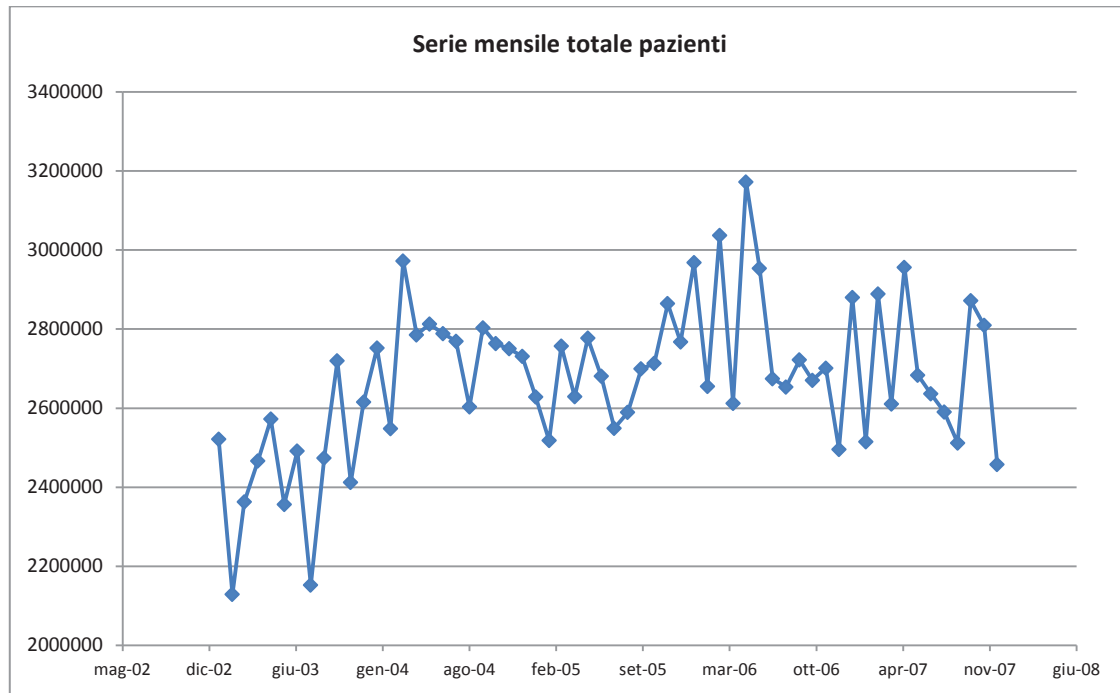


Grafico 4 - Serie Mensile Totale Pazienti

Analizziamo la serie storica contenente i dati di tutti i pazienti. Dall'analisi grafica notiamo la presenza di un trend crescente non lineare. Il tutto risulta più evidente scomponendo la serie con l'utilizzo del software R.

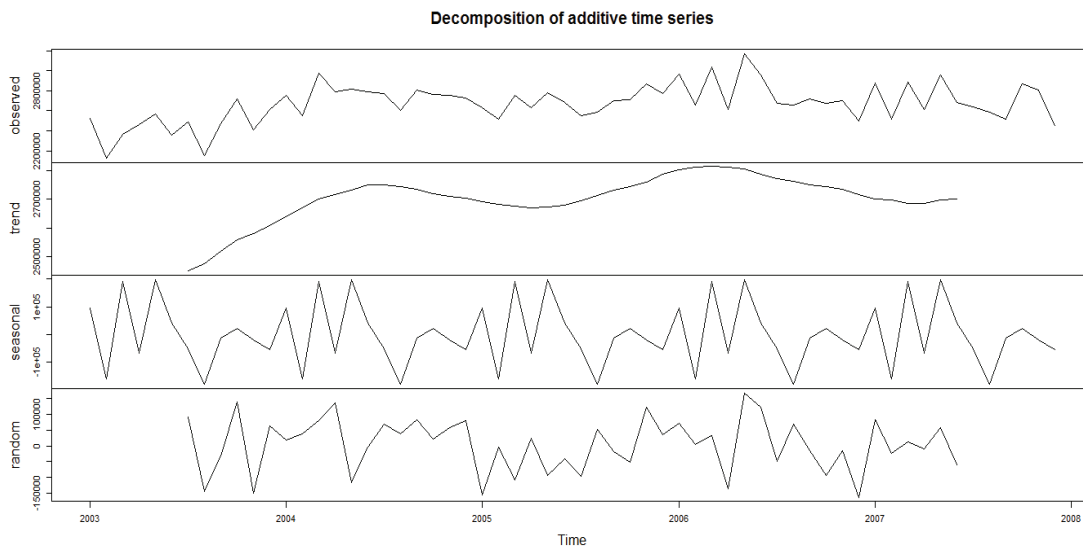


Grafico 5 - Scomposizione Serie Totale Pazienti

La scomposizione fornita da R suggerisce la presenza di stagionalità, questa è probabilmente dovuta agli eventi climatici stagionali, ad esempio il raffreddore o l'influenza per quanto riguarda i mesi invernali, le allergie ai pollini per i mesi primaverili mentre ad agosto il calo potrebbe essere dovuto alle vacanze estive.

Se dividiamo i pazienti per azienda sanitaria e creiamo una serie storica per ogni azienda, i risultati della decomposizione non cambiano e, approssimativamente, otteniamo lo stesso trend e la stessa stagionalità. Possiamo quindi dedurre che i fattori che hanno influenzato trend e stagionalità, non sono particolari di una data azienda ma hanno influito in tutte e 6.

## Azienda 1

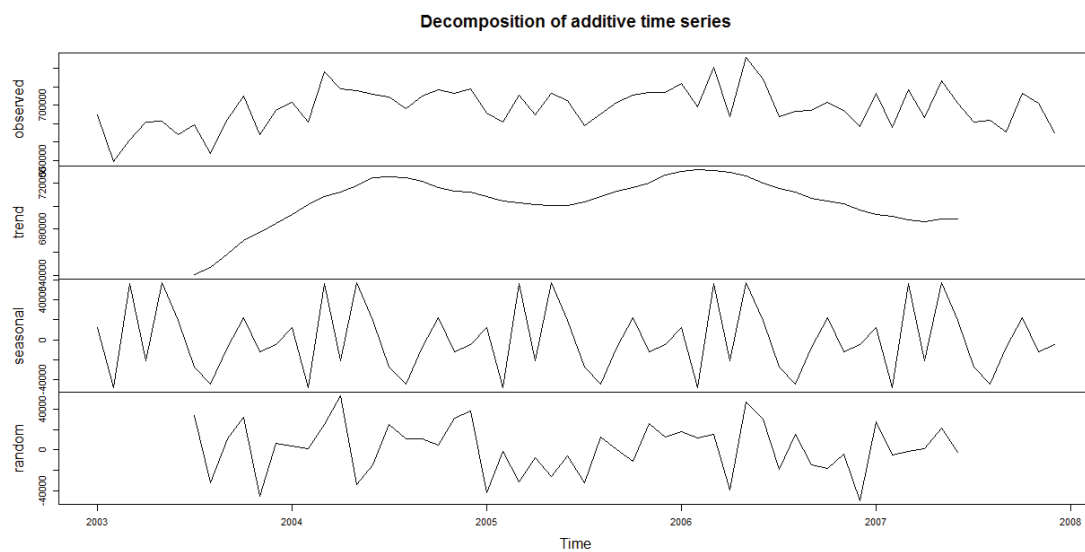


Grafico 6 - Scomposizione Serie Mensile Azienda 1



## Azienda 2

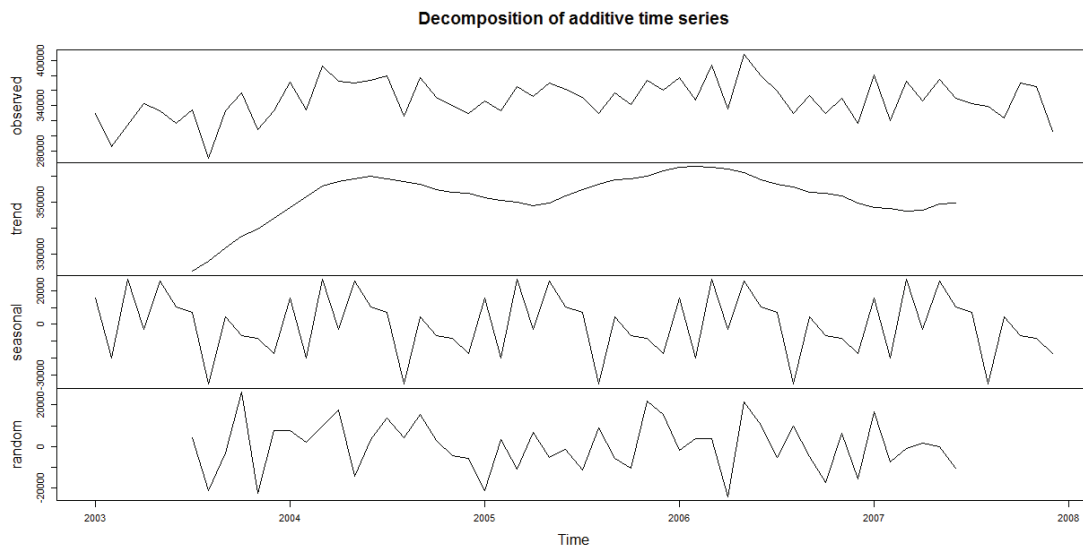


Grafico 7 - Scomposizione Serie Mensile Azienda 2

## Azienda 3

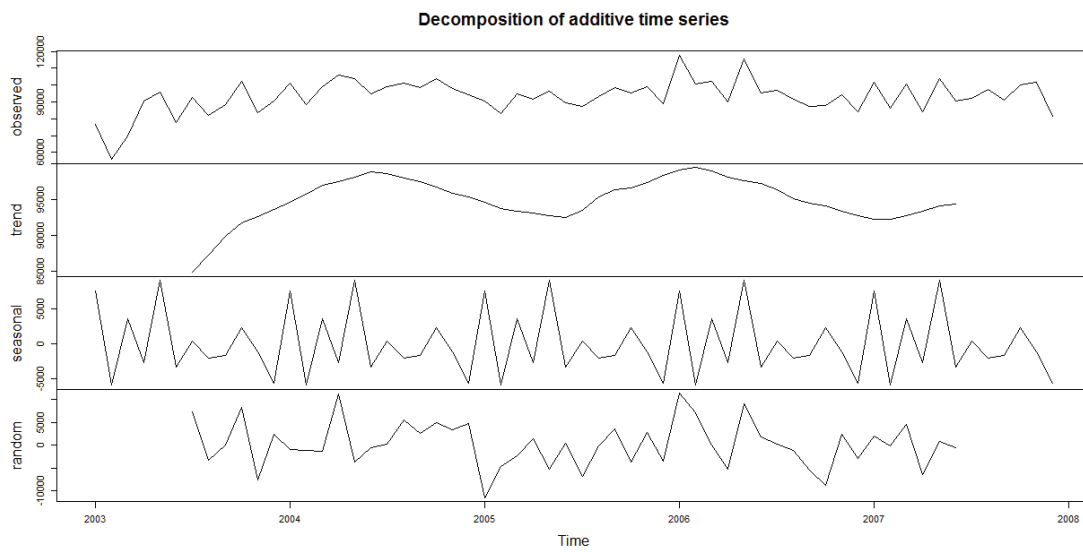


Grafico 8 - Scomposizione Serie Mensile Azienda 3

## Azienda 4

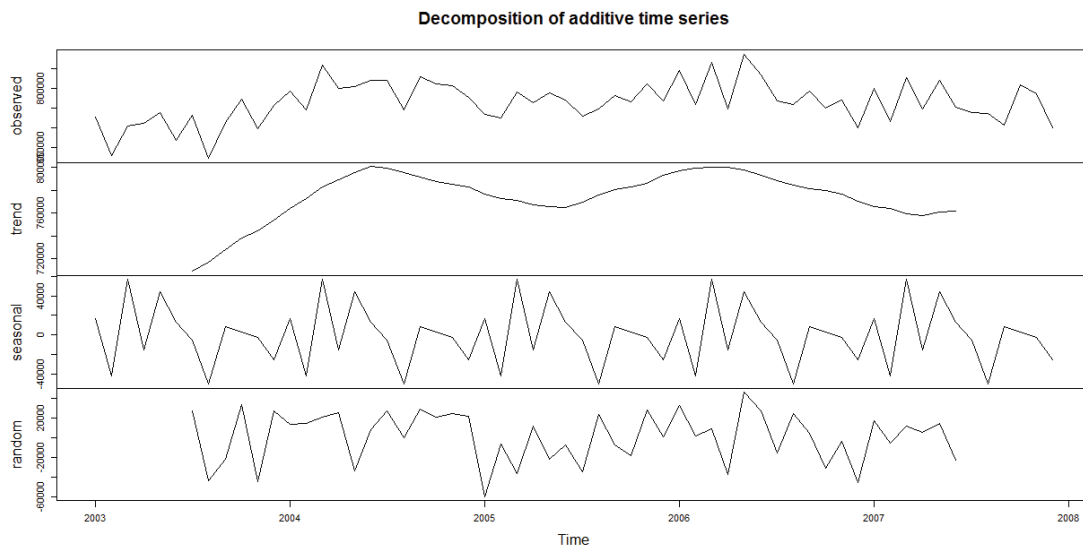


Grafico 9 - Scomposizione Serie Mensile Azienda 4

## Azienda 5

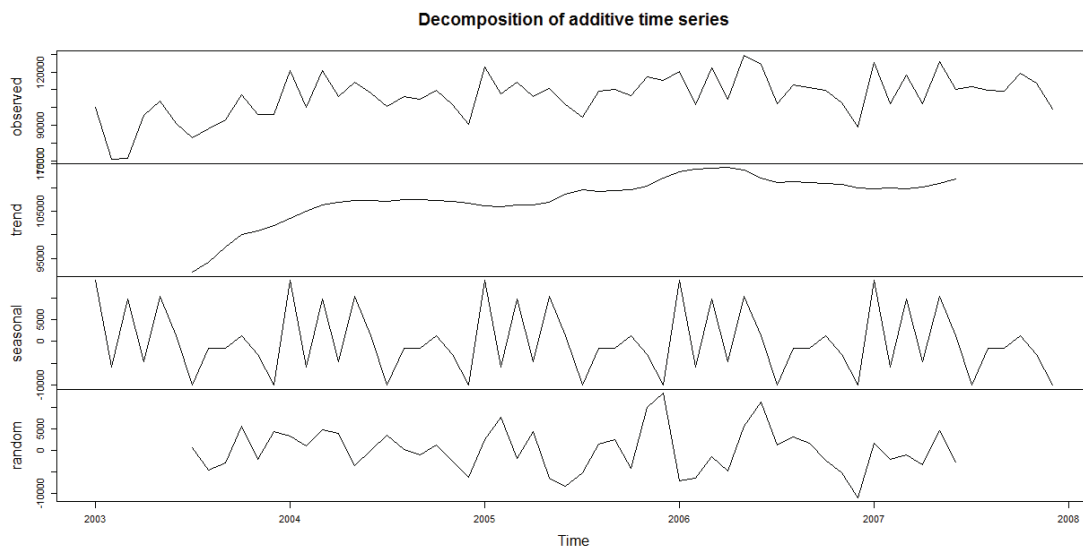
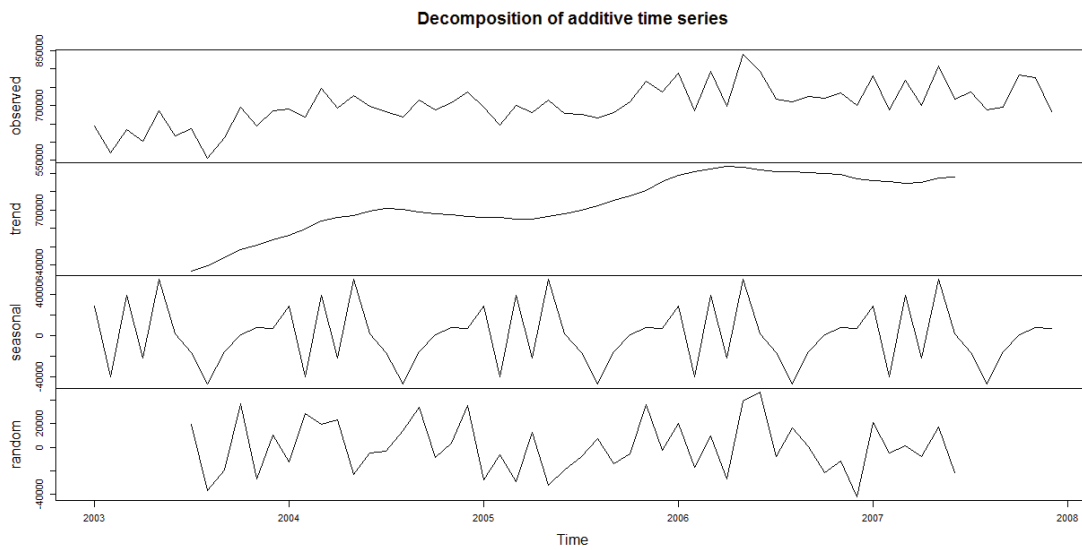


