

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN STATISTICA E GESTIONE DELLE IMPRESE

RELAZIONE FINALE:

**ANALISI E PREVISIONI DI SERIE STORICHE RELATIVE AL
TRAFFICO DI TELEFONIA MOBILE**

RELATORE: PROF. BISAGLIA LUISA

LAUREANDO: FRANCESCO VINCIS

ANNO ACCADEMICO: 2005-2006

INDICE

CAPITOLO 1 – PRESENTAZIONE

- | | | | |
|-----|--------------|----|---|
| 1.1 | Introduzione | p. | 5 |
| 1.2 | L'azienda | “ | 7 |

CAPITOLO 2 - VIDEOMESSAGGI

- | | | | |
|-----|---------------------------------------|---|---|
| 2.1 | Analisi del traffico di videomessaggi | “ | 9 |
|-----|---------------------------------------|---|---|

CAPITOLO 3 – TRAFFICO VOCE

- | | | | |
|-----|-----------------------------------|---|----|
| 3.1 | Introduzione | “ | 17 |
| 3.2 | Modelli ARIMA stagionali (SARIMA) | “ | 19 |
| 3.3 | Modelli di <i>Holt – Winters</i> | “ | 21 |
| 3.4 | Analisi del traffico voce | “ | 25 |
| 3.5 | Previsione traffico voce | “ | 37 |

CAPITOLO 4 - CONCLUSIONI

- | | | | |
|-----|-----------------------|---|----|
| 4.1 | Considerazioni finali | “ | 41 |
|-----|-----------------------|---|----|

1. PRESENTAZIONE

1.1 INTRODUZIONE

Questo lavoro consiste nell'analisi di due casi di studio differenti: il primo riguardante il traffico di videomessaggi inviati dai clienti della compagnia telefonica "3" nell'arco di un intero anno, il secondo il traffico voce dei clienti della stessa compagnia nell'arco di sei mesi.

Il primo studio, di tipo esclusivamente descrittivo, ha come scopo quello di comprendere come le campagne virali attuate dall'azienda influiscono sul traffico di videomessaggi inviati dai clienti. In questo caso il risultato dell'analisi porterà ad una scelta di tipo strategico: continuare con le campagne virali perché nel breve (o nel lungo periodo) influenzano positivamente il numero di videomessaggi inviati giornalmente, oppure smettere perché non hanno alcun effetto empirico. Nel caso del traffico di videomessaggi, questo tipo di analisi è molto utile, in quanto esso è un settore ancora in evoluzione ed è difficile trarre conclusioni sulle tendenze di fondo perché si hanno solo pochi anni come riferimento.

Il secondo studio consiste nell'analizzare il trend della serie storica giornaliera dei minuti di chiamate, e quindi fare previsioni per il mese successivo per il mese successivo. In un contesto aziendale come questo, una previsione ben redatta risulta essere uno strumento molto importante: una buona stima del traffico voce futuro riduce notevolmente il livello di incertezza con cui il management si troverà a prendere decisioni strategiche e fornisce un vantaggio competitivo nei confronti dei concorrenti.

1.2 L'AZIENDA

“3” è un operatore di telefonia mobile che opera in Italia, ma oltre il 90% della società è di proprietà di una multinazionale cinese già presente sul mercato un po' in tutta Europa.

Nel mercato italiano della telefonia mobile “3” fa il suo ingresso il 12 Marzo 2003. A differenza degli altri operatori (Tim, Vodafone e Wind) non entra direttamente nel segmento dei cellulari di seconda generazione, i GSM, ma punta tutto sugli apparecchi di terza generazione, gli UMTS. Questa scelta strategica è fatta soprattutto con l'intenzione di evitare la concorrenza dei *leader* italiani del settore, i quali, con l'ottica di poter sfruttare l'effetto rete, non sfruttabile invece da “3” in quanto nuovo entrante, ritardano il loro ingresso nel segmento di mercato della nuova tecnologia.

Proprio a causa di un'assenza di effetto rete, gli sforzi iniziali economici di “3” sono stati ingenti: la società ha dovuto, infatti, costruire da zero una rete di ripetitori in tutta Italia (attualmente il servizio copre il 75% della popolazione italiana), affidandosi al *roaming* degli altri gestori per le zone non coperte. Inoltre, ha dovuto farsi conoscere ed accettare dai clienti del settore. Allo stesso tempo, tuttavia, ha potuto espandersi, praticamente indisturbata, in un settore privo di concorrenti, in quanto Tim e Vodafone vi sono entrati solo da pochi mesi. I risultati si sono avuti, soprattutto grazie a strategie volte a favorire il passaggio di clienti, ma anche di manager, da Tim o Vodafone: oggi i clienti “3” in Italia sono più di un milione. Naturalmente, una volta acquisiti clienti, il passo successivo è stato quello di produrre servizi multimediali, usufruibili solo con la

tecnologia UMTS, tali da far incrementare i volumi di traffico dei propri clienti, e in pochi anni anche questo risultato è stato raggiunto. Attualmente “3” non possiede i volumi di traffico voce dei rivali, ma sta recuperando velocemente terreno, soprattutto nei confronti di Tim.

Naturalmente bisogna tener conto anche del fatto che l’Italia è un caso a parte nel mondo della telefonia mobile: è il primo paese europeo per tasso di penetrazione dei servizi di telefonia cellulare, si parla, infatti, di 41 milioni di abbonati nel Paese.

2. VIDEOMESSAGGI

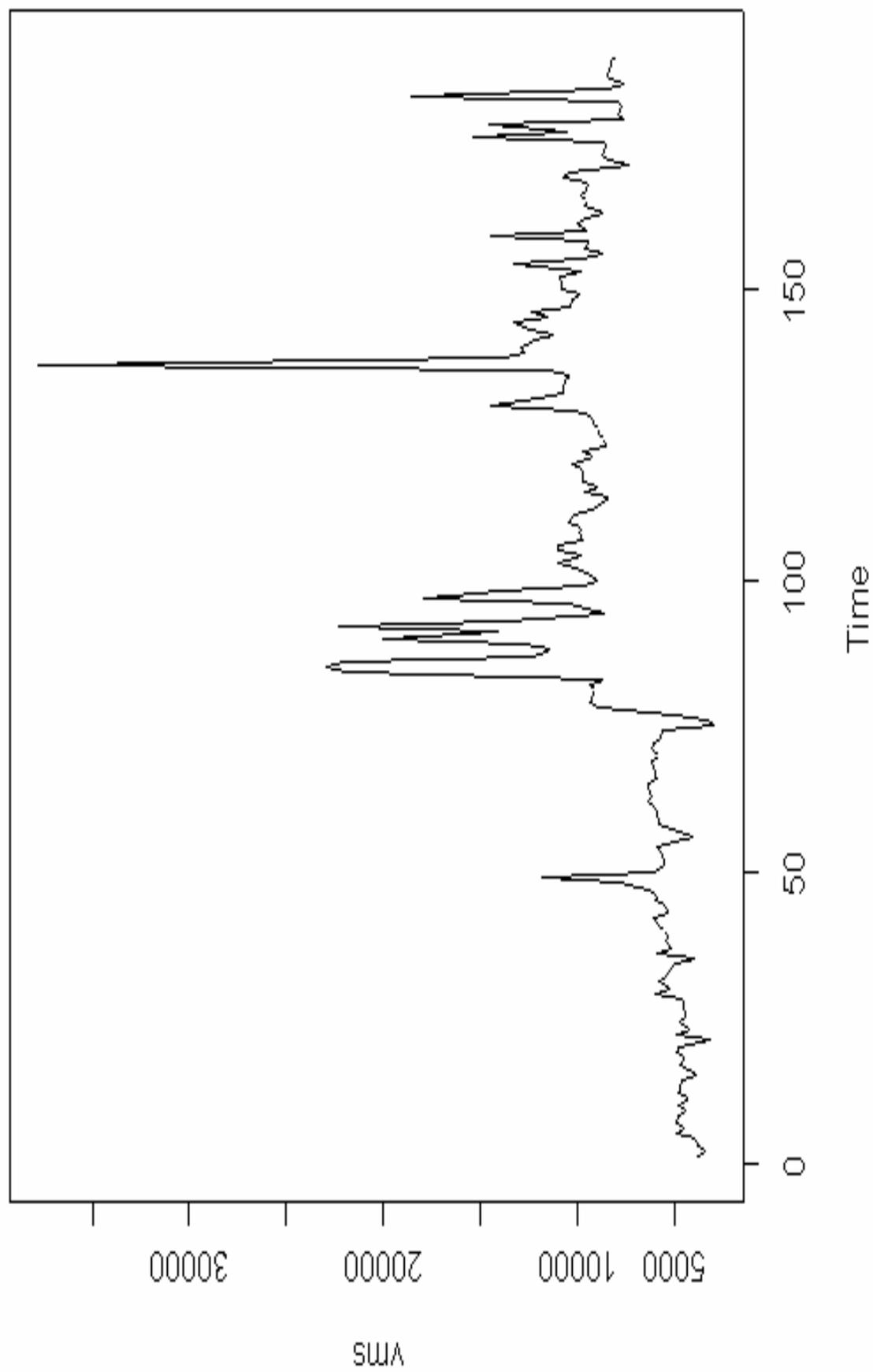
2.1 ANALISI DEL TRAFFICO DI VIDEOMESSAGGI

Lo studio riguardante il traffico di videomessaggi è stato effettuato su due serie storiche giornaliere per il periodo che va dal 01/10/04 al 08/04/05. La prima serie rappresenta il numero di videomessaggi inviati dai clienti di “3”, mentre la seconda riguarda i videomessaggi inviati da “3” ai propri clienti, in forma di azioni virali, messaggi informativi, o campagne pubblicitarie in collaborazione con altre aziende.

