



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE
ECONOMICHE ED AZIENDALI "M. FANNO"**

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

"MODELLI LINEARI E NON LINEARI PER SERIE STORICHE"

RELATRICE:

PROF.SSA LUISA BISAGLIA

DIPARTIMENTO DI AFFERENZA RELATRICE:

DIPARTIMENTO DI SCIENZE STATISTICHE

LAUREANDA: ANNA ZAVAN

MATRICOLA N. 1160969

ANNO ACCADEMICO 2019/2020

Il candidato, sottoponendo il presente lavoro, dichiara, sotto la propria personale responsabilità, che il lavoro è originale e che non sia stato già sottoposto, in tutto o in parte, dal candidato o da altri soggetti, in altre Università italiane o straniere ai fini del conseguimento di un titolo accademico. Il candidato dichiara altresì che tutti i materiali utilizzati ai fini della predisposizione dell'elaborato sono stati opportunamente citati nel testo e riportati nella sezione finale 'Riferimenti bibliografici' e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo al documento originale.

INDICE

1. INTRODUZIONE	4
2. SERIE STORICHE E TECNICHE DI LISCIAMENTO	7
2.1 Il Modello Moltiplicativo per l'Analisi delle Serie Storiche	8
2.2 Il Lisciamento di una Serie Storica	10
2.3 Il Metodo delle Medie Mobili	11
2.4 Il Metodo del Lisciamento Esponenziale	15
3. MODELLI LINEARI E NON LINEARI	19
3.1 Modello Lineare	19
3.2 Modelli Non Lineari	22
3.3 Il Modello Quadratico	22
3.4 Il Modello Esponenziale	24
4. MODELLI AUTOREGRESSIVI	26
4.1 Modelli Autoregressivi Lineari del p-esimo ordine	26
4.2 Modelli Autoregressivi Non Lineari (NLAR)	38
4.3 Modelli Autoregressivi a Soglia (SETAR)	40
5. CONCLUSIONI	42
6. BIBLIOGRAFIA	43

1. INTRODUZIONE

Spesso nella vita di tutti i giorni siamo costretti, o meglio, abbiamo la possibilità e la fortuna di effettuare delle scelte. A volte si tratta di cose di piccolo conto, come scegliere un percorso invece di un altro, o un luogo di vacanza nuovo rispetto alla meta dell'ultimo viaggio. A volte si tratta di scelte più importanti, quali la città in cui acquistare casa, la scuola da frequentare, per non parlare di scelte che coinvolgono la sfera affettiva di ciascuno.

Molte di queste scelte sono soggettive e del tutto incognite: non è dato prevedere le conseguenze della scelta che effettueremo e nella maggior parte dei casi non ci resta che affidarci al nostro istinto e al nostro buonsenso.

Ma nella vita di ciascuno di noi, può capitare di dover effettuare scelte strettamente legate all'ambito economico, sia per quanto attiene la sfera privata, come la scelta di un mutuo o di un piano di investimento per i nostri risparmi, sia per quanto riguarda il nostro ambito professionale, quali puntare su un determinato segmento di mercato piuttosto di un altro, o concentrare gli investimenti su una nuova tecnologia anziché migliorare l'esistente.

Per quanto l'istinto e il buonsenso rimangano dei principi guida importanti, è chiaro che di fronte a scelte di tipo economico un approccio più analitico al problema è da preferirsi, in particolar modo se ci consentirà di effettuare delle previsioni sulle conseguenze determinate dalle diverse opzioni che abbiamo a disposizione.

Il tema della previsione diventa quindi fondamentale al fine di effettuare delle scelte che si rivelino le più appropriate per il problema in esame. Pensiamo ad esempio alla fortuna di poter amministrare un nostro portafoglio titoli: è chiaro che diviene essenziale l'essere in grado di prevedere l'evoluzione del prezzo di un titolo al fine di valutare le migliori scelte per una proficua gestione del portafoglio. Naturalmente il problema è tutt'altro che banale, altrimenti chiunque con una conoscenza matematica e statistica più o meno approfondita potrebbe agevolmente ottenere risultati finanziari eccezionali semplicemente prevedendo l'andamento dei mercati. In realtà, come vedremo, l'analisi di fenomeni storici per prevederne il futuro andamento è uno dei problemi più complessi che si presentino ad un analista. Questo perché, anche in sistemi dinamici governati da equazioni molto semplici, può accadere che a fronte di una piccola variazione delle condizioni iniziali corrisponda un'evoluzione del fenomeno completamente diversa da quella che ci si aspetterebbe in mancanza della suddetta variazione. Si tratta del famoso "effetto farfalla" teorizzato per la prima volta agli inizi del XX secolo dal grande matematico Henri Poincaré.

Nonostante la complessità dei fenomeni sottostanti non è consentito arrendersi: è da considerarsi infatti preferibile una previsione imperfetta piuttosto che nessuna previsione.

L'obiettivo di questo lavoro è presentare alcune delle tecniche di previsione disponibili con particolare riferimento alle serie storiche; verranno brevemente illustrati i metodi di lisciamento, alcuni metodi di analisi lineari e non lineari, ed infine l'utilizzo dei modelli autoregressivi.

Nello svolgere questa presentazione ci riferiremo anche a due serie storiche reali che presentano caratteristiche utili allo scopo del lavoro: si tratta dei prezzi di chiusura giornalieri di due Società quotate sul mercato, facenti parte dell'indice FTSE MIB, ovvero Azimut Holding S.p.A. ed Enel S.p.A., considerate ciascuna in due periodi di 24 giorni successivi.

È stato considerato un periodo di tempo relativamente limitato di 24 giorni al fine di non appesantire la parte di calcolo e di rappresentazione grafica, ed inoltre i periodi sono stati selezionati in modo che le relative serie ottenute presentino caratteristiche utili allo studio. Per Azimut si è pertanto selezionato il periodo tra il 15 maggio e il 17 giugno 2020, per Enel il periodo tra il 6 aprile e il 7 maggio 2020.

Utilizzeremo quindi i vari modelli di analisi presentati per fare delle previsioni sul prezzo relativo al 25° giorno, ovvero rispettivamente il 18 giugno 2020 per Azimut e l'8 maggio 2020 per Enel. Nelle figure 1 e 2 sono rappresentati i due grafici relativi all'andamento dei prezzi nei periodi in esame, ricavati dal sito web Morning Star, dal quale sono stati ricavati anche i prezzi di chiusura giornalieri dei due titoli nei periodi considerati.

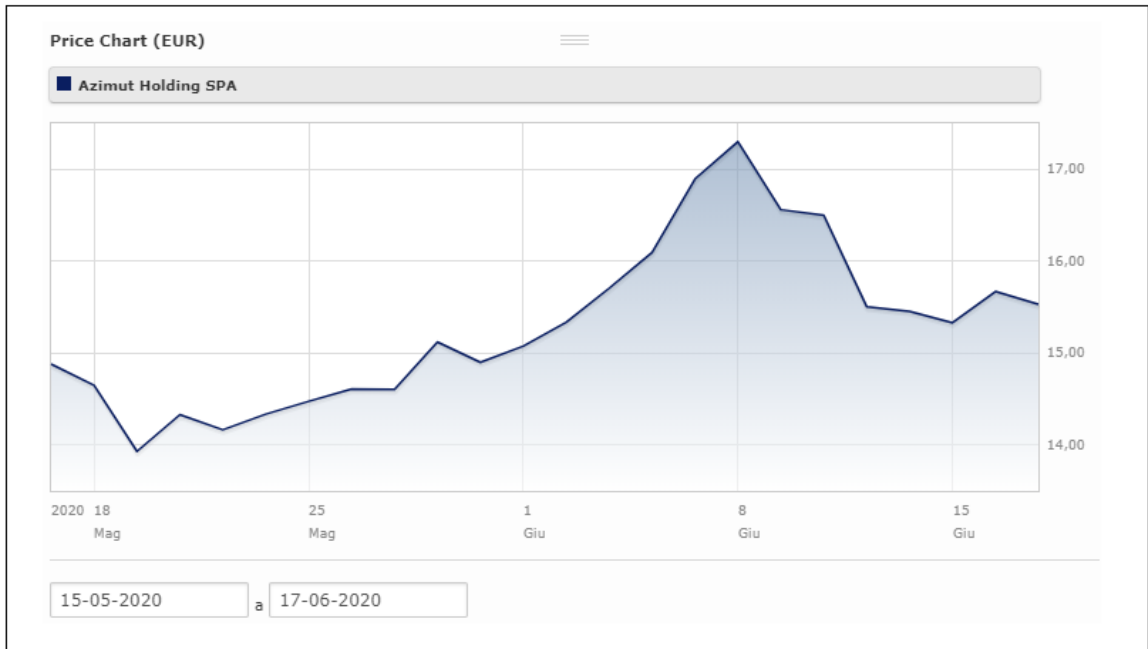


Fig. 1: Grafico relativo all'andamento dei prezzi di Azimut Holding S.p.A. nel periodo tra il 15 maggio e il 17 giugno 2020 (fonte: sito web Morning Star).



Fig. 2: Grafico relativo all'andamento dei prezzi di Enel S.p.A. nel periodo tra il 6 aprile e il 7 maggio 2020 (fonte: sito web Morning Star).

2. SERIE STORICHE E TECNICHE DI LISCIAMENTO

Per effettuare delle previsioni si può procedere sostanzialmente in due modi: da un punto di vista qualitativo oppure da un punto di vista quantitativo. In realtà i due approcci possono essere utilizzati proficuamente anche insieme, in quanto è spesso utile effettuare un'analisi qualitativa prima di spingersi oltre in un'analisi di tipo quantitativo.

Quando però non si dispone di dati storici, ovvero di rilevamenti effettuati o di stime affidabili per il fenomeno che si vuole analizzare e quindi prevedere, la strada che si deve percorrere è necessariamente quella dei metodi di previsione qualitativi. Un esempio può essere quello relativo alla previsione delle vendite di una nuova azienda che si affaccia per la prima volta sul mercato: anche in questo caso in realtà qualche considerazione di tipo quantitativo può essere fatta, magari considerando i dati dei principali competitor del mercato di riferimento.

Si tratta comunque di metodi che lasciano largo spazio alla soggettività e che esulano dall'argomento di questa tesi; ci limitiamo quindi a citare i più noti: *factor listing method*, *expert opinion* e *Delphi technique* (Hanke, Reitsch, 1998).

Al contrario, quando si dispone di dati storici, è possibile e spesso proficuo applicare metodi di previsione quantitativi, che spesso consentono di isolare la struttura sottostante del fenomeno che governa la serie dei dati, consentendo poi di utilizzare la comprensione e quantificazione dello stesso per effettuare delle previsioni sull'evoluzione futura della serie.

I metodi quantitativi si dividono in due macro-categorie: i metodi aleatori e i metodi basati su serie storiche (Levine, Krehbiel, Berenson, 2006).

La principale differenza tra i due approcci consiste nel fatto che, mentre nei metodi aleatori si cerca di determinare i fattori che influenzano la variabile di cui si vuole effettuare la previsione, nei metodi basati su serie storiche si cerca di utilizzare i valori passati e presenti assunti dalla variabile per effettuare previsioni sul suo possibile andamento futuro.

È opportuno a questo punto ricordare la definizione di serie storica: una serie storica è un insieme di dati numerici registrati ad intervalli regolari di tempo. (Bowerman, O'Connell, 1993).

Nel concreto e limitandoci all'ambito economico, possiamo trovare innumerevoli esempi di serie storiche: i prezzi di chiusura giornalieri di un'azione, il prodotto interno lordo di una nazione registrato con cadenza annuale, il fatturato trimestrale di una società, le vendite mensili di una catena di supermercati, il tasso di cambio giornaliero tra diverse valute.

2.1 Il Modello Moltiplicativo per l'Analisi delle Serie Storiche

L'analisi delle serie storiche si basa su un principio intuitivo secondo cui l'andamento di una serie nel passato, nel presente e nel futuro è governato sempre dai medesimi fattori, che esercitano quindi analoga influenza lungo tutto l'arco temporale di esistenza della serie storica. Pertanto se l'analista riesce ad individuare questi fattori, sarà poi in grado di effettuare delle previsioni sull'andamento futuro della serie.