Grafico 10 - Scomposizione Serie Mensile Azienda 5

## Azienda 6



*Grafico 11 - Scomposizione Serie Mensile Azienda 6*

### **3.3 PREVISIONI CON METODO HOLT-WINTERS**

Per implementare il metodo Holt-Winters stagionale, sia additivo che moltiplicativo, è stato utilizzato Excel. Le costanti di liscio,  $\alpha$ ,  $\beta$  e  $\gamma$ , sono state scelte per tentativi, provando varie combinazioni fino a quando non è stata trovata la terna che minimizzasse l'MQE e l'indice di Theil. Sono state applicate le formule di aggiornamento fino a dicembre 2006 e le formule di previsione per l'anno 2007. In questo modo è stato possibile confrontare le previsioni con i dati reali e valutare la bontà del metodo. Questa procedura è stata applicata ad ogni serie analizzata.

### 3.3.1 Previsioni con serie mensile totale pazienti

Nella tabella sono riportati i valori delle costanti di lisciamento, l'MQE e l'indice di Theil relativi alla terna scelta.

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,9	0,5	95627,94	0,035
Moltiplicativo	0,9	0,9	0,45	95427,22	0,035

Tabella 4 - H-W Serie mensile totale Pazienti

Come si può notare dalla tabella e, come si vedrà dal grafico, i due modelli hanno la stessa capacità predittiva. Si vede, infatti, che hanno pressoché lo stesso MQE e stesso valore per l'indice di Theil.

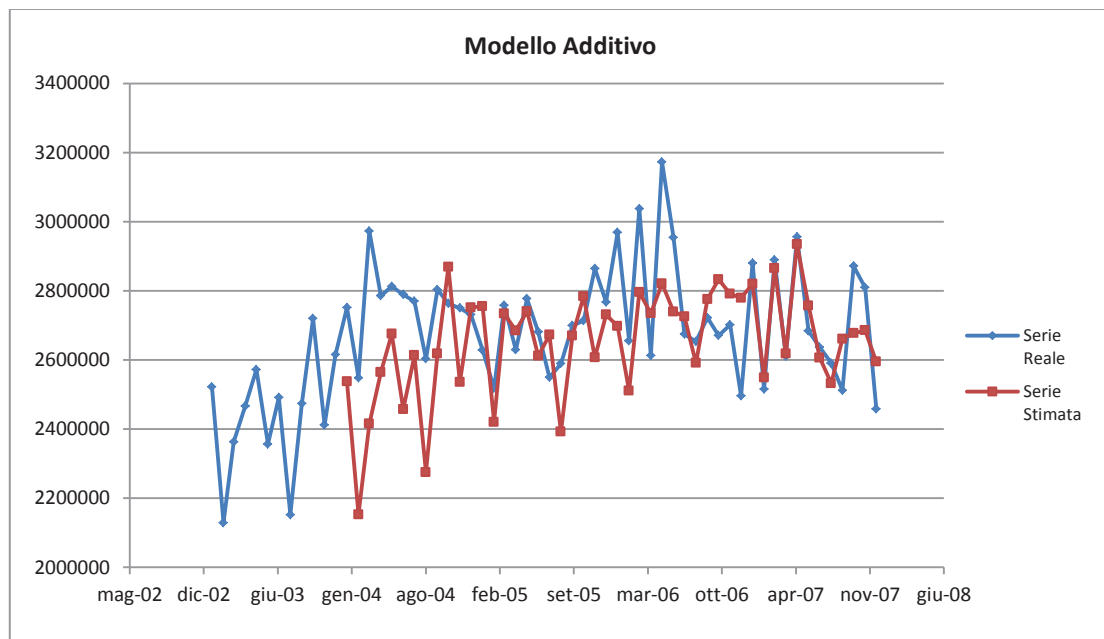


Grafico 12 - H-W Additivo Serie mensile totale Pazienti

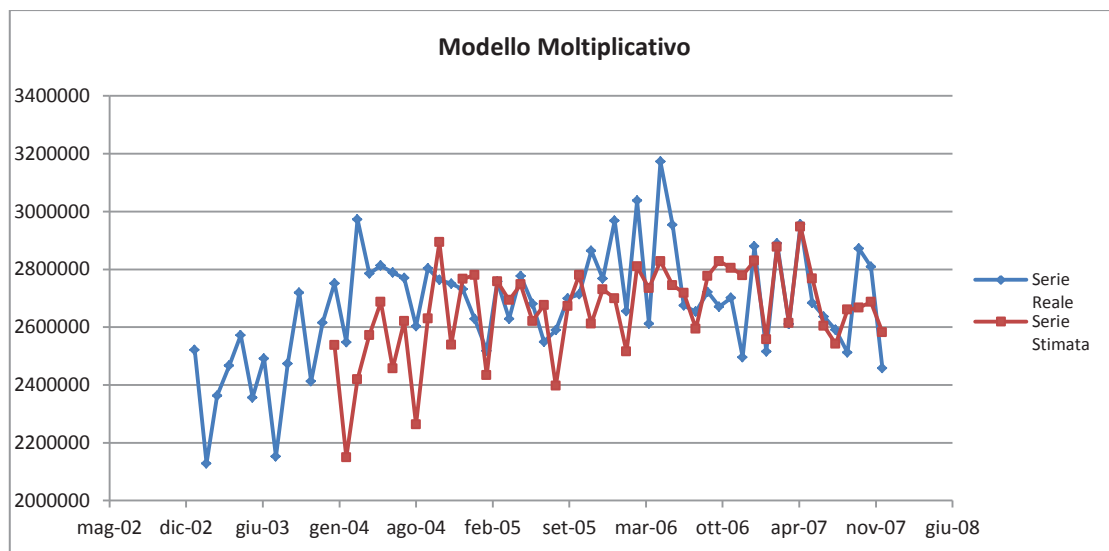


Grafico 13 - H-W Moltiplicativo Serie mensile totale Pazienti

Dai grafici notiamo che la serie perequata non segue precisamente l'andamento della serie reale, in alcuni casi si discosta molto dai valori reali, ad esempio da gennaio 2004 ad ottobre 2004. Tuttavia le previsioni dei primi mesi del 2007 sono molto coerenti con i dati reali, per il resto dell'anno si discostano lievemente, mantenendo comunque un buon adattamento.

### 3.3.2 Previsioni con serie mensile dei pazienti divisi per azienda

Analizziamo le serie storiche create dividendo i dati per azienda sanitaria.

#### Azienda 1

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,9	0,45	22508,21	0,033
Moltiplicativo	0,9	0,7	0,4	25733,62	0,038

Tabella 5 - H-W Serie mensile Azienda 1

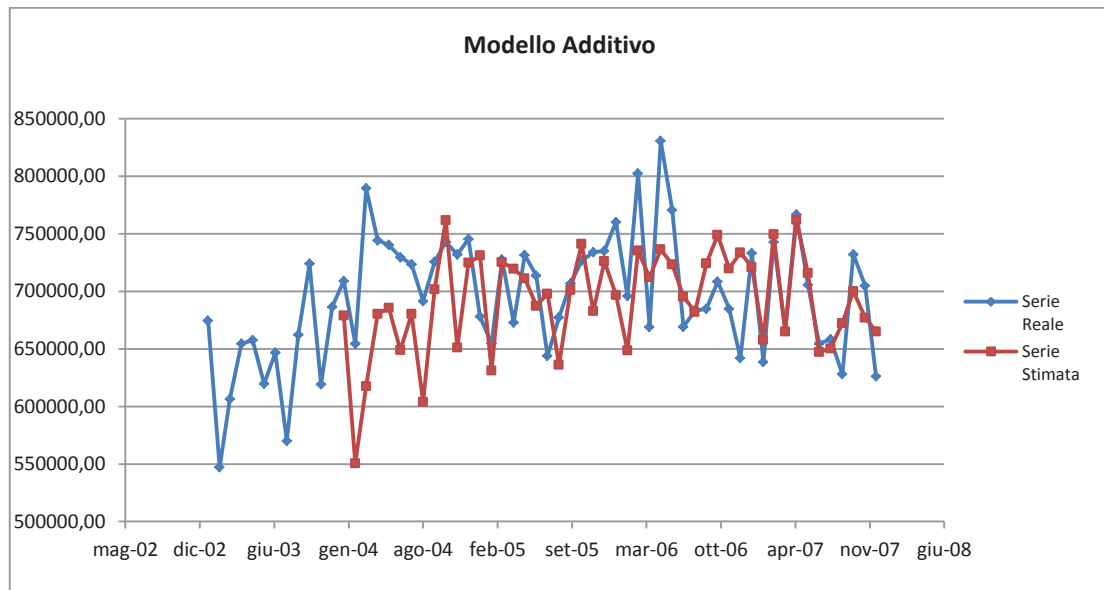


Grafico 14 - H-W Additivo Serie mensile Azienda 1

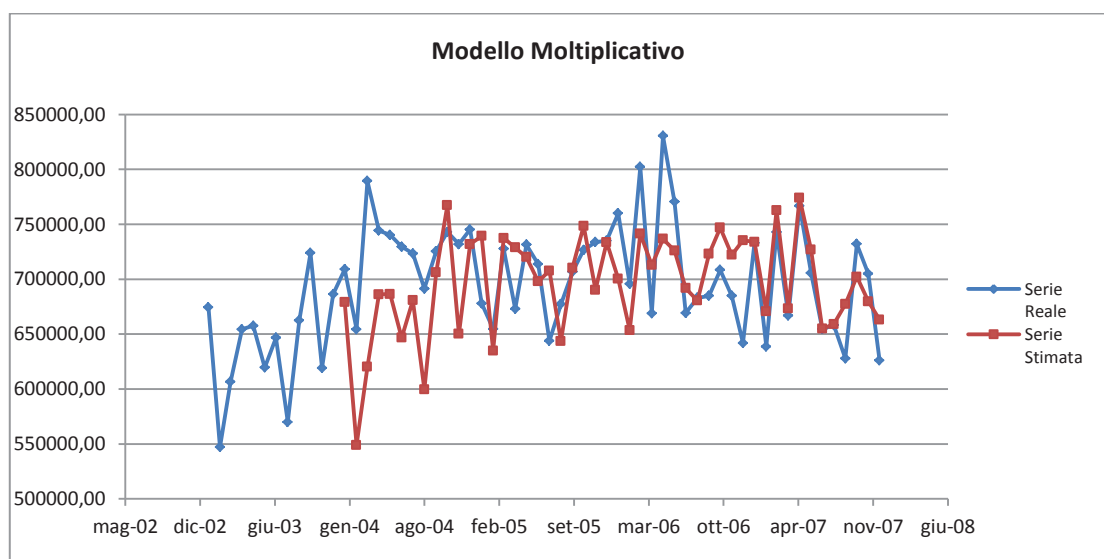


Grafico 15 - H-W Moltiplicativo Serie mensile Azienda 1

## Azienda 2

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,95	0,4	18568,57	0,053
Moltiplicativo	0,9	0,9	0,55	19054,82	0,055

Tabella 6 - H-W Serie mensile Azienda 2

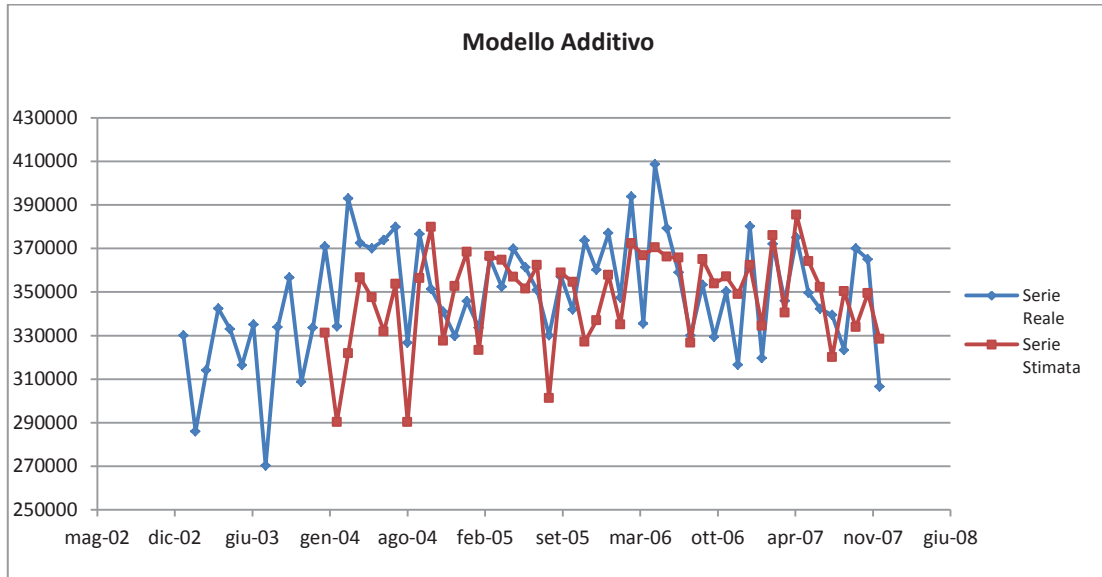


Grafico 16 - H-W Additivo Serie mensile Azienda 2

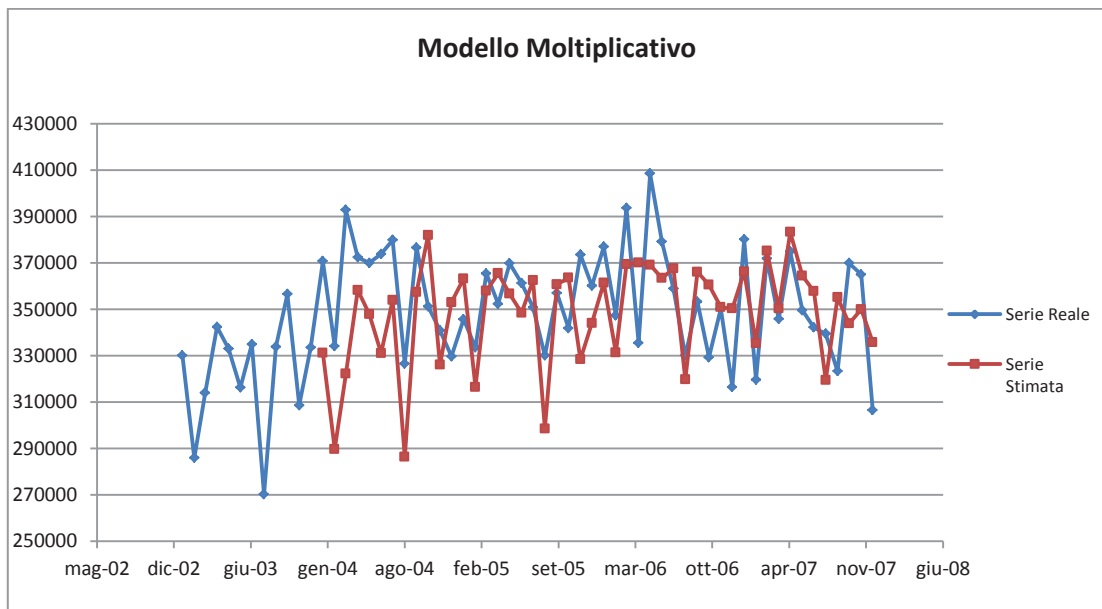


Grafico 17 - H-W Moltiplicativo Serie mensile Azienda 2

### Azienda 3

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,9	0,45	6515,91	0,069
Moltiplicativo	0,9	0,95	0,45	4803,95	0,051