Scopo dello studio è stato quello di capire se e come le campagne informative e soprattutto quelle virali promosse dall’azienda abbiano modificato l’andamento, nel breve e/o nel lungo periodo, del traffico di videomessaggi da parte dei clienti.

Da una prima analisi grafica della serie dei videomessaggi inviati dagli abbonati, si può subito notare come la serie subisca un cambio di livello intorno alla 77-esima osservazione, per poi stabilizzarsi intorno a questo nuovo livello. Si può, inoltre, osservare la presenza di alcuni picchi tra l’80-esima e la 100-esima osservazione, ed uno molto evidente verso la 135-esima osservazione (vedi grafico 1). Inoltre, dall’analisi delle funzioni di autocorrelazione (vedi grafico 2), si può notare come non ci sia la presenza di stagionalità.

Videomessaggi (Grafico 1)



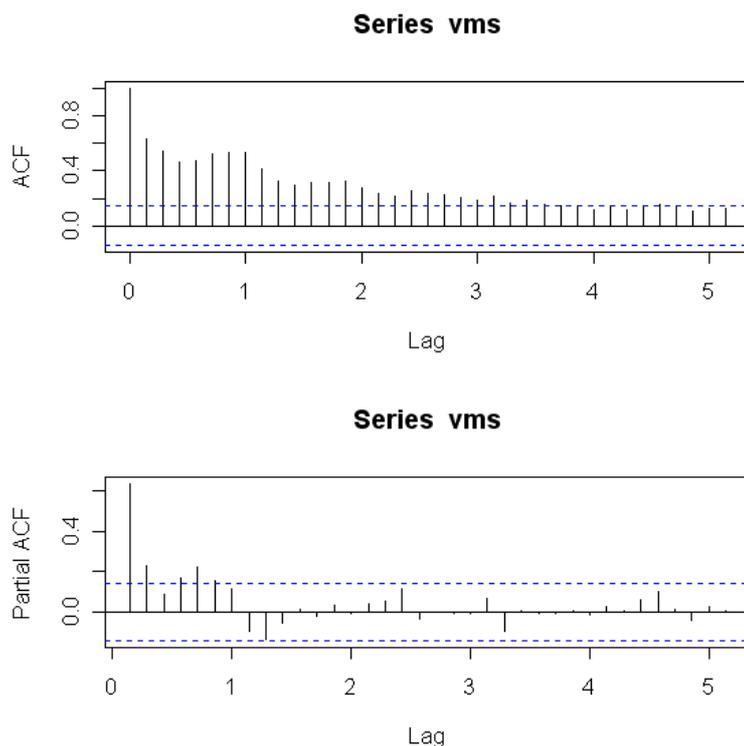


Grafico 2: funzioni di autocorrelazione empiriche della serie dei videomessaggi

Proviamo a capire, in maniera descrittiva, se effettivamente le campagne promosse da “3” raggiungono l’obiettivo prefissato. L’analisi empirica è stata effettuata tramite la procedura basata sul lavoro di Bai e Perron (1998). Essa consiste in una successione di verifiche d’ipotesi. Si comincia considerando l’assenza di *break* come ipotesi nulla e la presenza di *break* come ipotesi alternativa. Ogni volta che si trova un *break* il test si aggiorna e l’ipotesi nulla diventa la presenza di l punti di *break* contro $l+1$ *break*, dove $l = 1, 2, \dots, L$ con L numero massimo di *break*.

Il test è costruito tenendo conto della differente correlazione tra gli errori, delle diverse distribuzioni dei dati e degli errori tra regimi. La distribuzione asintotica relativa dipende dal parametro $\varepsilon = h/T$, dove T è la dimensione campionaria e h è la lunghezza minima accettabile per ogni regime.

Nella sua formulazione più generale, si considera una regressione multipla con m *break* ($m+1$ regimi):

$$y_t = x_t' \beta + z_t' \delta_j + u_t \quad t = T_{j-1}+1, \dots, T_j \quad (2.9)$$

per $j = 1, \dots, m+1$, dove

y_t è la variabile dipendente osservata al tempo t ;

x_t ($p \times 1$) e z_t ($q \times 1$) sono i vettori delle variabili indipendenti;

β e δ_j sono i corrispondenti vettori dei coefficienti;

u_t è l'errore al tempo t .

I valori (T_1, \dots, T_m) , ossia i punti di *break*, vengono considerati ignoti ($T_0 = 0$, $T_{m+1} = T$). Si tratta quindi di un modello a parziale cambiamento strutturale, poiché β non è soggetto ai salti di livello, mentre, se $p = 0$, si otterrà un puro modello a cambiamento strutturale, in quanto tutti i coefficienti variano con il *break*.

Lo scopo è stimare consistentemente i parametri $(\beta^0, \delta_1^0, \dots, \delta_m^0, T_1^0, \dots, T_m^0)$. Il metodo di stima è basato sul principio dei minimi quadrati, in cui per ogni m -esima partizione del campione (T_1, \dots, T_m) , le stime dei parametri si ottengono minimizzando la somma dei residui al quadrato (SSR) dell'equazione (2.9). La stima dei punti di *break* $(\hat{T}_1, \dots, \hat{T}_m)$ è data dal risultato dell'algoritmo:

$$(\hat{T}_1, \dots, \hat{T}_m) = \operatorname{argmin}_{T_1, \dots, T_m} S_T(T_1, \dots, T_m)$$

dove $S_T (T_1, \dots, T_m)$ è la SSR dell'equazione (2.9) per ogni m-esima partizione e la minimizzazione viene effettuata su tutte le partizioni (T_1, \dots, T_m) .

Se con la procedura descritta vengono individuati dei punti di cambiamento di regime corrispondenti ai periodi di attuazione delle campagne virali e informative, ciò può essere interpretato come un segnale di effettivo successo delle campagne stesse.

Nel nostro caso, la procedura identifica ben 5 punti di rottura, ed esattamente alle osservazioni: 77, 83, 92, 136 e 140.

All'osservazione 77 (ovvero il 16/12/04) non corrisponde nessuna azione da parte dell'azienda telefonica, ma assistiamo ad un semplice aumento dei volumi di traffico (peraltro già notato dopo l'analisi del grafico), situazione prevedibile in un mercato in pieno sviluppo come è quello dei videomessaggi.

Il secondo punto di rottura viene identificato all'osservazione 83, ovvero il 22 Dicembre, in questo caso proprio alla vigilia di una campagna virale, avvenuta il giorno seguente con l'invio da parte di "3" di 800.000 videomessaggi con gli auguri di Natale ai propri clienti. Naturalmente questa azione ha portato i volumi di scambi a livelli record, salendo per ben tre giorni (fino al giorno di Natale) oltre quota 20.000 videomessaggi, contro una media del periodo precedente di poco oltre i 9.000 videomessaggi al giorno. Il perché di una durata così limitata di tale effetto è chiaramente intuibile: gli auguri per il Natale non possono protrarsi oltre questo giorno.

Il terzo punto di rottura viene identificato alla 92-esima osservazione, ovvero il 31 Dicembre, due giorni dopo la seconda campagna virale, questa volta con gli auguri di Capodanno. Anche in questo caso è molto semplice intuire che il motivo di questo sfasamento temporale tra campagna virale ed effetto sia dovuto al fatto che di norma gli auguri di Capodanno si scambiano il 31 Dicembre o il primo Gennaio.

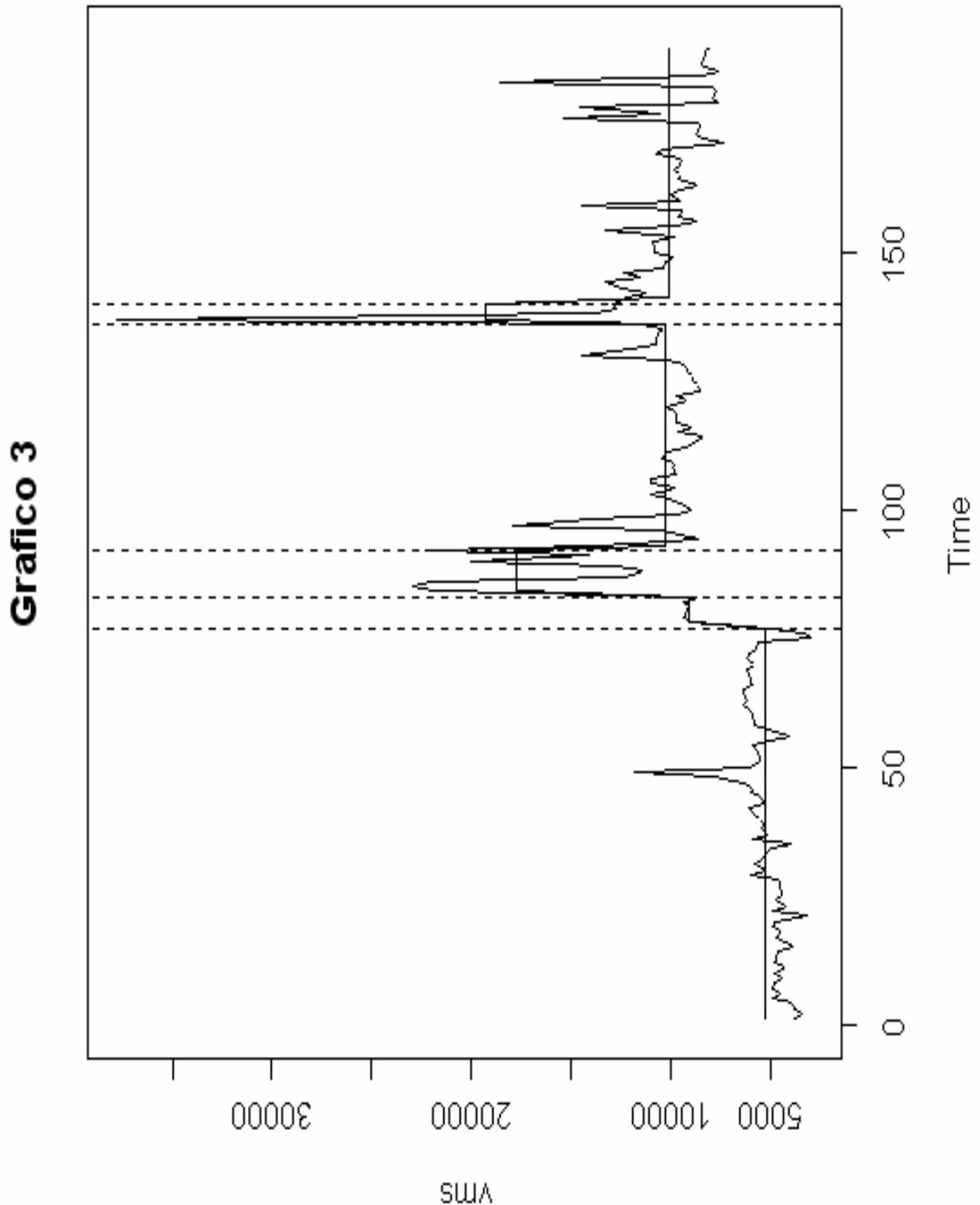
Nel periodo che va tra queste due campagne virali la media di videomessaggi inviati aumenta, assestandosi intorno ai 17700 videomessaggi al giorno. Successivamente la media scende, anche se si pone ad un livello maggiore di quella degli ultimi mesi del 2004: 10300 videomessaggi giornalieri contro i 5200 del trimestre Ottobre - Dicembre.

