

**UNIVERSITA' DEGLI STUDI DI PADOVA**

**FACOLTA' DI SCIENZE STATISTICHE**  
**CORSO DI LAUREA IN STATISTICA E GESTIONE DELLE IMPRESE**

**RELAZIONE FINALE:**  
**PREVISIONE DEL TRAFFICO VOCE DI UN'AZIENDA DI**  
**TELEFONIA MOBILE**

**RELATORE: PROF. BISAGLIA LUISA**

**LAUREANDO: BAZZANA LUCA**

**ANNO ACCADEMICO: 2004-2005**



# SOMMARIO

<b>INTRODUZIONE.....</b>	<b>1</b>
<b>CAPITOLO 1 – PRESENTAZIONE .....</b>	<b>3</b>
1.1 CARATTERISTICHE DELL’AZIENDA .....	3
1.1.1 Businnes Model.....	4
1.1.2 UMTS .....	5
<b>CAPITOLO 2 - I DATI.....</b>	<b>7</b>
2.1 IL CAMPIONE.....	7
<b>CAPITOLO 3 – MODELLAZIONE.....</b>	<b>13</b>
3.1 MODELLI ARIMA STAGIONALI (S)ARIMA .....	13
3.1.1 Teoria dei modelli.....	13
3.1.2 Modellazione dei minuti totali .....	14
3.1.3 Modellazione delle chiamate .....	21
3.1.4 Modellazione delle attivazioni .....	27
3.1.5 Scelta dei modelli migliori.....	33
<b>CAPITOLO 4 - LISCIAMENTO ESPONENZIALE .....</b>	<b>35</b>
4.1 PREVISIONI TRAMITE LISCIAMENTO ESPONENZIALE .....	35
4.1.1 I metodi di Holt-Winters stagionali.....	35
4.1.1.1 Metodo stagionale additivo.....	35
4.1.1.2 Metodo stagionale moltiplicativo .....	36
4.2 PREVISIONE DEI MINUTI.....	38
4.3 PREVISIONE DELLE CHIAMATE .....	39
4.4 PREVISIONE DELLE ATTIVAZIONI .....	40
<b>CAPITOLO 5 - MIGLIOR PREVISIONE .....</b>	<b>43</b>
5.1 PREVISIONE FINALE MINUTI.....	43
5.2 PREVISIONE FINALE CHIAMATE .....	44
5.3 PREVISIONE FINALE ATTIVAZIONI .....	44
5.4 CONSIDERAZIONI FINALI.....	45



## INDICE DELLE FIGURE

FIG. 1	MINUTI TOTALI.....	8
FIG. 2	CHIAMATE GIORNALIERE .....	8
FIG. 3	ATTIVAZIONI GIORNALIERE .....	10
FIG. 4	SERIE MINUTI DIFFERENZIATA.....	15
FIG. 5	OUTPUT MODELLO SARIMA(2,0,2)(0,1,1) <sub>7</sub> .....	16
FIG. 6	VALORI STIMATI MODELLO SARIMA(2,0,2)(0,1,1) <sub>7</sub> .....	16
FIG. 7	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(2,0,2)(0,1,1) <sub>7</sub> .....	17
FIG. 8	OUTPUT MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	17
FIG. 9	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	18
FIG. 10	VALORI STIMATI MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	19
FIG. 11	CONFRONTO PREVISIONI MINUTI.....	20
FIG. 12	SERIE CHIAMATE DIFFERENZIATA.....	21
FIG. 13	OUTPUT MODELLO SARIMA(2,0,2)(0,1,2) <sub>7</sub> .....	22
FIG. 14	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(2,0,2)(0,1,2) <sub>7</sub> .....	23
FIG. 15	VALORI STIMATI MODELLO SARIMA(2,0,2)(0,1,2) <sub>7</sub> .....	23
FIG. 16	OUTPUT MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	24
FIG. 17	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	25
FIG. 18	VALORI STIMATI MODELLO SARIMA(2,0,0)(2,1,0) <sub>7</sub> .....	25
FIG. 19	CONFRONTO PREVISIONI CHIAMATE .....	26
FIG. 20	SERIE ATTIVAZIONI DIFFERENZIATA.....	27
FIG. 21	OUTPUT MODELLO SARIMA(1,0,1)(1,1,1) <sub>7</sub> .....	28
FIG. 22	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(1,0,1)(1,1,1) <sub>7</sub> .....	29
FIG. 23	VALORI STIMATI MODELLO SARIMA(1,0,1)(1,1,1) <sub>7</sub> .....	29
FIG. 24	OUTPUT MODELLO SARIMA(5,0,0)(2,1,0) <sub>7</sub> .....	30
FIG. 25	VALORI STIMATI MODELLO SARIMA(5,0,0)(2,1,0) <sub>7</sub> .....	30
FIG. 26	CORRELOGRAMMA DEI RESIDUI MODELLO SARIMA(5,0,0)(2,1,0) <sub>7</sub> .....	31
FIG. 27	CONFRONTO PREVISIONI ATTIVAZIONI.....	32
FIG. 28	LISCIAMENTO MINUTI.....	38
FIG. 29	LISCIAMENTO CHIAMATE .....	39
FIG. 30	LISCIAMENTO ATTIVAZIONI.....	40



## INTRODUZIONE

In questa relazione abbiamo preso in considerazione alcune serie storiche relative al traffico voce di una delle cinque aziende italiane di telefonia mobile con licenza per telecomunicazioni UMTS.

L'analisi è stata condotta cercando di descrivere in modo analitico i minuti totali, il numero di chiamate e le attivazioni giornaliere, prevedendone l'andamento su un orizzonte temporale di quattro settimane.

Tra i metodi che abbiamo utilizzato per la modellazione dell'andamento delle serie storiche, quelli che ci hanno dato i risultati migliori sono i processi ARIMA stagionali, (S)ARIMA, e il metodo di Holt-Winters stagionale.

Abbiamo poi effettuato un'analisi quantitativa per valutare la bontà delle previsioni e scegliere il modello migliore, a fini previsivi, per ciascuna delle variabili in esame.





# Capitolo 1

## PRESENTAZIONE

### **1.1 CARATTERISTICHE DELL'AZIENDA**

L'azienda alla quale faremo riferimento nelle nostre analisi verrà chiamata, per motivi legati alla privacy, XXX.

La compagnia telefonica in questione è controllata da una multinazionale con sede a Hong Kong che, oltre ad essere titolare di dieci licenze UMTS nel mondo, è una protagonista assoluta dell'economia mondiale, con un valore di circa un terzo della capitalizzazione della Borsa di Hong Kong. Opera in diversi settori industriali: porti, energia, immobiliare, turismo, grande distribuzione e telecomunicazioni, dando lavoro ad oltre 120 mila dipendenti in 37 Paesi del mondo.

In Italia ha realizzato il più grande investimento estero dai tempi del piano Marshall e rappresenta il 15% degli investimenti diretti esteri arrivati nel nostro Paese negli ultimi cinque anni.

Ora punta a sviluppare una rete di terza generazione paneuropea: un progetto che consentirà di realizzare significative economie di scala nello sviluppo di prodotti e servizi, nell'accessibilità a competenze e partner tecnologici, nella completa disponibilità di una rete internazionale.

### **1.1.1 Business Model**

XXX occupa il centro della nuova "catena del valore" delle comunicazioni mobili di terza generazione. Questo significa puntare sulla realizzazione di nuovi standard d'innovazione ed eccellenza nella qualità del servizio e della rete, creare un'offerta all'insegna della trasparenza, della semplicità e della convenienza di prezzo, sviluppare un portafoglio di servizi di terza generazione tagliati su misura per le esigenze del mercato Individui e Aziende, realizzati "in casa" o creati da selezionati Content & Service Provider esterni.

Oltre a ciò, intende aprire la propria piattaforma tecnologica a tutti gli Internet player e ai fornitori di contenuti, giocando il ruolo cruciale di abilitatore tecnologico e di business, capace di offrire soluzioni integrate per la creazione e la fatturazione dei servizi multimediali mobili.

XXX sta creando una grande infrastruttura per l'Italia: una rete radiomobile a banda larga, interamente a commutazione di pacchetto, realizzata cioè secondo lo stesso standard tecnologico che è alla base della rete Internet. Questa rete sarà progettata per l'UMTS, per sfruttarne appieno le caratteristiche, e avrà una distribuzione omogenea sul territorio italiano, con l'obiettivo di raggiungere aree territoriali e segmenti di mercato non ancora serviti da infrastrutture a banda larga. La nuova rete avrà un elevato grado di flessibilità per gestire il mix e le variazioni dei volumi di traffico, puntando alla massima efficienza anche sotto il profilo dei costi.

Il sistema di fatturazione sarà il primo in Italia pensato per integrare tariffazione a pacchetto e tariffazione a servizi, quindi per bit e tempo di connessione. Inoltre punterà a minimizzare l'impatto ambientale, non solo grazie a livelli di emissione elettromagnetica da 5 a 10 volte inferiori a quelli delle tecnologie di seconda generazione, ma anche attraverso le nuove tecniche di "mimetismo" architettonico, rese possibili dalle ridotte dimensioni degli elementi trasmissivi.

### **1.1.2 UMTS**

UMTS è l'acronimo di Universal Mobile Telecommunications System, la prima tecnologia completamente digitale per la trasmissione di audio, video, dati, e terza tecnologia di comunicazione mobile, per questo detta 3G (Etacs e GSM sono i due predecessori). UMTS nasce come risposta a un'esigenza concreta: rendere multimediale la comunicazione mobile, uscendo dagli stretti e angusti confini di una semplice "telefonata".



## Capitolo 2

### I DATI

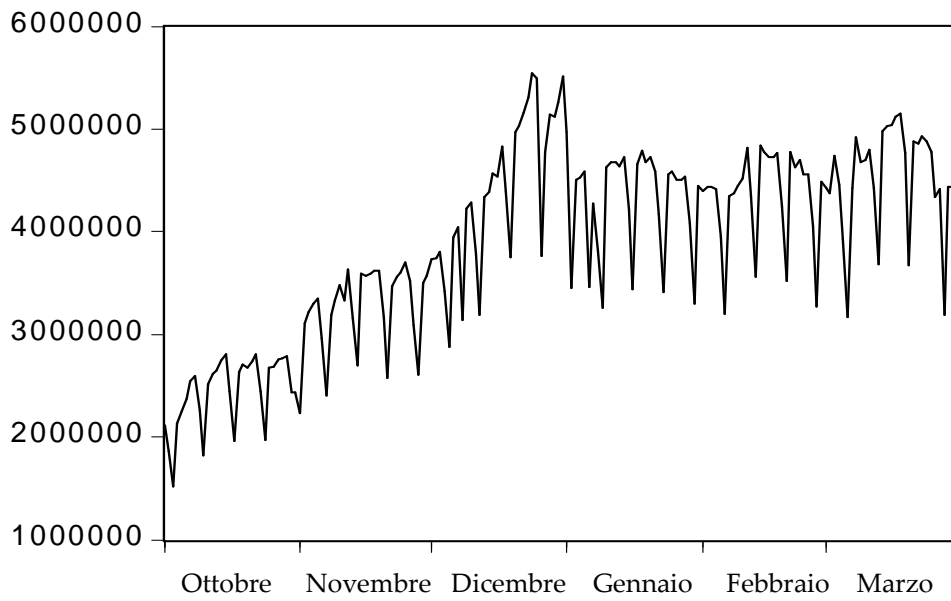
#### **2.1 IL CAMPIONE**

Per le analisi che abbiamo effettuato, ci è stato fornito un dataset contenente, oltre alle variabili che saranno oggetto di previsione, altre informazioni come le disattivazioni o la quantità di ricariche omaggio erogate.