Per poter effettuare questo tipo di ricerca è di grande aiuto l'analisi statistica che consente di disaggregare i fenomeni nelle loro componenti: i principali metodi statistici sviluppati dagli analisti a questo scopo sono il metodo additivo, il metodo moltiplicativo e il metodo misto, che è una combinazione dei due.

Nel modello additivo si ipotizza che le componenti temporali corrispondenti al trend generale, alle variazioni stagionali, a quelle cicliche e alle oscillazioni aleatorie, siano tra loro indipendenti, e cioè che non si influenzino a vicenda. Poiché nel caso dell'andamento dei prezzi di un titolo quest'ipotesi appare eccessivamente riduttiva, ci riferiremo nel seguito al modello moltiplicativo, che tiene conto delle possibili interazioni tra le diverse componenti della serie storica.

Consideriamo ad esempio la serie storica del prezzo del titolo Azimut nel periodo di tempo compreso fra il 15 maggio e il 17 giugno 2020, nel seguito identificato genericamente come "periodo considerato" (Figura 3).

Effettuando una semplice analisi qualitativa dei dati rappresentati graficamente, si può osservare che i valori presentano una tendenza all'aumento nei 24 giorni considerati: questo tipo di andamento osservabile in una serie storica, ovvero la tendenza di medio termine all'incremento o al decremento dei valori della serie, viene definito come "trend". Ovviamente il trend non è sufficiente a caratterizzare interamente una serie storica, fatto salvo il caso in cui la serie stessa non presenti un andamento del tutto lineare, rappresentabile quindi con una semplice linea retta.



Fig. 3 – Andamento del prezzo di mercato della Società Azimut nel periodo considerato.

In una serie storica infatti sono presenti altri fattori che rivestono grande importanza e contribuiscono a definire la serie. Il primo di questi fattori è la componente *ciclica*, che determina gli spostamenti verso l'alto o verso il basso rispetto al trend. Questi scostamenti determinati dalla componente ciclica possono avere diverse durate, ma solitamente coinvolgono periodi di tempo piuttosto lunghi, pari almeno ad alcune settimane o mesi, o addirittura anni se si considera ad esempio la serie storica dei dati relativi al PIL di una Nazione. Oltre che nella durata, questi scostamenti ciclici della serie dal trend differiscono anche nell'intensità e sono spesso strettamente legati a cicli economici più ampi, che esulano quindi da questo esame.

Il secondo fattore che contribuisce all'andamento della serie storica è la cosiddetta componente *irregolare*, altrimenti detta anche casuale, che riassume in sé tutti i fattori che appunto non presentano regolarità nella loro manifestazione e nel modo di influenzare l'andamento.

Esiste poi un terzo fattore, definito con il termine di *stagionalità*, del quale occorre tener conto quando si considerano serie storiche che abbracciano periodi più lunghi, ad esempio dati mensili o trimestrali di diversi anni che possono presentare caratteristiche ricorrenti a seconda del periodo dell'anno in cui viene registrato il dato.

Tutti questi fattori, nel modello moltiplicativo classico, contribuiscono ciascuno alla determinazione del valore *i*-esimo della serie storica, ciascun punto della quale può essere visto come il prodotto di queste quattro componenti.

Le equazioni (1) e (2) riassumono matematicamente questi concetti rispettivamente per serie storiche annuali ed infra-annuali:

$$Y_i = T_i \cdot C_i \cdot I_i \quad (1)$$

$$Y_i = T_i \cdot C_i \cdot I_i \cdot S_i \quad (2)$$

dove con riferimento all'i-esimo valore della serie Y_i , si ha:

T_i = valore della componente di trend

C_i = valore della componente ciclica

I_i = valore della componente irregolare

S_i = valore della componente stagionale

Tralasciando ovviamente i fattori ciclici e stagionali, che riguardano periodi più lunghi, questo metodo può essere applicato utilmente anche a serie storiche relative a periodi di tempo limitati, dove emergeranno sostanzialmente la componente di trend e quella irregolare.

2.2 Il Lisciamiento di una Serie Storica

Consideriamo ora la Tabella 1 e la Figura 4, nelle quali è rappresentata la serie storica del prezzo del titolo Enel nel periodo di osservazione compreso fra il 6 aprile e il 7 maggio 2020, nel seguito identificato genericamente come “periodo considerato”. In questo caso si nota subito come il titolo, nel periodo considerato, abbia manifestato un andamento molto irregolare, caratterizzato da forti oscillazioni giornaliere che non sembrano collocarsi lungo un'immaginaria direzione di trend lineare. Cercare quindi di individuare un andamento per la serie in esame potrebbe portare a risultati ingannevoli: conviene invece affidarsi a delle tecniche di lisciamiento, che consentono di cogliere la tendenza di medio periodo della serie senza forzature, ma anzi “smussando” e lisciando, appunto, le oscillazioni della serie storica.

Giorno	Prezzo	Giorno	Prezzo	Giorno	Prezzo	Giorno	Prezzo
06-apr-20	6,14	14-apr-20	6,45	22-apr-20	5,99	30-apr-20	6,24
07-apr-20	6,38	15-apr-20	6,21	23-apr-20	5,93	01-mag-20	6,24
08-apr-20	6,29	16-apr-20	6,25	24-apr-20	5,90	04-mag-20	6,01
09-apr-20	6,45	17-apr-20	6,21	27-apr-20	6,07	05-mag-20	6,09
10-apr-20	6,45	20-apr-20	6,19	28-apr-20	6,15	06-mag-20	6,07
13-apr-20	6,45	21-apr-20	6,01	29-apr-20	6,27	07-mag-20	6,19

Tab. 1 – Prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

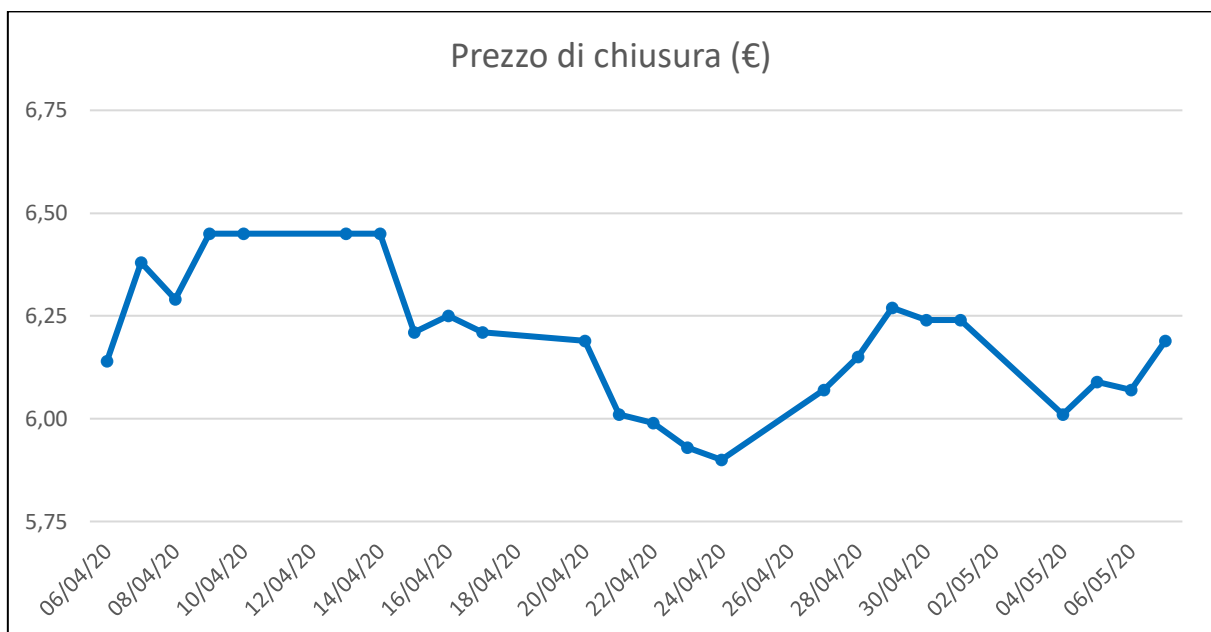


Fig. 4 – Rappresentazione grafica del prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

2.3 Il Metodo delle Medie Mobili

Definiamo anzitutto cosa si intende per media mobile. Una media mobile di periodo L consiste in una serie di medie aritmetiche calcolate su sequenze di valori osservati di lunghezza L (Brown, 1963). La notazione che identifica una media mobile di periodo pari a L è $MA(L)$.

Se ad esempio si vuole calcolare una media mobile con un periodo di 3 giorni su una serie di 13 giorni si deve procedere nel modo seguente. Essendo $L = 3$ giorni, le medie mobili consisteranno in una serie di medie aritmetiche che coinvolgono sequenze consecutive di 3 osservazioni. La prima di tali medie si ottiene quindi sommando i primi 3 valori della serie e dividendo per 3:

$$MA(3) = \frac{Y_1 + Y_2 + Y_3}{3}$$

La seconda media coinvolgerà invece i valori della serie dal secondo al quarto:

$$MA(3) = \frac{Y_2 + Y_3 + Y_4}{3}$$

Questo processo continua fino al calcolo dell'ultima media che sarà data da:

$$MA(3) = \frac{Y_{11} + Y_{12} + Y_{13}}{3}$$

Da questa semplice spiegazione matematica emergono chiaramente due caratteristiche del metodo di lisciamiento basato sulle medie mobili: la prima è che il periodo L ha un ruolo chiave nel metodo e la seconda è che la scelta di questo periodo è del tutto soggettiva.

L'analista, nella scelta del periodo L , deve tener conto di diversi fattori. Ad esempio, se all'interno della serie storica viene individuata una componente ciclica e/o stagionale, il periodo L deve essere scelto in modo da eliminare questa componente dall'analisi: in via preliminare sarà quindi necessario stimare la durata media dei cicli all'interno della serie e/o tener conto del periodo stagionale. Inoltre un periodo L troppo piccolo rispetto al numero dei dati disponibili effettuerà un lisciamiento troppo blando della serie dando luogo ad un'eccessiva ampiezza delle previsioni, viceversa un periodo L troppo grande porterà ad un'eccessivo smorzamento della serie che risulterà appiattita.

Per rappresentare graficamente le medie mobili, ciascun valore ottenuto come media deve essere inserito nel punto centrale della sequenza di tempi coinvolti nella media. Così nell'esempio precedente con periodo $L=3$ giorni, la prima media mobile sarà centrata nel secondo giorno, la seconda nel terzo e via di seguito fino all'ultima che si troverà in corrispondenza del dodicesimo giorno della serie. Risulta quindi evidente che con il metodo di lisciamiento basato sulle medie mobili, si perdono i dati della serie relativi al primo e all'ultimo

giorno. Più in generale si può quindi affermare che il metodo comporta la perdita dei dati relativi ai primi $(L-1)/2$ e agli ultimi $(L-1)/2$ periodi.

Contatore	Data	Prezzo di chiusura (€)	MA(3)	MA(7)
0	06-apr-20	6,14		
1	07-apr-20	6,38	6,27	
2	08-apr-20	6,29	6,37	
3	09-apr-20	6,45	6,40	6,37
4	10-apr-20	6,45	6,45	6,38
5	13-apr-20	6,45	6,45	6,36
6	14-apr-20	6,45	6,37	6,35
7	15-apr-20	6,21	6,30	6,32
8	16-apr-20	6,25	6,22	6,25
9	17-apr-20	6,21	6,22	6,19
10	20-apr-20	6,19	6,14	6,11
11	21-apr-20	6,01	6,06	6,07
12	22-apr-20	5,99	5,98	6,04
13	23-apr-20	5,93	5,94	6,03
14	24-apr-20	5,90	5,97	6,05
15	27-apr-20	6,07	6,04	6,08
16	28-apr-20	6,15	6,16	6,11
17	29-apr-20	6,27	6,22	6,13
18	30-apr-20	6,24	6,25	6,15
19	01-mag-20	6,24	6,16	6,15
20	04-mag-20	6,01	6,11	6,16
21	05-mag-20	6,09	6,06	
22	06-mag-20	6,07	6,12	
23	07-mag-20	6,19		

Tab. 2 - Medie mobili di ordine 3 e di ordine 7 calcolate sulla serie del prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

Per determinare in modo agevole le medie mobili di una serie di dati è utile procedere con l'ausilio di un programma di calcolo.