Gli ultimi punti di rottura corrispondono alla campagna virale effettuata in corrispondenza di San Valentino, dove per quattro giorni la media sale attorno ai 19000 videomessaggi giornalieri, per poi stabilizzarsi circa ai livelli pre-campagna (vedi grafico 3).

Tutto ciò fa pensare che le campagne virali abbiano avuto un loro effetto, in particolar modo le prime in corrispondenza delle festività natalizie. Queste infatti hanno portato ad un aumento della media della serie nel lungo periodo, mentre la campagna fatta per San Valentino ha un effetto limitato al breve periodo e non oltre. In effetti la media dei videomessaggi inviati è duplicata dall'ottobre 2004 all'aprile 2005, e ci si può aspettare un ulteriore aumento del traffico, come è stato per il traffico voce o degli sms, almeno finché il settore non sarà saturo.

Da sottolineare c'è, però, il fatto che la procedura non riesce ad individuare alcune campagne virali svolte dall'azienda come, per esempio, quelle attuate in prossimità dell'Epifania, del Carnevale, della Festa della

Donna, della Festa del Papà e del primo d'Aprile, in quanto festività minori. Se dovessimo prendere una decisione a livello manageriale su queste ultime campagne, forse sarebbe meglio sospenderle in quanto non offrono il ritorno sperato.



3. TRAFFICO VOCE

3.1 INTRODUZIONE

Nell'ottica delle imprese moderne, che si trovano a dover convivere in un mercato ad alta competitività, l'importanza del sistema previsionale è di notevole rilievo vista anche la maggior attenzione posta alle esigenze del cliente e alla loro soddisfazione.

Il lancio di nuovi articoli, la creazione di varianti di prodotti già esistenti, la gestione delle scorte, sono investimenti importanti e molto rischiosi per le aziende: ecco perché il ricorso a tecniche previsive sempre più precise rappresentano un punto di partenza indispensabile per ridurre la perdita di capitali.

Una parte delle informazioni si possono ottenere studiando il mercato di interesse ed i consumatori; un'altra parte deriva, invece, dallo studio statistico dell'andamento di alcune variabili di particolare interesse (come per esempio le vendite) nei mesi precedenti.

Nell'ambito di queste tematiche, il presente capitolo si propone di analizzare il traffico voce di "3".

Il settore analizzato è dotato di caratteristiche particolari, in quanto non si ha a che fare con previsioni per la gestione delle scorte o per il lancio di nuovi articoli, ma si assiste ad una serrata corsa all'accaparramento del maggior numero di clienti a scapito dei diretti concorrenti.

Offrendo un servizio a pagamento, l'aspetto più importante della strategia aziendale delle società di telefonia mobile, sarà quello di creare tariffe, promozioni, azioni virali tali da far aumentare il volume di traffico effettuato dai propri clienti.

Un ambito molto importante di questo settore, di cui analizzeremo l'andamento è il traffico voce, settore maturo e piuttosto facile da prevedere nella sua evoluzione.

A questo proposito vengono proposti alcuni metodi di previsione a breve termine, utilizzati anche in azienda, e dalla buona accuratezza previsionale. I modelli utilizzati sono quelli appartenenti alla classe dei modelli ARIMA stagionali (SARIMA), e *Holt - Winters* additivo e moltiplicativo, che verranno descritti nelle sezioni successive.

3.2 MODELLI ARIMA STAGIONALI (SARIMA)

Gran parte delle serie aziendali sono caratterizzate dalla presenza di una componente stagionale più o meno forte.

Mentre i metodi tradizionali si basano sull'assunzione che la componente stagionale sia deterministica e indipendente dagli altri movimenti della serie, la classe dei modelli presi in considerazione assume che la componente stagionale possa essere stocastica e correlata con le componenti non stagionali.

L'idea che sta alla base dei modelli SARIMA è che il processo deve poter descrivere due tipi di relazioni all'interno della serie osservata: la correlazione tra valori consecutivi, che è quella modellata dagli usuali modelli ARIMA, e la correlazione tra osservazioni che distano tra di loro di un multiplo del periodo.

Ciò comporta che le serie appartenenti alla classe dei modelli SARIMA siano caratterizzate da una forte correlazione seriale ai ritardi stagionali, oltre all'usuale autocorrelazione presente nei modelli ARIMA non stagionali.

3.3 MODELLI DI HOLT – WINTERS

I metodi di *Holt–Winters* nascono dal perfezionamento del metodo del lisciamiento esponenziale.

Esso prevede l'assegnazione di uno stesso peso ($1/k$) a ciascuna delle k osservazioni e consiste nell'attribuire pesi decrescenti allontanandosi nel passato.

Il metodo del lisciamiento esponenziale considera la previsione precedente e la corregge tenendo conto dell'errore verificatosi mediante un parametro α compreso tra 0 e 1.

Si deve, in un secondo momento, ad *Holt* l'introduzione del metodo non stagionale che rappresenta un'evoluzione del lisciamiento esponenziale introducendo una componente di crescita lineare in maniera tale da descrivere il trend.

Esso utilizza due parametri, α e β , vincolati all'intervallo 0-1.

Questo metodo copre campi di applicazione vasti, ma non può essere applicato a serie che presentano componenti stagionali, a meno di destagionalizzazioni preventive.

I metodi stagionali furono introdotti da *Winters* come estensione del precedente modello di *Holt*, con lo scopo di catturare direttamente la componente stagionale.

I due metodi, additivo e moltiplicativo utilizzano tre parametri α, β, γ , rispettivamente per il livello, per il trend e per la stagionalità.

Le formule dei vari metodi sono:

Lisciamento
esponenziale

$$F_{n,1} = F_{n-1,1} + (1 - \delta)(y_n - F_{n-1,1})$$

Metodo non
stagionale

$$F_{n,k} = \bar{y}_n + k \cdot \hat{T}_n \quad k \geq 1$$

$$\hat{y}_n = \alpha(\hat{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha)y_n$$

$$\hat{T}_n = \beta\hat{T}_{n-1} + (1 - \beta)(\hat{y}_n - \hat{y}_{n-1})$$

Metodo stagionale

$$F_{n,k} = \hat{y}_n + k \cdot \hat{T}_n + \hat{S}_{n+k-s} \quad 1 \leq k \leq s$$

additivo

$$F_{n,k} = \hat{y}_n + k \cdot \hat{T}_n + \hat{S}_{n+k-2s} \quad s+1 \leq k \leq 2s$$

$$\hat{y}_n = \alpha(\hat{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha)(y_n - S_{n-s})$$

$$\hat{T}_n = \beta\hat{T}_{n-1} + (1 - \beta)(\hat{y}_n - \hat{y}_{n-1})$$

$$\hat{S}_n = \gamma \cdot \hat{S}_{n-s} + (1 - \gamma)(y_n - \hat{y}_n)$$

Metodo stagionale

$$F_{n,k} = [\hat{y}_n + k \cdot \hat{T}_n] \cdot \hat{S}_{n+k-s} \quad 1 \leq k \leq s$$

moltiplicativo

$$F_{n,k} = [\hat{y}_n + k \cdot \hat{T}_n] \cdot \hat{S}_{n+k-2s} \quad s+1 \leq k \leq 2s$$

$$\hat{y}_n = \alpha(\hat{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha) \frac{y_n}{\hat{S}_{n-s}}$$

$$\hat{T}_n = \beta\hat{T}_{n-1} + (1 - \beta)(\hat{y}_n - \hat{y}_{n-1})$$

$$\hat{S}_n = \gamma \cdot \hat{S}_{n-s} + (1 - \gamma) \frac{y_n}{\hat{y}_n}$$

Dove: $F_{n,k}$ = previsione effettuata al tempo n per il periodo n+k

y_n = osservazione al tempo n

\hat{S}_n = stima della componente stagionale

\hat{T}_n = stima del trend

\hat{y}_n = stima del livello

α, β, γ = parametri dei modelli

Il modello stagionale fornisce delle previsioni accurate in presenza di trend e stagionalità evidenti e si basa sull'ipotesi che in prossimità di n la previsione con orizzonte di previsione k può essere fornita dalle rette:

$$F_{n,k} = \hat{y}_n + k \cdot \hat{T}_n + \hat{S}_{n+k-s} \quad 1 \leq k \leq s$$

$$F_{n,k} = \hat{y}_n + k \cdot \hat{T}_n + \hat{S}_{n+k-2s} \quad s+1 \leq k \leq 2s$$

Per il modello stagionale additivo, e dalle rette:

$$F_{n,k} = [\hat{y}_n + k \cdot \hat{T}_n] \cdot \hat{S}_{n+k-s} \quad 1 \leq k \leq s$$

$$F_{n,k} = [\hat{y}_n + k \cdot \hat{T}_n] \cdot \hat{S}_{n+k-2s} \quad s+1 \leq k \leq 2s$$

per il modello stagionale moltiplicativo.