Tabella 7 - H-W Serie mensile Azienda 3

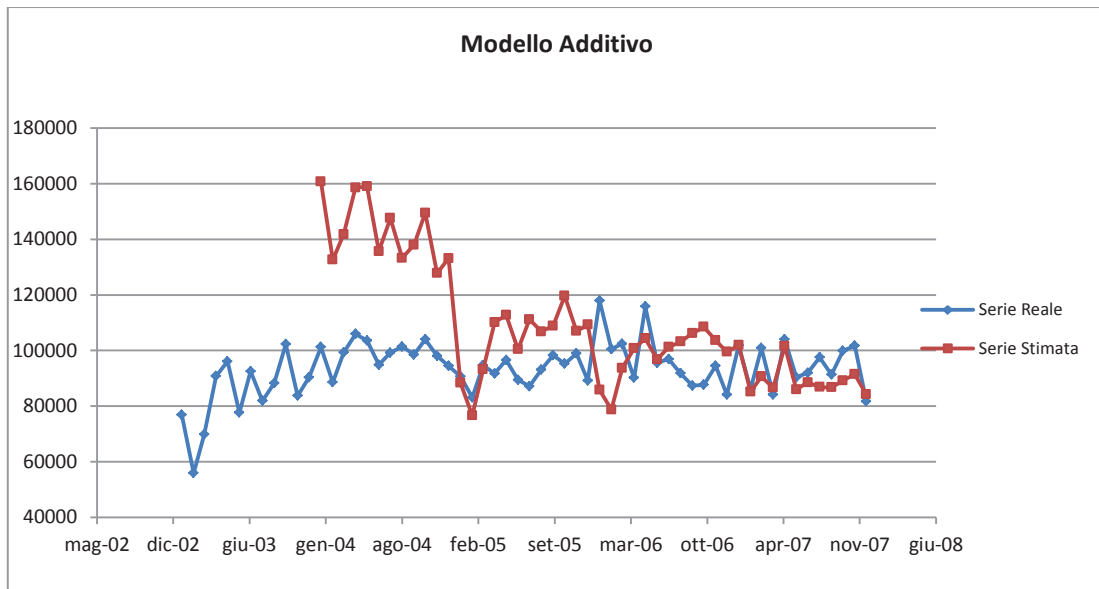


Grafico 18 - H-W Additivo Serie mensile Azienda 3

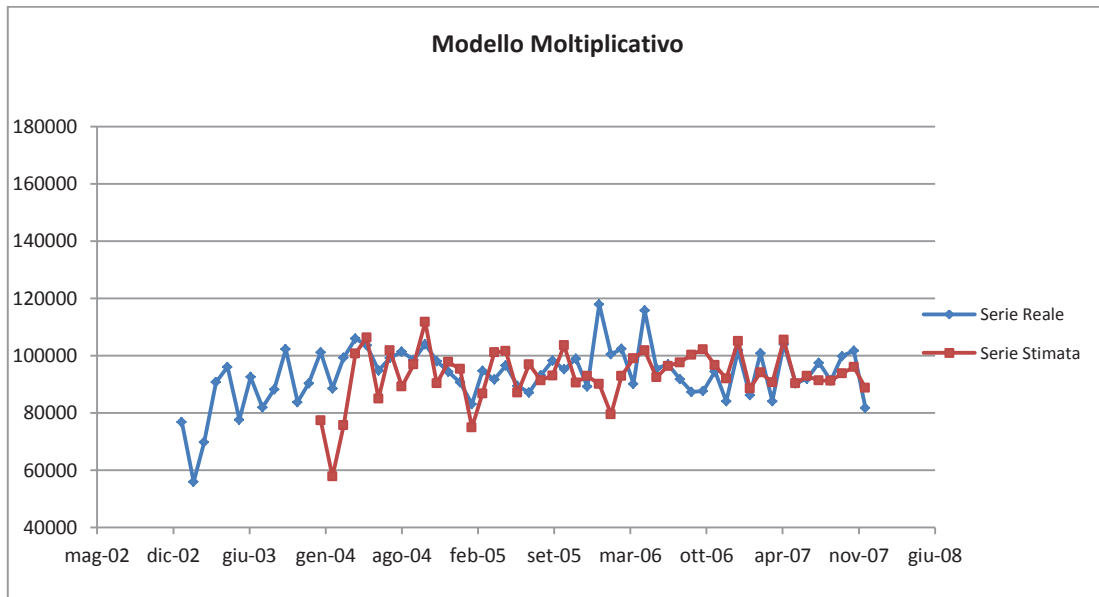


Grafico 19 - H-W Moltiplicativo Serie mensile Azienda 3



## Azienda 4

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,95	0,5	29497,71	0,039
Moltiplicativo	0,8	0,5	0,55	29909,22	0,039

Tabella 8 - H-W Serie mensile Azienda 4

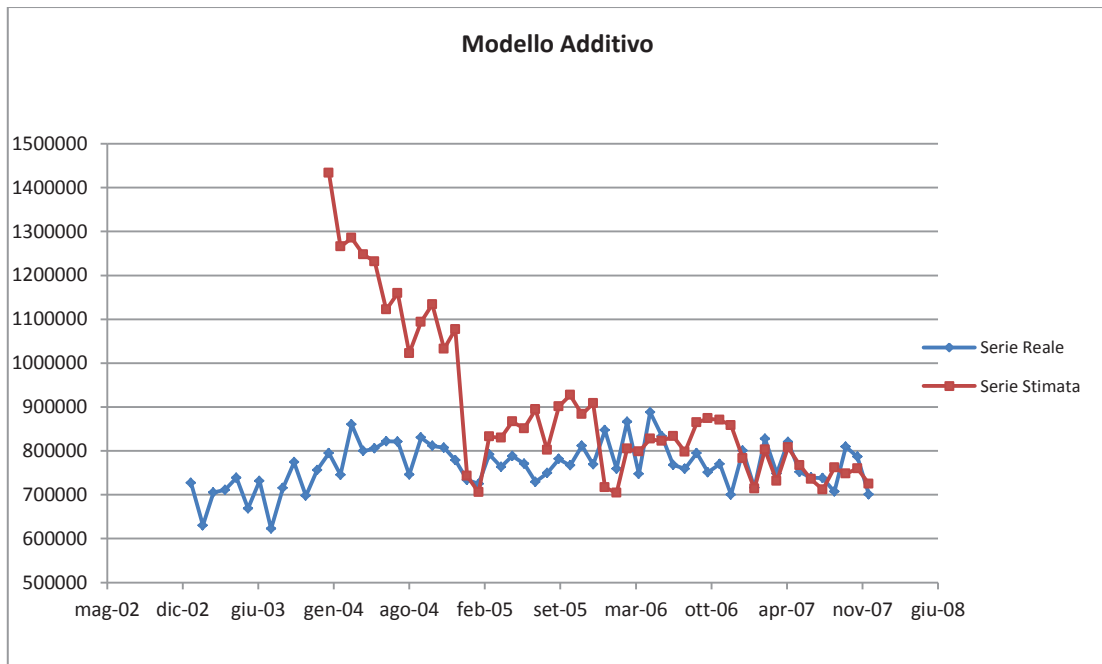


Grafico 20 - H-W Additivo Serie mensile Azienda 4

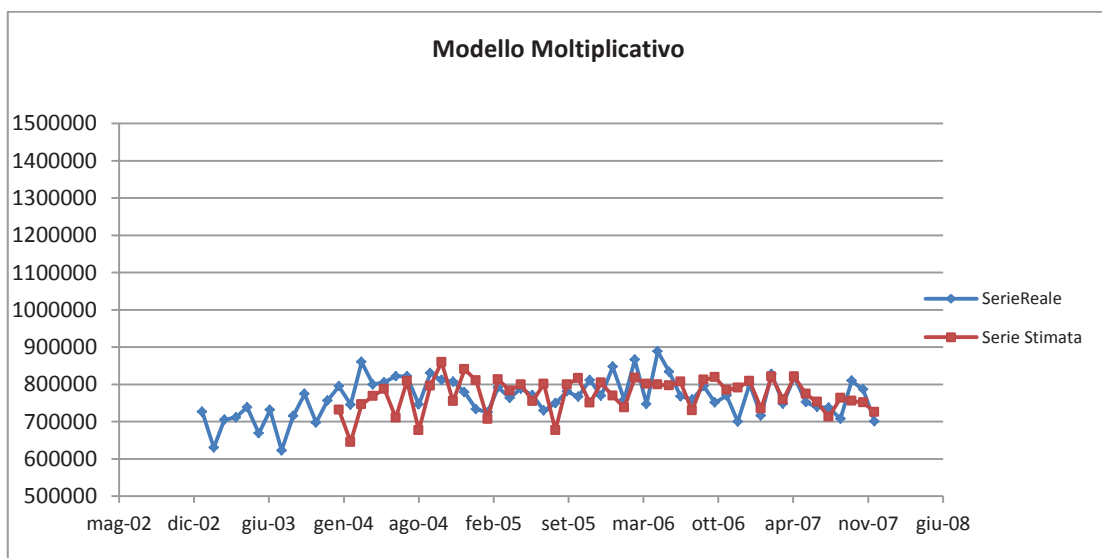


Grafico 21 - H-W Moltiplicativo Serie mensile Azienda 4

## Azienda 5

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,95	0,95	0,6	4711,92	0,042
Moltiplicativo	0,9	0,95	0,5	4954,41	0,045

Tabella 9 - H-W Serie mensile Azienda 5

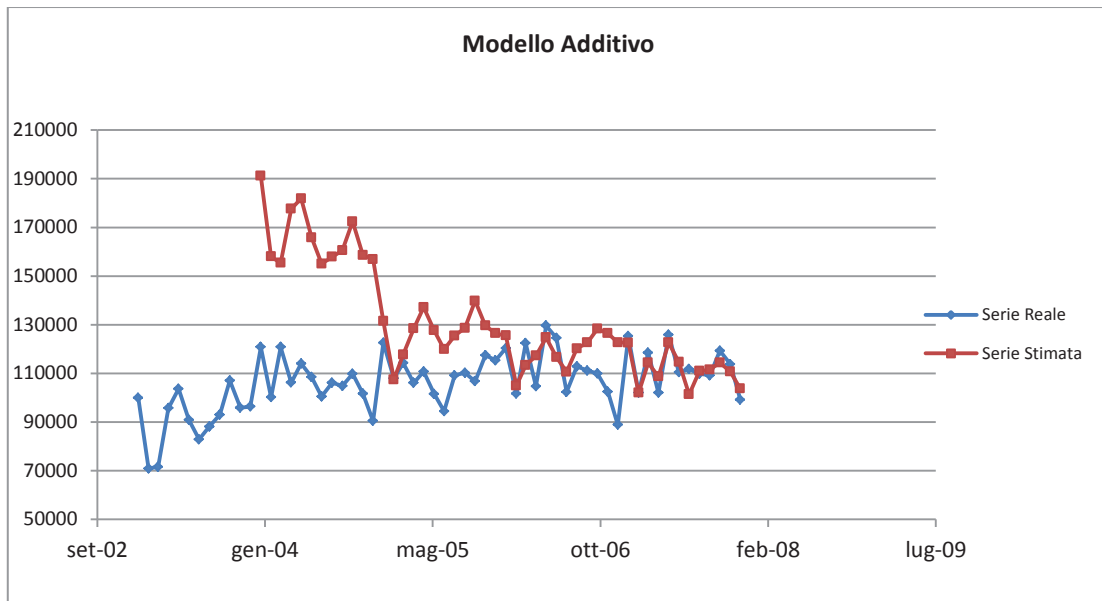


Grafico 22 - H-W Additivo Serie mensile Azienda 5

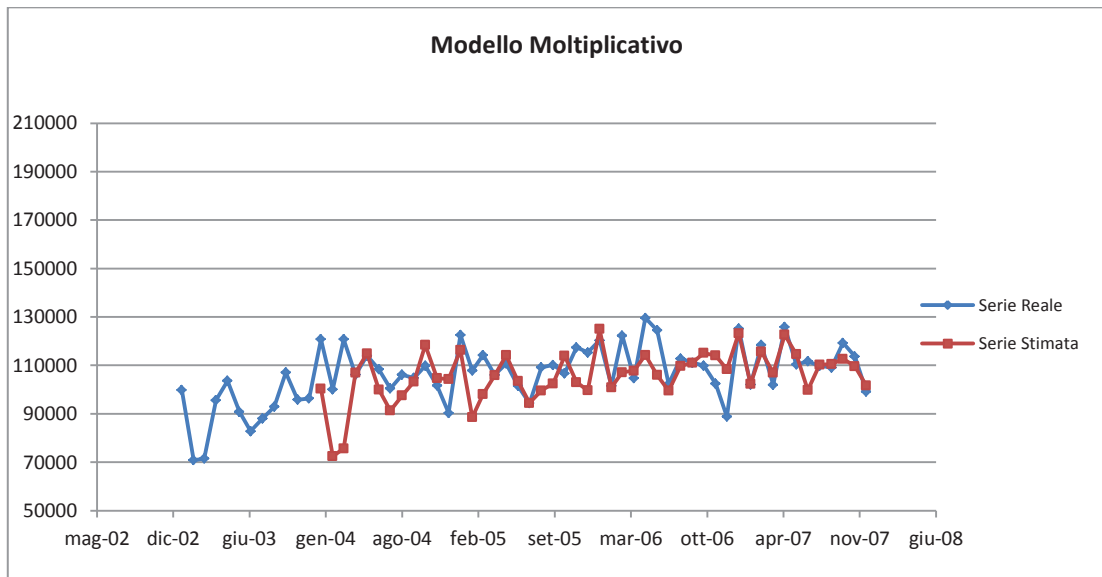


Grafico 23 - H-W Moltiplicativo Serie mensile Azienda 5

## Azienda 6

Modello	$\alpha$	$\beta$	$\gamma$	MQE	Indice di Theil
Additivo	0,9	0,95	0,5	35917,47	0,049
Moltiplicativo	0,2	0,3	0,05	34887,95	0,048

