



UNIVERSITA' DEGLI STUDI DI PADOVA

**FACOLTA' DI SCIENZE STATISTICHE
CORSO DI LAUREA IN SCIENZE STATISTICHE ED
ECONOMICHE**

RELAZIONE FINALE

**ANALISI DELLE VENDITE
DI
UNA AZIENDA
DI PRODOTTI DOLCIARI**

Relatrice: Prof. Francesca Bassi

Laureando: Arduino Beltrame

ANNO ACCADEMICO 2005-2006

Indice

Prefazione	p.3
Capitolo 1 Asolo Dolce s.a.s	p.4
1.1 Presentazione dell'azienda	p.4
1.2 La missione	p.5
1.3 Ricerca e sviluppo	p.5
1.4 Qualità	p.5
1.5 I marchi	p.6
Capitolo 2 Serie storiche economiche e previsioni	p.12
2.1 Introduzione alle serie storiche economiche	p.12
2.2 Obiettivi dell'analisi delle serie storiche	p.13
2.3.1 Approccio classico: le componenti di una serie storica	p.14
2.3.2 Analisi Classica delle diverse componenti delle serie storiche	p.19
2.4 Approccio moderno	p.21
2.5 Previsione	p.35
Capitolo 3	p.57
3.1 Breve panoramica sull'azienda	p.57
3.2 Il mercato dell'azienda	p.62
3.3 Rete distributiva dell'azienda	p.62
Capitolo 4	p.63
4.1 Analisi dei dati delle quantità vendute dal 2003 al 2005	p.64
4.2 Approccio classico	p.66
4.3 Approccio moderno	p.83
Conclusioni	p.109
Bibliografia	p.111

PREFAZIONE

Questo lavoro presenta un'applicazione dell'analisi delle serie storiche per le previsioni delle vendite ad un caso aziendale, relativo all' *Asolo Dolce sas*, ditta che produce pasta fresca per pasticceria dolce e salata.

Il lavoro prevede uno studio delle vendite tenendo conto delle quantità totali prodotte.

Il primo capitolo è interamente dedicato a presentare l'azienda: le sue caratteristiche, l'organizzazione interna, i suoi prodotti.

Nel secondo capitolo descrivo brevemente le serie storiche e le previsioni. Per quanto riguarda le serie storiche mi soffermo su cosa sono, come sono definite, le loro componenti, gli strumenti di analisi. Per le previsioni, brevemente, accenno "all'impianto logico" su cui si fondano, il loro uso che dovrebbero avere in azienda e i vari modi per calcolare (le previsioni).

Nel terzo capitolo invece faccio una breve panoramica sulle vendite dell'azienda e sui mercati nei quali si trova ad operare.

Nel quarto ed ultimo capitolo ci sarà la fase applicativa del mio lavoro. Affronterò l'analisi delle vendite sia con l'approccio moderno che con l'approccio classico. Per questo tipo di analisi ho usato programmi informatici che si chiamano Excel, R, Eviews. A seconda del tipo di lavoro che dovevo eseguire, usavo quello che, nel mio libero arbitrio, poteva servire meglio al mio obiettivo.

Ringrazio, alla fine, *Asolo Dolce* che mi ha fornito i dati necessari per poter sviluppare un'analisi reale.

Capitolo 1

Asolo Dolce s.a.s

1.1. Presentazione dell'azienda

Asolo Dolce nasce nel 1972 in un laboratorio artigianale collocato ai piedi di Asolo (Treviso), nel cuore della pedemontana veneta. L'azienda esprime pienamente lo spirito di questa terra ricca di storia, cultura e natura, coniugando così in modo originale il meglio della tradizione con la spinta all'innovazione.

Alla sua nascita la produzione giornaliera ammontava a circa 90 kg di pasta sfoglia, oggi *Asolo Dolce* è espressione di alta qualità pasticceria ed elevata produttività. Nel 1999 *Asolo Dolce* ha rilevato “*Il Pasticcere Fiorentino*”, azienda di Rignano sull'Arno che, su 2500 mq, produce cantuccini e amaretti. L'azienda fiorentina è la più grande in Italia nella produzione di Cantuccini con una redditività oraria di 600 kg. Nel 2001 *Asolo Dolce* ha proseguito la sua politica di acquisizione sul mercato, integrando nel gruppo, il marchio “*Tonello*” che produce pasta sfoglia su una superficie di 4000 mq con un'ottimale produttività oraria di 580 kg.

La bontà dei prodotti unita alla professionalità dei collaboratori hanno portato l'azienda a diventare oggi leader di mercato nel settore della sfoglia con una quota (I.R.I¹) del 28,5%.