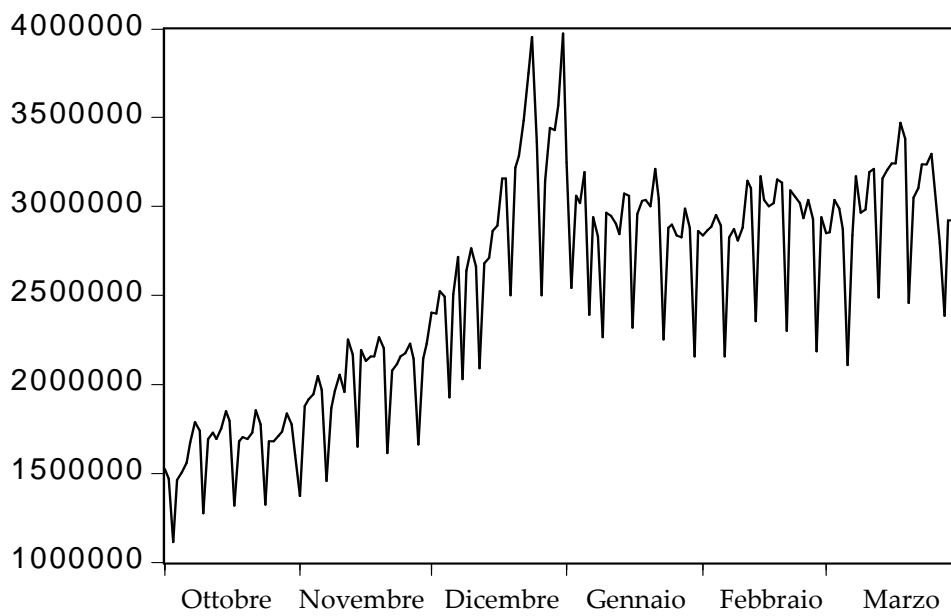
Tutte le serie storiche sono composte da dati giornalieri, riguardanti il periodo 1 ottobre 2004 – 31 marzo 2005, per un totale di 182 osservazioni.

Le tre serie di maggior interesse sono i minuti totali di chiamate, il numero di chiamate e le attivazioni. Da una prima analisi notiamo come tutte e tre le variabili presentino una stagionalità settimanale ed un cambiamento di regime a circa metà del periodo di osservazione, con un trend inizialmente positivo, che poi si annulla dopo la metà del periodo considerato nelle osservazioni.

Le variabili minuti e chiamate, come potevamo aspettarci, hanno andamenti simili:



**Fig.1 Minuti totali**



**Fig.2 Chiamate giornaliere**

Entrambe le serie storiche presentano una marcata periodicità settimanale, con un valore più basso rispetto agli altri, che corrisponde alla domenica. Questo è riconducibile ad un fattore stagionale, cioè al fatto che il telefono cellulare viene usato maggiormente per lavoro, quindi durante la settimana.

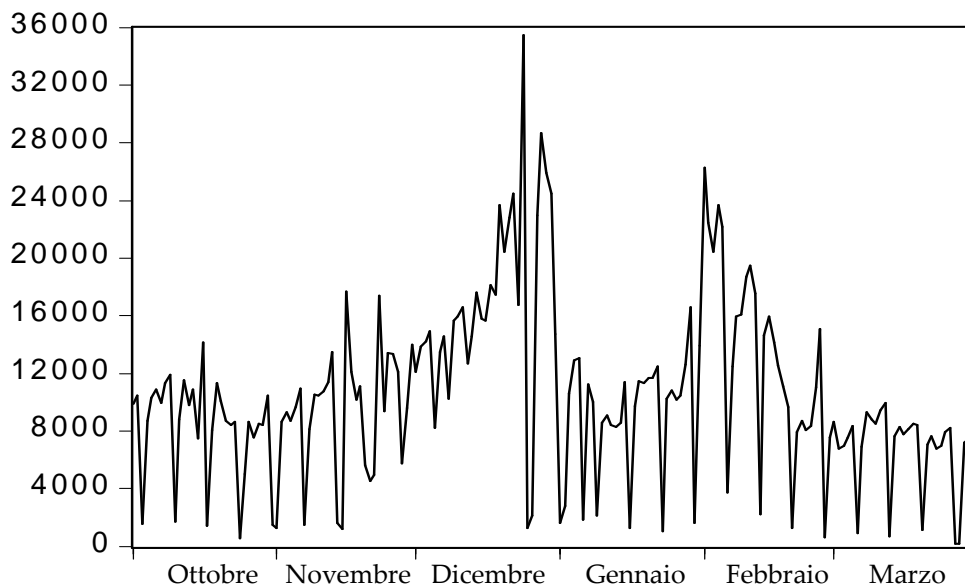
La stima della stagionalità, e l'eventuale depurazione della stessa, costituiscono un problema di fondamentale importanza, perché la componente stagionale può oscurare l'effettivo andamento della serie, soprattutto nei dati più recenti.

Inizialmente è presente un trend positivo che però si va ad annullare nei primi giorni di gennaio, mantenendo comunque la media degli ultimi tre mesi al livello raggiunto in corrispondenza delle osservazioni più elevate, quelle verso la fine della componente tendenziale.

Inoltre, con l'ausilio del grafico, abbiamo potuto individuare due valori anomali (*outlier*), corrispondenti alle osservazioni del 24 dicembre, vigilia di Natale, e del 31 dicembre, dove le chiamate sono state significativamente superiori rispetto alla media del periodo. L'identificazione di questi punti costituisce un problema di estrema delicatezza, poiché potrebbero causare problemi in fase di individuazione del meccanismo generatore della serie.

La terza variabile che sarà oggetto di previsione è quella relativa al numero di attivazioni giornaliere, cioè la quantità di nuovi numeri telefonici che vengono attivati nel corso di una giornata.

Con l'aiuto del grafico abbiamo potuto osservare come anche questa serie sia riconducibile ad un processo non stazionario:



**Fig. 3 Attivazioni giornaliere**

Anche qui si nota la stagionalità settimanale, con il picco negativo sempre in corrispondenza della domenica, giorno in cui i negozi sono chiusi e per questo risulta più difficile attivare una nuova sim.

In questo caso la media delle osservazioni sembra costante, tranne che in due periodi: dalla metà di novembre a fine dicembre, dove la serie è contraddistinta da un trend positivo, ed a fine gennaio, dove i dati subiscono un'impennata rispetto ai giorni precedenti ed a quelli immediatamente successivi.

All'interno del dataset erano presenti altre variabili, quali il numero di disattivazioni giornaliere, il cumulato delle attivazioni e delle



disattivazioni, e l'importo di ricariche bonus erogate. Quest'ultima è generata in questo modo: un cliente con un piano con meccanismo di autoricarica, "matura" nel corso del mese un bonus in base ai minuti di chiamate che riceve, poi il mese successivo riceve la ricarica pari all'importo maturato. Questo fatto ci ha fatto pensare subito che fosse una variabile importante nella costruzione del modello, cosa che però si è rivelata errata: la variabile bonus non è risultata significativa nella modellazione lineare, e non ha dato buoni risultati neanche utilizzandola come input in un modello dinamico a funzione di trasferimento.



## Capitolo 3

### MODELLAZIONE

#### **3.1 MODELLI ARIMA STAGIONALI (S)ARIMA**

##### **3.1.1 Teoria dei modelli**

Il processo che potrebbe adattarsi alla stima dei modelli di nostro interesse appartiene alla classe dei modelli SARIMA (Box e Jenkins 1976).

Gran parte delle serie aziendali, infatti, 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 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 alla usuale autocorrelazione presente nei modelli ARIMA non stagionali.

In questa fase della modellazione abbiamo deciso di utilizzare il programma statistico EViews. Questo ci ha permesso di analizzare i dati dando una nostra interpretazione sul processo generatore della serie in esame e, in seguito, di darne anche una previsione per l'ultimo mese.

Per ogni variabile abbiamo scelto i due modelli che si adattano meglio ai dati, per poi far cadere la scelta su quello che tramite un'analisi quantitativa ha dato risultati migliori in fase di previsione.

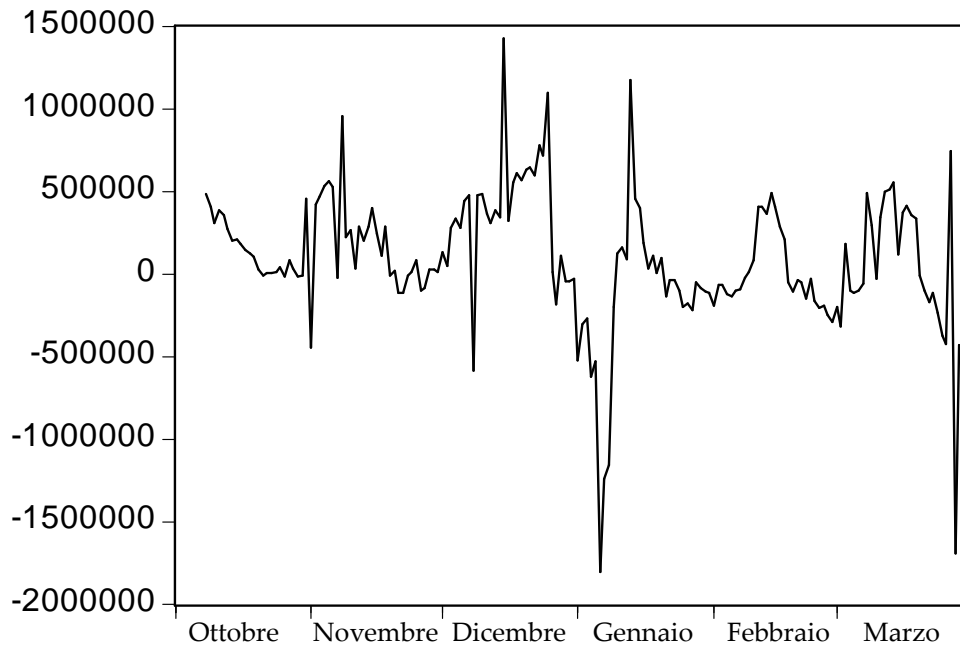
La ricerca dei modelli è stata quindi effettuata considerando il periodo ottobre - febbraio, lasciando l'ultimo mese per confrontare le previsioni con i dati che si sono verificati nella realtà.

### **3.1.2 Modellazione dei minuti totali**

Dalle considerazioni precedenti, per la variabile minuti è stata necessaria una differenziazione stagionale di ordine sette.

A questo punto la nostra serie di riferimento non sarà più quella iniziale, bensì quella differenziata ossia  $D(\text{minuti}, 0, 7)$ , dove  $D$  rappresenta la differenziazione e 7 rappresenta il suo ordine.

Il nuovo grafico mostra come sia scomparsa la stagionalità, ed anche la stazionarietà in media sia migliorata:



**Fig. 4 Serie minuti differenziata**

Dall'analisi della funzione di autocorrelazione, si può ipotizzare che il processo generatore dei dati contenga due componenti autoregressive di ordine uno e due, più un'altra componente stagionale al ritardo sette.