Tabella 10 - H-W Serie mensile Azienda 6

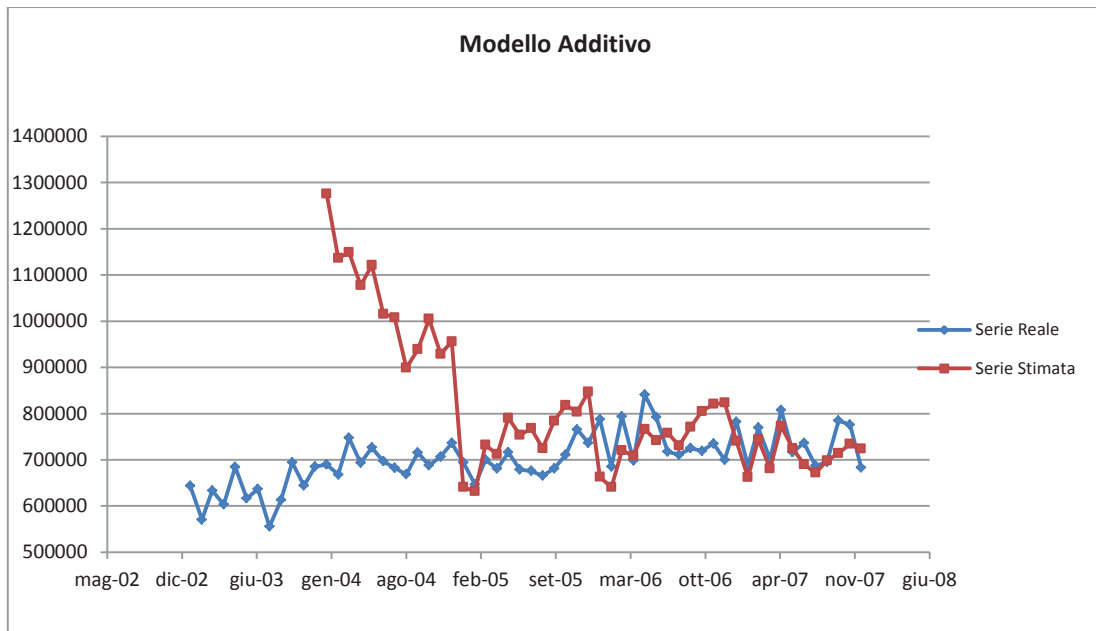


Grafico 24 - H-W Additivo Serie mensile Azienda 6

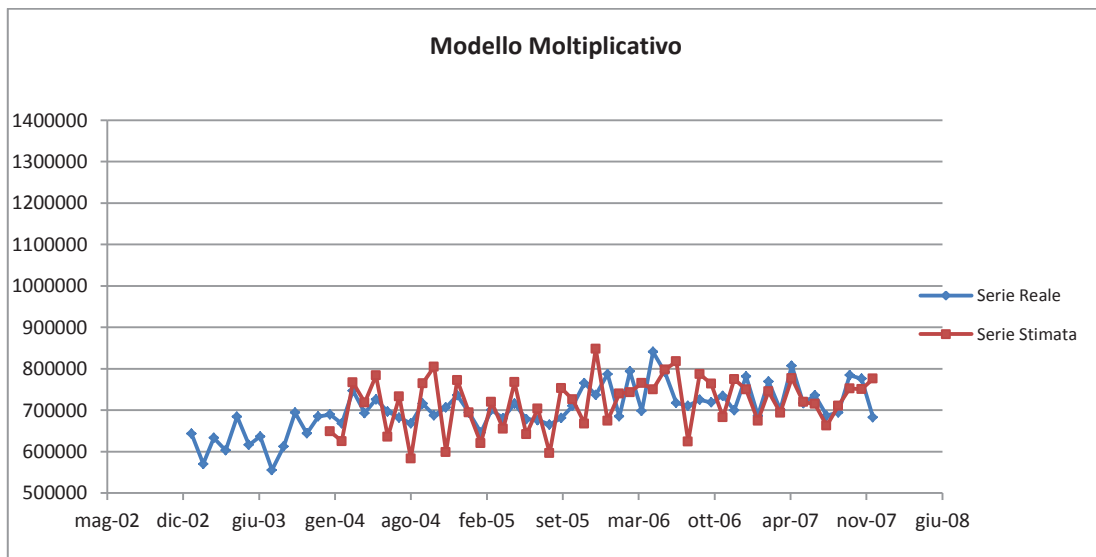


Grafico 25 - H-W Moltiplicativo Serie mensile Azienda 6

Riusciamo ad ottenere delle previsioni coerenti anche dividendo il dataset per azienda sanitaria. Infatti l'indice di Theil è molto vicino a 0 per tutte le serie. Le

differenze tra modello additivo e moltiplicativo sono lievi per quanto riguarda le previsioni, a volte l'MQE è a favore del metodo additivo, mentre aumentano se prendiamo in considerazione l'andamento generale della serie. Questo è evidente per le aziende 3, 4, 5 e 6 dove possiamo osservare, con il metodo additivo, una notevole divergenza tra serie perequata e serie reale, soprattutto per il 2004. La serie perequata con il metodo moltiplicativo, invece, segue bene l'andamento della serie reale. Le anomalie che riscontriamo durante il 2004 nella serie prevista sono dovute all'andamento insolito della serie reale nel 2006. In ultima analisi è quindi da preferire il metodo moltiplicativo.

## **3.4 APPLICAZIONE MODELLI SARIMA**

### **3.4.1 Serie totale Pazienti**

Dato che il procedimento utilizzato è generale e può essere utilizzato per tutte le serie, saranno mostrati tutti i passaggi solo per la prima serie, ossia quella complessiva, per le rimanenti invece saranno mostrati solo i risultati. Per poter procedere all'identificazione del modello dobbiamo rendere il processo stazionario. Dalle analisi effettuate precedentemente sulla serie abbiamo notato la presenza di un trend e di stagionalità, che vanno quindi eliminati. La serie per ciò risulta non stazionaria in media. Questo risulta evidente anche dai correlogrammi riportati di seguito.

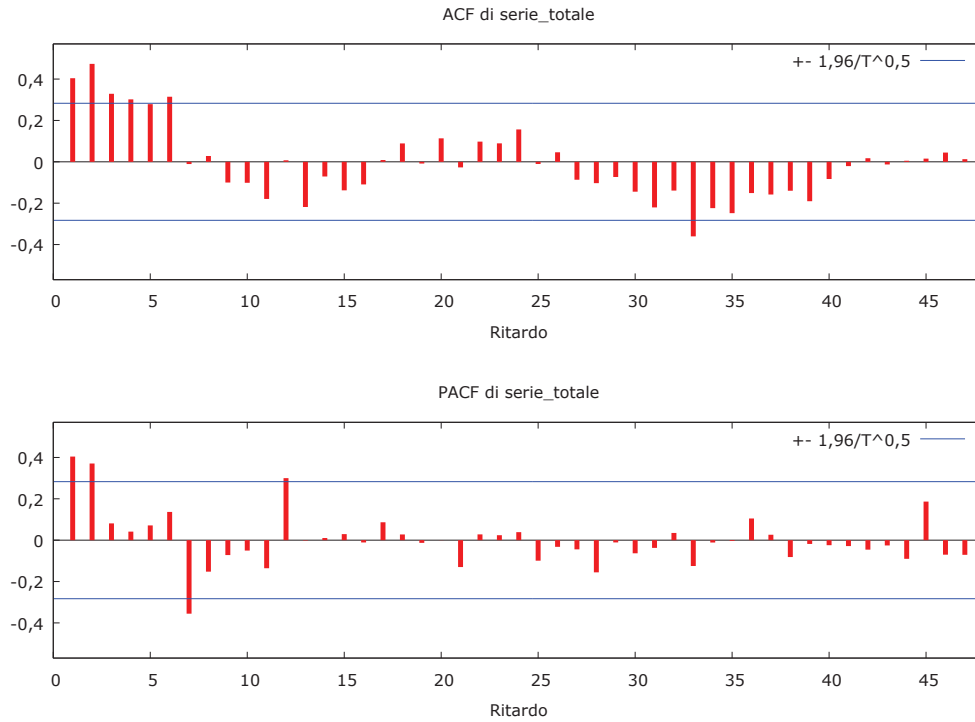


Grafico 26 - Correlogramma Serie totale Pazienti

Un metodo per eliminare il trend consiste nella differenziazione della serie. In seguito a questa modifica la serie si presenta come nel grafico 27.

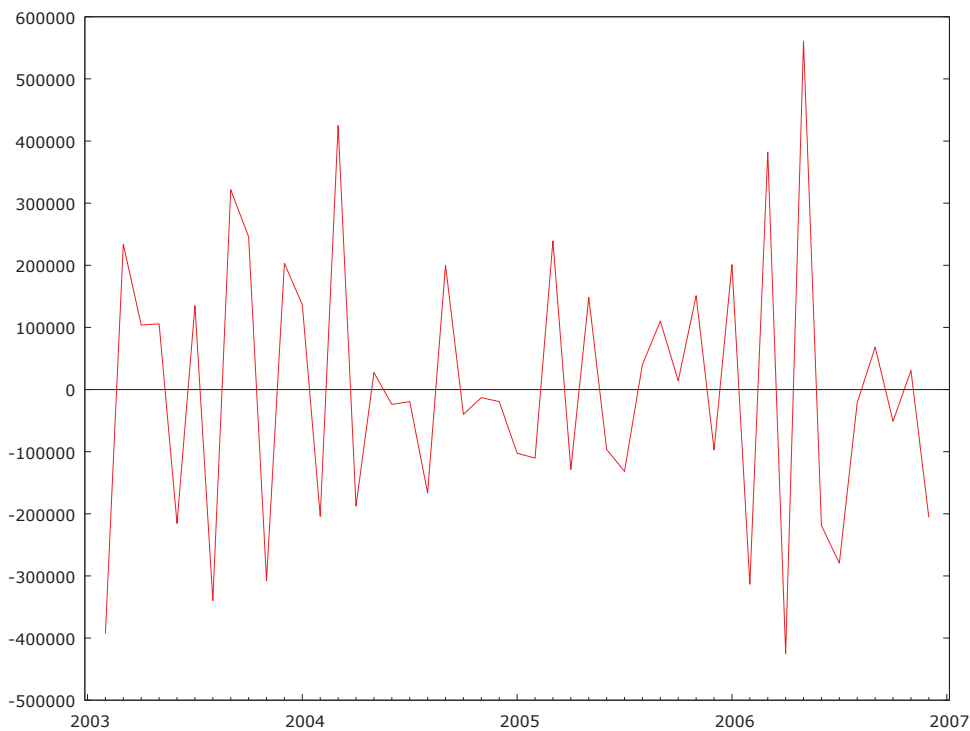


Grafico 27 - Serie totale Pazienti differenziata

Ora la serie è resa stazionaria e possiamo passare alla prossima fase.

## Identificazione

In questa fase si cerca di identificare l'ordine dei parametri  $p$ ,  $d$  e  $q$  (anche  $P$ ,  $D$ ,  $Q$  in caso di modello stagionale). Per fare ciò dobbiamo utilizzare i grafici dell'autocorrelazione (ACF) e dell'autocorrelazione parziale (PACF) stimati sul campione.

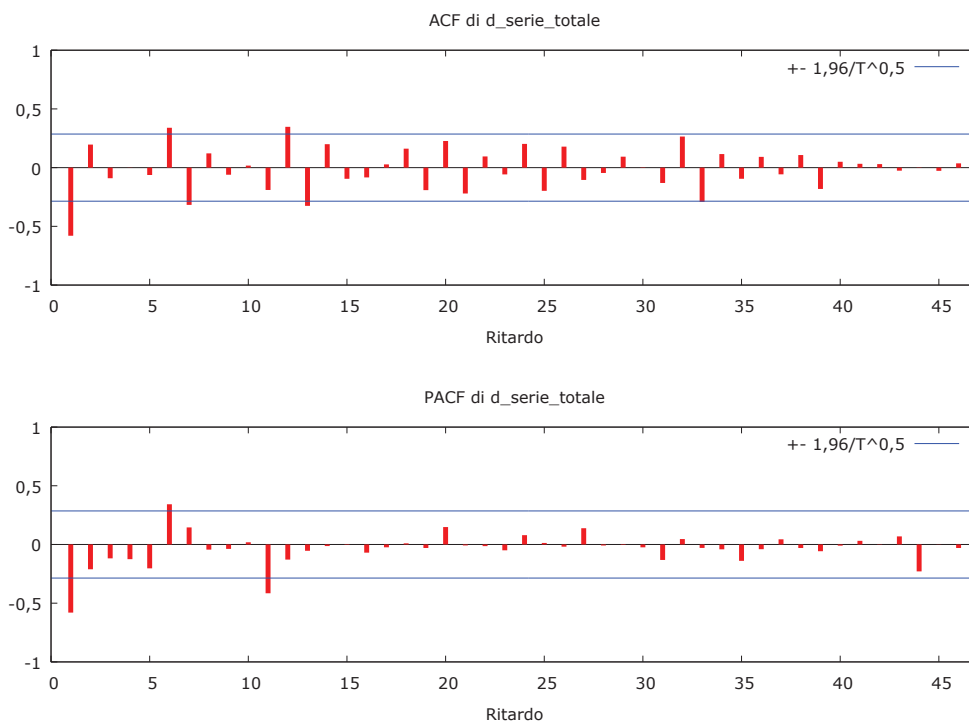


Grafico 28 - Correlogramma Serie totale Pazienti differenziata

Il ritardo 1 è significativo sia nell'ACF sia nell'PACF e tendono a 0 lentamente. Ipotizziamo perciò un modello ARMA(1,1). Per poter eseguire delle previsioni confrontabili con il metodo Holt-Winters applichiamo il modello alla serie non differenziata. Il modello quindi sarà un ARIMA(1,1,1). Utilizziamo il software gretl per stimarne i parametri e per verificarne la correttezza.