Asolo Dolce si avvale di circa 50 collaboratori ed è certificata **ISO 9001** con progettazione dei prodotti all'interno dell'azienda stessa.

1.2 *La missione*

¹ Azienda di ricerca di mercato, sondaggi di opinione e ricerca sociale. www.infores.com

Asolo Dolce crede in una “mission” aziendale che fa la differenza in particolare nel suo metodo e stile di lavoro. *Asolo Dolce* crede infatti nelle risorse umane e opera quindi per creare un ambiente che valorizzi lo sviluppo e la crescita di tutti i collaboratori.

I valori in cui si identifica l'azienda sono infatti quelli dell'entusiasmo, dell'onestà, della sincerità dell'apertura, della qualità, della professionalità, del rigore e della collaborazione. Un punto di vista mai ovvio né scontato che avvalora tutte le scelte aziendali.

1.3 Ricerca e sviluppo

Pronta a rispondere alle esigenze del mercato e ai nuovi stili di vita, *Asolo Dolce* è dotata di un efficiente reparto “Ricerca e Sviluppo”, a cui possono contribuire tutti i collaboratori. Nel 2001 le idee di miglioramento dei collaboratori sono state 280 con un ROI (*return of investment*) di 158 milioni di lire.

Lo staff di Ricerca e Sviluppo ogni anno propone e realizza almeno tre referenze nuove. Come la gamma dei prodotti anche la strategia dell'azienda viene continuamente adeguata a nuovi e stimolanti obiettivi anche se il principale rimane sempre la soddisfazione dei propri consumatori.

1.4 Qualità

Asolo Dolce è la garanzia di prodotti sempre fragranti e genuini perché ottenuti perseguendo l'obiettivo "qualità" nella scelta degli ingredienti, come le confetture al 45\% di frutta, le creme anidre senz'acqua e i grassi vegetali (Unilever) a basso contenuto di colesterolo dell'innovativa e affidabile tecnologia produttiva Rheon, del sistema-azienda certificato ISO 9001.

1.5 I marchi

Il processo di produzione di un determinato bene o servizio presenta economie di scala se i costi unitari di produzione dell'impresa diminuiscono all'aumentare della produzione di quel determinato bene o servizio. Affinché il costo medio (Cme ossia costo unitario del prodotto) diminuisca all'aumentare della produzione, il costo marginale (Cma ossia costo dell'ultima unità prodotta) deve essere inferiore al costo medio. L'economie di scala vengono definite, di solito, in termini di una funzione dei costi medi decrescenti.

Si parla, invece, di economie di scopo quando i costi unitari di produzione dell'impresa diminuiscono all'aumentare della varietà di beni o servizi prodotti. L'economie di scopo si definiscono, in genere, in termini di costo totale relativo per la produzione di una varietà di beni e servizi in una stessa impresa anziché separatamente in due o più imprese. Ossia, per una azienda è meno costoso produrre entrambi i beni X e Y che per una azienda produrre X e per un'altra produrre Y. L'economie di scala sono collegate all'economia di scopo e, a volte, i termini vengono usati in modo intercambiabile.

Negli ultimi anni *Asolo Dolce* ha compiuto importanti scelte aziendali con l'acquisizione di nuovi e significativi marchi del settore dolciario, realizzando così sia economie di scopo sia economie di scala.

Nel 1999 ha rilevato, come sopra detto, “*Il Pasticcere Fiorentino*”, azienda di Rignano sull'Arno (Firenze). L'indirizzo internet è www.pasticcerEFIorentino.it.

Nel 2001 ha integrato nel gruppo il marchio “*Tonello*”, di Casoni di Mussolente (Vicenza), che produce pasta sfoglia su 4000 mq. Il suo indirizzo internet è www.tonello.com.

Nel complesso le tre unità produttive realizzano 360.000 kg al giorno di prodotto di elevata qualità che viene preferito quotidianamente da 327.000 consumatori.

Asolo Dolce unisce la migliore tradizione pasticceria ai sistemi di automazione più avanzati. Dalle speciali impastatrici alle macchine per la laminazione della

pasta, l'intero processo produttivo si avvale di tecnologie dell'ultima generazione che regolate dal *know-how* aziendale assicurano al prodotto i massimi requisiti di bontà e igienicità.