Dopo alcuni tentativi, il modello che ci ha fornito risultati migliori in base ai criteri di Akaike e di Schwarz, è stato un SARIMA(2,0,2)(0,1,1)<sub>7</sub>:

**Dependent Variable: D(MINUTI,0,7)**

Method: Least Squares

Sample(adjusted): 10/10/2004 2/28/2005

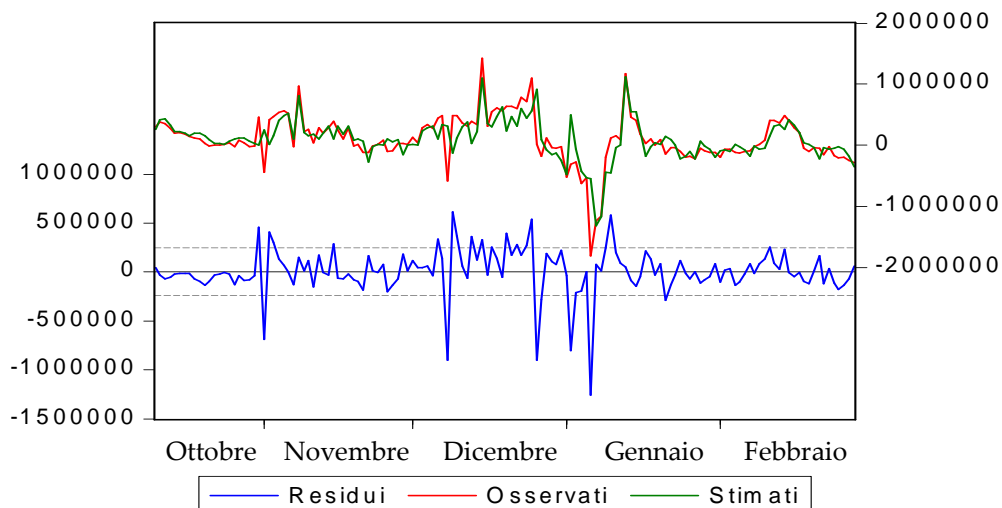
Included observations: 142 after adjusting endpoints

Convergence achieved after 12 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.499641	0.083090	6.013265	0.0000
AR(2)	0.457623	0.086267	5.304715	0.0000
MA(2)	-0.186075	0.097450	-1.909431	0.0583
SMA(7)	-0.851317	0.041985	-20.27670	0.0000
R-squared	0.620847	Mean dependent var		102642.1
Adjusted R-squared	0.612605	S.D. dependent var		393149.0
S.E. of regression	244700.2	<b>Akaike info criterion</b>		<b>27.68122</b>
Sum squared resid	8.26E+12	<b>Schwarz criterion</b>		<b>27.76448</b>
Log likelihood	-1961.367	Durbin-Watson stat		2.024535

**Fig. 5 Output modello SARIMA(2,0,2)(0,1,1)<sub>7</sub>**

In questo modello tutte le componenti risultano significative con  $\alpha=0.05$ , tranne la media mobile di ordine due ( $p=0.058$ ), che però non togliamo poiché è molto utile per annullare una dipendenza visibile nella funzione di autocorrelazione al ritardo due. Così facendo, anche il correlogramma dei residui risulta ora soddisfacente, dimostrando che la stima si adatta bene ai dati:



**Fig. 6 Valori stimati modello SARIMA(2,0,2)(0,1,1)<sub>7</sub>**

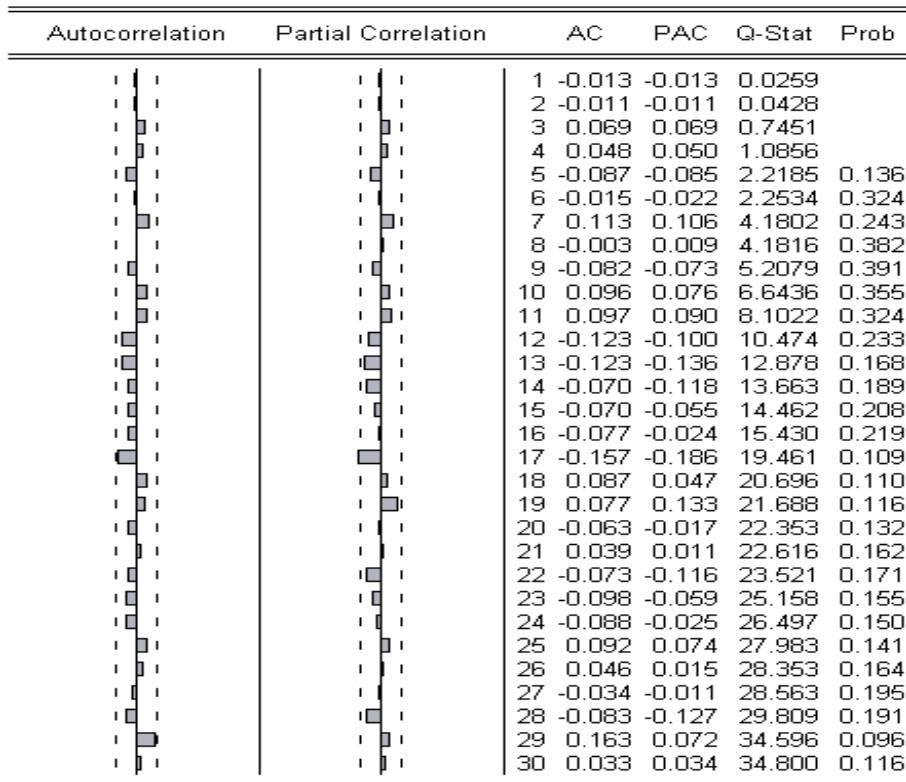


Fig. 7 Correlogramma dei residui modello SARIMA(2,0,2)(0,1,1)<sub>7</sub>

Un altro modello che sembra adattarsi bene ai dati è il SARIMA(2,0,0)(2,1,0)<sub>7</sub>:

**Dependent Variable: D(MINUTI,0,7)**

Method: Least Squares  
Sample(adjusted): 10/24/2004 2/28/2005  
Included observations: 128 after adjusting endpoints  
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.534650	0.085260	6.270819	0.0000
AR(2)	0.333279	0.084691	3.935222	0.0001
SAR(7)	-0.520490	0.086762	-5.999030	0.0000
SAR(14)	-0.342697	0.086158	-3.977537	0.0001
R-squared	0.576479	Mean dependent var		95988.44
Adjusted R-squared	0.566232	S.D. dependent var		411534.7
S.E. of regression	271041.1	<b>Akaike info criterion</b>		<b>27.88868</b>
Sum squared resid	9.11E+12	<b>Schwarz criterion</b>		<b>27.97781</b>
Log likelihood	-1780.876	Durbin-Watson stat		2.060182

Fig. 8 Output modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>

In questo modello, tutte le componenti risultano altamente significative, ma i criteri di Akaike e di Schwarz sono leggermente superiori a quelli del modello precedente.

Il correlogramma dei residui è abbastanza soddisfacente, anche se presenta ancora qualche correlazione ai ritardi stagionali:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.037	-0.037	0.1828	
		2	-0.075	-0.076	0.9211	
		3	0.060	0.054	1.3959	
		4	0.046	0.045	1.6847	
		5	-0.035	-0.023	1.8491	0.174
		6	0.063	0.065	2.3930	0.302
		7	-0.044	-0.050	2.6622	0.447
		8	0.015	0.022	2.6923	0.611
		9	-0.040	-0.051	2.9126	0.713
		10	0.111	0.111	4.6399	0.591
		11	0.068	0.076	5.2949	0.624
		12	-0.092	-0.077	6.5064	0.591
		13	-0.078	-0.078	7.3978	0.596
		14	-0.100	-0.151	8.8686	0.545
		15	0.015	0.014	8.9020	0.631
		16	-0.027	-0.041	9.0065	0.702
		17	-0.163	-0.150	12.985	0.449
		18	0.050	0.055	13.370	0.498
		19	0.133	0.128	16.082	0.377
		20	-0.013	0.039	16.108	0.445
		21	-0.166	-0.199	20.392	0.255
		22	-0.055	-0.098	20.874	0.286
		23	-0.059	-0.060	21.426	0.314
		24	-0.080	-0.059	22.449	0.317
		25	0.090	0.109	23.752	0.305
		26	0.077	0.072	24.717	0.311
		27	-0.042	0.051	25.006	0.350
		28	-0.158	-0.206	29.172	0.214
		29	0.203	0.109	36.125	0.070
		30	0.035	-0.027	36.329	0.086

Fig. 9 Correlogramma dei residui modello SARIMA(2,0,0)(2,1,0)



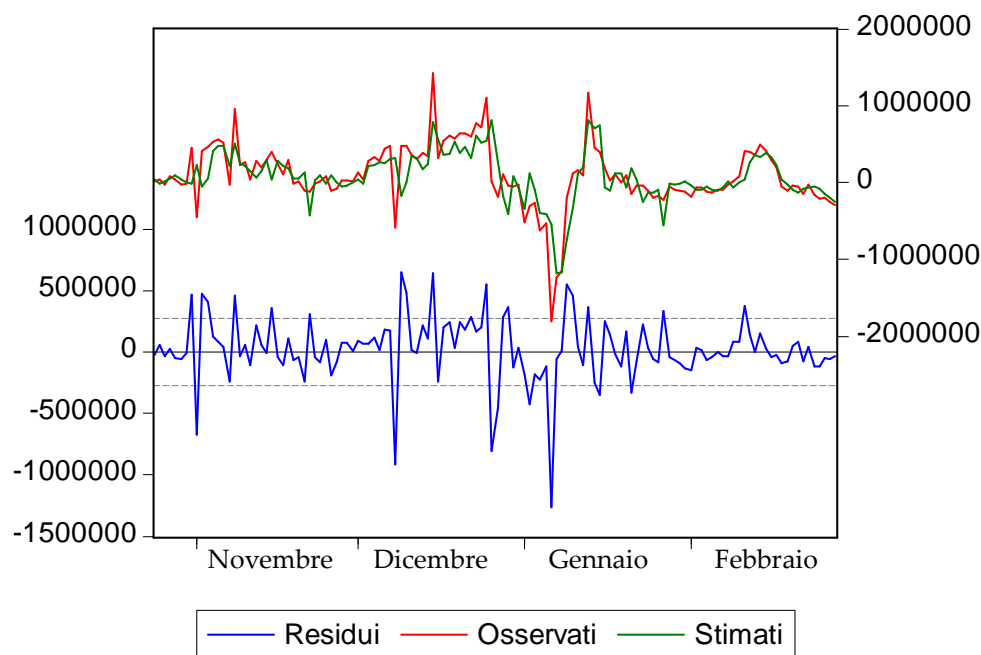


Fig. 10 Valori stimati modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>

A questo punto dobbiamo scegliere quale modello funziona meglio in fase previsiva. Per fare questo utilizziamo le previsioni e le confrontiamo con i valori realizzati attraverso due indici quantitativi, l'EQM e l'EMA, e scegliamo il modello per cui tali indici sono inferiori.