Nella Tabella 2 è rappresentato il prezzo di chiusura giornaliero del titolo Enel nel periodo di tempo compreso fra il 6 aprile e il 7 maggio 2020 insieme con le corrispondenti medie mobili di ampiezza 3 e 7. Le stesse sono state anche rappresentate graficamente nella Figura 5.

Come sottolineato prima, si può osservare che nella media mobile di ordine 3 mancano il primo e l'ultimo valore, mentre in quella di ordine 7 mancano i primi 3 valori e gli ultimi 3.

Il grafico evidenzia come il lisciamiento ottenuto con la media mobile di ampiezza 7 sia notevolmente maggiore rispetto a quello ottenuto con la media mobile di ordine 3, rendendo meno “rumorosa” la rappresentazione della tendenza della serie storica. D'altro canto comporta una maggior perdita di valori significativi (sei contro due). Si può quindi affermare che vi è una compensazione tra l'efficacia del lisciamiento e la rappresentatività della serie liscciata.

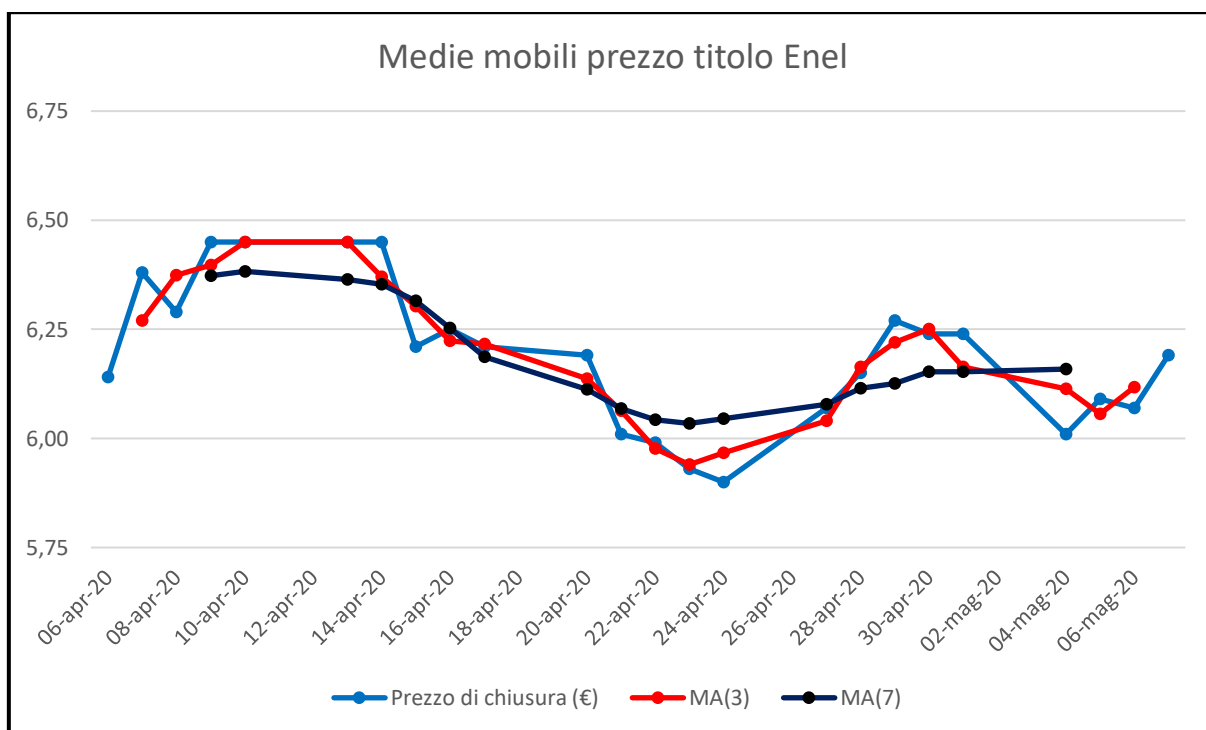


Fig. 5 – Rappresentazione grafica delle medie mobili di ordine 3 e di ordine 7 calcolate sulla serie del prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

2.4 Il Metodo del Lisciamento Esponenziale

Un altro dei metodi che possono essere utilizzati per smussare una serie storica al fine di ottenere un'indicazione sull'andamento di medio termine della serie è quello del lisciamento (o smorzamento) esponenziale.

Questo metodo, analogamente a quello delle medie mobili, consente di interpretare i dati di una serie storica anche nell'eventualità in cui questa non presenti un trend evidente, come nel caso relativo alla serie del prezzo del titolo Enel presentata nella Tabella 1 e nella Figura 2, rendendo possibile effettuare previsioni di breve termine. Inoltre il principale vantaggio di questo metodo rispetto a quello delle medie mobili è legato al fatto che non si perde alcun valore della serie.

Dal punto di vista prettamente matematico il metodo del lisciamento esponenziale può essere descritto come una media mobile ponderata con pesi che decrescono esponenzialmente. A differenza di quanto accade con il sistema di analisi basato sulle medie mobili semplici, nel caso del lisciamento esponenziale ciascun valore della serie smussata viene reso dipendente da tutti i valori precedenti, ognuno dei quali viene pesato in modo diverso, assegnando dei pesi decrescenti via via che si passa dai valori più recenti a quelli più remoti.

In sostanza, al periodo i viene assegnato un peso maggiore rispetto al valore relativo al periodo $i-1$, che a sua volta avrà un peso maggiore rispetto al valore osservato nel periodo $i-2$, con pesi via via decrescenti fino ad arrivare al primo valore osservato della serie, il quale presenterà il peso con valore minore.

Il calcolo del lisciamento esponenziale di una serie storica può essere effettuato sulla base di tre soli termini, ovvero il valore corrente della serie Y_i , il valore della serie smussata calcolato per il periodo precedente, E_{i-1} , e il peso introdotto nella spiegazione precedente e assegnato soggettivamente, che più correttamente si indica con il termine di "costante di lisciamento", W (Brown, 1963). Per ogni periodo i si ha quindi la seguente formula per la determinazione della serie smussata:

$$E_i = W \cdot Y_i + (1 - W) \cdot E_{i-1} \quad (3)$$

dove:

E_i = valore della serie lisciata esponenzialmente relativo al periodo i

E_{i-1} = valore della serie lisciata esponenzialmente relativo al periodo $i-1$

Y_i = valore osservato della serie storica nel periodo i

W = costante di lisciamento assegnata soggettivamente ($0 < W < 1$)

$$E_1 = Y_1$$

È evidente dall'equazione 3 che la costante di liscio gioca un ruolo fondamentale nel risultato del liscio che si otterrà per la serie e di conseguenza anche sulla previsione dei valori futuri. Pertanto la scelta di tale valore è cruciale nello sviluppo del metodo e, pur essendo soggettiva, può essere fatta secondo le seguenti considerazioni: un valore basso della costante di liscio W consente di ottenere uno smussamento molto efficace della serie, garantendo l'attenuazione delle componenti cicliche e irregolari ed evidenziando meglio le tendenze di medio periodo della serie storica; occorre tuttavia tener presente che per valori di W prossimi a zero il contributo delle nuove osservazioni che entrano nel modello è pressoché nullo e la nuova previsione tende a coincidere con la precedente; viceversa un valore alto, ovvero prossimo a 1, della costante di liscio W si rivela efficace nel caso in cui lo scopo principale del lavoro sia effettuare un'accurata previsione di breve termine.

Per meglio comprendere questa tecnica di analisi, procediamo ora al liscio esponenziale della serie storica del prezzo della Società Enel analizzata in precedenza con il metodo delle medie mobili. Anche per il calcolo del liscio esponenziale procederemo utilizzando le funzionalità di un programma di calcolo, utilizzando due fattori di smorzamento, $W_1=0,5$ e $W_2=0,25$.

Vediamo quindi quali sono i passaggi matematici da svolgersi per il calcolo della serie smussata con costante di liscio W_1 . Come punto di partenza consideriamo il primo valore osservato $Y_0 = 6,14$, che coincide con il primo valore della serie smussata E_0 . Quindi, utilizzando il valore osservato della serie nel periodo 1, è possibile ottenere anche il secondo valore della serie smussata, con l'applicazione della semplice formula:

$$E_1 = W_1 \cdot Y_1 + (1 - W_1) \cdot E_0 = (0,50) \cdot (6,38) + (1 - 0,50) \cdot (6,14) = 6,26$$

Per i giorni successivi si procede iterativamente:

$$E_2 = W_1 \cdot Y_2 + (1 - W_1) \cdot E_1 = (0,50) \cdot (6,29) + (1 - 0,50) \cdot (6,26) = 6,28$$

In questo modo è quindi possibile calcolare tutti i valori della serie smussata, che sono riportati nella seguente Tabella 3 per la serie relativa al prezzo di chiusura giornaliero del titolo Enel nel periodo di tempo considerato, calcolati rispettivamente con pesi pari a $W_1=0,5$ e $W_2=0,25$.

Contatore	Data	Prezzo di chiusura (€)	EXP W(0,50)	EXP W(0,25)
0	06-apr-20	6,14	6,14	6,14
1	07-apr-20	6,38	6,26	6,20
2	08-apr-20	6,29	6,28	6,22
3	09-apr-20	6,45	6,36	6,28
4	10-apr-20	6,45	6,41	6,32
5	13-apr-20	6,45	6,43	6,35
6	14-apr-20	6,45	6,44	6,38
7	15-apr-20	6,21	6,32	6,34
8	16-apr-20	6,25	6,29	6,31
9	17-apr-20	6,21	6,25	6,29
10	20-apr-20	6,19	6,22	6,26
11	21-apr-20	6,01	6,11	6,20
12	22-apr-20	5,99	6,05	6,15
13	23-apr-20	5,93	5,99	6,09
14	24-apr-20	5,90	5,95	6,04
15	27-apr-20	6,07	6,01	6,05
16	28-apr-20	6,15	6,08	6,08
17	29-apr-20	6,27	6,17	6,12
18	30-apr-20	6,24	6,21	6,15
19	01-mag-20	6,24	6,22	6,18
20	04-mag-20	6,01	6,12	6,13
21	05-mag-20	6,09	6,10	6,12
22	06-mag-20	6,07	6,09	6,11
23	07-mag-20	6,19	6,14	6,13

Tab. 3 – Lisciamento esponenziale della serie relativa al prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

Nella Figura 6 le due serie livellate così ottenute sono state rappresentate graficamente insieme alla serie originaria.

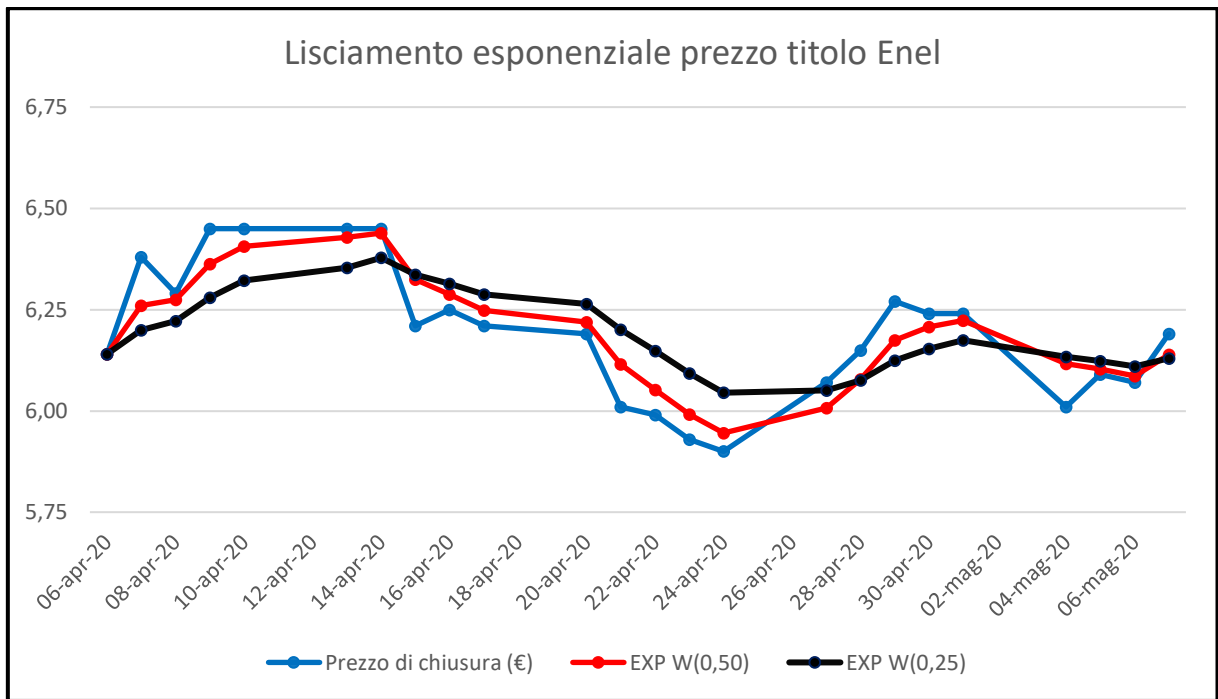


Fig. 6 – Grafico delle serie smussate con costanti di lisciamento pari a 0.5 e 0.25 calcolate sul prezzo di chiusura giornaliero del titolo Enel nel periodo considerato.