\hat{y}_n rappresenta l'intercetta, \hat{T}_n la pendenza della retta ed \hat{S}_{n+k-2s} il fattore di stagionalità. L'equazione della stima del livello è praticamente identica a quella del lisciamiento esponenziale semplice, se escludiamo il termini della stima del trend (\hat{T}_n) e della stagionalità (\hat{S}_{n-s}).

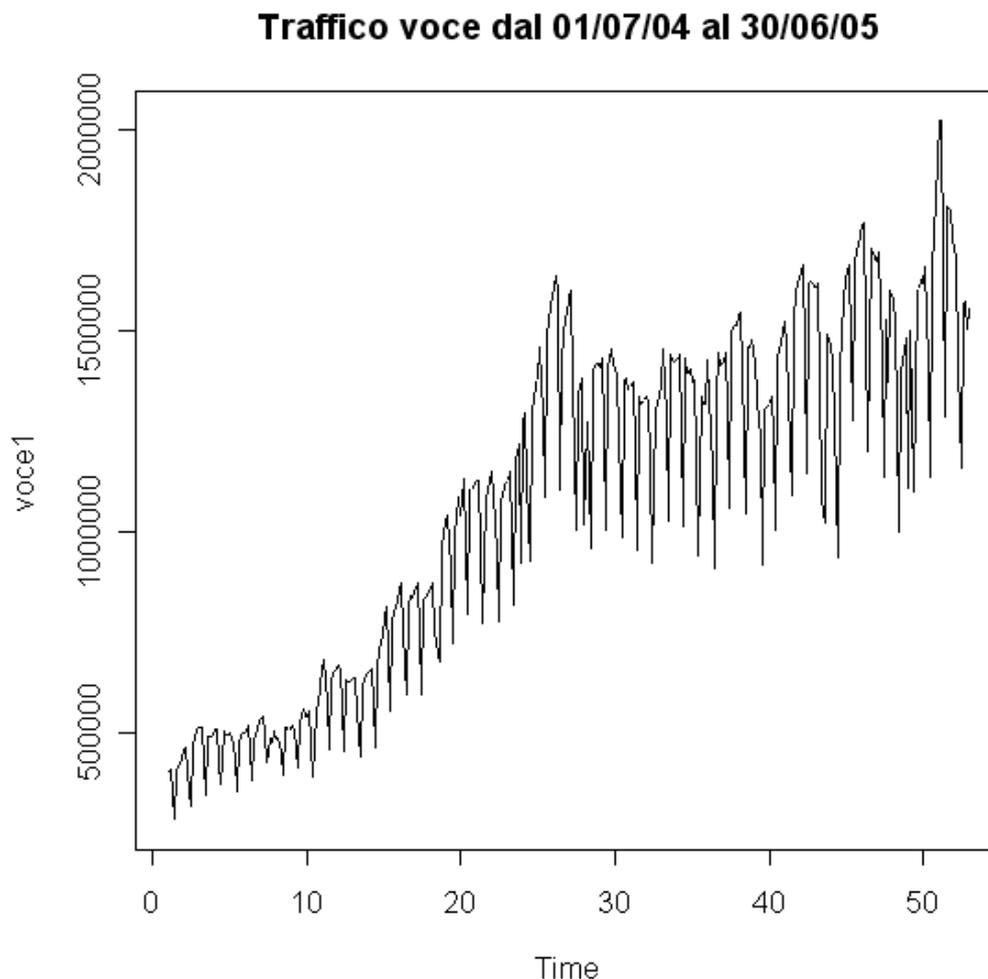
Il primo risulta essere una media ponderata della stima del trend al tempo precedente e della differenza tra due successivi valori del livello liscio esponenzialmente ($\hat{y}_n - \hat{y}_{n-1}$).

Il fattore di stagionalità, invece, è una media ponderata con i pesi γ e $1-\gamma$ del vecchio valore \hat{S}_{n-s} e di una stima locale della stagionalità $y_n - \hat{y}_n$.

3.4 ANALISI DEL TRAFFICO VOCE

La serie che analizziamo è la serie giornaliera dei i minuti di chiamate effettuate dai clienti di “3” nell’arco di tempo che va dal 01/07/2004 al 31/07/2005. L’obiettivo dell’analisi è quello di individuare l’andamento della serie nell’arco dell’anno e quindi stimare un modello che ci permetta di fornire delle previsioni il più precise possibili. A questo scopo, utilizziamo il primo anno di osservazioni (fino al 30/06/2005) per la stima del modello e teniamo l’ultimo mese per le previsioni.

Iniziamo lo studio analizzando il grafico della serie. Si nota la presenza di un notevole trend crescente soprattutto nella prima parte della serie, trend che successivamente si attenua a partire all’incirca dalla 22-esima settimana (vedi grafico).

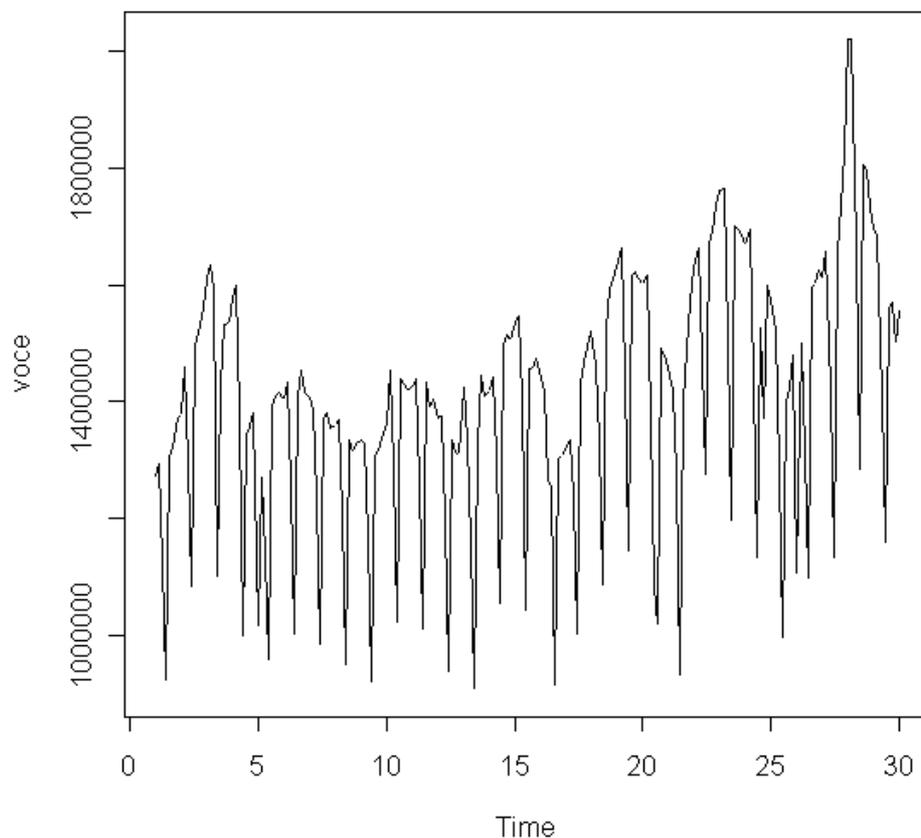


Poiché il non tener conto di questo cambiamento di regime, potrebbe inficiare la stima del modello, abbiamo utilizzato la procedura di *Bai* e *Perron*. In particolare, quello che vogliamo è verificare se effettivamente intorno alla 22-esima settimana c'è un *break*. In tal caso, per la stima del modello adeguato, considereremo solo la parte della serie corrispondente al secondo regime.