L'EQM, errore quadratico medio, si calcola facendo la radice quadrata della media dei quadrati degli errori della previsione rispetto alla serie reale.

L'EMA è invece la media dei valori assoluti degli errori.

$$EQM = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \qquad EMA = \frac{1}{n} \sum_{t=1}^n |e_t|$$

dove  $e_t$  = valore vero - valore previsto

Attraverso un grafico vediamo visivamente come si adattano le previsioni ai dati.

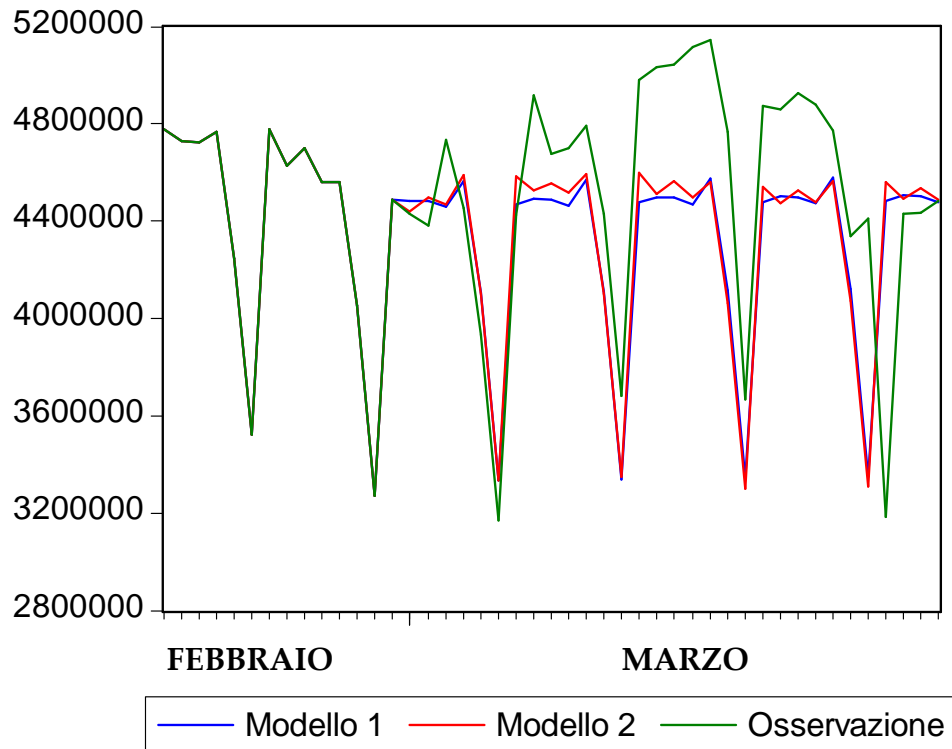


Fig. 11 Confronto previsioni minuti

Ora confrontiamo i due indici che valutano la bontà delle previsioni:

	Modello 1	Modello 2
EQM	<b>452751,9</b>	456318,8
EMA	352815,6	<b>350575,1</b>

Questo è un caso particolare, poiché l'errore quadratico medio ci farebbe optare per il primo modello, mentre l'errore medio assoluto per il

secondo. Dato che i valori sono molto vicini, potremmo considerare i due modelli equivalenti.

Scegliamo quindi il primo modello, quello che risultava migliore nell'adattamento ai dati, con i criteri di Akaike e Schwarz minori.

### 3.1.3 Modellazione delle chiamate

Anche per la variabile relativa alle chiamate giornaliere è stata necessaria una differenziazione stagionale di ordine sette.

La nuova serie si rivela molto simile a quella dei minuti che abbiamo osservato in precedenza:

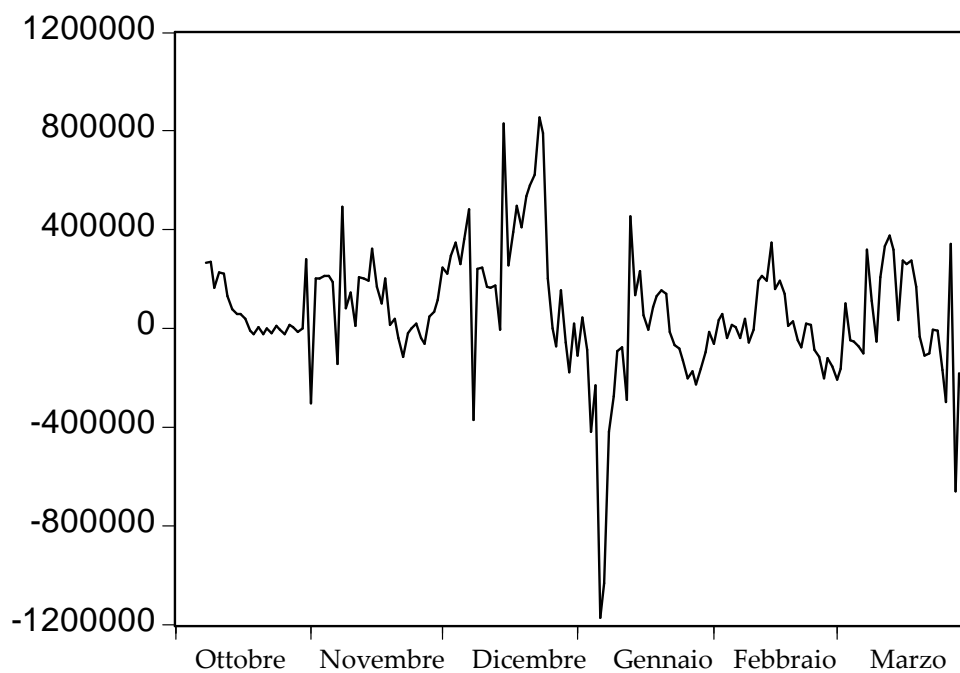


Fig. 12 Serie chiamate differenziata

Dall'analisi dell'autocorrelogramma della serie delle chiamate differenziate,  $D(\text{chiamate},0,7)$ , si evince, come potevamo aspettarci, che non vi è notevole differenza rispetto a quello relativo alla serie dei minuti.

Per questo motivo anche le stime dei modelli sono state praticamente identiche, con l'aggiunta di una componente a media mobile stagionale di ordine 14 al modello migliore:

**Dependent Variable: D(CHIAMATE,0,7)**

Method: Least Squares  
Sample(adjusted): 10/10/2004 2/28/2005  
Included observations: 142 after adjusting endpoints  
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.483060	0.083308	5.798479	0.0000
AR(2)	0.475267	0.086248	5.510502	0.0000
MA(2)	-0.190796	0.098663	-1.933817	0.0552
SMA(7)	-0.667175	0.082248	-8.111767	0.0000
SMA(14)	-0.273261	0.080253	-3.404985	0.0009
R-squared	0.626721	Mean dependent var		65208.08
Adjusted R-squared	0.615823	S.D. dependent var		259551.3
S.E. of regression	160875.2	<b>Akaike info criterion</b>		<b>26.84922</b>
Sum squared resid	3.55E+12	<b>Schwarz criterion</b>		<b>26.95330</b>
Log likelihood	-1901.295	Durbin-Watson stat		2.027420
Inverted AR Roots	.97	-.49		
Inverted MA Roots	.99	.75 -.36i	.75+.36i	.62+.78i
	.62 -.78i	.44	.19+.82i	.19 -.82i
	-.22+.97i	-.22 -.97i	-.44	-.52+.65i
	-.52 -.65i	-.84	-.89 -.43i	-.89+.43i

**Fig. 13 Output modello SARIMA(2,0,2)(0,1,2)<sub>7</sub>**

Il modello migliore, rispetto ai criteri di Akaike e di Schwarz, ha tutte le componenti significative, tranne l'MA(2) che, come al punto precedente non risulta significativo al livello del 5%. Il correlogramma dei residui mostra che non restano correlazioni al di fuori delle bande, quindi possiamo considerare il modello soddisfacente:

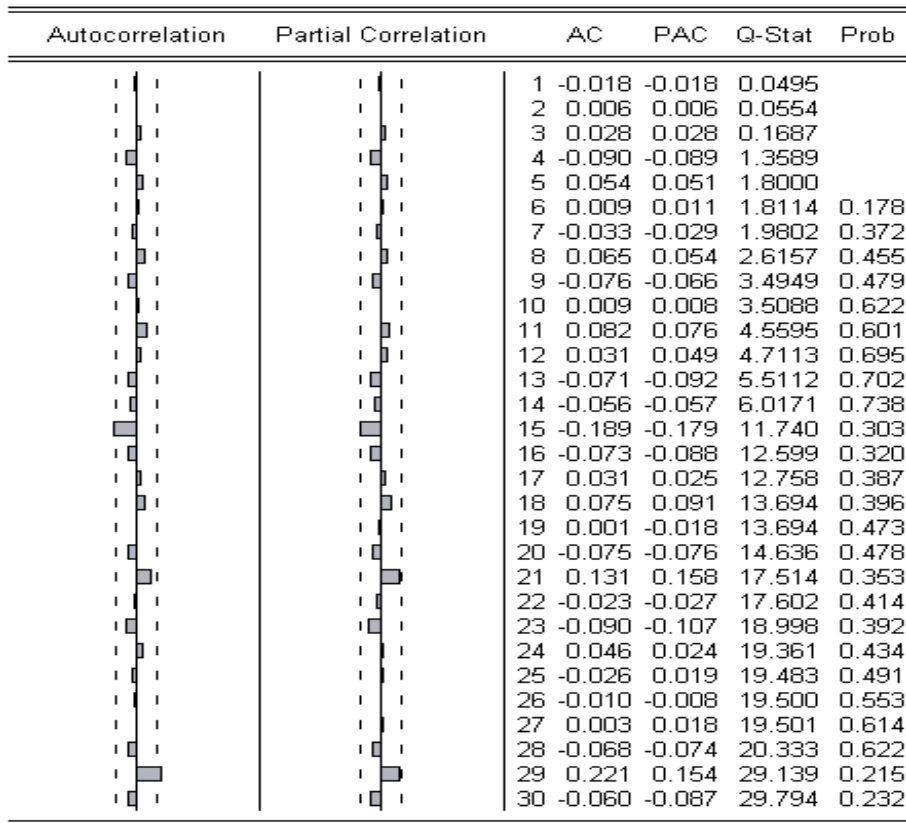


Fig. 14 Correlogramma dei residui modello SARIMA(2,0,2)(0,1,2)<sub>7</sub>

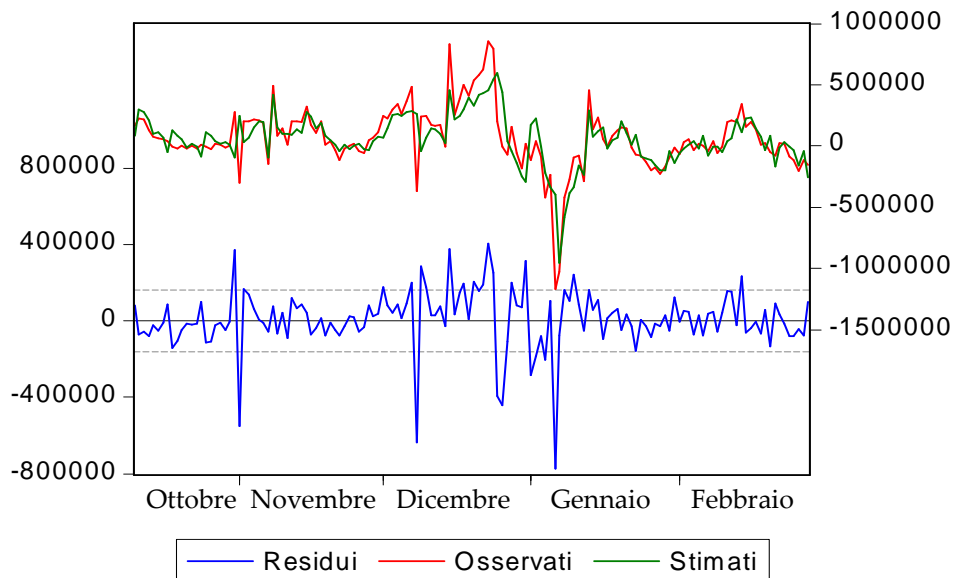


Fig. 15 Valori stimati modello SARIMA(2,0,2)(0,1,2)<sub>7</sub>

L'altro modello che abbiamo usato per le previsioni è un SARIMA(2,0,0)(2,1,0)<sub>7</sub> :