## Stima dei parametri e controllo diagnostico

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)					
Variabile dipendente: (1-L) serie_tot					
Errori standard basati sull'Hessiana					
	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>z</i>	<i>p-value</i>	
phi_1	-0,345254	0,201977	-1,7094	0,08738	*
theta_1	-0,468034	0,186801	-2,5055	0,01223	**
Media var. dipendente	-546,0000	SQM var. dipendente		219002,4	
Media innovazioni	8308,045	SQM innovazioni		165019,1	
Log-verosimiglianza	-631,6764	Criterio di Akaike		1269,353	
Criterio di Schwarz	1274,903	Hannan-Quinn		1271,441	
	<i>Reale</i>	<i>Immaginario</i>	<i>Modulo</i>	<i>Frequenza</i>	
AR					
Radice 1	-2,8964	0,0000	2,8964	0,5000	
MA					
Radice 1	2,1366	0,0000	2,1366	0,0000	



Grafico 29 - Serie totale Pazienti reale e stimata

I due parametri stimati risultano entrambi significativi. Il grafico 29 ci mostra i valori stimati dal modello con i valori reali. Dobbiamo ora analizzare i residui del modello.

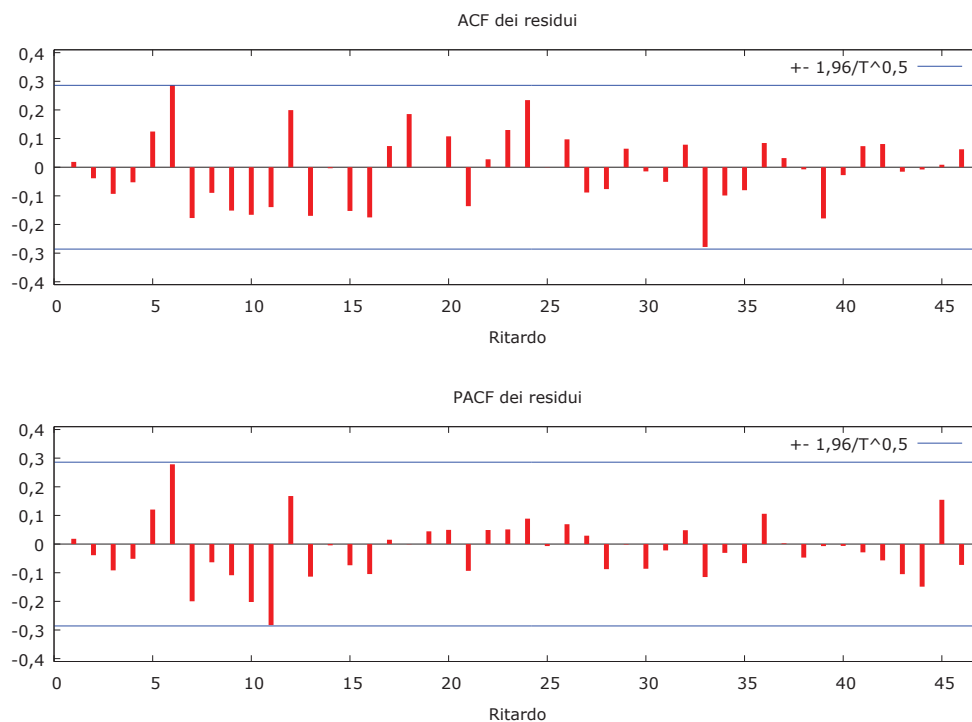


Grafico 30 - Correlogramma residui ARIMA(1,1,1)

I residui sono incorrelati, ma alcuni ritardi le autocorrelazioni globali e parziali sono significativamente diverse da 0. Pertanto proviamo ad introdurre una componente autoregressiva stagionale, stimando quindi un modello SARIMA(1,1,1)(1,0,0)<sub>12</sub>.



Modello 4: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)

Variabile dipendente: (1-L) serie\_tot

Errori standard basati sull'Hessiana

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>z</i>	<i>p-value</i>	
phi_1	-0,345786	0,209012	-1,6544	0,09805	*
Phi_1	0,503903	0,168972	2,9822	0,00286	***
theta_1	-0,333104	0,197315	-1,6882	0,09138	*
Media var. dipendente	-546,0000	SQM var. dipendente		219002,4	
Media innovazioni	-4917,653	SQM innovazioni		149098,1	
Log-verosimiglianza	-628,5600	Criterio di Akaike		1265,120	
Criterio di Schwarz	1272,521	Hannan-Quinn		1267,905	
	<i>Reale</i>	<i>Immaginario</i>	<i>Modulo</i>	<i>Frequenza</i>	
AR					
Radice 1	-2,8920	0,0000	2,8920	0,5000	
AR (stagionale)					
Radice 1	1,9845	0,0000	1,9845	0,0000	
MA					
Radice 1	3,0021	0,0000	3,0021	0,0000	

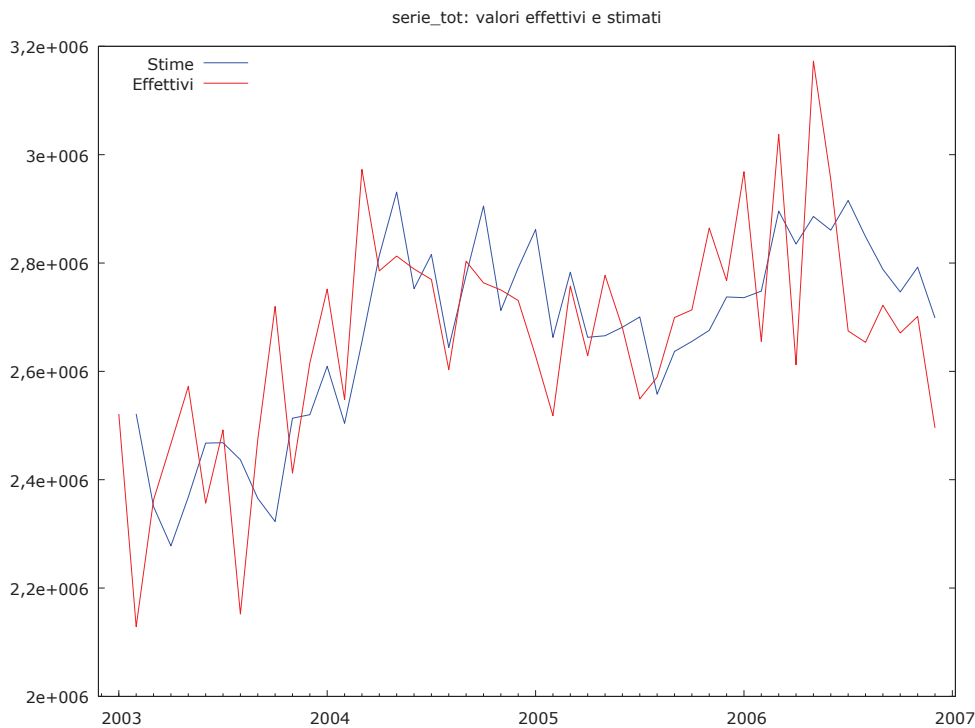


Grafico 31 - Serie totale Pazienti reale e stimata con SARIMA(1,1,1)(1,0,0)

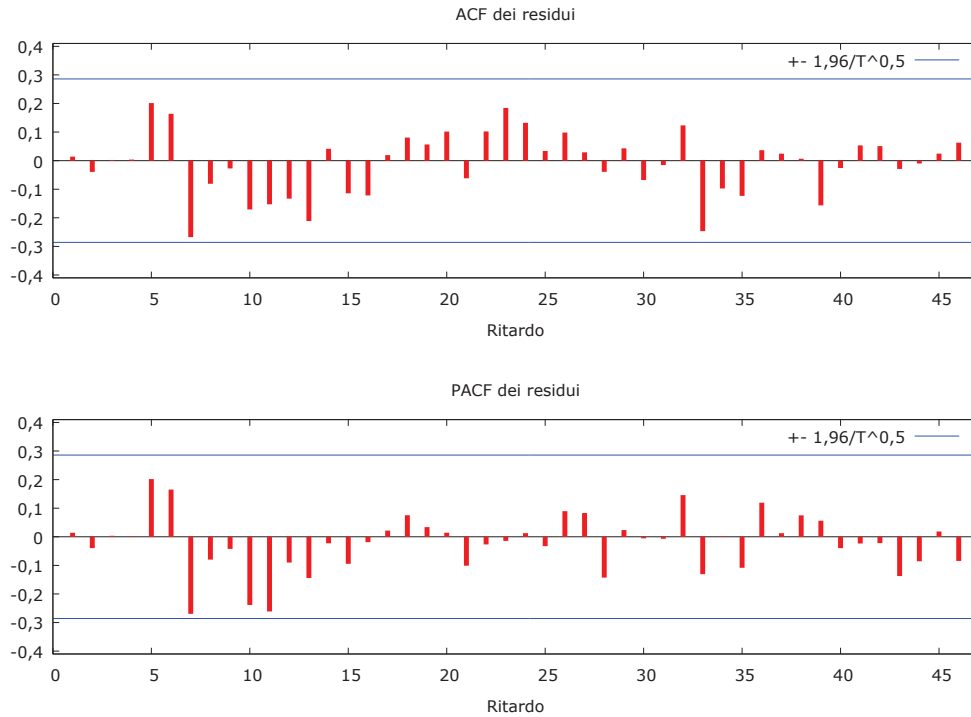


Grafico 32 - Correlogramma residui SARIMA(1,1,1)(1,0,0)

I parametri sono anche in questo caso tutti significativi. I residui ora sono tutti incorrelati e i valori delle autocorrelazioni rientrano tutti nelle barre di confidenza. Per entrambi i modelli stimati osserviamo lo scarso adattamento con la serie reale, visibile dai grafici 29 e 31.

Modello	AIC	SC
ARIMA(1,1,1)	1269,353	1274,903
SARIMA(1,1,1)(1,0,0) <sub>12</sub>	1265,120	1272,521

Tabella 11 - AIC e SC Serie totale Pazienti

La tabella 11 riassume i criteri utilizzati per il confronto tra modelli. Entrambi i criteri utilizzati portano, come ci si aspetterebbe, alla scelta del secondo modello.

## Previsioni

Utilizziamo ora i modelli stimati per effettuare delle previsioni per l'anno 2007, le quali saranno confrontate con i valori reali.



Grafico 33 - Previsioni Serie totale Pazienti ARIMA(1,1,1)



Grafico 34 - Previsioni Serie totale Pazienti SARIMA(1,1,1)(1,0,0)

Dai grafici 33 e 34 è evidente come la componente AR(1) stagionale introdotta nel secondo modello permetta di raggiungere dei risultati migliori in fase di predizione. Da notare che in tutti e due i casi le barre di confidenza per i valori predetti sono molto ampie. Questo significa che la varianza dell'errore è elevata. A conferma di quanto dedotto graficamente, riportiamo nella tabella 12 gli indici di valutazione delle previsioni.

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
ARIMA(1,1,1)	178740	0,066
SARIMA(1,1,1)(1,0,0) <sub>12</sub> .	153900	0,057

Tabella 12 - MQE e indice di Theil Serie totale Pazienti

In conclusione, dati i risultati riportati nelle tabelle 11 e 12 siamo propensi a scegliere il modello SARIMA(1,1,1)(1,0,0)<sub>12</sub>, il quale è migliore sia in fase di modellazione che in quella predittiva.

### **3.4.2 SERIE MENSILI DIVISE PER AZIENDA**

Per le restanti serie saranno presentati solo i risultati, mostrando per ogni serie le stime dei parametri, correlogramma dei residui, il grafico riportante valori reali e valori stimati e le previsioni. A richiesta saranno forniti tutti i passaggi intermedi che hanno portato alla scelta dei vari modelli.

## Azienda 1

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_1  
 Errori standard basati sull'Hessiana

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>z</i>	<i>p-value</i>	
Phi_1	0,511625	0,164579	3,1087	0,00188	***
theta_1	-0,571285	0,116933	-4,8856	<0,00001	***
Media var. dipendente	-695,0494	SQM var. dipendente		63581,71	
Media innovazioni	-1942,260	SQM innovazioni		44671,32	
Log-verosimiglianza	-571,9410	Criterio di Akaike		1149,882	
Criterio di Schwarz	1155,433	Hannan-Quinn		1151,971	

	<i>Reale</i>	<i>Immaginario</i>	<i>Modulo</i>	<i>Frequenza</i>
AR (stagionale)				
Radice 1	1,9546	0,0000	1,9546	0,0000
MA				
Radice 1	1,7504	0,0000	1,7504	0,0000

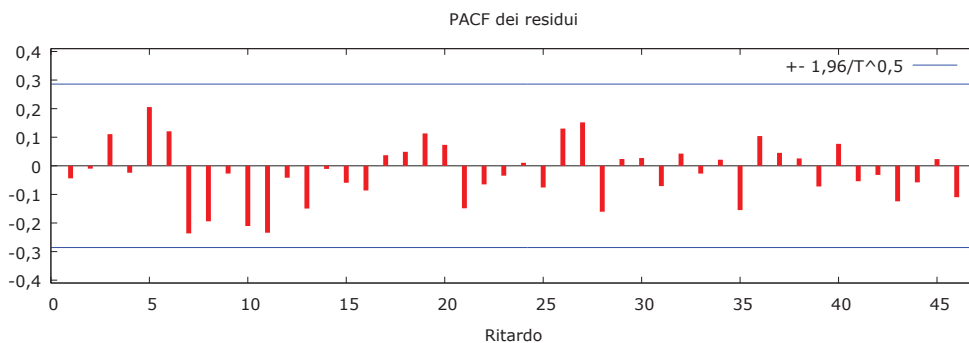
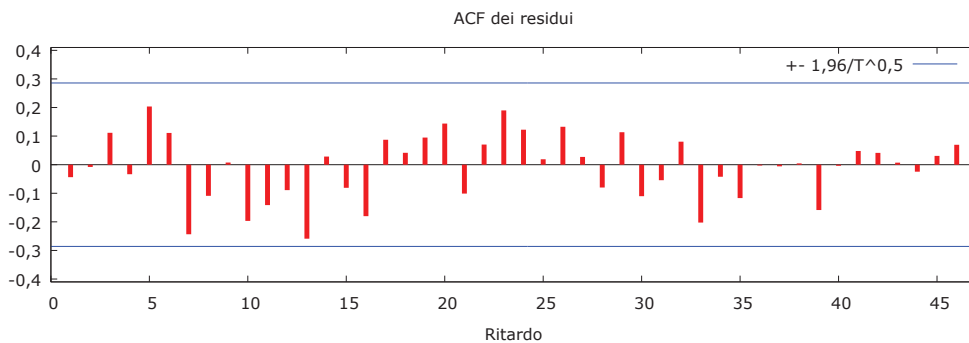