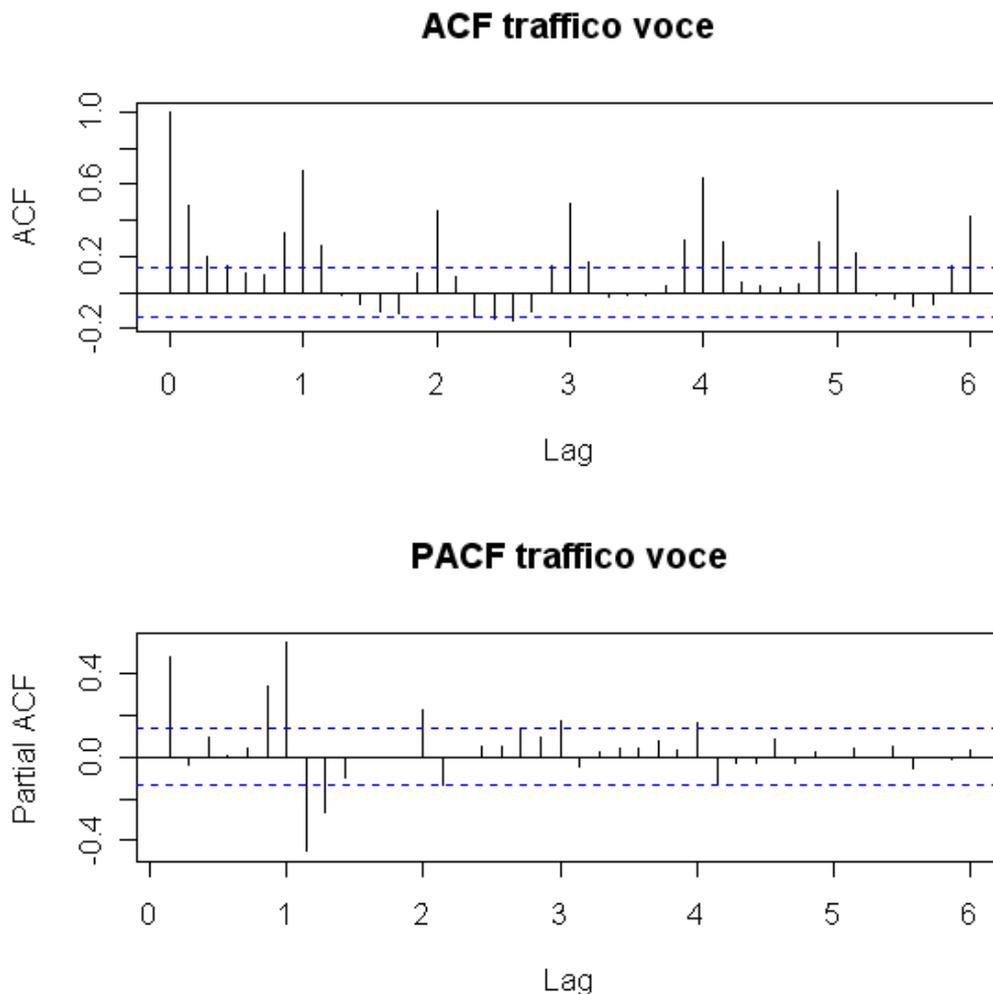
L'analisi effettuata attraverso la procedura di *Bai* e *Perron* per l'individuazione di punti di cambiamento strutturali, conferma appieno la nostra ipotesi: viene individuato un punto di rottura alla 161-esima osservazione, che corrisponde all'8 Dicembre 2004.

D'ora in poi andremo, quindi, ad analizzare la serie dalla 162-osservazione. Il grafico di questa parte di serie è:

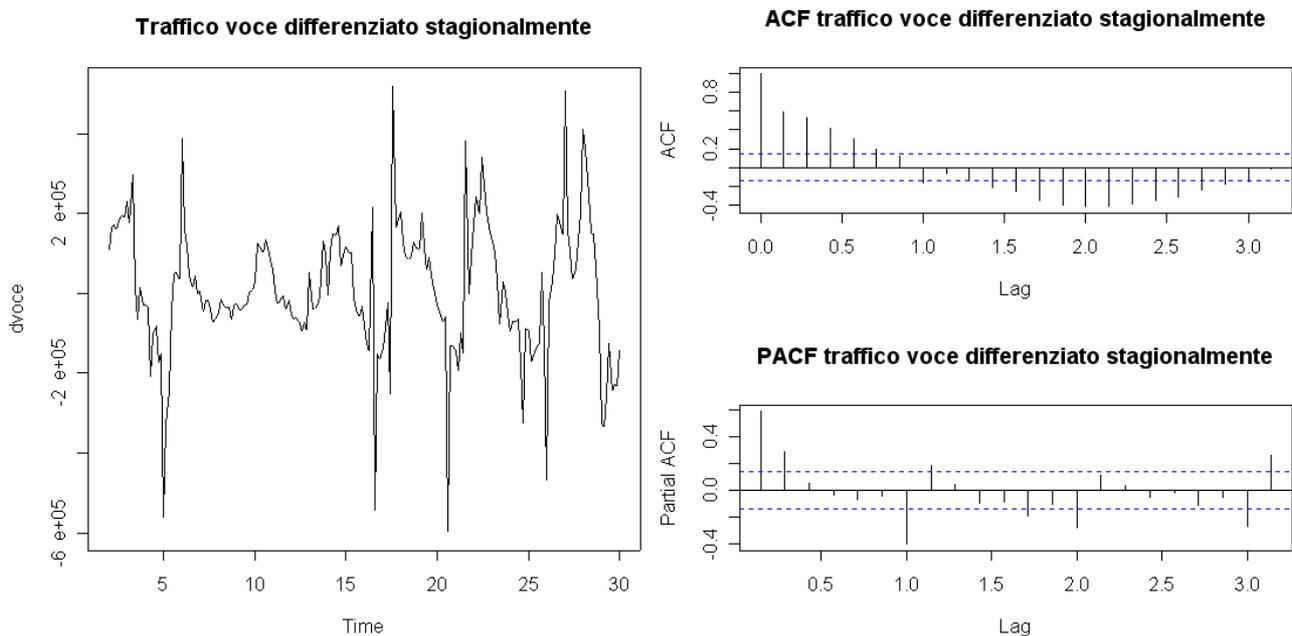
Traffico voce dal 08/12/04 al 30/06/05



Ora il trend crescente è molto meno evidente e non sembrerebbe nemmeno necessario effettuare una differenziazione di primo ordine, andiamo comunque ad osservare il grafico dei correlogrammi.



Sia l'ACF che la PACF sembrano confermare il fatto che la differenziazione di primo ordine non sia indispensabile, mentre sembra utile effettuare una differenziazione di tipo stagionale. Procediamo con la differenziazione e osserviamo il grafico della serie differenziata e quello dei suoi correlogrammi.



Osservando attentamente il grafico dei correlogrammi si può notare la presenza di almeno due componenti autoregressive non stagionali, ed una o più componenti a media mobile stagionali.

Tutto ciò ci può far ipotizzare che un buon punto di partenza sia il modello SARIMA(2,0,0)(0,1,0)₇ oppure SARIMA(2,0,0)(0,1,1)₇.

Proviamo a stimare dapprima il modello più semplice, ovvero con meno parametri, tra quelli ipotizzabili nel nostro caso: il SARIMA(2,0,0)(0,1,0)₇. Stimando il modello tramite il pacchetto statistico “R”, abbiamo ottenuto i seguenti risultati:

*Arima (x = voce, order = c(2, 0, 0), seasonal = list(order = c(0, 1, 0),
freq = 7))*

Coefficienti:

ar1 ar2

0.4208 0.2884

s.e. 0.0678 0.0683

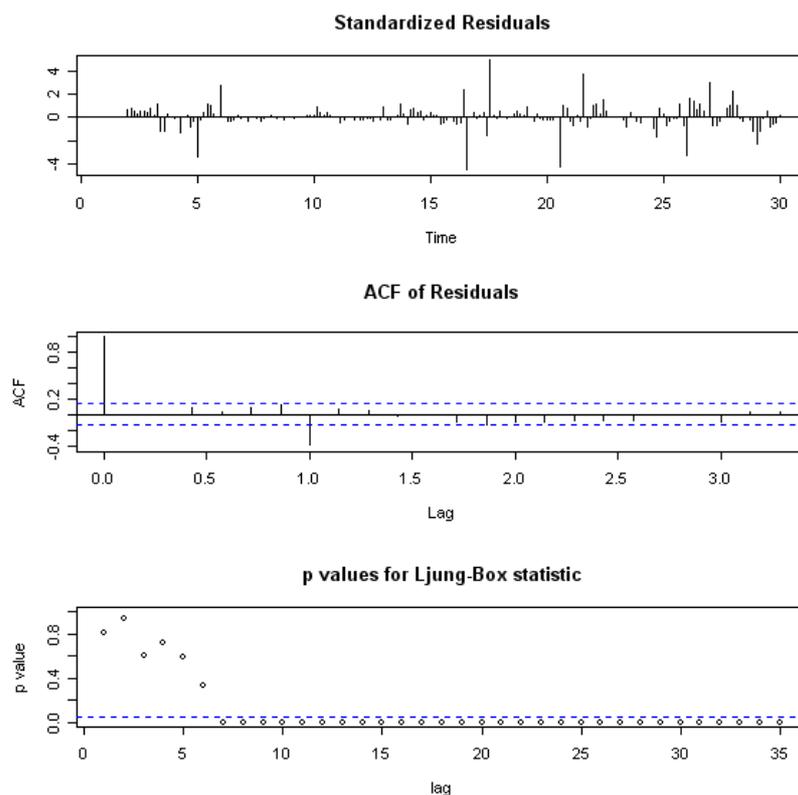
*sigma^2 estimated as 1.671e+10: log likelihood = -2598.44, aic =
5202.87*

statistica t:

ar1 ar2

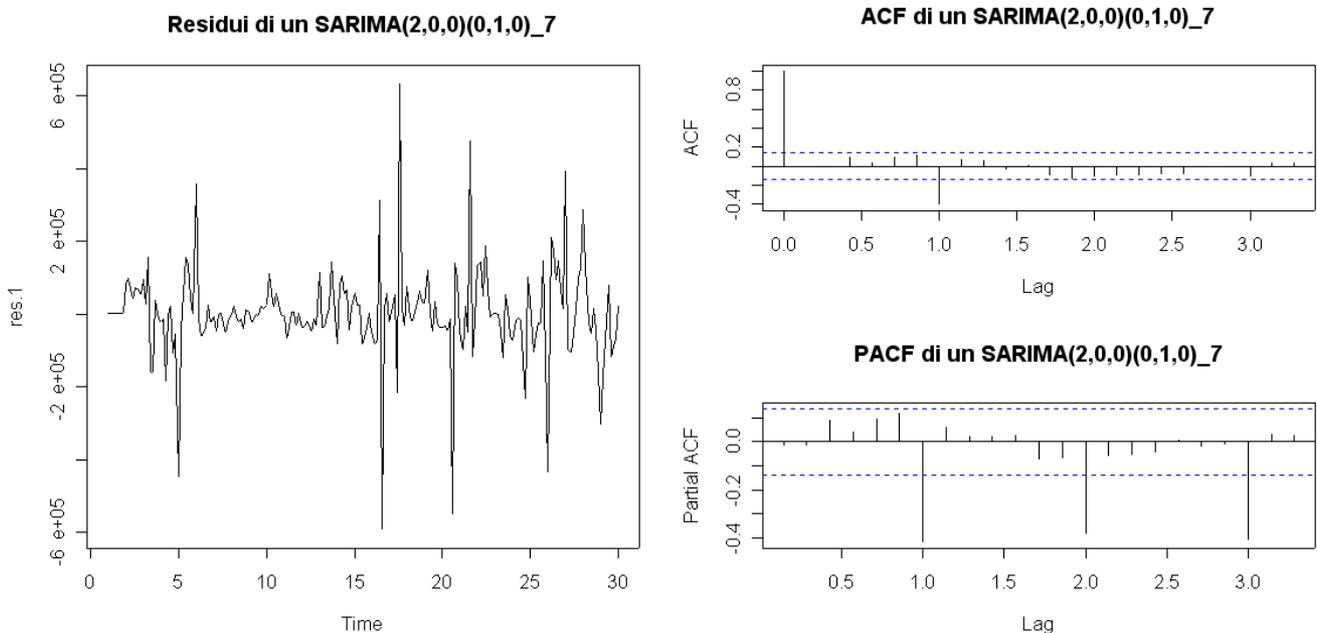
6.203189 4.223708

Entrambi i coefficienti stimati risultano significativi.



Tuttavia la funzione di autocorrelazione presenta un coefficiente significativo al ritardo 7 (una settimana) ed anche il test di *Ljung – Box* che verifica l'ipotesi nulla di indipendenza dei residui, non può accettare tale ipotesi dopo il sesto ritardo.

Evidentemente è necessario individuare un'ulteriore componente che possa spiegare la correlazione fra i residui del modello SARIMA(2,0,0)(0,1,0)_7.



Come già ipotizzato sembra esserci la presenza di una componente a media mobile stagionale. Proviamo, quindi, a stimare il modello SARIMA(2,0,0)(0,1,1)_7.

La procedura di stima porta ai seguenti risultati:

*Arima (x = voce, order = c(2, 0, 0), seasonal = list(order = c(0, 1, 1),
freq = 7))*

Coefficienti:

<i>ar1</i>	<i>ar2</i>	<i>sma1</i>
0.4906	0.3350	-0.9070

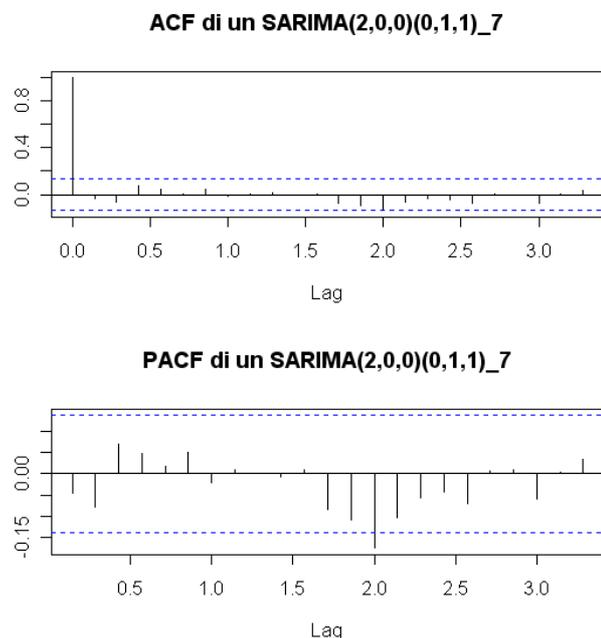
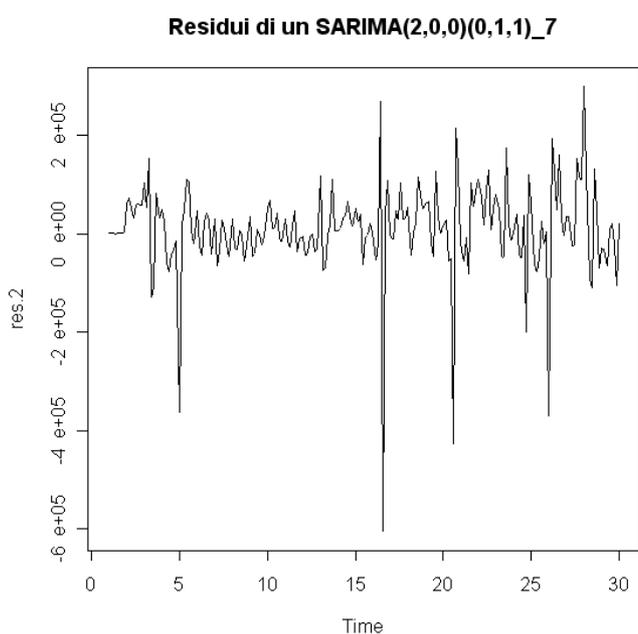
s.e. **0.0681** **0.0681** **0.0529**

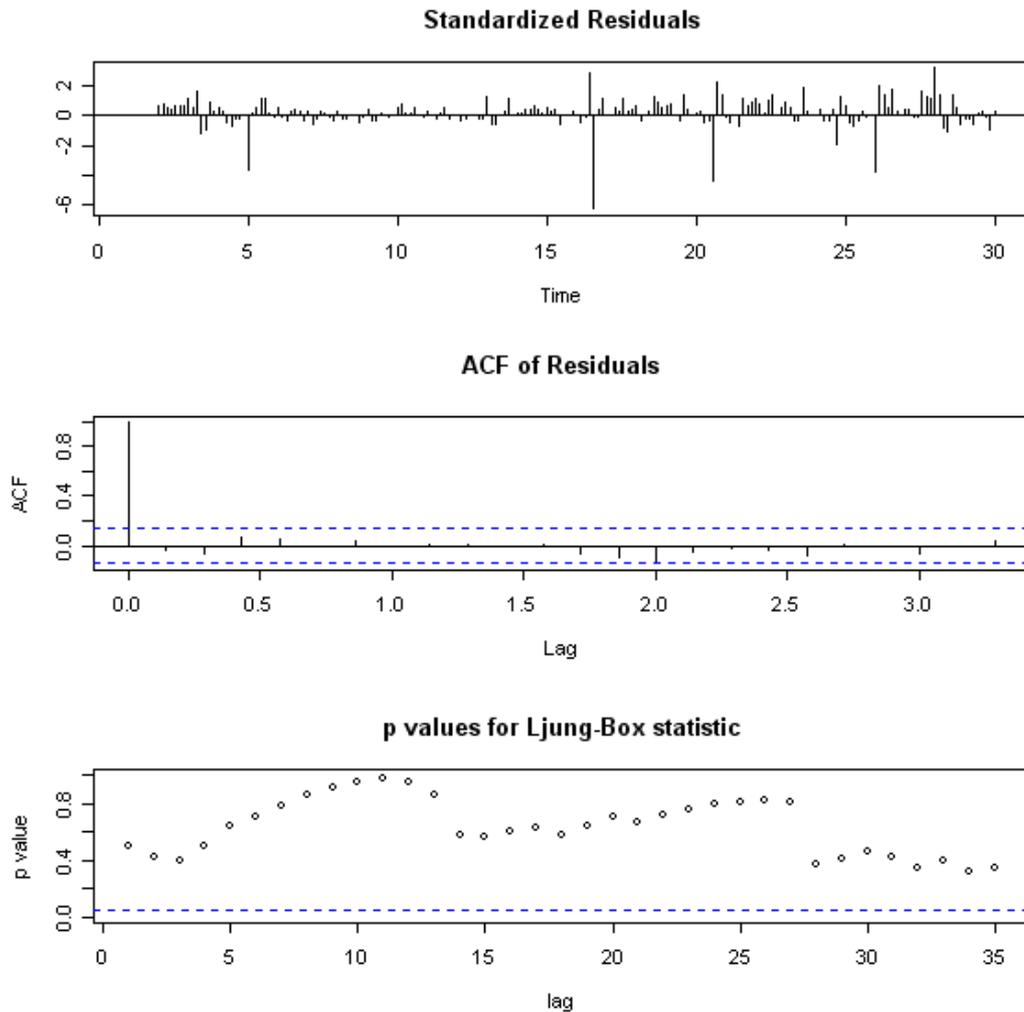
*sigma*² *estimated as* **9.223e+09**: *log likelihood* = **-2545.73**, *aic* = **5099.45**

statistica t:

<i>ar1</i>	<i>ar2</i>	<i>sma1</i>
7.200674	4.916712	-17.158088

I parametri sono tutti significativi, ed inoltre, il valore del criterio di *Akaike* diminuisce rispetto a quello del modello precedente, segno che il parametro introdotto contribuisce ad un miglior adattamento del modello alla serie in esame.