**Dependent Variable: D(CHIAMATE,0,7)**

Method: Least Squares  
Sample(adjusted): 10/24/2004 2/28/2005  
Included observations: 128 after adjusting endpoints  
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.520638	0.084099	6.190757	0.0000
AR(2)	0.361457	0.084005	4.302776	0.0000
SAR(7)	-0.497920	0.084162	-5.916188	0.0000
SAR(14)	-0.410337	0.084346	-4.864911	0.0000
R-squared	0.577582	Mean dependent var		65280.26
Adjusted R-squared	0.567362	S.D. dependent var		272023.8
S.E. of regression	178924.2	<b>Akaike info criterion</b>		<b>27.05806</b>
Sum squared resid	3.97E+12	<b>Schwarz criterion</b>		<b>27.14719</b>
Log likelihood	-1727.716	Durbin-Watson stat		2.048282
Inverted AR Roots	.92	.90 -.26i	.90+.26i	.77+.54i
	.77 -.54i	.36+.87i	.36 -.87i	.05 -.94i
	.05+.94i	-.39	-.45+.82i	-.45 -.82i
	-.70+.63i	-.70 -.63i	-.93+.16i	-.93 -.16i

**Fig. 16 Output modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>**

Questo modello ha tutte le variabili significative, ed il correlogramma dei residui mostra che le correlazioni rimangono all'interno delle bande. L'andamento dei residui non è quello tipico di un White Noise, ma questo è dovuto alla presenza degli outlier e dal cambiamento di regime della serie.

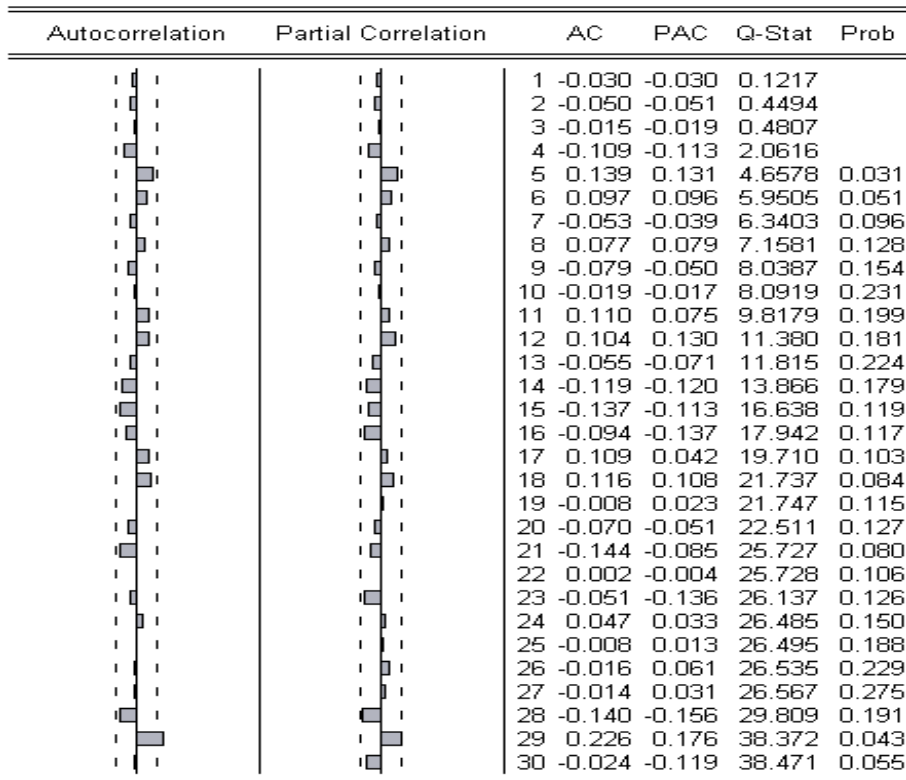


Fig. 17 Correlogramma dei residui modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>

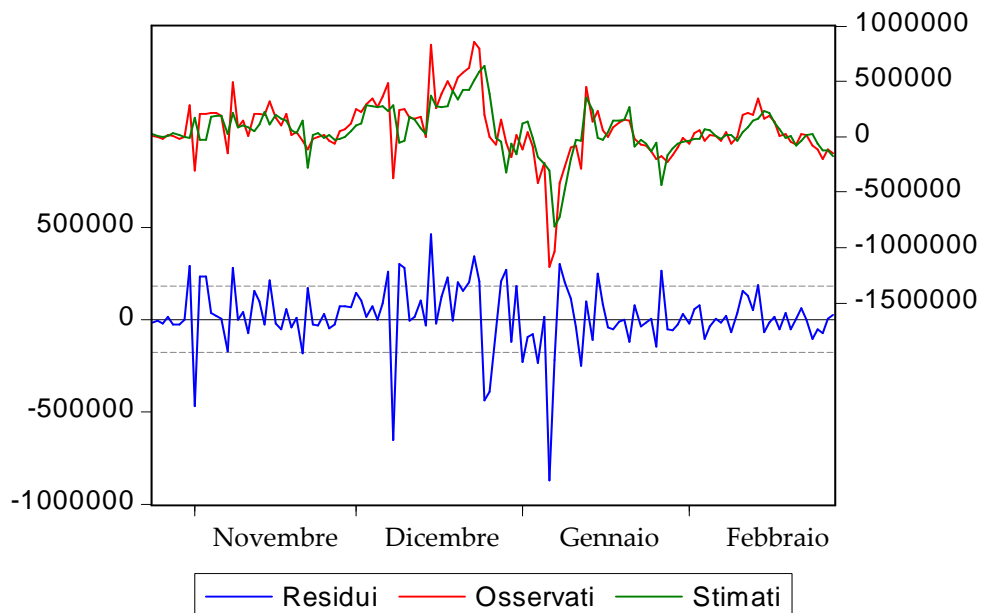
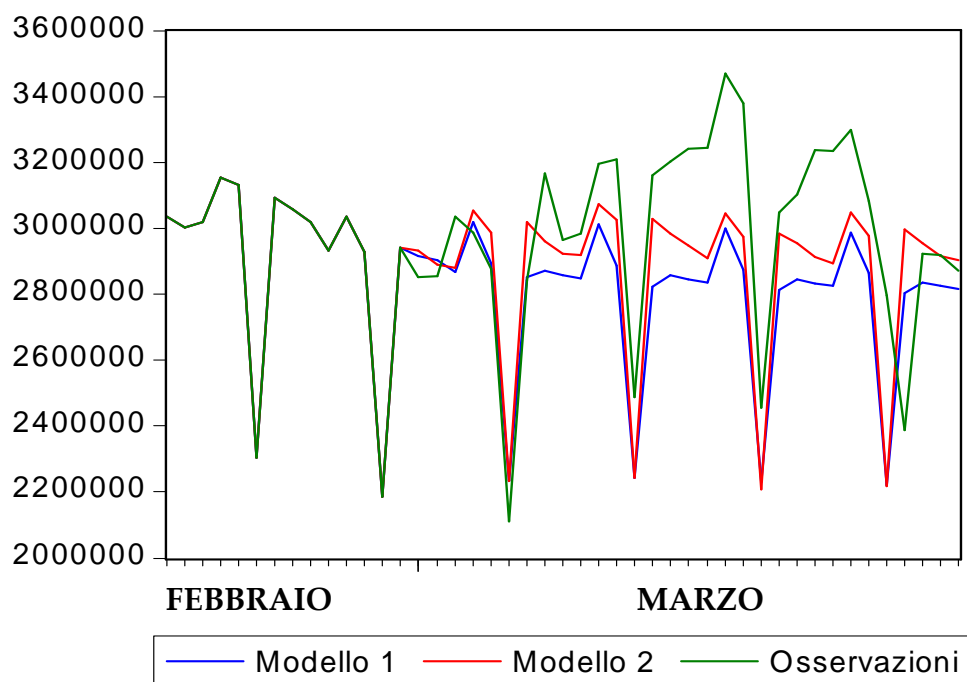


Fig. 18 Valori stimati modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>

Anche in questo caso, utilizzeremo le previsioni prodotte da ciascun modello e le confronteremo per scegliere il migliore.

Per quest'operazione abbiamo modellato i dati fino al 28 febbraio, ed abbiamo stimato poi l'ultimo mese, marzo, per poter confrontare le stime ottenute con i valori che si sono effettivamente verificati.

Come possiamo vedere dal grafico, entrambi i modelli sottostimano i valori effettivamente realizzati, anche se il secondo sembra avvicinarsi di più ai dati.



**Fig. 19 Confronto previsioni chiamate**

Confrontando gli indici quantitativi per valutare la bontà delle previsioni, scopriamo che, in questo caso, il modello "peggiore" in fase di modellazione produce previsioni migliori:



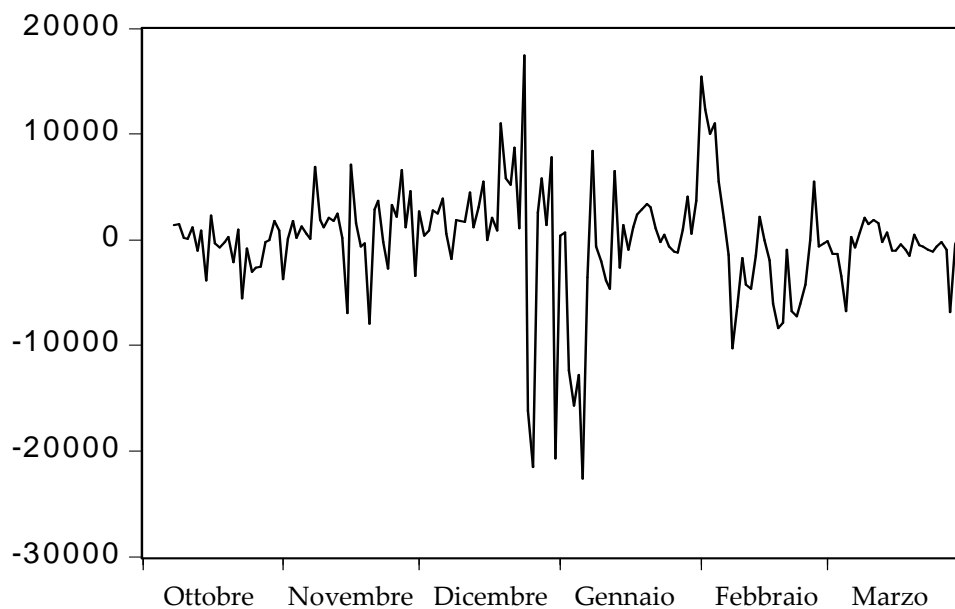
	Modello 1	Modello 2
EQM	288105,3	<b>251428,2</b>
EMA	241965,3	<b>199091,8</b>

Il secondo modello è preferibile al primo, dato che nella fase di previsione, dove interessa a noi, si comporta meglio, adattandosi in modo migliore alle osservazioni.