Grafico 35 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 1

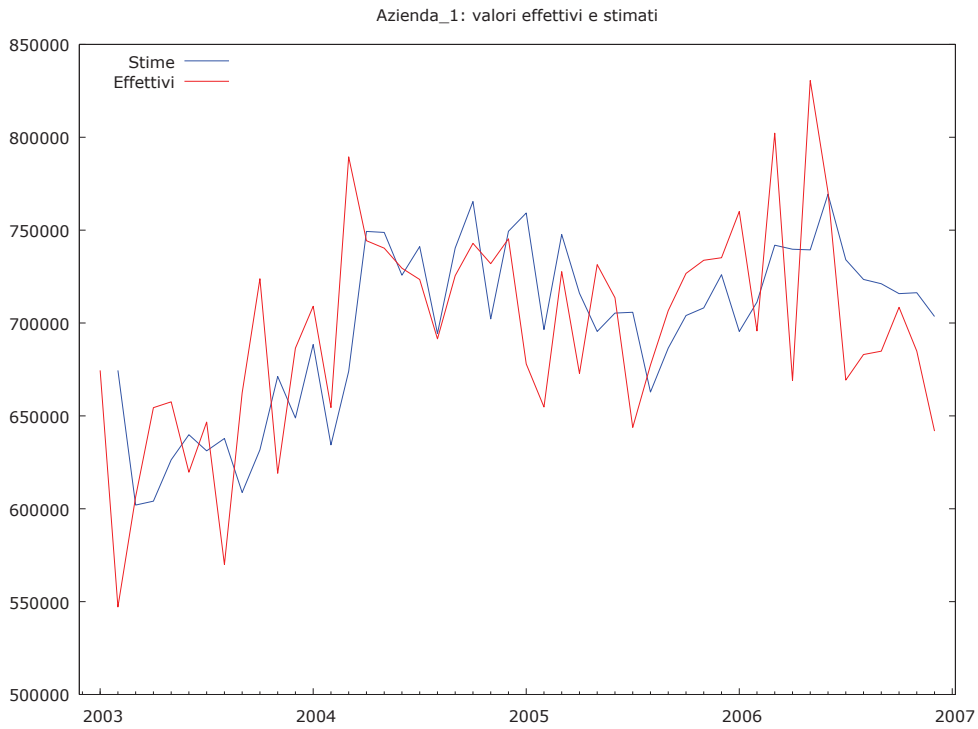


Grafico 36 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 1

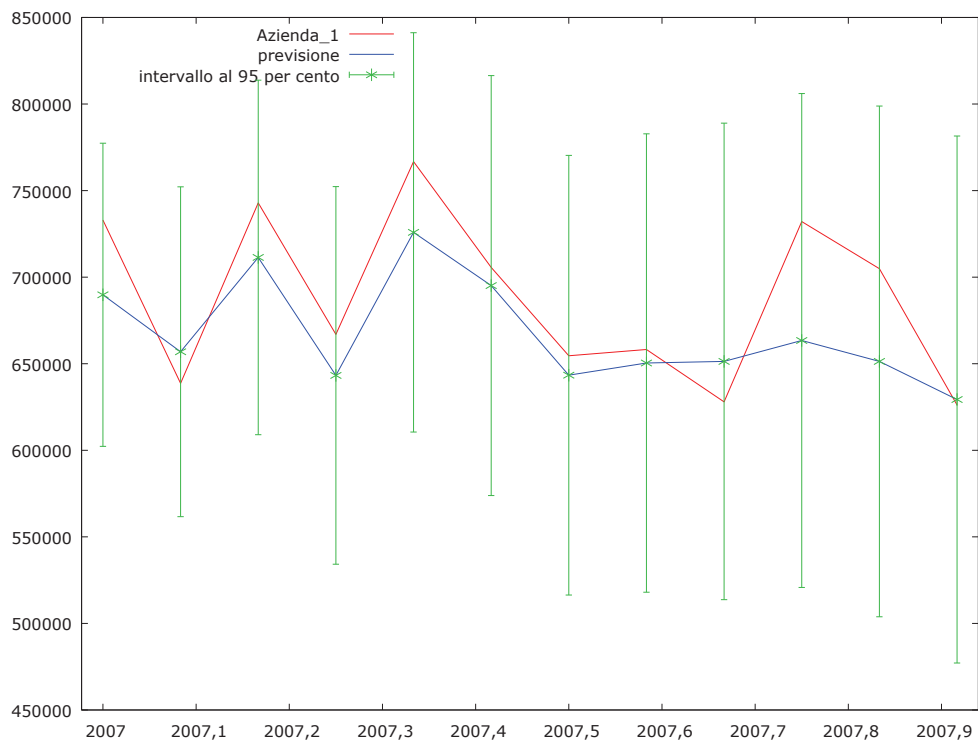


Grafico 37 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 1

Modello	MQE	Indice di Theil
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	33980	0,049

Tabella 13 - Indici Previsioni SARIMA(0,1,1)(1,0,0)

## Azienda 2

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_2  
 Errori standard basati sull'Hessiana

	Coefficiente	Errore Std.	z	p-value	
theta_1	-0,592318	0,11806	-5,0171	<0,00001	***
Theta_1	0,438032	0,18011	2,4320	0,01501	**
Media var. dipendente	-288,4984	SQM var. dipendente		33060,62	
Media innovazioni	-245,4740	SQM innovazioni		22843,07	
Log-verosimiglianza	-539,8924	Criterio di Akaike		1085,785	
Criterio di Schwarz	1091,335	Hannan-Quinn		1087,873	

	Reale	Immaginario	Modulo	Frequenza
MA				
Radice 1	1,6883	0,0000	1,6883	0,0000
MA (stagionale)				
Radice 1	-2,2829	0,0000	2,2829	0,5000

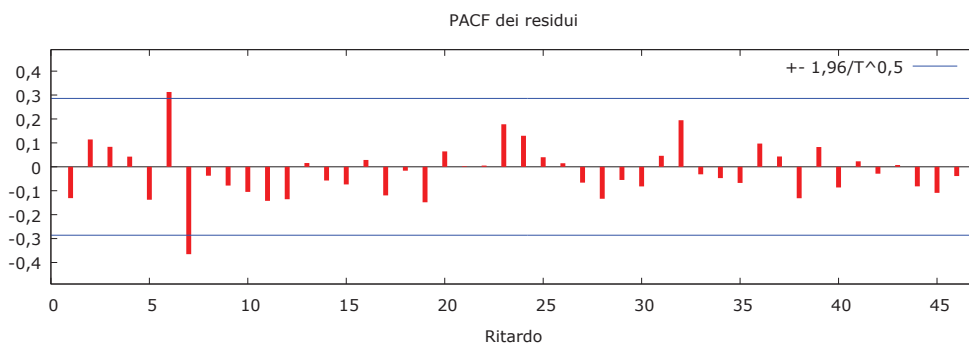
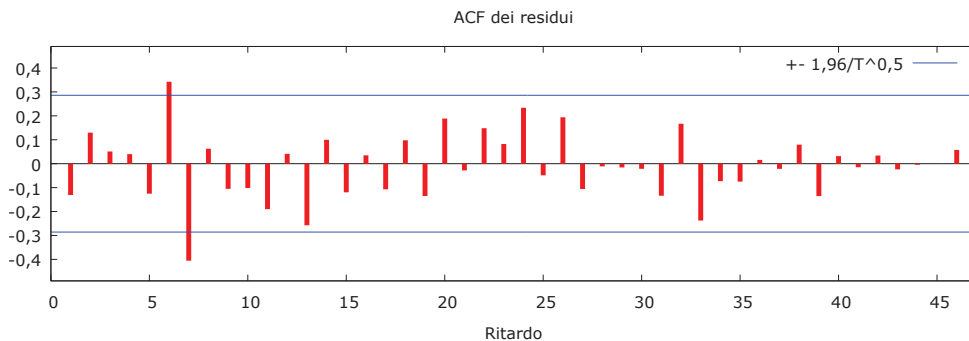


Grafico 38 - Correlogramma residui SARIMA(0,1,1)(0,0,1) Azienda 2

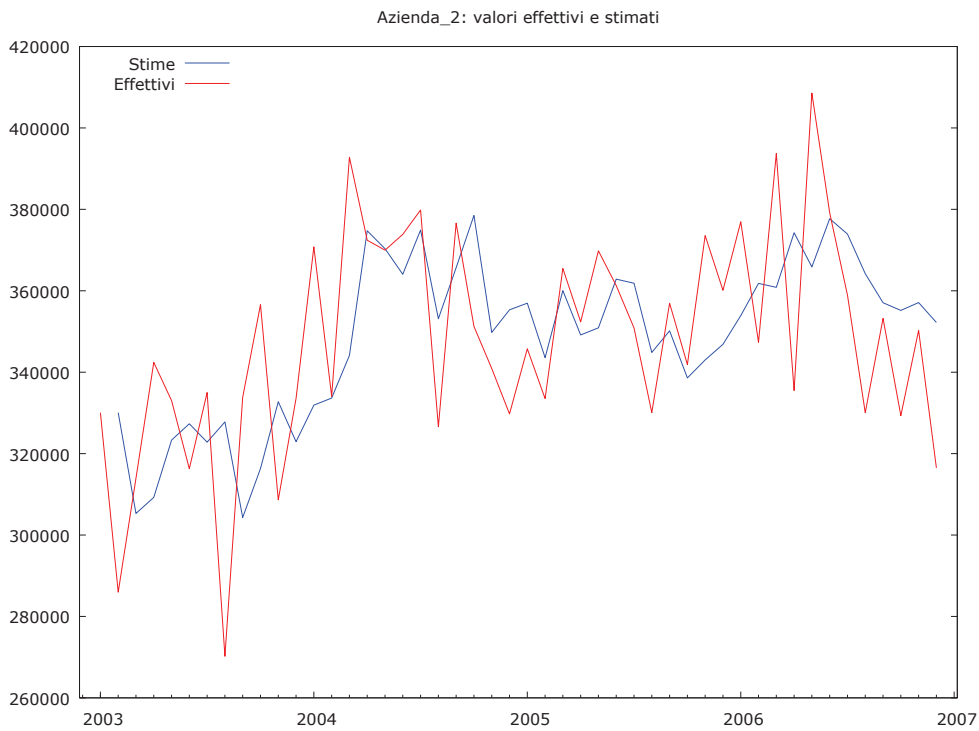


Grafico 39 - Serie reale e stimata SARIMA(0,1,1)(0,0,1) Azienda 2

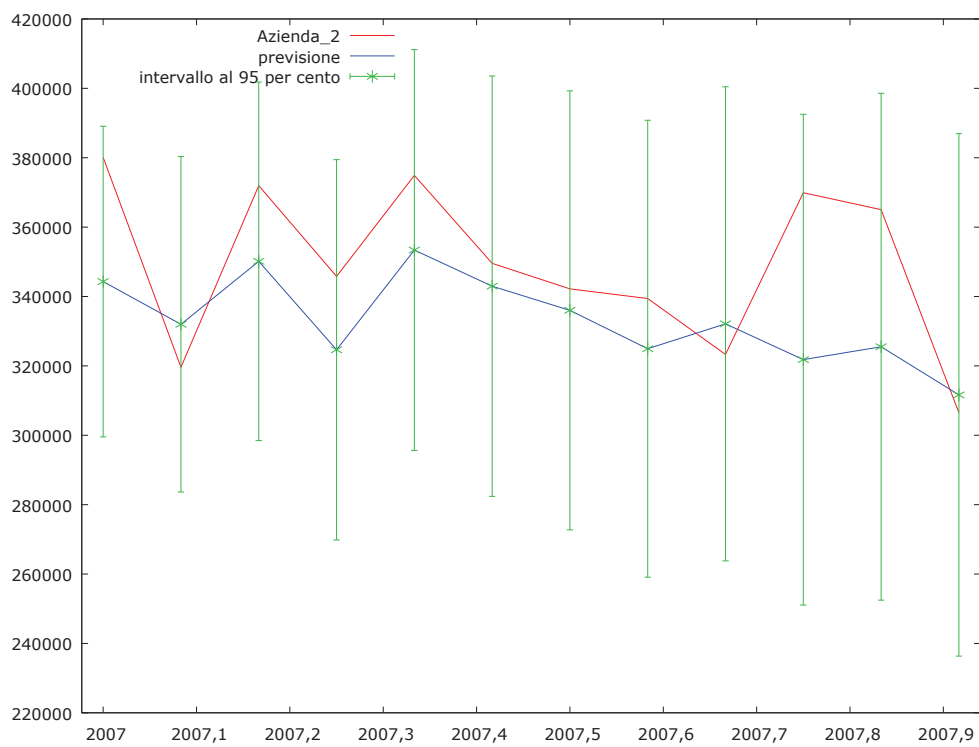


Grafico 40 - Previsioni SARIMA(0,1,1)(0,0,1)

Modello	MQE	Indice di Theil
SARIMA(0,1,1)(0,0,1) <sub>12</sub>	24314	0,070

Tabella 14 - Indici previsioni SARIMA(0,1,1)(0,0,1)



### Azienda 3

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_3  
 Errori standard basati sull'Hessiana

	Coefficiente	Errore Std.	z	p-value	
Phi_1	0,399888	0,198754	2,0120	0,04422	**
theta_1	-0,564446	0,116431	-4,8479	<0,00001	***
Media var. dipendente	153,6474	SQM var. dipendente		11475,26	
Media innovazioni	256,6071	SQM innovazioni		8808,443	
Log-verosimiglianza	-494,8499	Criterio di Akaike		995,6998	
Criterio di Schwarz	1001,250	Hannan-Quinn		997,7884	

	Reale	Immaginario	Modulo	Frequenza
AR (stagionale)				
Radice 1	2,5007	0,0000	2,5007	0,0000
MA				
Radice 1	1,7716	0,0000	1,7716	0,0000

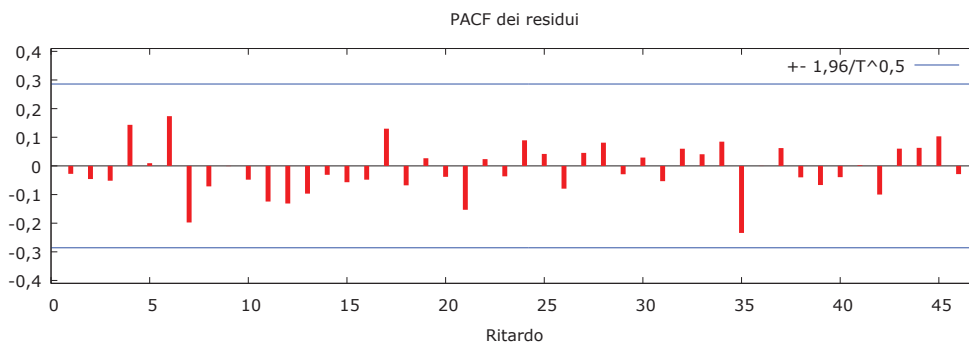
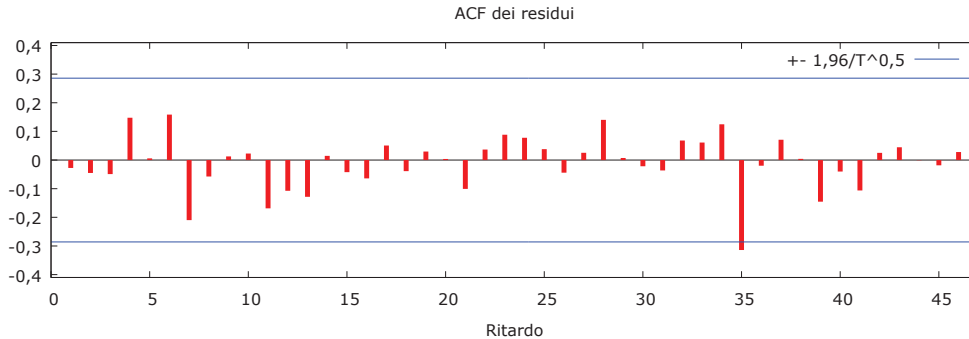