Ora per procedere ad una previsione di breve periodo della serie storica, è sufficiente considerare il valore smussato della serie livellata relativo al periodo i : tale valore infatti coincide con la previsione della serie relativa al periodo $i+1$, secondo la notazione seguente:

$$\hat{Y}_{i+1} = E_i \quad (4)$$

Così, ad esempio, per prevedere il prezzo del titolo Enel nel periodo 24, ovvero nel 25° giorno della serie, corrispondente all'8 maggio, possiamo utilizzare il valore smussato ottenuto per il periodo 23, ovvero il 7 maggio:

$$\hat{Y}_{24} = E_{23} = W_1 \cdot Y_{23} + (1-W_1) \cdot E_{22}$$

Previsione corrente = Valore corrente smussato =

$$(W) \cdot (\text{valore corrente osservato}) + (1-W) \cdot (\text{precedente valore smussato})$$

Nel ns. caso si ottengono i seguenti valori a seconda della costante di lisciamento considerata:

$$W_1: \hat{Y}_{24} = E_{23} = (0,5) \cdot (6,19) + (1-0,5) \cdot (6,09) = 6,14$$

$$W_2: \hat{Y}_{24} = E_{23} = (0,25) \cdot (6,19) + (1-0,25) \cdot (6,09) = 6,13$$

Nel caso di serie storiche che presentano un'eventuale componente stagionale, l'approccio seguito fino ad ora può essere generalizzato per tener conto anche di questa componente.

3. MODELLI LINEARI E NON LINEARI

Come accennato nel capitolo precedente, il primo passo da compiere nell'analisi di una serie storica è quello di rappresentarne graficamente i valori ponendo in ascissa gli intervalli di tempo nella scala corretta. Questo consente spesso di rendersi conto con una semplice analisi qualitativa dei dati se la serie presenta un trend di medio periodo, osservazione che si rivela particolarmente importante in quanto il trend è sicuramente la componente più rilevante da tenere in considerazione nell'analisi di una serie storica. Infatti l'individuazione di un andamento di medio periodo, sia esso lineare o non lineare, consente, come vedremo nel seguito, di effettuare previsioni sull'andamento della serie nel breve e medio periodo. Per individuare il tipo di trend manifestato da una serie utilizzeremo in questo capitolo il metodo dei minimi quadrati.

3.1 Modello Lineare

Consideriamo dapprima il caso più semplice, ovvero quello in cui la serie in esame presenta un trend di tipo lineare. Applicando il metodo dei minimi quadrati è possibile individuare una retta avente la seguente notazione analitica:

$$\hat{Y}_i = b_0 + b_1 \cdot X_i \quad (5)$$

dove Y rappresenta la variabile dipendente, ovvero i valori della serie, e X la variabile indipendente, ovvero gli intervalli di tempo in cui vengono misurati i valori.

Questa retta sarà tale per cui i valori dei due coefficienti b_0 e b_1 risulteranno determinati in modo da minimizzare la somma delle differenze al quadrato fra il valore osservato della serie e il valore della retta interpolante la serie stessa, ovvero in notazione analitica:

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{minimo}$$

L'equazione (5) della retta interpolante così individuata può essere successivamente utilizzata per effettuare una previsione dei valori della variabile dipendente Y in corrispondenza di valori della X diversi da quelli osservati, semplicemente sostituendo a X il valore in corrispondenza del quale si vuole prevedere la Y .

Occorre ricordare nuovamente che quando si applica il metodo dei minimi quadrati all'analisi di una serie storica per la determinazione del trend, la variabile indipendente è il tempo. Al fine di rappresentare correttamente la serie storica nel diagramma cartesiano e soprattutto al fine di individuare con le tecniche di calcolo disponibili i valori di b_0 e b_1 corretti, è necessario posizionare l'origine dell'asse dei tempi rappresentato dalle ascisse in corrispondenza del primo periodo per il quale sono disponibili i dati misurati e quindi di considerare, nel nostro caso, il primo giorno come il periodo zero ($X = 0$). Se ad esempio stiamo lavorando con una serie di 24 giorni, come quelle esemplificate per le Società Azimut e Enel, al primo giorno verrà assegnato il valore 0, al secondo il valore 1 e così via fino al ventiquattresimo giorno a cui sarà assegnato il valore 23.

Riprendiamo la serie storica rappresentata nella Figura 1 relativa al prezzo di mercato della Società Azimut nei 24 giorni del periodo considerato.

Una volta codificati i valori della variabile X da 0 a 23 è possibile ottenere facilmente l'espressione della retta interpolante, che rappresenta il trend di medio periodo della serie, utilizzando il software del programma di calcolo, che ne fornisce anche un'agevole rappresentazione grafica, visibile nella figura 7:

$$\hat{Y} = 0,0874 \cdot X_i + 14,1990$$

dove l'origine è rappresentata dal giorno zero, il 15 maggio, e le unità della variabile X sono di un giorno.

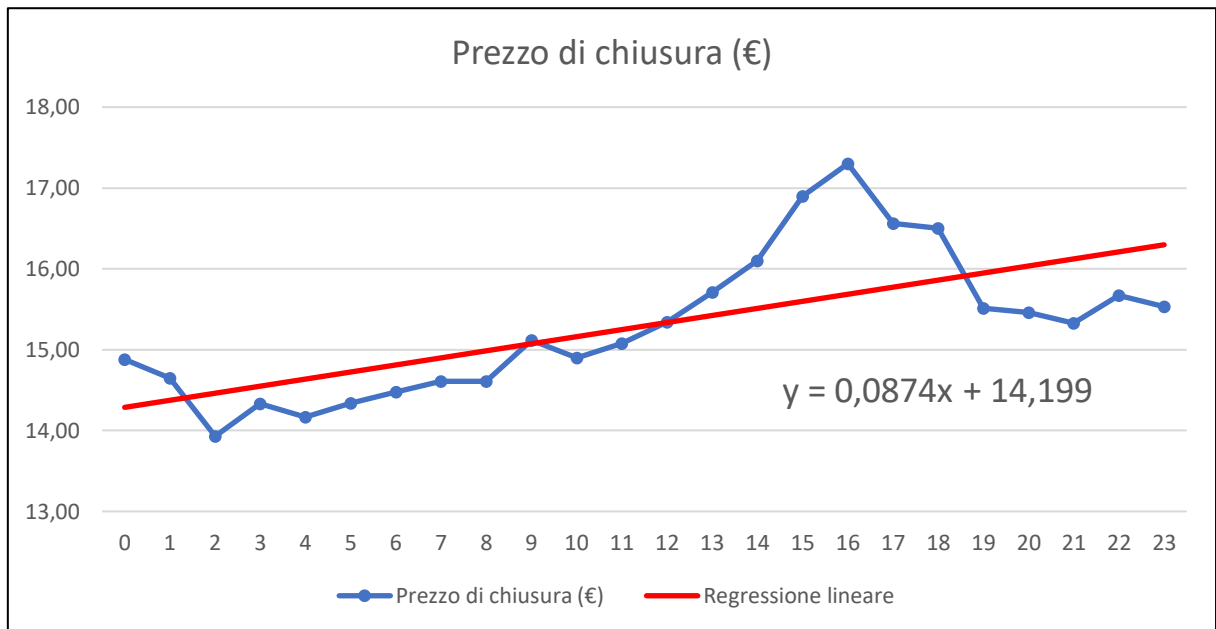


Fig. 7 – Interpolazione della serie del prezzo di mercato della Società Azimut nel periodo considerato.

Il significato matematico dei coefficienti della retta di interpolazione può essere illustrato come segue:

- L'intercetta $b_0 = 14,1990$ rappresenta il valore del trend nel primo giorno della serie, ovvero il prezzo stimato del titolo Azimut il giorno 15 maggio.
- L'inclinazione $b_1 = 0,0874$ rappresenta l'aumento giornaliero previsto sul prezzo del titolo.

Una volta che è stata individuata l'equazione della retta che rappresenta il trend, diventa immediato anche effettuare una previsione del prezzo per i giorni successivi. Ad esempio per calcolare il valore stimato per il giorno 25, ovvero il 18 giugno 2020, è sufficiente sostituire nell'equazione della retta al posto della X il valore corrispondente $X_{25}=24$. Di conseguenza la nostra previsione per il 18 giugno sarà:

$$\hat{Y}_{25} = (0,0874) \cdot (24) + 14,1990 = 16,30$$

Osservando meglio la Figura 5 si può notare che in alcune parti del grafico i valori della serie storica si allontanano in maniera significativa dalla retta interpolante. In particolare, nella parte

di destra del grafico, si nota come il trend stimato presenti un notevole incremento di medio periodo della serie, nonostante la stessa nella realtà sembri aver preso una direzione diversa con una serie di valori più o meno posizionati attorno al valore di 15,50 caratteristico degli ultimi giorni della serie. È lecito quindi sospettare che il trend lineare ipotizzato per la serie non sia adeguato e sia invece opportuno analizzarla utilizzando dei trend di tipo non lineare.

3.2 Modelli Non Lineari

Come si è visto una semplice analisi lineare della serie storica in esame può essere fuorviante per effettuare previsioni sui dati futuri. In particolare dalla Figura 5, si può vedere come a fronte di una tendenza al ribasso del prezzo negli ultimi giorni della serie, l'interpolazione lineare effettuata fornisca una stima per il 25° giorno di un prezzo vicino ai valori massimi registrati dalla serie nei giorni precedenti.

Analizzeremo ora due modelli di tipo non lineare per valutare se questi siano in grado di meglio interpretare i dati a disposizione: il primo cercherà di interpretare la serie con un trend quadratico, il secondo con un trend esponenziale.

3.3 Il Modello Quadratico

Il modello quadratico si basa su un polinomio di secondo grado ed è il più semplice fra i modelli non lineari. Il trend quadratico si ottiene nuovamente applicando il metodo dei minimi quadrati all'espressione seguente:

$$\hat{Y}_i = b_0 + b_1 \cdot X_i + b_2 \cdot X_i^2 \quad (6)$$

dove:

b_0 = intercetta stimata di Y

b_1 = effetto lineare stimato della variabile X sulla variabile Y

b_2 = effetto non lineare stimato della variabile X sulla variabile Y

Ancora una volta possiamo utilizzare il software Excel per i calcoli necessari alla determinazione del trend quadratico.

La formulazione della (6) che si ottiene per il caso in esame è:

$$\hat{Y} = -0,0049 \cdot X_i^2 + 0,2089 \cdot X_i + 13,6730$$

dove l'origine è rappresentata dal giorno 0, 15 maggio, e l'unità di misura della variabile X è il giorno. Come nel caso del trend lineare, l'equazione così ottenuta per il trend quadratico può essere utilizzata a scopi previsivi sostituendo il valore di X assegnato al giorno per il quale interessa una previsione della serie e calcolando il corrispondente valore di \hat{Y} . Per esempio, se vogliamo prevedere il prezzo del titolo Azimut per il giorno 18 giugno ($X_{25} = 24$), si ottiene:

$$\hat{Y}_{25} = -0,0049 \cdot (24)^2 + 0,2089 \cdot (24) + 13,6730 = 15,86$$

Nella Figura 8 è riportato l'output ottenuto in Excel della regressione quadratica relativa al prezzo di mercato della Società Azimut nel periodo considerato. Analizzando la figura, si può notare come il modello quadratico sia in grado di interpolare la serie meglio di quanto non faccia quello lineare, in particolare nella parte finale del grafico, più utile ai fini previsivi.