### **3.1.4 Modellazione delle attivazioni**

La serie delle attivazioni, come le precedenti, non è stazionaria, abbiamo bisogno quindi di una differenziazione.

La nuova variabile  $D(\text{attivazioni}, 0,7)$  ha un andamento migliore:



**Fig. 20 Serie attivazioni differenziata**

Da un'analisi del correlogramma della nuova serie, abbiamo notato la presenza di una forte correlazione al ritardo sette, e di una componente autoregressiva di ordine uno. Dopo alcuni tentativi, il modello migliore, secondo i criteri di Akaike e di Schwarz, e in base alla correlazione dei residui è stato ottenuto combinando una media mobile e una componente autoregressiva, entrambe di ordine uno, e una media mobile e una componente autoregressiva stagionali di ordine sette. L'output della stima è il seguente:

**Dependent Variable: D(ATTI,0,7)**

Method: Least Squares  
Sample(adjusted): 10/16/2004 2/28/2005  
Included observations: 136 after adjusting endpoints  
Convergence achieved after 13 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<b>AR(1)</b>	<b>0.884652</b>	<b>0.068969</b>	<b>12.82679</b>	<b>0.0000</b>
<b>SAR(7)</b>	<b>0.177384</b>	<b>0.093544</b>	<b>1.896262</b>	<b>0.0601</b>
<b>MA(1)</b>	<b>-0.590513</b>	<b>0.115193</b>	<b>-5.126301</b>	<b>0.0000</b>
<b>SMA(7)</b>	<b>-0.937341</b>	<b>0.025045</b>	<b>-37.42664</b>	<b>0.0000</b>
R-squared	0.432497	Mean dependent var		-18.35294
Adjusted R-squared	0.419600	S.D. dependent var		6031.215
S.E. of regression	4594.822	<b>Akaike info criterion</b>		<b>19.73222</b>
Sum squared resid	2.79E+09	<b>Schwarz criterion</b>		<b>19.81788</b>
Log likelihood	-1337.791	Durbin-Watson stat		1.901994
Inverted AR Roots	.88	.78	.49 -.61i	.49+.61i
	-.17 -.76i	-.17+.76i	-.70 -.34i	-.70+.34i
Inverted MA Roots	.99	.62 -.77i	.62+.77i	.59
	-.22 -.97i	-.22+.97i	-.89+.43i	-.89 -.43i

**Fig. 21 Output modello SARIMA(1,0,1)(1,1,1)<sub>7</sub>**

La componente autoregressiva stagionale di ordine sette non risulta significativa al livello del 5%, ma non la togliamo dal modello perché è fondamentale per l'annullamento di una dipendenza stagionale al ritardo sette. La funzione di autocorrelazione dei residui del modello è dunque:

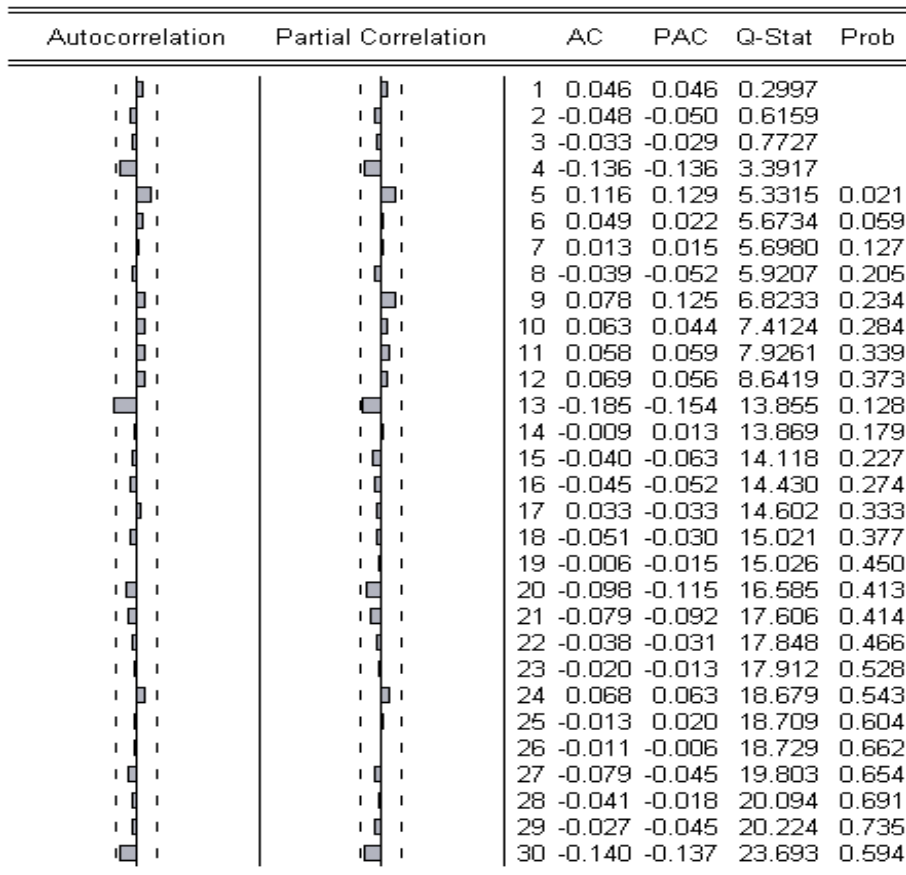


Fig. 22 Correlogramma dei residui modello SARIMA(1,0,1)(1,1,1)<sub>7</sub>

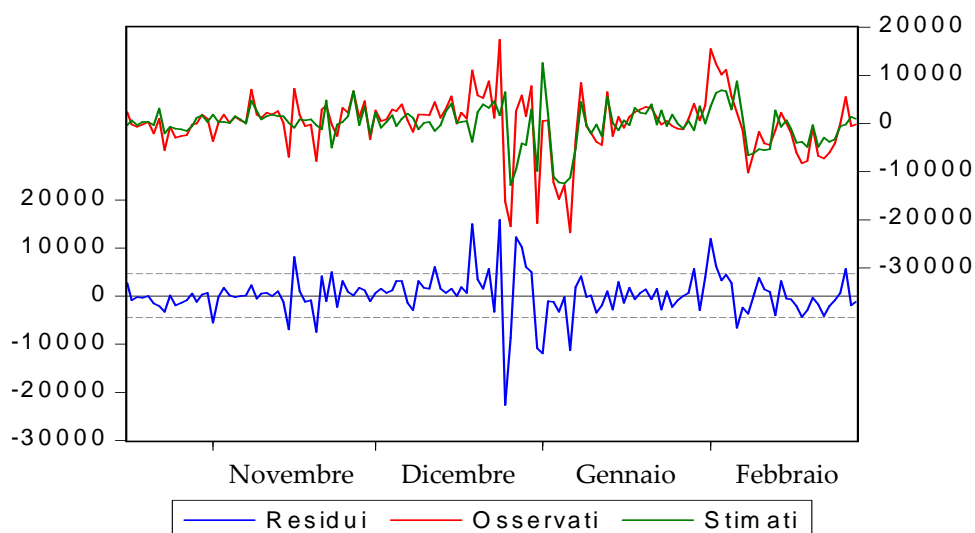


Fig. 23 Valori stimati modello SARIMA(1,0,1)(1,1,1)<sub>7</sub>

Anche in questo caso abbiamo scelto un modello alternativo al precedente, che risulta essere un SARIMA(5,0,0)(2,1,0)<sub>7</sub>, dove sono significative solo le componenti AR(1) e AR(5).

Dependent Variable: D(ATTL0,7)

Method: Least Squares

Sample(adjusted): 10/27/2004 2/28/2005

Included observations: 125 after adjusting endpoints

Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.456747	0.079187	5.767944	0.0000
AR(5)	0.226258	0.080082	2.825339	0.0055
SAR(7)	-0.486998	0.092027	-5.291931	0.0000
SAR(14)	-0.224074	0.091779	-2.441452	0.0161
R-squared	0.314234	Mean dependent var		75.42400
Adjusted R-squared	0.297231	S.D. dependent var		6254.658
S.E. of regression	5243.362	<b>Akaike info criterion</b>		<b>19.99879</b>
Sum squared resid	3.33E+09	<b>Schwarz criterion</b>		<b>20.08930</b>
Log likelihood	-1245.924	Durbin-Watson stat		2.094107
Inverted AR Roots	.86	.86 -.27i	.86+.27i	.74+.50i
	.74 -.50i	.33 -.84i	.33+.84i	.32+.68i
	.32 -.68i	.07 -.90i	.07+.90i	-.45 -.78i
	-.45+.78i	-.53 -.43i	-.53+.43i	-.66+.61i
	-.66 -.61i	-.89+.13i	-.89 -.13i	