Il test di *Ljung – Box* offre dei risultati di gran lunga più accettabili rispetto al precedente modello stimato, ed anche i correlogrammi sono buoni. Tuttavia osservando attentamente il grafico dei correlogrammi si può individuare almeno un altro parametro a media mobile stagionale. Proviamo, quindi, a stimare il modello SARIMA(2,0,0)(0,1,2)₇:

Arima (x = voce, order = c(2, 0, 0), seasonal = list(order = c(0, 1, 2), freq = 7))

Coefficienti:

ar1 ar2 sma1 sma2

0.4902 0.3351 -0.8937 -0.0174

s.e. 0.0683 0.0681 0.0856 0.0873

sigma^2 estimated as 9.212e+09: log likelihood = -2545.71, aic = 5101.41

statistica t:

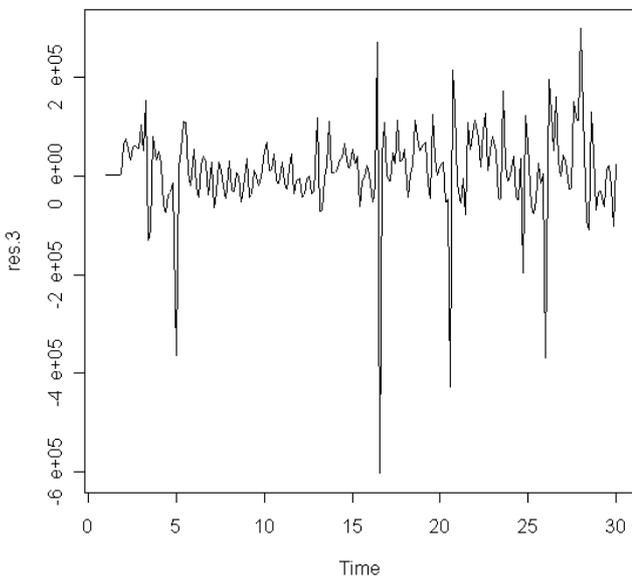
ar1 ar2 sma1 sma2

7.1786875 4.9181566 -10.4463574 -0.1989307

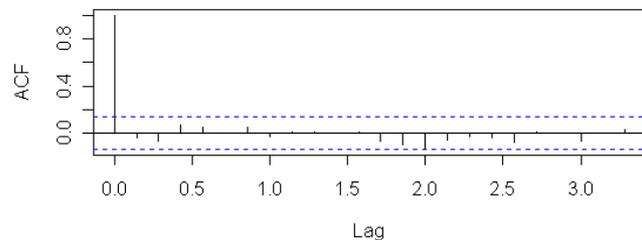
La seconda componente a media mobile stagionale inserita, risulta essere non significativa, come si può notare anche dalla statistica t.

Andiamo comunque ad analizzare il grafico dei residui ed i correlogrammi.

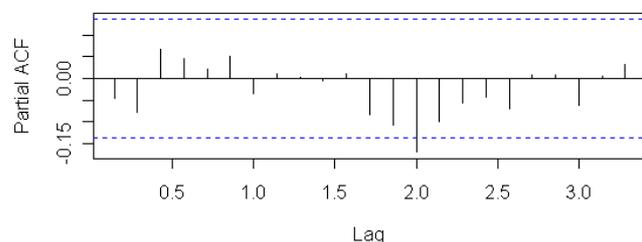
Residui di un SARIMA(2,0,0)(0,1,2)_7



ACF di un SARIMA(2,0,0)(0,1,2)_7



PACF di un SARIMA(2,0,0)(0,1,2)_7



La componente stagionale è ancora evidente dal grafico dei correlogrammi, proviamo quindi a stimare un SARIMA(2,0,0)(0,1,3)₇.

Arima (x = voce, order = c(2, 0, 0), seasonal = list(order = c(0, 1, 3), freq = 7))

Coefficienti:

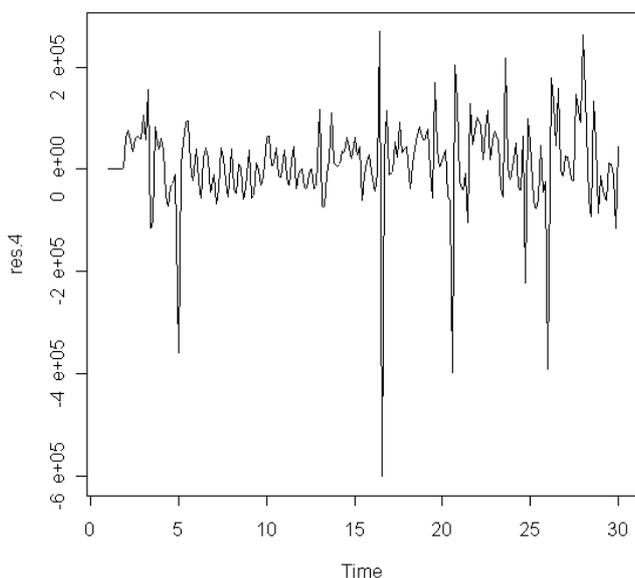
<i>ar1</i>	<i>ar2</i>	<i>sma1</i>	<i>sma2</i>	<i>sma3</i>
0.4610	0.3370	-0.9033	-0.0666	0.1281
<i>s.e.</i> 0.0699	<i>s.e.</i> 0.0685	<i>s.e.</i> 0.0716	<i>s.e.</i> 0.0839	<i>s.e.</i> 0.0746

sigma^2 estimated as 9.166e+09: log likelihood = -2544.15, aic = 5100.29

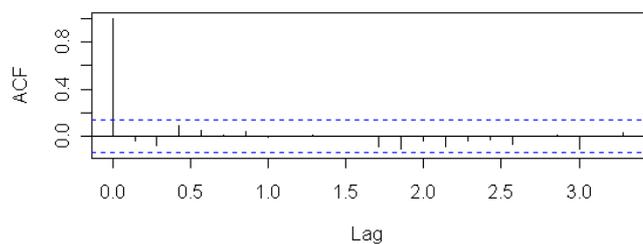
Statistica t:

<i>ar1</i>	<i>ar2</i>	<i>sma1</i>	<i>sma2</i>	<i>sma3</i>
6.5972135	4.9186602	-12.6235156	-0.7941206	1.7174903

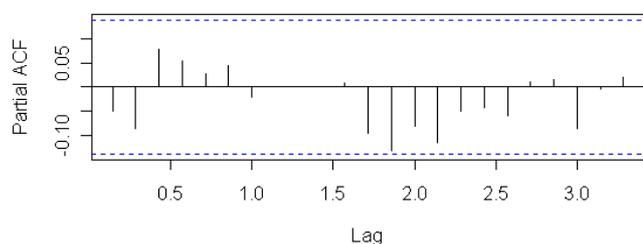
Residui di un SARIMA(2,0,0)(0,1,3)₇



ACF di un SARIMA(2,0,0)(0,1,3)₇



PACF di un SARIMA(2,0,0)(0,1,3)₇



In questo caso la componente residua dei correlogrammi viene ridotta, ma entrambi i coefficienti a media mobile stagionale risultano non significativi, ed il valore della statistica test, inferiore ad 1,96 per entrambi i parametri (posto $\alpha = 0.05$) conferma la non significatività.

Pertanto per poter scegliere il modello che meglio si adatta alla nostra serie andiamo a confrontare i risultati ottenuti.

	AR1 (stat t)	AR2 (stat t)	SMA1 (stat t)	SMA2 (stat t)	SMA3 (stat t)	AIC
Sarima(2,0,0) (0,1,0)_7	0.4208 6.2031	0.2884 4.2237				5202.87
Sarima(2,0,0) (0,1,1)_7	0.4906 7.2006	0.3350 4.9167	-0.9070 -17.1580			5099.45
Sarima(2,0,0) (0,1,2)_7	0.4902 7.1786	0.3351 4.9181	-0.8937 -10.4463	<u>-0.0174</u> -0.1989		5101.41
Sarima(2,0,0) (0,1,3)_7	0.4610 6.5971	0.3370 4.9186	-0.9033 -12.6235	<u>-0.0666</u> -0.7941	<u>-0.0746</u> 1.7149	5100.29

Dei quattro modelli stimati gli ultimi due mostrano dei coefficienti non significativi, dunque la scelta si riduce ai primi due.

Di questi quello che scegliamo è modello il SARIMA(2,0,0)(0,1,1)₇ in quanto presenta un valore del criterio di *Akaike* inferiore rispetto al SARIMA(2,0,0)(0,1,0)₇.

Andremo quindi ad effettuare le nostre previsioni per il mese di Luglio del 2005 utilizzando questo modello.

3.5 PREVISIONI TRAFFICO VOCE

Per effettuare le previsioni per il mese di Luglio utilizzeremo tre modelli di stima: il modello SARIMA identificato nel paragrafo precedente e i due modelli di *Holt - Winters* stagionali additivo e moltiplicativo. Per prima cosa dobbiamo confrontare questi ultimi due modelli, per decidere quale sia più adatto alla previsione richiesta: per fare ciò andiamo a confrontare il valore dell'EQM relativo ai due modelli di *Holt - Winters* stimati fino al 30/06/04. Le costanti di lisciamiento che minimizzano l'EQM dei due metodi risultano essere: $\alpha = 0,4$ $\beta = 0,4$ $\gamma = 0,95$ per il metodo additivo, e $\alpha = 0,8$ $\beta = 0,95$ $\gamma = 0,95$ per il metodo moltiplicativo. Il valore dell'errore quadratico medio (EQM) relativo ai due modelli è:

	Metodo Additivo	Metodo Moltiplicativo
EQM	116548.421	118908.738

I due valori non si discostano di molto (anche se sembra preferibile il modello additivo), pertanto le previsioni per il mese di Luglio verranno calcolate con entrambi i modelli.