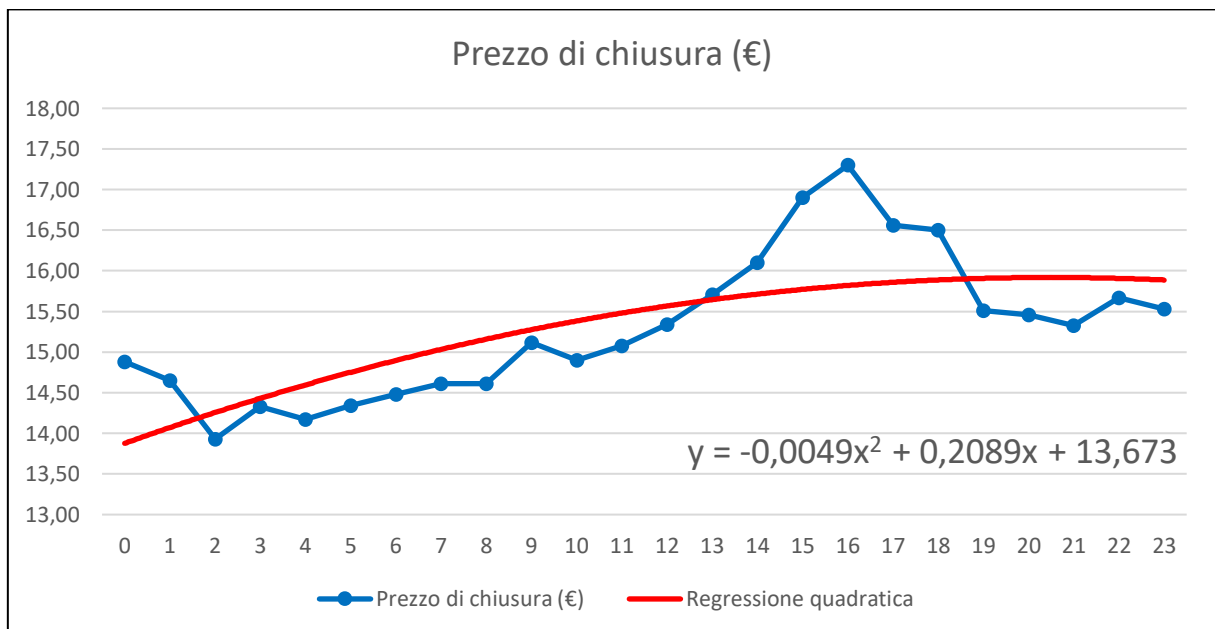


Fig. 8 – Interpolazione quadratica della serie del prezzo di mercato della Società Azimut nel periodo considerato.

Naturalmente incrementando i gradi del polinomio, ad esempio al terzo o al quarto grado, è possibile ottenere modelli interpretativi che seguono in modo ancora migliore l'evoluzione

della serie, ma non è detto che questo aspetto si traduca automaticamente in previsioni più affidabili.

3.4 Il Modello Esponenziale

Vediamo ora un altro tipo di modello non lineare, ovvero il modello esponenziale. Questo modello risulta particolarmente proficuo in quei casi in cui i valori di una serie aumentano con un tasso crescente, tale per cui la differenza percentuale fra le osservazioni sia costante nel tempo. Così a prima vista, questo non sembra essere il caso della serie storica del titolo Azimut che stiamo esaminando, ma verificheremo questo aspetto cercando di applicare comunque il modello ai dati.

L'equazione che rappresenta il modello esponenziale è la seguente:

$$\hat{Y}_i = b_0 \cdot b_1^{X_i} \quad (7)$$

dove:

b_0 = intercetta stimata di Y

$(b_1 - 1) \cdot 100\%$ = stima del tasso di crescita percentuale nell'unità di tempo

L'equazione (7) può essere ridotta alla notazione analitica rappresentata dall'equazione (8) per mezzo di una trasformazione logaritmica:

$$\log \hat{Y}_i = \log b_0 + X_i \cdot \log b_1 \quad (8)$$

Mediante questo passaggio abbiamo trasformato l'equazione esponenziale (7) nella sua forma lineare espressa dalla (8). Di conseguenza è possibile applicare il metodo dei minimi quadrati alla variabile $\log Y_i$ e quindi ottenere la stima dell'inclinazione ($\log b_1$) e dell'intercetta ($\log b_0$). I calcoli si possono effettuare ancora una volta con l'ausilio del software Excel ottenendo il seguente risultato:

$$\log \hat{Y} = 0,0025 \cdot X_i + 1,1526$$

dove il giorno iniziale è sempre il 15 maggio e l'unità di misura dell'asse delle ascisse è il giorno. I valori di b_0 e b_1 si ottengono calcolando l'antilogaritmo dei coefficienti stimati della regressione:

$$b_0 = \text{antilog}(1,1526) = (10)^{1,1526} = 14,2102$$

$$b_1 = \text{antilog}(0,0025) = (10)^{0,0025} = 1,0058$$

Quindi sostituendo i valori individuati nell'equazione (7) si ottiene il trend esponenziale stimato, dato dalla seguente espressione:

$$\hat{Y} = (14,2102) \cdot (1,0058)^{X_i}$$

dove come di consueto il giorno iniziale è il 15 maggio e l'unità dell'asse delle ascisse è il giorno.

L'intercetta $b_0 = 14,2102$ rappresenta il valore stimato del trend nel giorno iniziale (15 maggio); mentre il valore $(b_1 - 1) \cdot 100\% = 0,58\%$ rappresenta la stima del tasso di crescita giornaliero percentuale nella serie del prezzo di mercato della Società Azimut nel periodo considerato.

Come si è visto per i modelli precedentemente analizzati, anche nel caso del modello esponenziale per ottenere la previsione della serie in un istante futuro è sufficiente sostituire il valore di X assegnato al giorno in una delle equazioni (7) o (8) e calcolare il corrispondente valore della serie stimata. Per esempio, se vogliamo prevedere il prezzo del titolo per il 25° giorno ($X_{25} = 24$), corrispondente al 18 giugno, dobbiamo effettuare il seguente calcolo:

$$\hat{Y} = (14,2102) \cdot (1,0058)^{24} = 16,33$$

Il trend esponenziale stimato è stato rappresentato nella Figura 9 insieme con la serie originaria. Dall'analisi del grafico appare confermata l'ipotesi inizialmente fatta circa l'inadeguatezza del modello, visto che nella parte finale del grafico il trend esponenziale punta verso l'alto, contrariamente a quanto accade per la serie storica reale. Tra i tre modelli analizzati sembra quindi essere il meno adeguato sia a rappresentare il trend della serie sia ad effettuare previsioni sul futuro andamento.

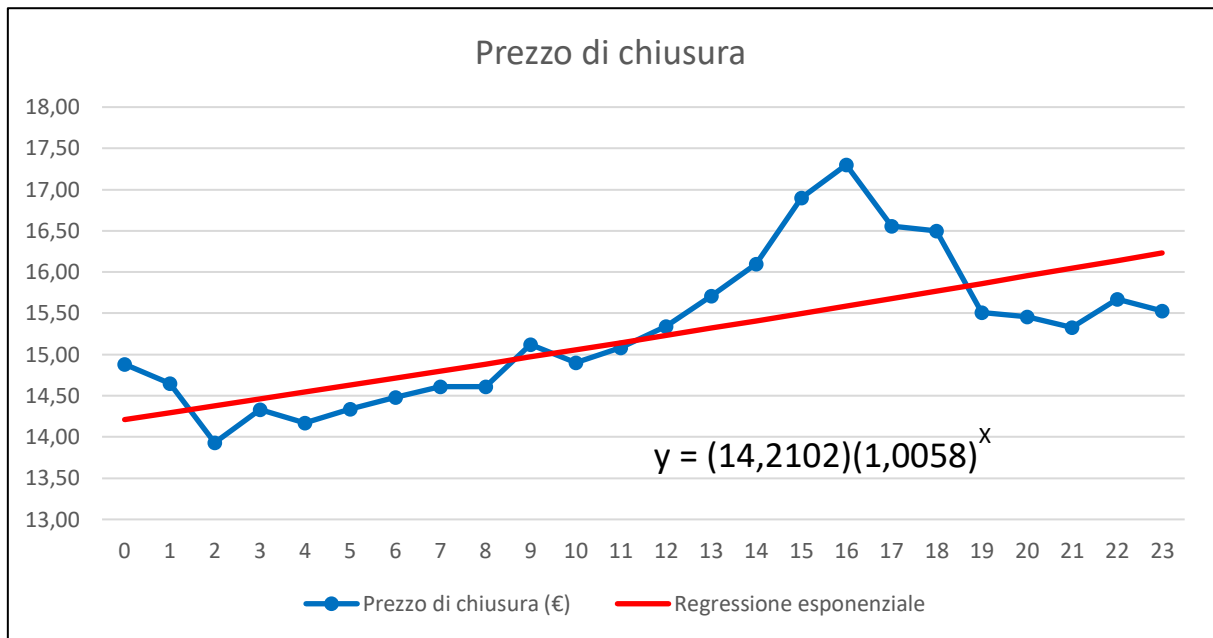


Fig. 9 – Interpolazione esponenziale della serie del prezzo di mercato della Società Azimut nel periodo considerato.

4. MODELLI AUTOREGRESSIVI

Affrontiamo ora un diverso approccio al problema di fornire un modello predittivo per i dati di una serie storica. Questo diverso approccio al problema nasce dal fatto che spesso nella realtà dei fenomeni dinamici, economici o di altro tipo, si può notare una più o meno forte correlazione tra i valori consecutivi di una serie storica: si evidenzia in altre parole un fenomeno di autocorrelazione tra i valori della serie.

Questi legami di dipendenza tra i valori di una serie storica possono essere sfruttati per studiare l'andamento futuro della serie, costruendo dei modelli diversi da quelli analizzati in precedenza: questi modelli si definiscono autoregressivi.

4.1 Modelli Autoregressivi Lineari del p-esimo ordine

L'autocorrelazione può essere del primo ordine quando è relativa a valori adiacenti, del secondo ordine quando è relativa a valori posti a distanza di due periodi l'uno dall'altro, o più in generale del p-esimo ordine se la distanza tra i valori considerati è pari a p periodi.

Nelle equazioni (9), (10) e (11) è rappresentata la notazione analitica rispettivamente per modelli autoregressivi del primo, secondo e p-esimo ordine (Levine, Krehbiel, Berenson, 2006):

$$Y_i = A_0 + A_1 Y_{i-1} + \delta_i \quad (9)$$

$$Y_i = A_0 + A_1 Y_{i-1} + A_2 Y_{i-2} + \delta_i \quad (10)$$

$$Y_i = A_0 + A_1 Y_{i-1} + A_2 Y_{i-2} + \dots + A_p Y_{i-p} + \delta_i \quad (11)$$

dove:

Y_i = valore osservato della serie al tempo i

Y_{i-1} = valore osservato della serie al tempo $i - 1$

Y_{i-2} = valore osservato della serie al tempo $i - 2$

Y_{i-p} = valore osservato della serie al tempo $i - p$

A_0 = costante da stimare con il metodo dei minimi quadrati

A_1, A_2, \dots, A_p = parametri autoregressivi da stimare con il metodo dei minimi quadrati

δ_i = componente di errore non autocorrelata, di media nulla e con varianza costante

A questo punto è bene sottolineare come l'equazione del modello autoregressivo del primo ordine (equazione 9) sia del tutto analoga a quella del modello di regressione lineare semplice (equazione 5), mentre l'equazione del modello autoregressivo del p-esimo ordine (equazione 11) può essere interpretata come equazione di un modello di regressione lineare multipla.