Fig. 24 Output modello SARIMA(5,0,0)(2,1,0)<sub>7</sub>

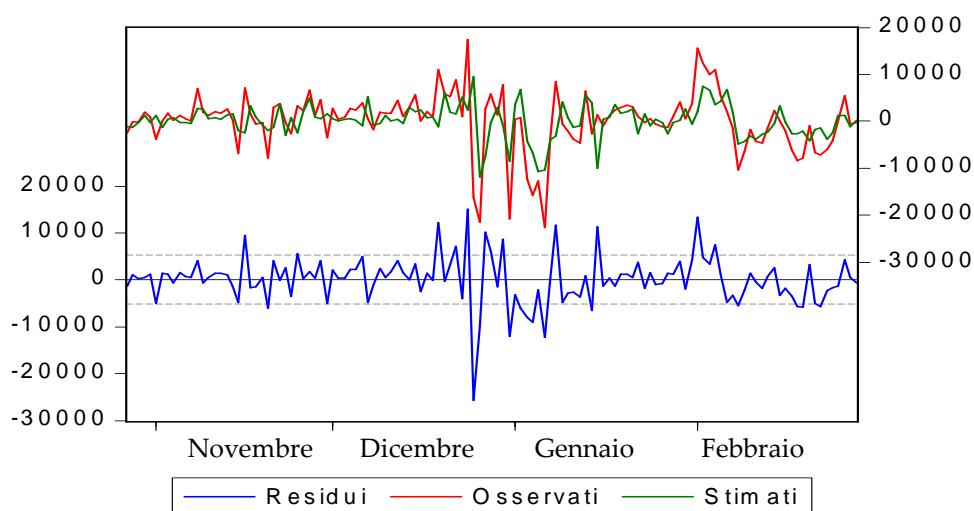


Fig. 25 Valori stimati modello SARIMA(5,0,0)(2,1,0)<sub>7</sub>

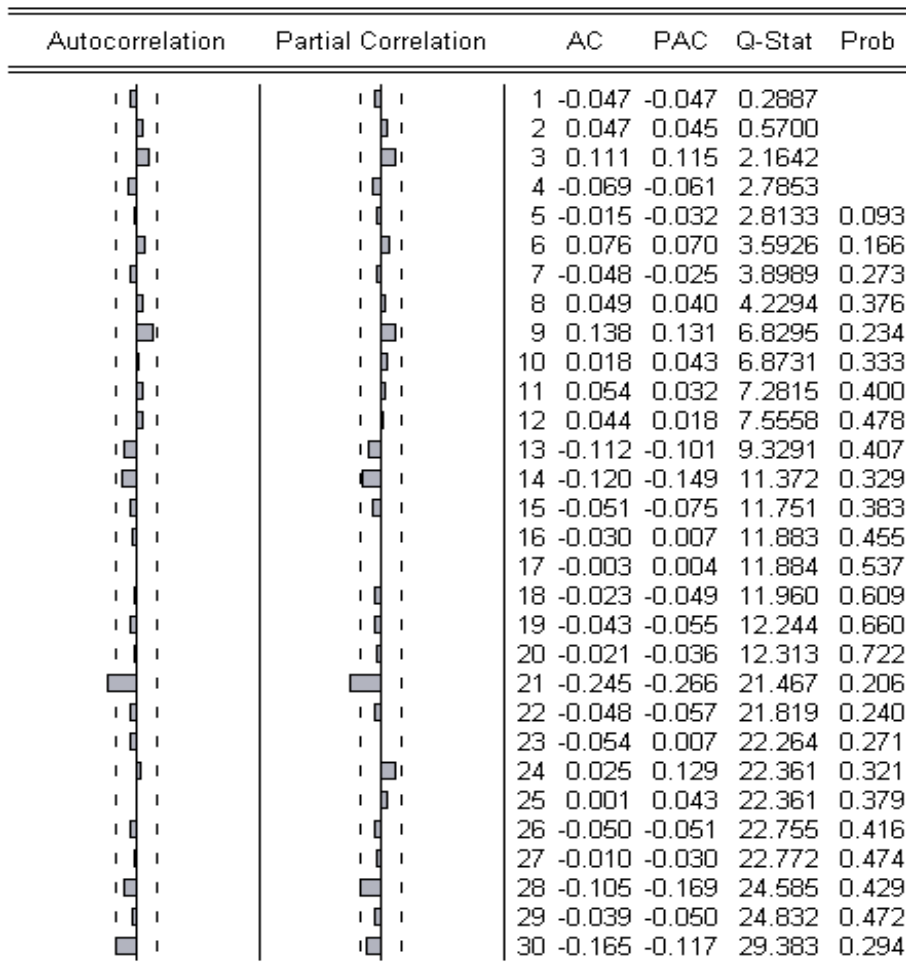


Fig. 26 Correlogramma dei residui modello SARIMA(5,0,0)(2,1,0)<sub>7</sub>

A questo punto, per scegliere il modello più adatto al nostro problema, vediamo quale dei due si comporta meglio in fase di previsione.

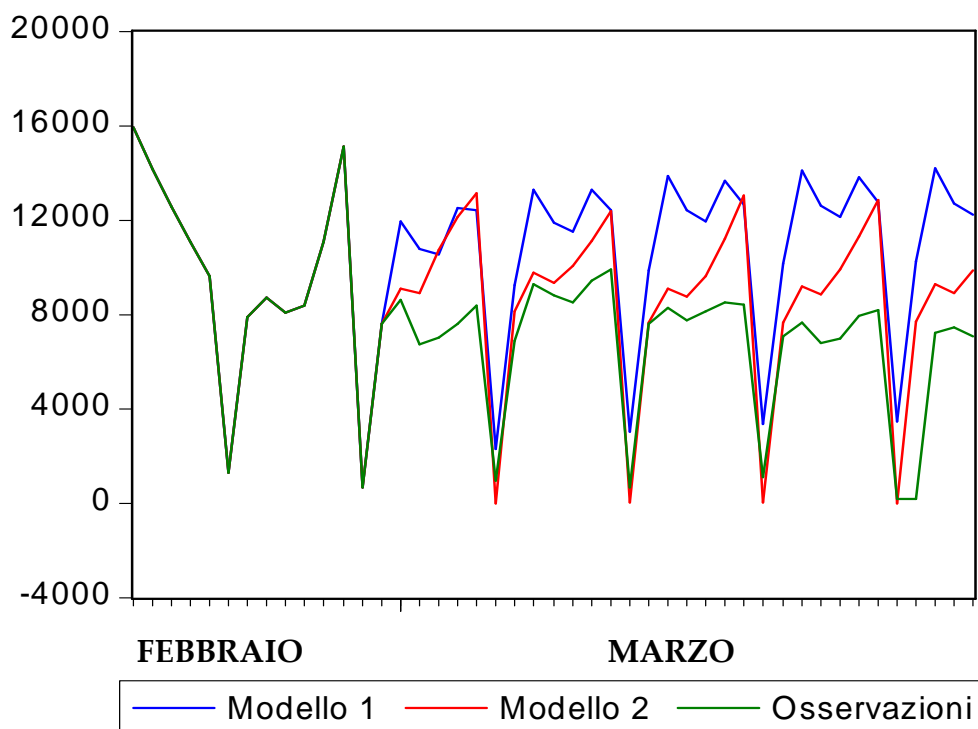


Fig. 27 Confronto previsioni attivazioni

Le previsioni sovrastimano le effettive osservazioni, a causa del picco presente nella settimana che precede l'inizio del mese di previsione. Graficamente il secondo modello sembra avvicinarsi di più ai dati reali, come confermano gli indici quantitativi:

	Modello 1	Modello 2
EQM	4592,4	<b>2727,2</b>
EMA	4261,7	<b>2137,9</b>

Il modello che prevede meglio è dunque il secondo, quello che si comporta "peggio" nella modellazione.

### **3.1.5 Scelta dei modelli migliori**

I modelli che hanno previsto meglio le osservazioni sono stati:

- per i minuti  
modello SARIMA(2,0,2)(0,1,1)<sub>7</sub>  
con componente MA(1) non significativa  
EQM = 452751,9  
EMA = 352815,6
- per le chiamate  
modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>  
EQM = 251428,2  
EMA = 199091,8
- per le attivazioni  
modello SARIMA(5,0,0)(2,1,0)<sub>7</sub>  
con componenti AR(2), AR(3), AR(4) non significative  
EQM = 2727,2  
EMA = 2137,9





## Capitolo 4

### LISCIAMENTO ESPONENZIALE

#### 4.1 PREVISIONI TRAMITE LISCIAMENTO ESPONENZIALE

Il metodo del liscio esponenziale è molto usato all'interno delle aziende per fare previsioni. È una tecnica utilizzata per smussare una serie storica al fine di evidenziare i movimenti di lungo periodo della serie stessa.

In questo capitolo descriveremo solo i metodi che considerano anche la componente stagionale, visto le caratteristiche dei nostri dati.

##### 4.1.1 I metodi di Holt-Winters stagionali

###### 4.1.1.1 Metodo stagionale additivo

Il metodo di Holt-Winters stagionale additivo, si basa sull'ipotesi che in prossimità di  $n$  la previsione con orizzonte temporale  $k$  può essere fornita dalla retta:

$$F_{n,k} = \bar{y}_n + T_n(n+k-n) + S_n \equiv \bar{y}_n + T_n k + S_n$$

con  $S_n$  fattore stagionale di periodo  $s$ .

Le stime di  $\bar{y}_n$ ,  $T_n$  e  $S_n$  sono date dalle seguenti formule di aggiornamento:

$$\bar{y}_n = \alpha(\hat{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha)(y_n - \hat{S}_{n-s}), \quad 0 < \alpha < 1$$

$$\hat{T}_n = \beta\hat{T}_{n-1} + (1 - \beta)(\hat{y}_n - \hat{y}_{n-1}), \quad 0 < \beta < 1$$

$$\hat{S}_n = \gamma\hat{S}_{n-s} + (1 - \gamma)(y_n - \hat{y}_n), \quad 0 < \gamma < 1$$

Con  $\alpha$ ,  $\beta$ ,  $\gamma$  costanti di liscio che possono essere determinate soggettivamente o minimizzando una qualche misura dell'errore di previsione (EQM, EMA).

La previsione si effettua considerando gli ultimi valori della serie liscia:

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

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

#### 4.1.1.2 Metodo stagionale moltiplicativo

Il metodo di Holt-Winters stagionale moltiplicativo, si basa sull'ipotesi che in prossimità di  $n$  la previsione con orizzonte temporale  $k$  può essere fornita dalla retta:

$$F_{n,k} = [\bar{y}_n + T_n(n+k-n)]S_n \equiv (\bar{y}_n + T_n k)S_n$$

con  $S_n$  fattore stagionale di periodo  $s$ .

Le stime di  $\bar{y}_n$ ,  $T_n$  e  $S_n$  sono date dalle seguenti formule di aggiornamento:

$$\bar{y}_n = \alpha(\hat{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha) \frac{y_n}{\hat{S}_{n-s}}, \quad 0 < \alpha < 1$$

$$\hat{T}_n = \beta \hat{T}_{n-1} + (1 - \beta)(\hat{y}_n - \hat{y}_{n-1}), \quad 0 < \beta < 1$$

$$\hat{S}_n = \gamma \hat{S}_{n-s} + (1 - \gamma) \frac{y_n}{\hat{y}_n}, \quad 0 < \gamma < 1$$

Anche in questo caso  $\alpha$ ,  $\beta$  e  $\gamma$  sono costanti di lisciamiento che possono essere determinate soggettivamente o minimizzando gli indici di errore di previsione (EQM, EMA).

La previsione viene effettuata considerando gli ultimi valori della serie lisciata:

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

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

Per ciascuna serie abbiamo utilizzato sia il metodo additivo, sia quello moltiplicativo, e abbiamo scelto quello che prevede meglio l'ultimo mese delle osservazioni.

## 4.2 PREVISIONE DEI MINUTI

La scelta della combinazione migliore delle costanti di liscio è stata condotta cercando la terna che minimizza gli indici di errore di previsione (EQM ed EMA), sulla stima dei primi cinque mesi dei dati.

Utilizzando il metodo additivo i valori più adatti al nostro scopo sono stati  $\alpha=0.45$ ,  $\beta=0.95$  e  $\gamma=0.50$ , mentre con il metodo moltiplicativo sono stati  $\alpha=0.40$ ,  $\beta=0.95$  e  $\gamma=0.70$ .