Per quanto riguarda il modello SARIMA scelto per effettuare le previsioni andiamo ad osservare il grafico (Grafico 1) dei valori ottenuti, per avere una prima idea della bontà previsioni.

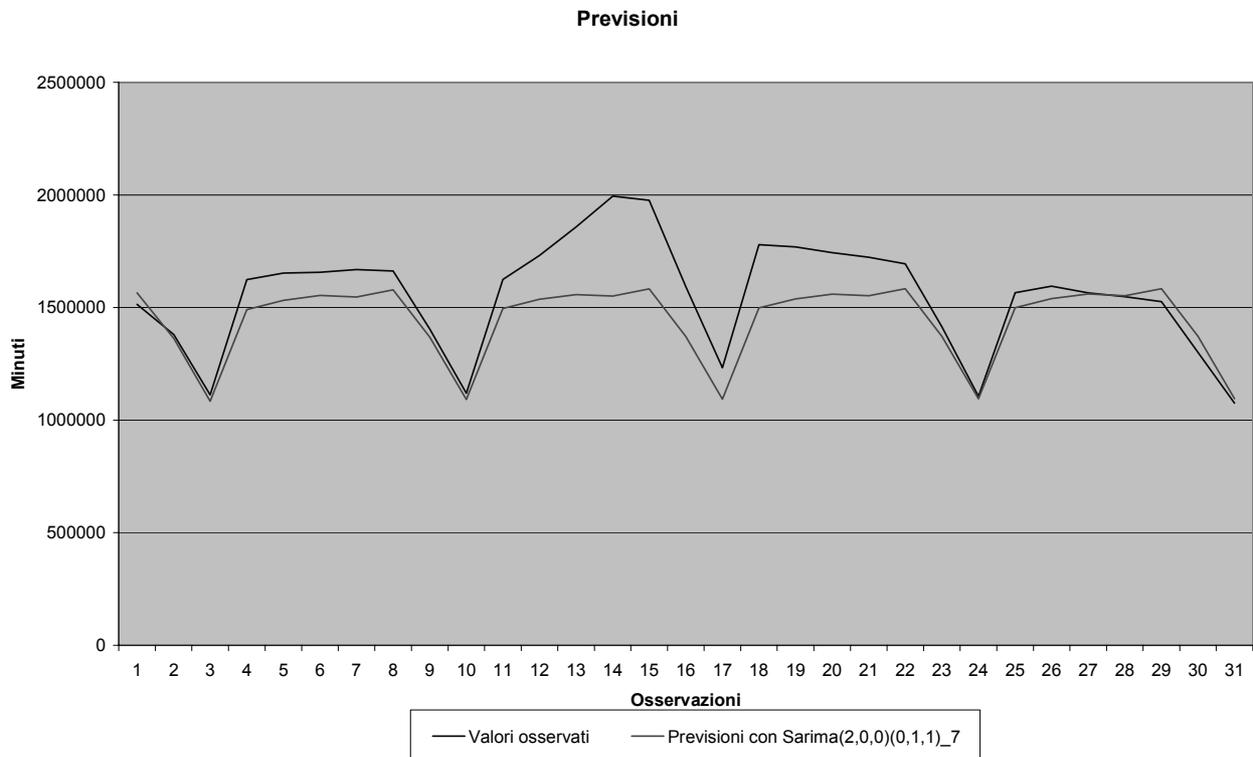


Grafico 1: Confronto tra i valori osservati per il mese di Luglio 2005 e le previsioni effettuate con il SARIMA(2,0,0)(0,1,1)_7

Dal grafico dei valori previsti sembra che il modello fornisca delle previsioni accettabili, anche se come per gli altri due si assiste ad una sottostima dei valori osservati.

Per avere un confronto ancora più chiaro tra i metodi scelti, osserviamo il grafico che mette a confronto le tre previsioni con i valori realmente osservati per il mese di Luglio, successivamente andiamo a confrontare il

valore dell'EQM e dell'EMA (errore medio assoluto: $EMA = \frac{1}{n} \sum_{t=1}^n |e_t|$) per

ogni modello, per poter definire in maniera analitica quale tra essi offra le previsioni più attendibili

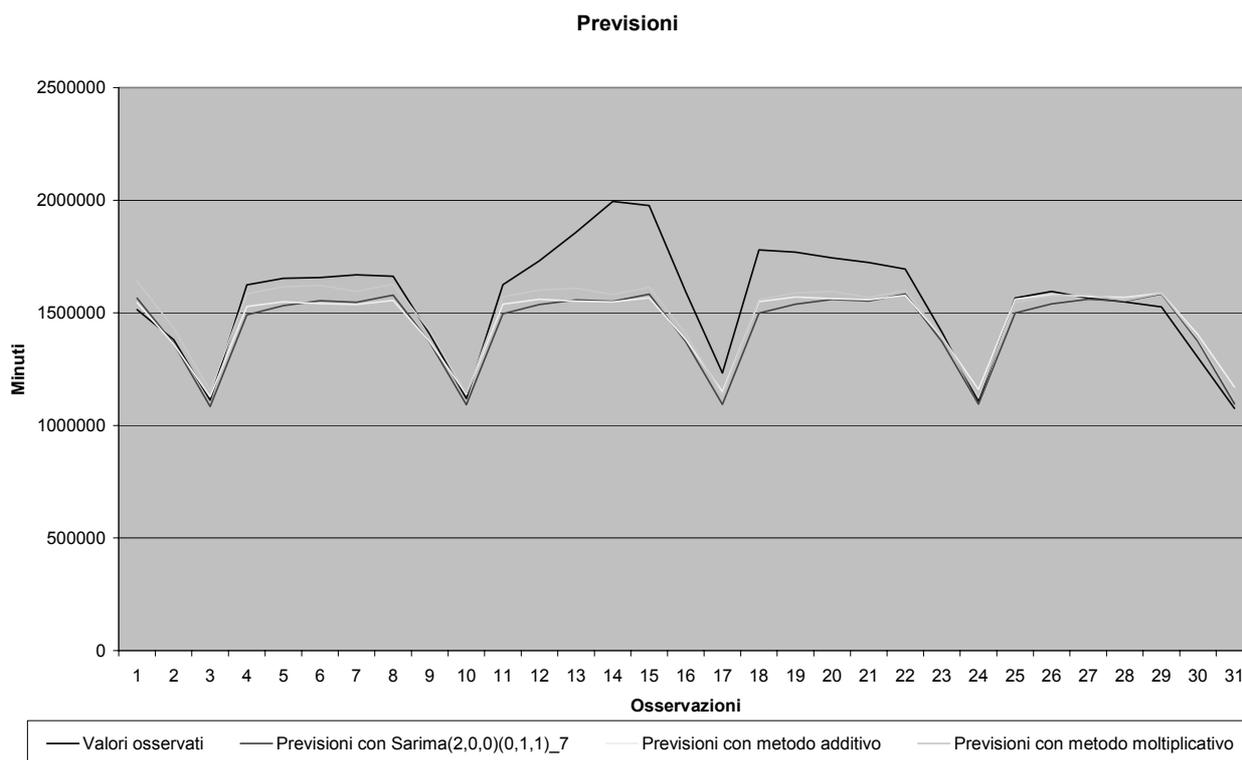


Grafico 2: Confronto tra le previsioni effettuate con i modelli di Holt – Winters additivo e moltiplicativo, il SARIMA(2,0,0)(0,1,1)₇ ed i valori osservati.

	Sarima(2,0,0) (0,1,1) ₇	Holt – Winters Additivo	Holt – Winters Moltiplicativo
EMA	124326.5614	118039.688	98510.140
EQM	166663.51	161417.966	140316.199

Dal grafico le previsioni più fedeli ai veri valori realizzati sembrano essere quelle del modello di *Holt – Winters* stagionale, anche se di poco, in quanto sono quelle che sottostimano in maniera minore i valori realmente osservati.

Andando poi a confrontare gli indici di bontà delle previsioni, la nostra ipotesi trova conferma: sia per quanto riguarda l'EMA che l'EQM i risultati migliori sono forniti proprio dal modello di *Holt – Winters* stagionale moltiplicativo.

4. CONCLUSIONI

4.1 CONSIDERAZIONI FINALI

I fenomeni analizzati in questo lavoro presentano delle caratteristiche di non linearità che non sono semplici da descrivere.

Nonostante tutto, i risultati ottenuti sono veramente interessanti, in particolar modo per quanto riguarda l'analisi effettuata sul traffico di videomessaggi. L'individuazione, infatti, di una reale connessione causa-effetto tra campagne virali e volumi di traffico è un risultato che apre la strada a numerose altre analisi che non possiamo svolgere in questa tesi.

Risultati come quelli che abbiamo ottenuto nella prima parte del nostro lavoro, sono un tipico esempio di come un'analisi statistica possa fornire importanti informazioni volte a migliorare la gestione degli investimenti da parte delle imprese.

Per quanto riguarda, invece, i risultati ottenuti dalla previsione effettuata sulla serie del traffico voce, essi rappresentano uno degli esempi più diffusi dell'utilizzo delle tecniche statistiche in azienda.

Il possesso di previsioni di buon livello, come quelle che abbiamo ottenuto in questo elaborato, è uno degli scopi primari per ogni azienda che intende massimizzare i propri sforzi verso obiettivi che possono garantirle dei ritorni in termini economici.

BIBLIOGRAFIA

- (1) Bai J. , Perron P. , (1998) “ Estimating and testing linear models with multiple structural changes”, *Econometrica* 66,47-78.
- (2) Di Fonzo T. , Lisi F. (2005) “Serie storiche economiche”, Carocci.
- (3) “Time series analysis: univariate and multivariate methods”, Wei W. W., 1990.