Nel seguito vedremo come si possono stimare i parametri indicati con A_0, A_1, \dots, A_p del modello: le relative stime così ottenute saranno indicate con le corrispondenti lettere minuscole a_0, a_1, \dots, a_p .

Il primo interrogativo che nasce dall'analisi delle equazioni individuate sopra per il modello autoregressivo lineare è naturalmente come individuare l'ordine appropriato del modello.

Come abbiamo visto questa scelta è legata al tipo di relazioni esistenti tra i valori della serie storica: si tratta in sostanza di stabilire quale deve essere l'ampiezza di queste relazioni fra i valori osservati per la serie.

Se si decide di considerare le sole relazioni fra valori consecutivi della serie storica si adotterà un modello del primo ordine, se invece si vuole tenere conto anche dei legami tra valori ritardati di due periodi occorrerà un modello del secondo ordine. Questo ragionamento può essere esteso fino a considerare tutti i valori che distano tra loro 1,2, ..., p periodi, adottando quindi un modello del p-esimo ordine.

Nell'individuare il corretto valore di p ci possono venire in aiuto due considerazioni:

- 1) se è vero che un modello di ordine superiore è generalmente più accurato di un modello di ordine inferiore, è anche vero che può essere molto più complicato da gestire a livello computazionale, specialmente nel caso di modelli autoregressivi non lineari, cui accenneremo nel seguito;
- 2) inoltre il valore di p non deve essere eccessivamente alto rispetto al numero dei valori della serie (n); infatti per stimare il coefficiente della p-esima variabile autoregressiva A_p , si tiene conto del numero di valori individuati da $(n - p)$ e quindi un valore troppo alto di p comporta la perdita di un numero eccessivo di valori della serie rispetto al possibile vantaggio che potrebbe offrire un modello di ordine superiore.

Dopo aver identificato l'ordine del modello e individuato i parametri A_0, A_1, \dots, A_p del modello applicando il metodo dei minimi quadrati tramite un opportuno programma di calcolo, è necessario verificare che l'ordine individuato per il modello sia adeguato al fenomeno da descrivere.

Per fare questa analisi è possibile agire per esclusione: ad esempio se dall'analisi qualitativa precedente emerge che il fenomeno potrebbe essere descritto adeguatamente da un modello del secondo ordine, conviene adottare inizialmente un modello del terzo ordine e poi procedere alla verifica della significatività del parametro di ordine massimo (3) del modello per capire se vada mantenuto o se ci si possa limitare ad un modello di secondo ordine.

In altre parole occorre risolvere un problema di verifica di ipotesi sulla significatività dei parametri che via via si vengono a trovare in corrispondenza dell'ultimo ordine del modello (Frees, 1996).

In alternativa a questa metodologia per identificare l'ordine p del modello possono essere utilizzate le funzioni di autocorrelazione o i criteri di informazione automatica AIC, Akaike Information Criterion (Akaike, 1973), o BIC, Bayesian Information Criterion (Schwarz, 1978), che non prenderemo in considerazione in questa trattazione.

In un modello autoregressivo di ordine p sarà quindi necessario impostare le seguenti ipotesi sul parametro A_p (parametro autoregressivo di ordine massimo):

$H_0: A_p = 0$ (il parametro di massimo ordine è uguale a zero)

$H_1: A_p \neq 0$ (il parametro di massimo ordine è diverso da zero)

Il test statistico per la verifica delle due ipotesi è dato dall'equazione (12):

$$t = \frac{a_p - A_p}{s_{a_p}} \quad (12)$$

dove:

a_p = stima del parametro autoregressivo di ordine massimo A_p

s_{a_p} = deviazione standard di a_p

Si può dimostrare che questo test segue una distribuzione t di Student con $(n - 2p - 1)$ gradi di libertà. Infatti per ottenere i gradi di libertà dobbiamo sottrarre a n il numero di parametri del modello (p parametri di inclinazione + 1 intercetta) e il numero p di confronti fra le osservazioni che vengono "persi" (Frees, 1996).

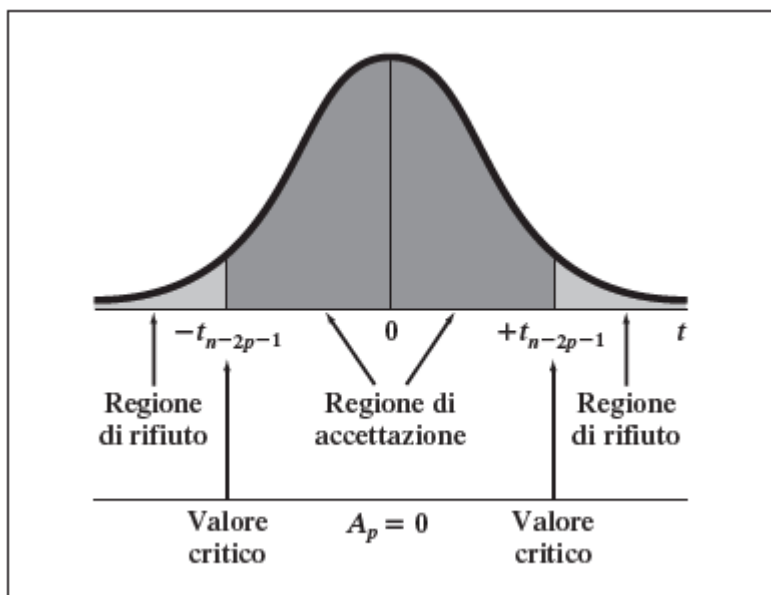


Fig. 10 - Regioni di rifiuto e di accettazione di un test di significatività a due code sul parametro autoregressivo di ordine massimo A_p (Levine, Krehbiel, Berenson, 2006).

Quindi, fissato il livello di significatività α , l'ipotesi nulla deve essere rifiutata se il valore osservato della statistica test è maggiore in modulo del valore critico della distribuzione t di Student corrispondente, arrivando quindi alla seguente regola decisionale:

rifiutare H_0 se $t > t_{n-2p-1}$ oppure $t < -t_{n-2p-1}$

accettare H_0 altrimenti

Nella Figura 10 sono riportate le regioni di rifiuto e di accettazione del test descritto. Se il valore ottenuto dalla statistica test porta ad accettare l'ipotesi nulla $A_p = 0$, significa che la componente autoregressiva di ordine massimo deve essere scartata, ovvero che il modello utilizzato contiene un numero troppo alto di parametri. Il test va poi ripetuto sul parametro A_{p-1} , che rappresenta ora il nuovo parametro autoregressivo di massimo ordine.

Viceversa, se il test sul parametro A_p di ordine massimo porta a rifiutare l'ipotesi nulla significa che l'ordine selezionato per il modello è adeguato. Naturalmente in questo caso è anche possibile chiedersi se sia il caso di aumentare ulteriormente l'ordine del modello e verificare la significatività del nuovo parametro di ordine massimo: tuttavia è conveniente svolgere questa valutazione dopo aver verificato l'efficacia del modello adottato nell'interpretazione dei valori della serie storica.

La procedura di controllo della significatività del parametro di ordine massimo va ripetuta fino a quando l'ipotesi nulla viene rifiutata, in questo caso infatti si ha la certezza della significatività dell'ultimo parametro autoregressivo e quindi il modello selezionato può essere utilizzato a scopi previsivi.

Una volta individuato il numero ottimale di componenti autoregressive con il metodo sopra descritto, è possibile procedere alla stima dei parametri, ottenendo quindi la seguente equazione per la rappresentazione del modello del p-esimo ordine stimato (Levine, Krehbiel, Berenson, 2006):

$$\hat{Y}_i = a_0 + a_1\hat{Y}_{i-1} + a_2\hat{Y}_{i-2} + \dots + a_p\hat{Y}_{i-p} \quad (13)$$

dove:

\hat{Y}_i = valore stimato della serie al tempo i

\hat{Y}_{i-1} = valore stimato (o osservato) della serie al tempo i - 1

\hat{Y}_{i-2} = valore stimato (o osservato) della serie al tempo $i - 2$

\hat{Y}_{i-p} = valore stimato (o osservato) della serie al tempo $i - p$

$a_0, a_1, a_2, \dots, a_p$ = parametri stimati

Supponendo ora di trovarci nell'istante n , il modello così individuato può essere utilizzato per effettuare delle previsioni k passi avanti inserendo l'istante $(n + k)$ nell'equazione 13, che assume quindi la seguente forma:

$$\hat{Y}_{n+k} = a_0 + a_1 \hat{Y}_{n+k-1} + a_2 \hat{Y}_{n+k-2} + \dots + a_p \hat{Y}_{n+k-p} \quad (14)$$

dove:

k = numero di periodi nel futuro

\hat{Y}_{n+k-p} = previsione effettuata all'istante n per l'istante Y_{n+k-p} se $(k - p) > 0$

\hat{Y}_{n+k-p} = valore osservato di Y_{n+k-p} se $(k - p) \leq 0$

$a_0, a_1, a_2, \dots, a_p$ = parametri stimati

È opportuno sottolineare che, quando il modello autoregressivo del p -esimo ordine viene utilizzato a scopi previsivi, il numero di osservazioni che entrano in gioco nella previsione è sempre pari a p , a prescindere dalla distanza k nel futuro del valore che vogliamo prevedere. Pertanto se adottiamo un modello del terzo ordine ($p = 3$), una previsione k passi in avanti rispetto all'istante n , si baserà unicamente sui valori osservati nei periodi $n+k-1$, $n+k-2$, $n+k-3$.

Per chiarire meglio questo concetto, supponiamo che $k=1$ e che vogliamo cioè stimare il valore della serie all'istante $n+1$; applicando l'equazione (14) otteniamo la seguente previsione all'istante $n+1$, basata sui valori osservati negli istanti n , $n-1$, $n-2$:

$$\hat{Y}_{n+1} = a_0 + a_1 Y_n + a_2 Y_{n-1} + a_3 Y_{n-2}$$

La previsione all'istante $n+1$ entra poi in gioco nella determinazione della previsione all'istante $n+2$:

$$\hat{Y}_{n+2} = a_0 + a_1 \hat{Y}_{n+1} + a_2 Y_n + a_3 Y_{n-1}$$

Procedendo iterativamente si ottengono le previsioni agli istanti $n+3$ e $n+4$ successivi:

$$\hat{Y}_{n+3} = a_0 + a_1 \hat{Y}_{n+2} + a_2 \hat{Y}_{n+1} + a_3 Y_n$$

$$\hat{Y}_{n+4} = a_0 + a_1 \hat{Y}_{n+3} + a_2 \hat{Y}_{n+2} + a_3 \hat{Y}_{n+1}$$

Questo procedimento può essere ripetuto a livello teorico senza limiti. È intuitivo però comprendere che via via che le previsioni si allontanano nel tempo dai valori osservati della serie storica, queste diventeranno via via meno affidabili, essendo basate sulle previsioni precedentemente effettuate e non sui valori reali osservati della serie storica.

Occorre adesso osservare che, se il vantaggio dei modelli autoregressivi è quello di sfruttare il fenomeno della correlazione presente tra i valori di una serie storica, il loro limite principale consiste nel fatto che sono applicabili solo in presenza di processi stocastici stazionari.

Un processo stocastico si definisce stazionario se le variabili casuali risultano in equilibrio dinamico rispetto alla variabile indipendente tempo t , in termini di valore atteso e/o di varianza. In questo caso si può dire che il processo stocastico presenta una distribuzione di equilibrio quando $t \rightarrow \infty$, ovvero i valori osservati presentano una certa “omogeneità temporale” di natura stocastica.

Più precisamente si parla di processo stocastico stazionario in senso stretto o forte qualora la distribuzione delle variabili casuali $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$ non sia funzione di (t_1, t_2, \dots, t_k) per ogni $k \geq 1$, ovvero spostare l'origine dei tempi di un ritardo j non ha alcun effetto sulla distribuzione osservata delle variabili casuali. Formalmente (Chatfield, 1996):

$$(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}) = (Y_{t_1+j}, Y_{t_2+j}, \dots, Y_{t_k+j})$$

$$\forall (t_1, t_2, \dots, t_k) \text{ e } \forall j$$

Per $k = 1$, ne consegue che $Y_t = Y_{t+j}$ e quindi tutte le variabili casuali del processo sono identicamente distribuite, e avranno quindi uguale media e varianza:

$$\mu(t) = E(Y_t) = \mu \quad \sigma^2(t) = Var(Y_t) = \sigma^2 \quad \forall t$$

con μ e σ^2 entrambe costanti che non dipendono dalla variabile tempo t .