Le previsioni generate sono state le seguenti:

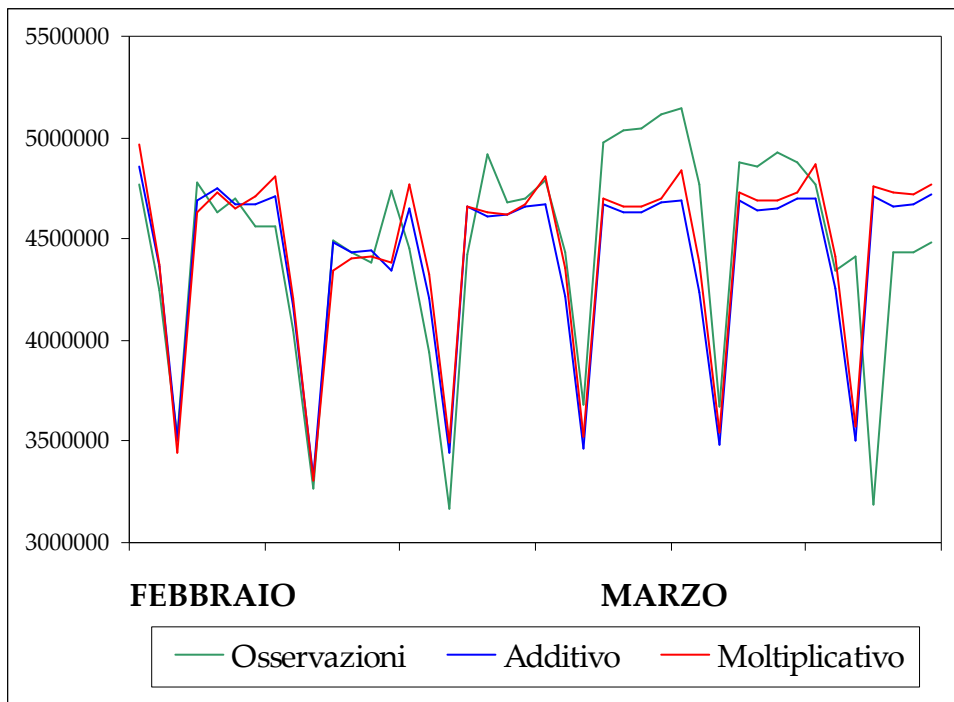


Fig. 28 Liscio minuti

Le due previsioni sembrano abbastanza simili, per scegliere uno dei modelli andiamo a vedere quale dei due metodi ha ottenuto un EQM minore:

	H-W Additivo	H-W Moltiplicativo
EQM	426742,9	<b>418953,6</b>
EMA	314108,4	<b>296869,4</b>

Per i minuti utilizzeremo quindi il metodo stagionale moltiplicativo.

### 4.3 PREVISIONE DELLE CHIAMATE

Anche per le chiamate, abbiamo utilizzato entrambi i metodi per poi scegliere quello che prevede meglio.

La terna migliore di costanti di liscio per il modello additivo è stata  $\alpha=0.50$ ,  $\beta=0.95$  e  $\gamma=0.60$ , mentre con il metodo moltiplicativo è stata  $\alpha=0.45$ ,  $\beta=0.95$  e  $\gamma=0.80$ .

Questi valori hanno prodotto le seguenti previsioni:

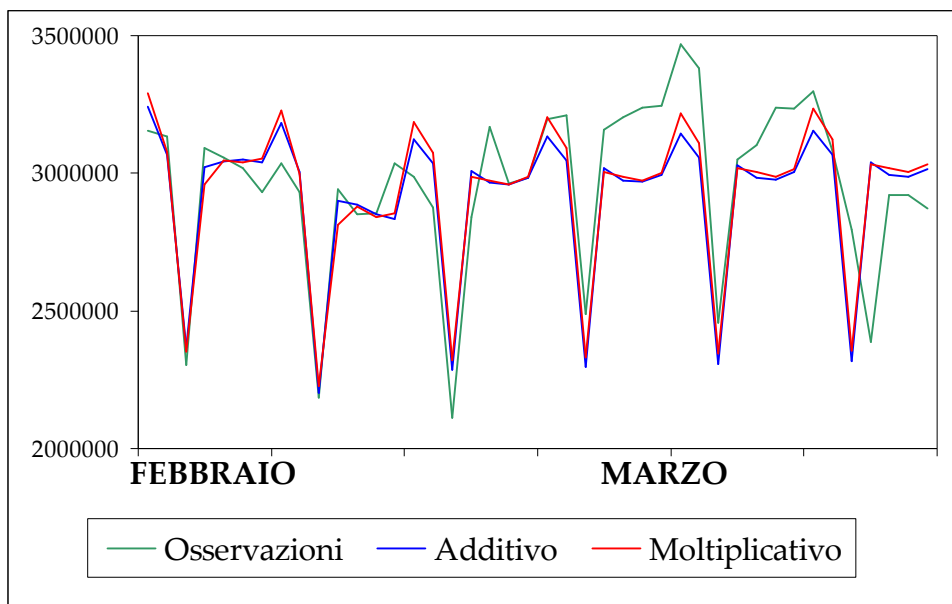


Fig. 29 Liscio chiamate

Anche in questo caso le due previsioni sembrano abbastanza simili, anche se il metodo moltiplicativo sembra adattarsi meglio ai dati.

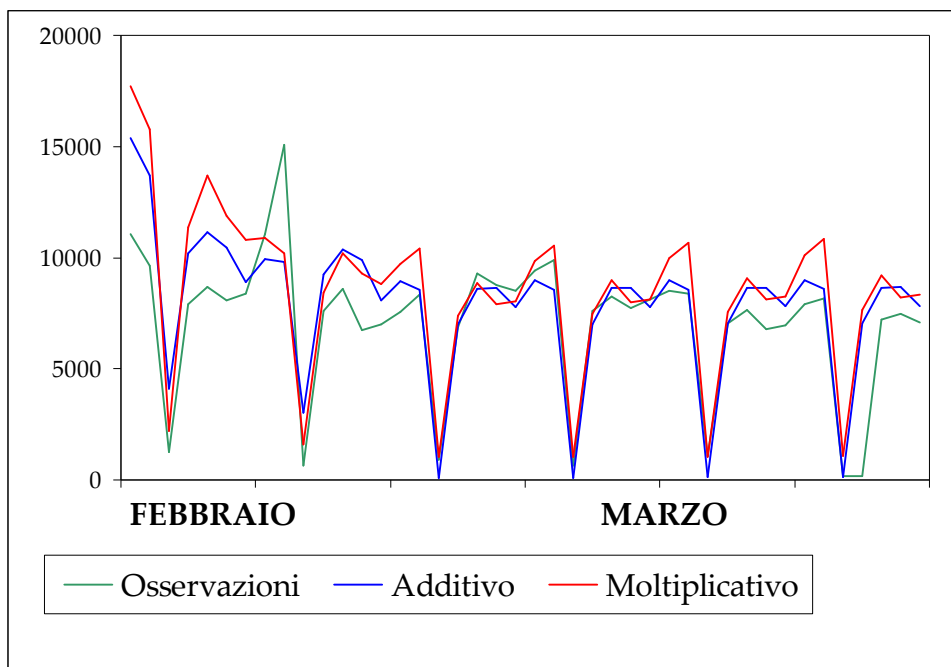
Andiamo a vedere cosa dicono gli indici di errore:

	H-W Additivo	H-W Moltiplicativo
EQM	231247,9	<b>219854,1</b>
EMA	184607,7	<b>174530,4</b>

Anche per le chiamate utilizzeremo il metodo stagionale moltiplicativo.

#### **4.4 PREVISIONE DELLE ATTIVAZIONI**

Con la variabile relativa alle attivazioni abbiamo notato subito come il metodo additivo si adattasse bene ai dati:



**Fig. 30** Lisciamento attivazioni

Anche gli indici di errore confermano quello che abbiamo visto dal grafico:

	H-W Additivo	H-W Moltiplicativo
EQM	<b>1540,9</b>	1884,2
EMA	<b>925,1</b>	1240,9

Per le attivazioni scegliamo quindi il modello additivo.





## Capitolo 5

### MIGLIOR PREVISIONE

Per completare l'analisi, andiamo a scegliere, per ogni serie in esame, il metodo che è andato meglio in fase di previsione.

Per fare questo utilizziamo ancora i due indici di errore, l'EQM e l'EMA, e confronteremo quelli del modello che ha effettuato la previsione migliore con il metodo (S)ARIMA, e quelli ottenuti tramite il lisciamiento esponenziale stagionale.

#### **5.1 PREVISIONE FINALE MINUTI**

I modelli che hanno ottenuto i migliori risultati sono stati il modello SARIMA(2,0,2)(0,1,1)<sub>7</sub>, e il modello di Holt-Winters stagionale moltiplicativo con costanti di lisciamiento  $\alpha=0.40$ ,  $\beta=0.95$  e  $\gamma=0.70$ .

	SARIMA	H-W Moltiplicativo
EQM	452751,9	<b>418953,6</b>
EMA	352815,6	<b>296869,4</b>

Gli indici di bontà delle previsioni ci portano a preferire il modello di Holt-Winters stagionale moltiplicativo.

## 5.2 PREVISIONE FINALE CHIAMATE

Per la serie delle chiamate, i risultati migliori li abbiamo ottenuti con il modello SARIMA(2,0,0)(2,1,0)<sub>7</sub>, e con il metodo di Holt-Winters stagionale moltiplicativo con costanti di liscio  $\alpha=0.45$ ,  $\beta=0.95$  e  $\gamma=0.80$ .

	SARIMA	H-W Moltiplicativo
EQM	251428,2	<b>219854,1</b>
EMA	199091,8	<b>174530,4</b>

Anche in questo appare evidente come il metodo di Holt-Winters stagionale moltiplicativo sia quello che porta a previsioni migliori.

## 5.3 PREVISIONE FINALE ATTIVAZIONI

La variabile relativa alle attivazioni è stata modellata tramite un SARIMA(5,0,0)(2,1,0)<sub>7</sub>, ed il metodo di Holt-Winters stagionale additivo con costanti di liscio  $\alpha=0.45$ ,  $\beta=0.95$  e  $\gamma=0.80$ .

	SARIMA	H-W Additivo
EQM	2727,2	<b>1540,9</b>
EMA	2137,9	<b>925,1</b>

Come per le serie precedenti, il metodo di Holt-Winters stagionale additivo ha fornito previsioni migliori rispetto al modello SARIMA.

## **5.4 CONSIDERAZIONI FINALI**

Le tre variabili che sono state oggetto d'analisi, presentano delle caratteristiche molto complicate da modellare.

Questo perché nel settore all'interno del quale opera l'azienda da noi presa in considerazione, partecipano una serie di fattori esogeni impossibili da modellare e da prevedere, quali le promozioni sull'acquisto di una nuova sim da parte dell'azienda XXX, le promozioni da parte delle altre compagnie telefoniche, il lancio di nuovi piani telefonici...

Senza dubbio però, questa ricerca ha dato alcune linee guida per effettuare programmazioni di mercato ad alto livello.

Quello che è parso evidente è che il metodo di Holt-Winters stagionale, porta ad ottenere previsioni migliori rispetto agli altri metodi utilizzati.

Molte aziende, infatti, effettuano le loro previsioni attraverso questo metodo, che, pur nella sua semplicità, in molti casi porta ad ottenere buoni risultati.



## **BIBLIOGRAFIA**

- (1) "Introduzione all'analisi delle serie storiche economiche", Di Fonzo, Lisi.
- (2) "Analisi moderna delle serie storiche", Domenico Piccolo, Franco Angeli 1983
- (3) "Complementi di statistica economica", Di Fonzo, Lisi, Cleup ed. Padova 2001
- (4) "Econometria delle serie storiche", James D. Hamilton, Monduzzi Bologna 1995

## **BIBLIOGRAFIA DEI SITI INTERNET**

- (1) <http://telefonino.leonardo.it/index.html>
- (2) <http://www.cellularitalia.com>
- (3) <http://www.mondo3.com>