Grafico 41 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 3

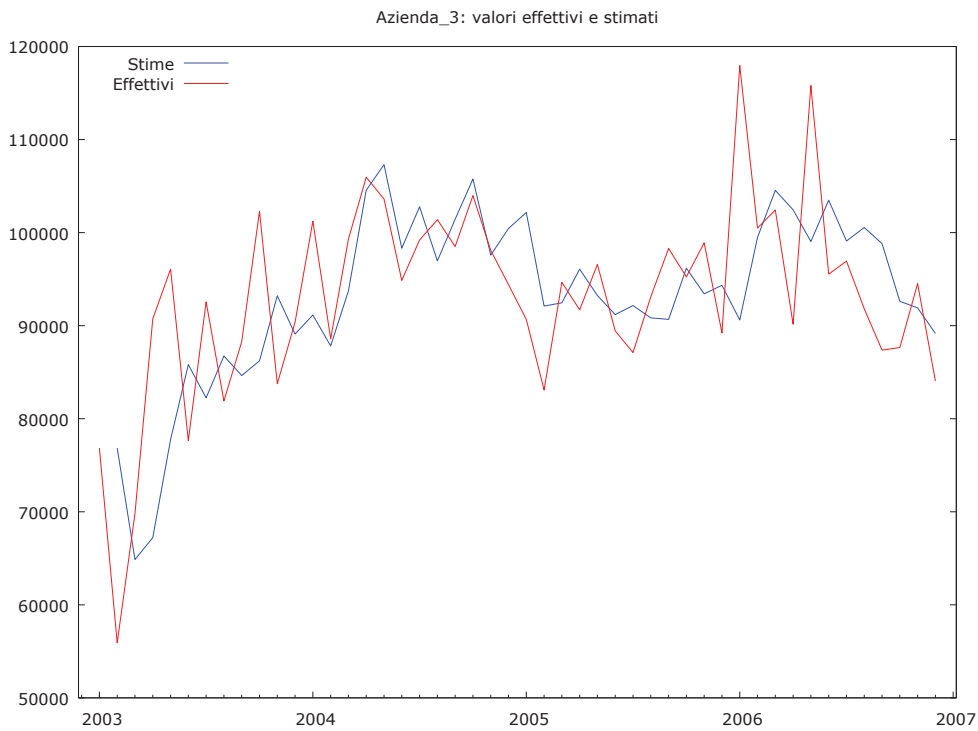


Grafico 42 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 3

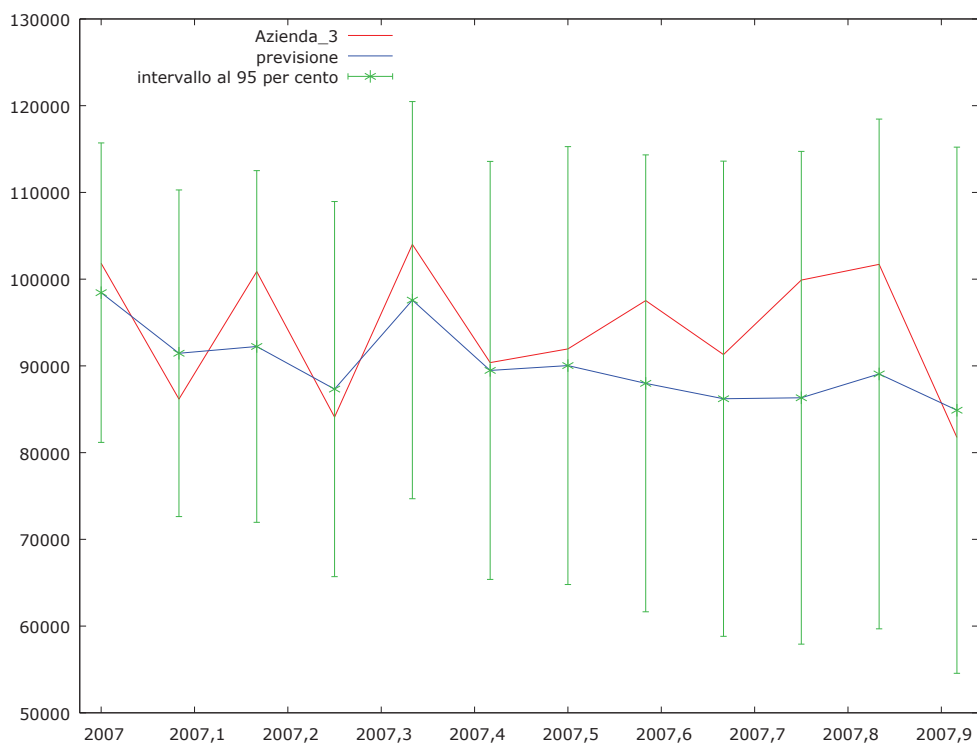


Grafico 43 - Previsioni SARIMA(0,1,1)(1,0,0)

Modello	MQE	Indice di Theil
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	7308,7	0,077

Tabella 15 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 3

## Azienda 4

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_4  
 Errori standard basati sull'Hessiana

	Coefficiente	Errore Std.	z	p-value	
phi_1	-0,453566	0,187623	-2,4174	0,01563	**
Phi_1	0,350557	0,183624	1,9091	0,05625	*
theta_1	-0,29402	0,193096	-1,5227	0,12784	
Media var. dipendente	-566,7754	SQM var. dipendente		63179,90	
Media innovazioni	-849,7446	SQM innovazioni		43585,59	
Log-verosimiglianza	-569,8391	Criterio di Akaike		1147,678	
Criterio di Schwarz	1155,079	Hannan-Quinn		1150,463	

		Reale	Immaginario	Modulo	Frequenza
AR	Radice 1	-2,2048	0,0000	2,2048	0,5000
AR (stagionale)	Radice 1	2,8526	0,0000	2,8526	0,0000
MA	Radice 1	3,4011	0,0000	3,4011	0,0000

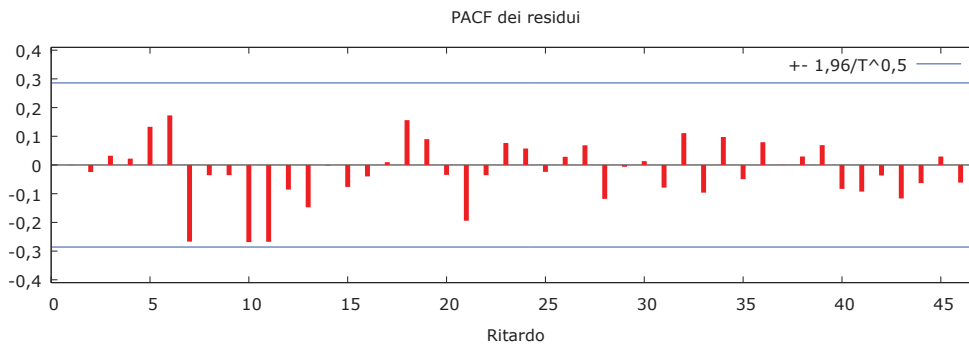
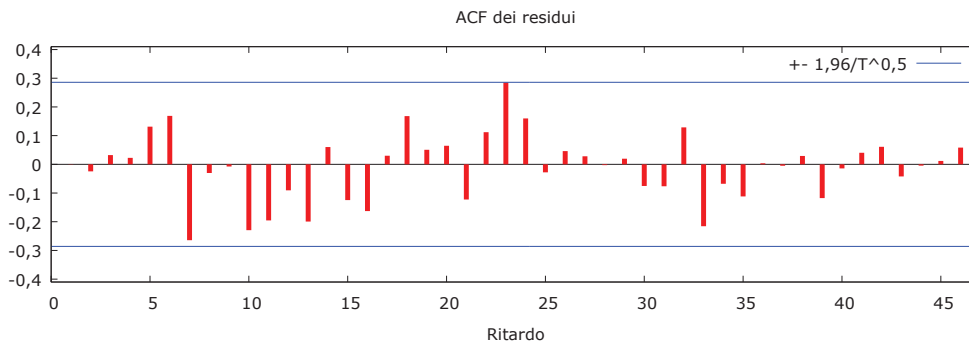


Grafico 44 - Correlogramma residui SARIMA(1,1,1)(1,0,0) Azienda 4

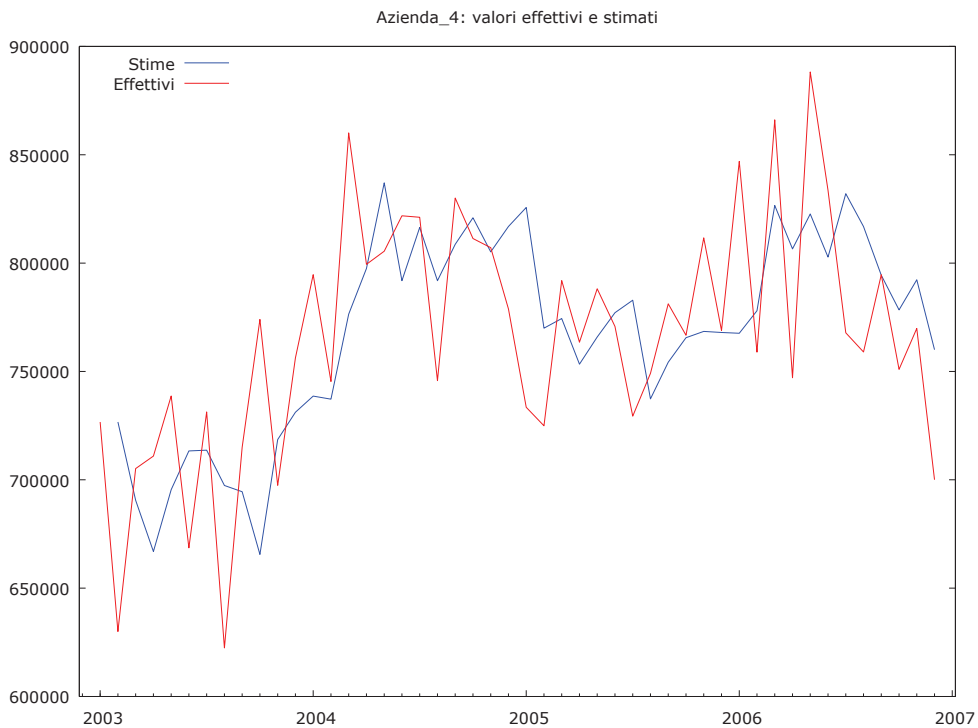


Grafico 45 - Serie reale e stimata SARIMA(1,1,1)(1,0,0) Azienda 4

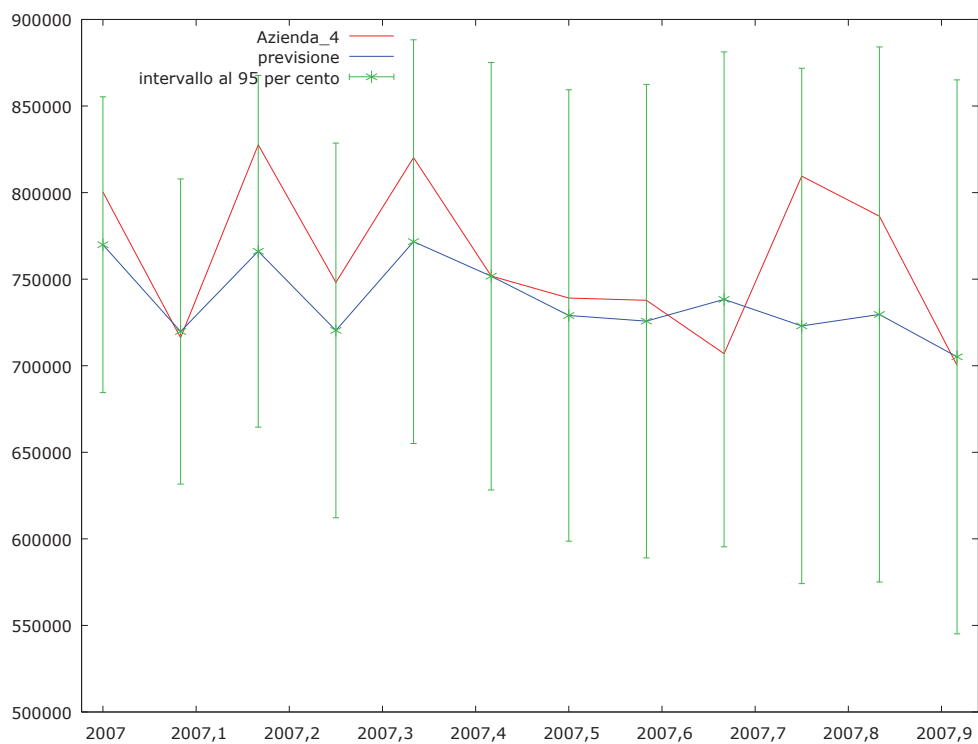


Grafico 46 - Previsioni SARIMA(1,1,1)(1,0,0) Azienda 4

Modello	MQE	Indice di Theil
SARIMA(1,1,1)(1,0,0) <sub>12</sub>	40627	0,053

Tabella 16 - Indici previsioni SARIMA(1,1,1)(1,0,0) Azienda 4

## Azienda 5

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_5  
 Errori standard basati sull'Hessiana

	Coefficiente	Errore Std.	z	p-value	
Phi_1	0,400837	0,175812	2,2799	0,02261	**
theta_1	-0,625394	0,133073	-4,6996	<0,00001	***
Media var. dipendente	-234,9815	SQM var. dipendente		13716,33	
Media innovazioni	-284,0376	SQM innovazioni		10311,05	
Log-verosimiglianza	-502,3133	Criterio di Akaike		1010,627	
Criterio di Schwarz	1016,177	Hannan-Quinn		1012,715	

	Reale	Immaginario	Modulo	Frequenza
AR (stagionale)				
Radice 1	2,4948	0,0000	2,4948	0,0000
MA				
Radice 1	1,5990	0,0000	1,5990	0,0000

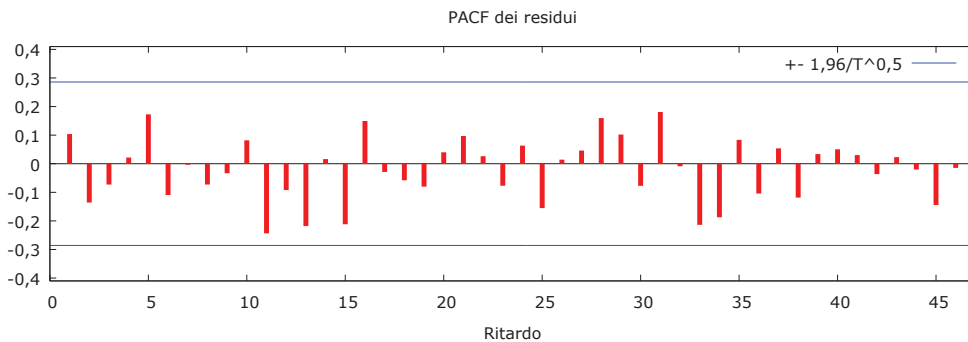
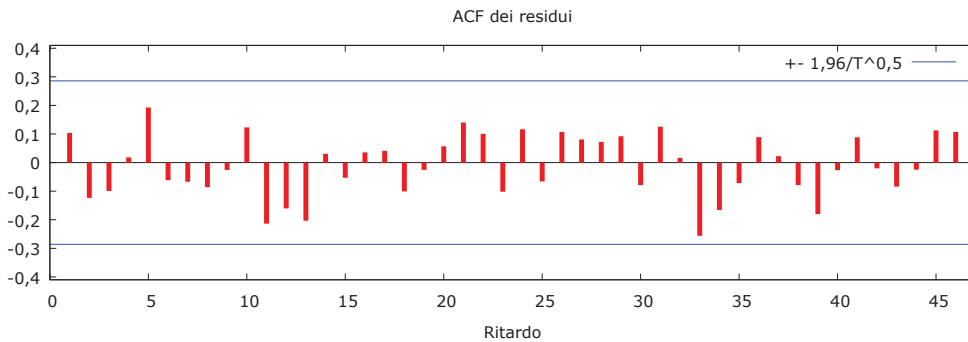