Per $k = 2$, si ha invece $(Y_{t1}, Y_{t2}) = (Y_{t1+j}, Y_{t2+j})$ e quindi la distribuzione congiunta dipende solamente da $(t2 - t1)$ e non dalla traslazione j .

Ne consegue che se si considerano le componenti (Y_t, Y_{t+h}) la covarianza dipende solo dal ritardo h :

$$Cov(Y_t, Y_{t+h}) = E\{[Y_t - E(Y_t)] \cdot [Y_{t+h} - E(Y_{t+h})]\} = E[(Y_t - \mu) \cdot (Y_{t+h} - \mu)] = \gamma_h$$

La funzione γ_h appena introdotta viene denominata funzione di autocovarianza del processo e in un processo stocastico stazionario in senso stretto è funzione solo del ritardo $h = 0, \pm 1, \pm 2, \dots$

Per $h = 0$, la covarianza coincide con la varianza di Y_t :

$$\gamma_0 = Cov(Y_t, Y_t) = Var(Y_t) = \sigma^2(t) = \sigma^2$$

Si definisce analogamente anche una funzione di autocorrelazione, che misura la correlazione tra Y_t e Y_{t+h} , come segue:

$$\rho_h = \gamma_h / \gamma_0$$

con $\rho_0 = 1$.

Per quanto possa sembrare sorprendente che esistano processi per i quali la distribuzione delle variabili Y_t sia la stessa per qualsiasi valore di t , in realtà anche intuitivamente si può comprendere come esistano diversi processi stocastici che tendono ad una distribuzione di equilibrio per t che tende ad infinito: infatti dopo che un processo si è evoluto per un certo tempo, allontanandosi dalle condizioni iniziali, la distribuzione delle sue variabili stocastiche tende a stabilizzarsi. Proprio questo fenomeno costituisce la stazionarietà del processo.

Tuttavia la verifica dell'ipotesi di stazionarietà in senso stretto è nella maggior parte dei casi reali quasi impossibile, ci si limita dunque spesso a controllare che siano verificate delle condizioni meno forti e riguardanti solo i momenti fino al secondo ordine (media, varianza, covarianza). Un processo che rispetti tali proprietà è definito processo stocastico stazionario in senso lato o debole. Generalmente si considera solo quest'ultimo tipo di stazionarietà nelle

applicazioni, riconducendo la verifica alle proprietà di media, varianza e autocovarianza. In particolare diremo che un processo è stazionario in senso lato se verifica le seguenti condizioni:

$$E(Y_t) = \mu \quad \forall t$$

$$E(Y_t - \mu)^2 = \sigma^2 \quad \forall t$$

$$E[(Y_t - \mu) \cdot (Y_s - \mu)] = \gamma_{|t-s|} \quad \forall \text{ coppia } (t,s)$$

La prima condizione richiede che il valor medio del processo sia costante e pari a μ al variare di t ; la seconda impone che il processo abbia varianza σ^2 finita e costante al variare di t ; l'ultima condizione infine implica che per ogni t e s la funzione di autocovarianza fra le variabili Y_t e Y_s dipenda solo dalla distanza tra t e s , cioè $|t-s|$.

Tipicamente l'andamento dei prezzi di un titolo non è un processo stocastico stazionario in senso lato o debole, né tantomeno in senso stretto o forte e quindi l'analisi autoregressiva non può essere applicata. Esiste però un sistema per aggirare questo problema trasformando i valori della serie storica in modo da renderla stazionaria in senso debole. Un metodo per eseguire questa trasformazione è quello di considerare la serie generata dalle differenze prime del prezzo del titolo e applicare il modello autoregressivo a questa serie. Una volta ottenuta la stima della differenza prima è poi possibile risalire alla stima del prezzo.

Riprendiamo ora l'esempio della serie del prezzo di mercato della Società Enel nel periodo di tempo compreso fra il 6 aprile e il 7 maggio 2020. Nella Tabella 4 sono riportati, oltre al prezzo di chiusura, le differenze prime e i predittori per un modello autoregressivo del terzo ordine. Notiamo che nel passaggio dalla serie originaria alla serie delle differenze prime, si perde il primo valore rispetto alle 24 osservazioni disponibili, e che nelle variabili ritardate a 1, 2 e 3 periodi vengono persi ulteriormente 1, 2 e 3 valori rispetto alla serie delle differenze prime.

La Tabella 4 serve come punto di partenza per la determinazione di un modello autoregressivo del terzo ordine. I parametri stimati a_0 , a_1 , a_2 e a_3 , si possono ottenere con il consueto supporto di un programma di calcolo, in questo caso il software statistico open-source Gretl, ottenendo i seguenti risultati:

$$a_0 = 0,00365467 \quad Sa_0 = 0,0237361$$

$$a_1 = -0,0237634 \quad Sa_1 = 0,196091$$

$$a_2 = 0,324219 \quad Sa_2 = 0,202646$$

$$a_3 = -0,221051 \quad Sa_3 = 0,212677$$

Contatore	Data	Prezzo di chiusura (€)	Differenze prime	Y_{i-1}	Y_{i-2}	Y_{i-3}
0	06-apr-20	6,14	0	0	0	0
1	07-apr-20	6,38	0,24	0	0	0
2	08-apr-20	6,29	-0,09	0,24	0	0
3	09-apr-20	6,45	0,16	-0,09	0,24	0
4	10-apr-20	6,45	0,00	0,16	-0,09	0,24
5	13-apr-20	6,45	0,00	0,00	0,16	-0,09
6	14-apr-20	6,45	0,00	0,00	0,00	0,16
7	15-apr-20	6,21	-0,24	0,00	0,00	0,00
8	16-apr-20	6,25	0,04	-0,24	0,00	0,00
9	17-apr-20	6,21	-0,04	0,04	-0,24	0,00
10	20-apr-20	6,19	-0,02	-0,04	0,04	-0,24
11	21-apr-20	6,01	-0,18	-0,02	-0,04	0,04
12	22-apr-20	5,99	-0,02	-0,18	-0,02	-0,04
13	23-apr-20	5,93	-0,06	-0,02	-0,18	-0,02
14	24-apr-20	5,90	-0,03	-0,06	-0,02	-0,18
15	27-apr-20	6,07	0,17	-0,03	-0,06	-0,02
16	28-apr-20	6,15	0,08	0,17	-0,03	-0,06
17	29-apr-20	6,27	0,12	0,08	0,17	-0,03
18	30-apr-20	6,24	-0,03	0,12	0,08	0,17
19	01-mag-20	6,24	0,00	-0,03	0,12	0,08
20	04-mag-20	6,01	-0,23	0,00	-0,03	0,12
21	05-mag-20	6,09	0,08	-0,23	0,00	-0,03
22	06-mag-20	6,07	-0,02	0,08	-0,23	0,00
23	07-mag-20	6,19	0,12	-0,02	0,08	-0,23

Tab. 4 - Predittori di modelli autoregressivi del primo, secondo e terzo ordine sulla serie delle differenze prime ricavata dal prezzo di mercato della Società Enel nel periodo considerato.

Una volta ottenuti i parametri stimati è quindi possibile scrivere il modello predittivo del terzo ordine secondo la forma (13):

$$\hat{Y}_{n+1} = 0,00365467 - 0,0237634Y_n + 0,324219Y_{n-1} - 0,221051Y_{n-2}$$

e procedere al test di significatività sul parametro di ordine massimo a_3 , per il quale otteniamo:

$$t = \frac{a_3 - A_3}{Sa_3} = \frac{-0,221051 - 0}{0,212677} = -1,039$$

essendo poi i gradi di libertà della distribuzione t di student pari a:

$$n - 2p - 1 = 23 - 6 - 1 = 16$$

Poiché la scelta di un modello di ordine superiore garantisce una maggior potenza interpretativa della serie storica, nella scelta del livello di significatività adotteremo un valore $\alpha = 10\%$ in modo da estendere l'area di rifiuto dell'ipotesi nulla così da non correre il rischio di scartare un parametro che avrebbe potuto essere effettivamente significativo. Con $\alpha = 10\%$ si ottiene quindi:

$$t_{16} = 1,337$$

ed essendo quindi $t > -t_{16}$ si accetta l'ipotesi nulla per cui il parametro a_3 non è significativo. Occorre pertanto considerare un modello autoregressivo di ordine 2 e, ripetendo il procedimento visto prima con il software Gretl, si ottengono i nuovi parametri seguenti:

$$a_0 = 0,00731507 \quad Sa_0 = 0,0286329$$

$$a_1 = -0,0911412 \quad Sa_1 = 0,189814$$

$$a_2 = 0,336027 \quad Sa_2 = 0,207832$$

Procediamo quindi a scrivere il modello predittivo del secondo ordine secondo la forma (13):

$$\hat{Y}_{n+1} = 0,00731507 - 0,0911412Y_n + 0,336027Y_{n-1}$$

e ad effettuare il test di significatività sul parametro di ordine massimo a_2 , per il quale otteniamo:

$$t = \frac{a_2 - A_2}{Sa_2} = \frac{0,336027 - 0}{0,207832} = 1,617$$

essendo poi i gradi di libertà della distribuzione t di student pari a:

$$n - 2p - 1 = 23 - 4 - 1 = 18$$

e impostando un livello di significatività $\alpha = 10\%$ si ottiene:

$$t_{18} = 1,330$$

ed essendo quindi $t > t_{18}$ si rifiuta l'ipotesi nulla: pertanto il parametro a_2 è significativo. Possiamo quindi concludere che la serie delle differenze prime del titolo Enel è correttamente rappresentata da un modello lineare autoregressivo di ordine due avente la seguente notazione matematica:

$$\hat{Y}_{n+1} = 0,00731507 - 0,0911412Y_n + 0,336027Y_{n-1}$$

Nel grafico alla figura 11 è rappresentata la serie delle differenze prime a confronto con la serie stimata dal modello autoregressivo di secondo ordine ottenuto.

Dal grafico appare evidente che il modello rappresenta in modo efficace la serie delle differenze prime e può essere pertanto utilizzato a scopi previsivi. Ad esempio per il 25° giorno, corrispondente all'8 maggio, avente contatore 24, si ottiene il seguente valore per la differenza prima:

$$\hat{Y}_{n+1} = 0,00731507 - 0,0911412 \cdot 0,12 + 0,336027 \cdot (-0,02) = -0,01$$

E conseguentemente il prezzo stimato per Enel al 25° giorno sarà pari al prezzo osservato al 24° giorno più la differenza prima stimata, ovvero:

$$6,19 + (-0,01) = 6,18$$

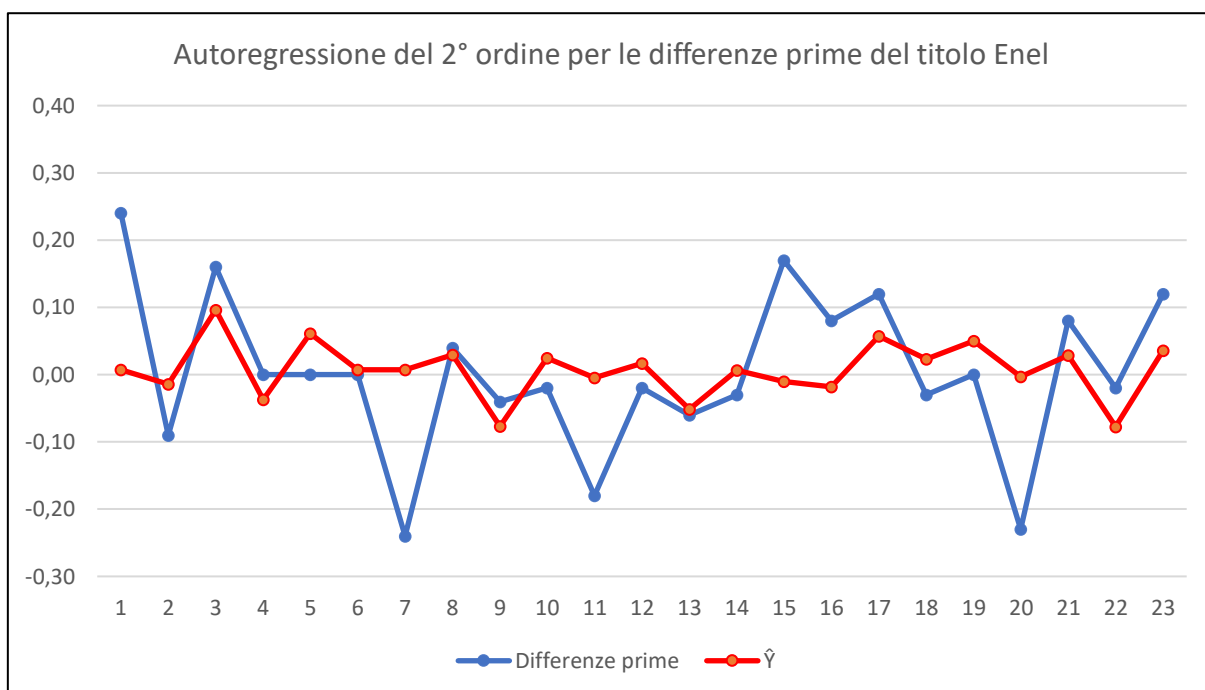


Fig. 11 – Rappresentazione grafica della serie delle differenze prime per il titolo Enel a confronto con le stime ottenute con il modello autoregressivo del secondo ordine.