Grafico 47 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 5

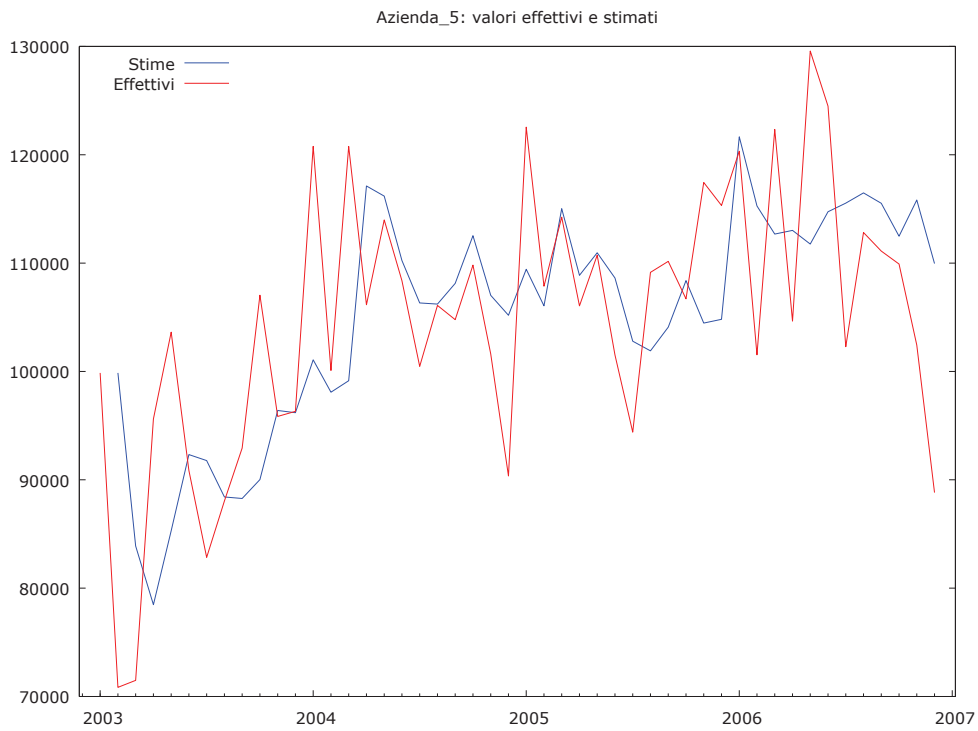


Grafico 48 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda5

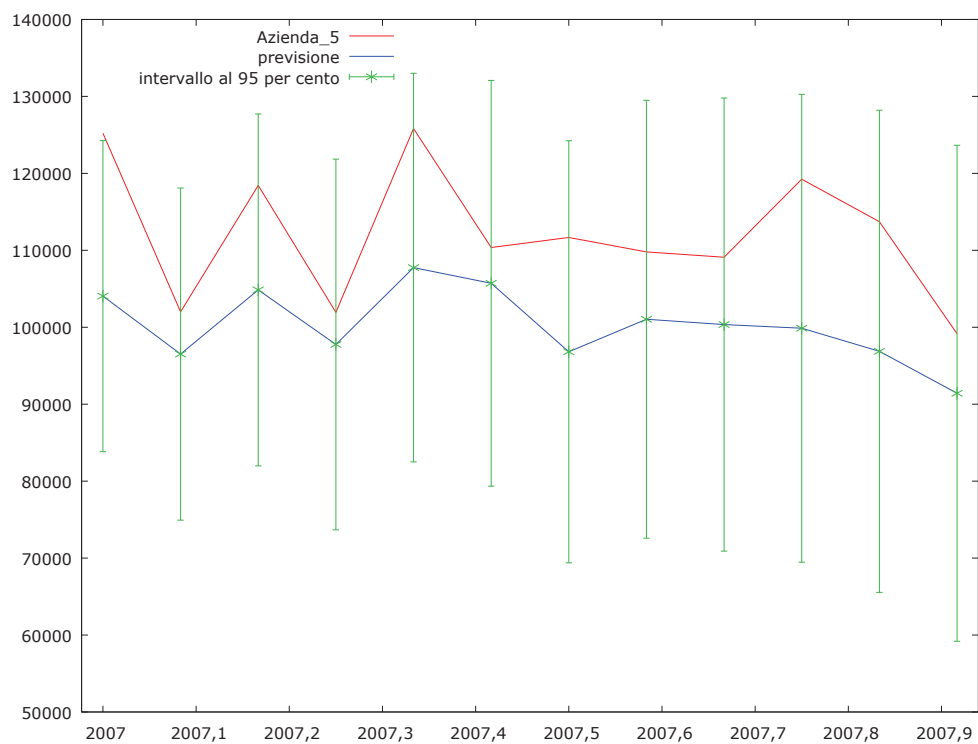


Grafico 49 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 5

Modello	MQE	Indice di Theil
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	13295	0,118

Tabella 17 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 5

## Azienda 6

Modello 1: ARIMA, usando le osservazioni 2003:02-2006:12 (T = 47)  
 Variabile dipendente: (1-L) Azienda\_6  
 Errori standard basati sull'Hessiana

	Coefficiente	Errore Std.	z	p-value	
Phi_1	0,478597	0,170853	2,8012	0,00509	***
theta_1	-0,590056	0,11575	-5,0977	<0,00001	***
Media var. dipendente	1187,743	SQM var. dipendente		54841,48	
Media innovazioni	1522,577	SQM innovazioni		38971,30	
Log-verosimiglianza	-565,2814	Criterio di Akaike		1136,563	
Criterio di Schwarz	1142,113	Hannan-Quinn		1138,651	

	Reale	Immaginario	Modulo	Frequenza
AR (stagionale)				
Radice 1	2,0894	0,0000	2,0894	0,0000
MA				
Radice 1	1,6948	0,0000	1,6948	0,0000

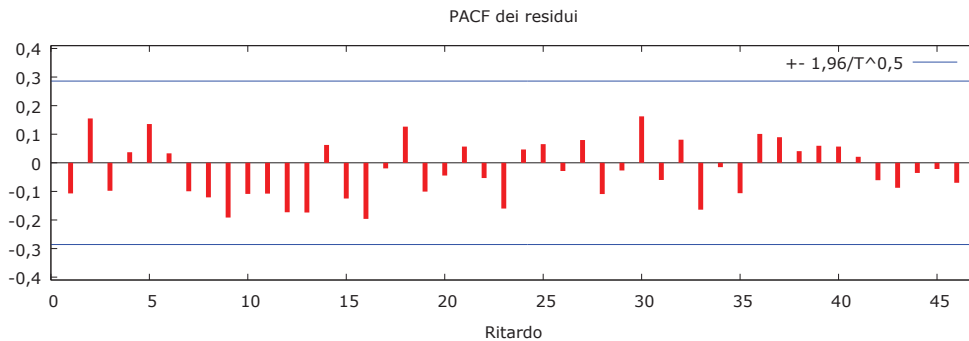
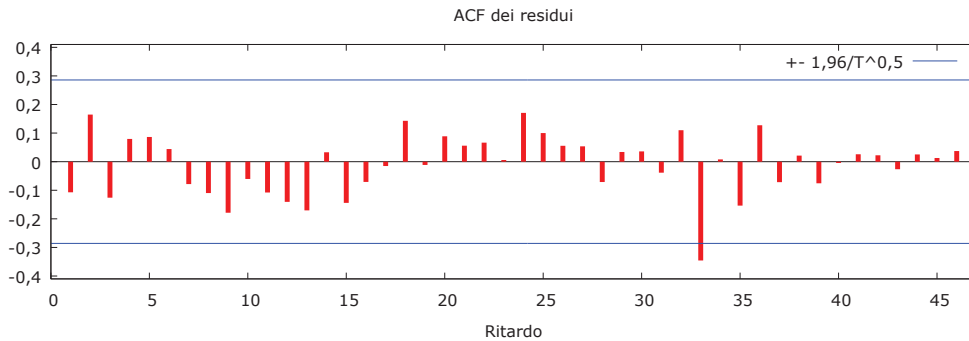


Grafico 50 - Correlogramma residui SARIMA(0,1,1)(1,0,0) Azienda 6

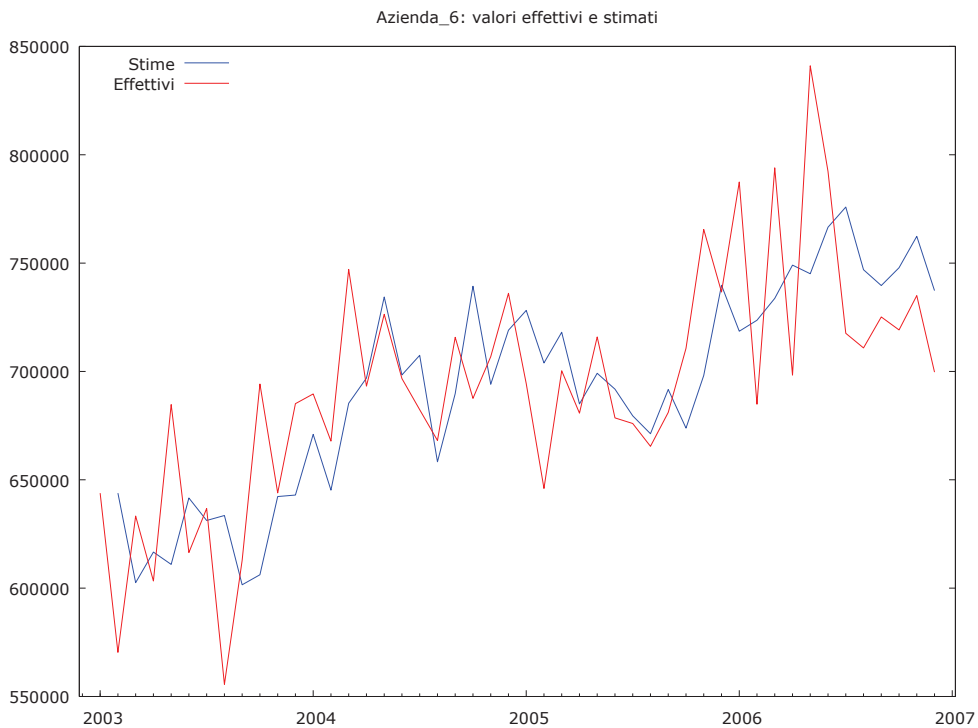


Grafico 51 - Serie reale e stimata SARIMA(0,1,1)(1,0,0) Azienda 6



Grafico 52 - Previsioni SARIMA(0,1,1)(1,0,0) Azienda 6

Modello	MQE	Indice di Theil
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	34355	0,047

Tabella 18 - Indici previsioni SARIMA(0,1,1)(1,0,0) Azienda 6



I modelli stimati per le Aziende 1, 3, 5 e 6 sono identici. Come è stato detto in precedenza, le serie hanno tutte un andamento simile quindi è normale aspettarsi un risultato di questo tipo. Per quanto riguarda le previsioni, le consideriamo accettabili tutte tranne che per l'Azienda 5. In questo caso il modello sottostima parecchio la serie reale. Come per la serie totale anche le previsioni divise per azienda presentano degli intervalli di confidenza piuttosto ampi.

## Capitolo 4

### CONSIDERAZIONI FINALI

Come precisato nell'introduzione di questo elaborato, le previsioni congiunturali dei costi sono importanti anche per un'Azienda sanitaria. In questo elaborato abbiamo proposto principalmente due metodi di previsione, uno più semplice ed immediato ed uno più complesso. Tutti i modelli identificati tendono a sottostimare i costi soprattutto nella parte finale del 2007. Le previsioni sono risultate più accurate utilizzando il metodo Holt-Winters. Questo deriva dal metodo con cui è stato costruito il modello, dando più importanza alla capacità predittiva a scapito della modellazione. Con i modelli SARIMA invece si cerca di modellare la componente stocastica della serie. Si sono ottenuti dei modelli meno potenti in fase predittiva ma migliori in quella di modellazione. In conclusione riportiamo tutti i risultati ottenuti.

#### Serie totale Pazienti

Modello	MQE	Indice di Theil
H-W Stagionale Additivo	95627,94	0,035
H-W Stagionale Moltiplicativo	95427,22	0,035
SARIMA(1,1,1)(1,0,0) <sub>12</sub>	153900	0,057

Tabella 19 – Riepilogo previsioni serie totale Pazienti

#### Azienda 1

Modello	MQE	Indice di Theil
H-W Stagionale Additivo	22508,21	0,033
H-W Stagionale Moltiplicativo	25733,62	0,038
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	33980	0,049

Tabella 20 - Riepilogo previsioni Azienda 1

### **Azienda 2**

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
H-W Stagionale Additivo	18568,57	0,053
H-W Stagionale Moltiplicativo	19054,82	0,055
SARIMA(0,1,1)(0,0,1) <sub>12</sub>	24314	0,070

Tabella 21 - Riepilogo previsioni Azienda 2

### **Azienda 3**

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
H-W Stagionale Additivo	6515,91	0,069
H-W Stagionale Moltiplicativo	4803,95	0,051
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	7308,7	0,077

Tabella 22 - Riepilogo previsioni Azienda 3

### **Azienda 4**

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
H-W Stagionale Additivo	29497,71	0,039
H-W Stagionale Moltiplicativo	29909,22	0,039
SARIMA(1,1,1)(1,0,0) <sub>12</sub>	40627	0,053

Tabella 23 - Riepilogo previsioni Azienda 4

### **Azienda 5**

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
H-W Stagionale Additivo	4711,92	0,042
H-W Stagionale Moltiplicativo	4954,41	0,045
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	13295	0,118

Tabella 24 - Riepilogo previsioni Azienda 5

## **Azienda 6**

<b>Modello</b>	<b>MQE</b>	<b>Indice di Theil</b>
H-W Stagionale Additivo	35917,47	0,049
H-W Stagionale Moltiplicativo	34887,95	0,048
SARIMA(0,1,1)(1,0,0) <sub>12</sub>	34355	0,047

Tabella 25 - Riepilogo previsioni Azienda 6

## **BIBLIOGRAFIA**

**Di Fonzo T, Lisi F.,** *Serie storiche economiche. Analisi statistiche e applicazioni.* Carrocci Editore, 2005

<http://www.regione.fvg.it>

## **PACCHETTI STATISTICI**

Gretl 1.9.11 (2012-11-21) <http://gretl.sourceforge.net/>

R version 2.15.2 (2012-10-26)

Copyright (C) 2012 The R Foundation for Statistical Computing

ISBN 3-900051-07-0 <http://www.r-project.org/>