Nei due paragrafi seguenti accenneremo brevemente a due diverse aree di sviluppo dei modelli autoregressivi, che cercano di superare i limiti insiti nel modello autoregressivo lineare mediante due diversi approcci. La prima area di sviluppo è orientata all'adozione di modelli non lineari, la seconda invece si concentra sulla possibilità di individuare aree di applicabilità dei modelli lineari anche a serie storiche con andamento non lineare.

4.2 Modelli Autoregressivi Non Lineari (NLAR)

Come abbiamo già riscontrato nell'analisi dei modelli basati sulla variabile indipendente X (tempo), anche nel caso dei modelli autoregressivi l'assunzione dell'ipotesi di linearità dei fenomeni appare troppo semplicistica rispetto alla realtà degli stessi.

Quando si osserva il comportamento erratico di una serie finanziaria, sembra naturale interpretare i dati come se fossero generati da un processo puramente casuale, ovvero di natura stocastica, guidato da interventi esogeni al processo stesso, siano essi di natura informativa o strutturale.

In realtà però questo approccio, per quanto possa sembrare naturale e scontato, non è l'unico adatto a spiegare la complessità degli andamenti dei prezzi finanziari. Infatti anche sistemi deterministici non lineari possono dare origine a osservazioni estremamente complesse e apparentemente indistinguibili da quelle conseguenti a un processo di tipo stocastico. Questi sistemi sembrano generare alcune delle caratteristiche qualitative dei fenomeni casuali e possono quindi rivelarsi particolarmente utili nell'interpretazione dei dati delle serie storiche finanziarie.

La rappresentazione di questi sistemi richiede l'adozione di modelli autoregressivi di tipo non lineare (Jones, 1978), che si basano su una forma generalizzata della notazione matematica utilizzata nell'analisi dei sistemi autoregressivi lineari:

$$Y_i = f(A_0, A_1, A_2, \dots, A_p, Y_{i-1}, Y_{i-2}, \dots, Y_{i-p}, \delta_i) \quad (15)$$

Y_i = valore osservato della serie al tempo i

Y_{i-1} = valore osservato della serie al tempo $i - 1$

Y_{i-2} = valore osservato della serie al tempo $i - 2$

Y_{i-p} = valore osservato della serie al tempo $i - p$

A_0 = costante da stimare con il metodo dei minimi quadrati

A_1, A_2, \dots, A_p = parametri autoregressivi da stimare con il metodo dei minimi quadrati

δ_i = componente di errore non autocorrelata, di media nulla e con varianza costante

È evidente che questa notazione generalizzata rispetto a quella relativa ai modelli lineari (11) lascia aperta una vastissima gamma di alternative e possibilità; del resto è facile intuire che la classe dei modelli non lineari è molto più estesa di quella dei modelli lineari.

Purtroppo questo significa anche che il rifiuto dell'assunzione di linearità, di per sé, è generalmente assai poco informativo sulla forma del modello f da adattare ai dati. È la medesima situazione che si verifica nell'usuale test di ipotesi statistiche, in cui il rigetto di un'ipotesi può essere estremamente significativo, ma non fornisce in generale informazioni sulle modifiche da apportare all'ipotesi nulla per renderla maggiormente adatta ai dati oggetto di esame. Ancora una volta quindi la validità del modello, e conseguentemente la scelta dello stesso, sarà determinata dalla capacità di interpretazione dei dati della serie storica da parte dello stesso.

È bene evidenziare come la (15) nella sua notazione implicita nasconda di fatto la complessità di calcolo connessa con l'utilizzo di modelli autoregressivi non lineari. A titolo di esempio proviamo a rappresentare un modello autoregressivo di tipo quadratico di ordine p , che rappresenta uno dei modelli non lineari più semplici che si possono adottare. La notazione matematica in forma esplicita diviene la seguente:

$$\hat{Y}_i = a_0 + a_1\hat{Y}_{i-1} + b_1(\hat{Y}_{i-1})^2 + a_2\hat{Y}_{i-2} + b_2(\hat{Y}_{i-2})^2 + \dots + a_p\hat{Y}_{i-p} + b_p(\hat{Y}_{i-p})^2 \quad (16)$$

\hat{Y}_i = valore stimato della serie al tempo i

\hat{Y}_{i-1} = valore stimato (o osservato) della serie al tempo $i - 1$

\hat{Y}_{i-2} = valore stimato (o osservato) della serie al tempo $i - 2$

\hat{Y}_{i-p} = valore stimato (o osservato) della serie al tempo $i - p$

$a_0, a_1, a_2, \dots, a_p$ = parametri stimati

$b_0, b_1, b_2, \dots, b_p$ = parametri stimati

Già in questo semplice modello è possibile osservare come sia necessario introdurre un numero di parametri stimati superiore rispetto a quello dei modelli autoregressivi lineari e come le interazioni tra i valori stimati e osservati della serie storica siano di natura più complessa. Queste caratteristiche, a fronte di una maggior capacità descrittiva dei modelli autoregressivi non lineari, generano difficoltà di calcolo che richiedono l'utilizzo di software per l'analisi statistica più completi rispetto al semplice foglio di calcolo utilizzato finora, la cui trattazione esula dallo scopo di questa breve presentazione.

4.3 Modelli Autoregressivi a Soglia (SETAR)

I modelli a soglia nascono da un'osservazione in sé molto semplice. Riprendendo uno dei concetti introdotti inizialmente, si è visto come il primo e fondamentale strumento di analisi di una serie storica consista nella rappresentazione grafica dei dati. Osservando un grafico di una qualsiasi serie storica di dati si può notare come esistano degli intervalli più o meno estesi della variabile indipendente riportata in ascissa all'interno dei quali l'andamento dei dati della serie può essere efficacemente rappresentato con un trend di tipo lineare.

I valori che delimitano questi intervalli vengono definiti "soglie" e la creazione di questi intervalli di linearità tra una soglia e l'altra è uno dei modi più semplici di ridurre la non linearità della funzione f rappresentata dalla (15), assumendone in sostanza la linearità a tratti. Questa

assunzione rappresenta anche un tentativo per superare i limiti dei modelli lineari, ma è facile intuire che la classe di modelli che ne deriva è assai vasta, visto che funzioni lineari a tratti sono in grado di approssimare arbitrariamente qualunque altra funzione continua.

Un'interessante evoluzione di questi primi modelli a soglia consiste nell'applicarvi il concetto di autocorrelazione già discusso in precedenza, parametrizzando le funzioni lineari componenti di f in funzione dei valori passati della serie stessa (Tong, 1983).

I modelli risultanti vengono definiti SETAR (Self Exciting Threshold Autoregressive) e possono essere rappresentati in notazione analitica nel modo seguente (Bisaglia, Gerolimetto, 2015):

$$\hat{Y}_i = \sum_{j=1}^k \left\{ \phi_0^{(j)} + \phi_1^{(j)} \hat{Y}_{i-1} + \dots + \phi_{p_j}^{(j)} \hat{Y}_{i-p_j} + \sigma^{(j)} a_i \right\} I(\hat{Y}_{i-d} \in A_j) \quad (17)$$

dove:

\hat{Y}_i = valore stimato della serie al tempo i

\hat{Y}_{i-1} = valore stimato (o osservato) della serie al tempo $i - 1$

\hat{Y}_{i-2} = valore stimato (o osservato) della serie al tempo $i - 2$

\hat{Y}_{i-p_j} = valore stimato (o osservato) della serie al tempo $i - p_j$

\hat{Y}_{i-d} = variabile di soglia, rappresentata dai valori della serie stessa

A_j = partizione di R tale per cui $\bigcup_{j=1}^k A_j = R$ e $A_i \cap A_j = \emptyset$ per ogni $i \neq j$

$\sigma^{(j)}, \phi_0^{(j)}, \phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}$ = parametri non noti da stimare

a_i = componente di errore non autocorrelata, di media nulla e con varianza costante

d, p_1, \dots, p_j = valori interi positivi che determinano l'ordine dei j modelli lineari e le soglie di variazione del regime di linearità

Un modello autoregressivo a soglia consiste quindi in un sistema di modelli autoregressivi lineari ciascuno con applicazione a un intervallo limitato del fenomeno, delimitato dai valori di soglia del processo. Chiaramente all'interno delle aree di linearità si possono applicare i concetti già visti al paragrafo 4.1 per i modelli autoregressivi lineari.

5. CONCLUSIONI

Questa breve analisi delle serie storiche e dei loro modelli interpretativi mi ha consentito di affrontare meglio alcuni argomenti introdotti durante il corso di statistica, consentendomi anche di sperimentare alcune applicazioni pratiche della teoria su alcune serie storiche di titoli azionari reali, mediante l'utilizzo di programmi di calcolo a larga diffusione.

La presentazione delle serie storiche, dei modelli autoregressivi e delle tecniche di analisi lineare e non lineare, credo sarà un primo passo utile per approfondire queste tematiche in maniera più approfondita, anche dal punto di vista computazionale, nel prosieguo dei miei studi in tema economico, finanziario e statistico.

6. BIBLIOGRAFIA

Akaike, H., *Information theory and an extension of the maximum likelihood principle* (Second International Symposium on Information Theory. Petrov B.N. and Csaki F., Budapest, 1973).

Bisaglia L., Gerolimetto M., *Testing for (non)linearity in economic time series: a Montecarlo comparison* (Journal of Methodological and Applied Statistics, volume 16-2014, 1st ed., Liguori Editore, 2015):

Bowerman B.L., O'Connell R.T., *Forecasting and Time Series* (3rd ed., North Scituate, MA: Duxbury Press, 1993).

Brown R.G., *Smoothing, Forecasting, and Prediction* (Englewood Cliffs, NJ: Prentice Hall, 1963).

Chatfield C., *The Analysis of Time Series: an Introduction* (5th ed. Chapman & Hall/CRC, 1996).

Frees E.W., *Data Analysis Using Regression Models: The Business Perspective* (Upper Saddle River, NJ: Prentice Hall, 1996).

Jones D., *Nonlinear autoregressive processes* (Proc. R. Soc. London, A360:71-95, 1978).

Hanke J.E., Reitsch A.G., *Business Forecasting* (6th ed. Upper Saddle River, NJ: Prentice Hall, 1998).

Levine D.M., Krehbiel T.C., Berenson M.L., *Business Statistics: A First Course* (4th ed. Pearson Education Inc., 2006)

Schwarz, G., *Estimating the dimension of a model* (The Annals of Statistics, 6, 461-464, 1978).

Tong H., *Threshold Models in Non-linear Time Series Analysis* (Springer, Heidelberg. Lecture Notes in Statistics, No. 21, 1983)

Zivot E., Wang J., Nonlinear Time Series Models. In: Modeling Financial Time Series with S-Plus® (Springer, New York, NY, 2006)