Particolare attenzione viene dedicata al confezionamento del prodotto avvalendosi appositamente di macchinari esclusivi ed innovativi. Seguendo un percorso completamente automatizzato che esclude qualsiasi contatto manuale, il prodotto finito viene ospitato in vaschette di *OPS* e racchiuso con un film per alimenti che mantiene le proprietà organolettiche e assicura freschezza totale. Il passaggio sulla pesatrice automatica garantisce la fedeltà del peso dichiarato.

Nel grande magazzino di *Asolo Dolce* le operazioni sono completamente automatizzate. Il robot, automatizzato di ventosa, preleva le singole confezioni e le inserisce nei robusti cartoni che, attraverso binari di scorrimento, vanno a formare i *pallets*. Prima di essere spediti, i *pallets* vengono avvolti con il film plastico per maggiore stabilità.

I prodotti del gruppo sono:

Asolette glassate

Asoline

Asolo Ligth(sugar free)

Amaretti

Amaretti morbidi

Bon Boncin glassati

Cantuccini(gusto classico o gusto al cioccolato)

Cocco ciok

Fogliette dolci

Juppy

Le pasticcine al cioccolato

Linea snack

I Biscotti di prato

Strudel(alla confettura di mela,albicocca,frutti di bosco o ligth mela)

Tindy (con crema di limone o crema di nocciola)

Melody (alla confettura di mela)

Jumbo(con crema alla nocciola)

Sfizzì gusto pizza o rosmarino

Sfoglia Ciock

Sfogliatine glassate

Sfogliotti (alla confettura di mela, albicocca o frutti di bosco)

Strudelini

Tindy

Tre Basi

Ventaglietti

Sfizzì

Pasticceria Duse

La struttura organizzativa descrive il modo in cui un'impresa utilizza una determinata divisione del lavoro per assolvere compiti organizzativi, specificare il modo in cui il personale deve svolgere i propri compiti e facilitare il flusso di informazioni interne ed esterne. Alfred Chandler in *Strategy and structure* "sosteneva che gli alti dirigenti delle grandi imprese le strutturano in modo da consentire loro di perseguire la strategia economica prescelta ossia la struttura segue la strategia, si adatta ad essa. In pratica, i cambiamenti nella struttura organizzativa sono motivati da cambiamenti nella strategia, che a loro volta sono associati a cambiamenti nelle condizioni esterne che le imprese devono affrontare. La tesi di Chandler è che la struttura si adatta alla strategia.

Asolo Dolce è una tipica azienda di tipo padronale, dove la proprietà è concentrata e tendenzialmente stabile ed appartiene ad uno o a pochi individui e alle loro famiglie, attualmente si tratta della famiglia Gallina. La presenza dell'imprenditore, Sergio Gallina e della moglie Giusy Contarin all'interno dell'azienda è molto visibile.

Il modello di impresa padronale, è il più diffuso al mondo. Si applica in maniera privilegiata alle imprese di minore dimensione e si distingue per alcune specifiche **virtù**:

Vis imprenditoriale. Alla base del successo di queste imprese c'è una straordinaria miscela di intuito, volontà, intraprendenza e leadership, che si trovano di regola concentrate in uno o pochissimi individui, allo stesso tempo proprietari e gestori dell'impresa.

Dedizione. Nell'impresa padronale tra imprenditore e impresa si stabilisce un rapporto quasi filiale, che si impernia su basi non solo razionali (l'impresa come strumento di guadagno) ma anche emotive (l'impresa come mezzo per tramandare il proprio nome, come eredità da trasmettere ai figli, come luogo di realizzazione personale), e che moltiplica le energie dell'impresa e genera forte spirito di appartenenza nei collaboratori.

Agilità. La coincidenza tra proprietari e direzione consente decisioni rapide e riduce i pericoli di burocratizzazione dell'organizzazione. Per questo le imprese padronali denotano di regola una particolare abilità nel saper modificare la propria offerta e i propri processi, cogliendo con grande rapidità le opportunità derivanti dall'emergere di nuovi bisogni o da nuove tecnologie.

A fronte di questi vantaggi, l'impresa di tipo padronale presenta una serie di **debolezze**

Debolezza finanziaria. Dal punto di vista finanziario, l'impresa si trova costretta a far dipendere le proprie opportunità di crescita dalla consistenza del patrimonio del suo azionista di controllo. Di conseguenza essa può dovere rinunciare ad opportunità di investimento che potrebbero risultare profittevoli.

Debolezza manageriale. Sotto il profilo manageriale, la debolezza dell'impresa padronale si manifesta nella sua congenita difficoltà ad attrarre e trattenere manager di adeguata qualità professionale e ad adottare strutture organizzative e sistemi di gestione evoluti.

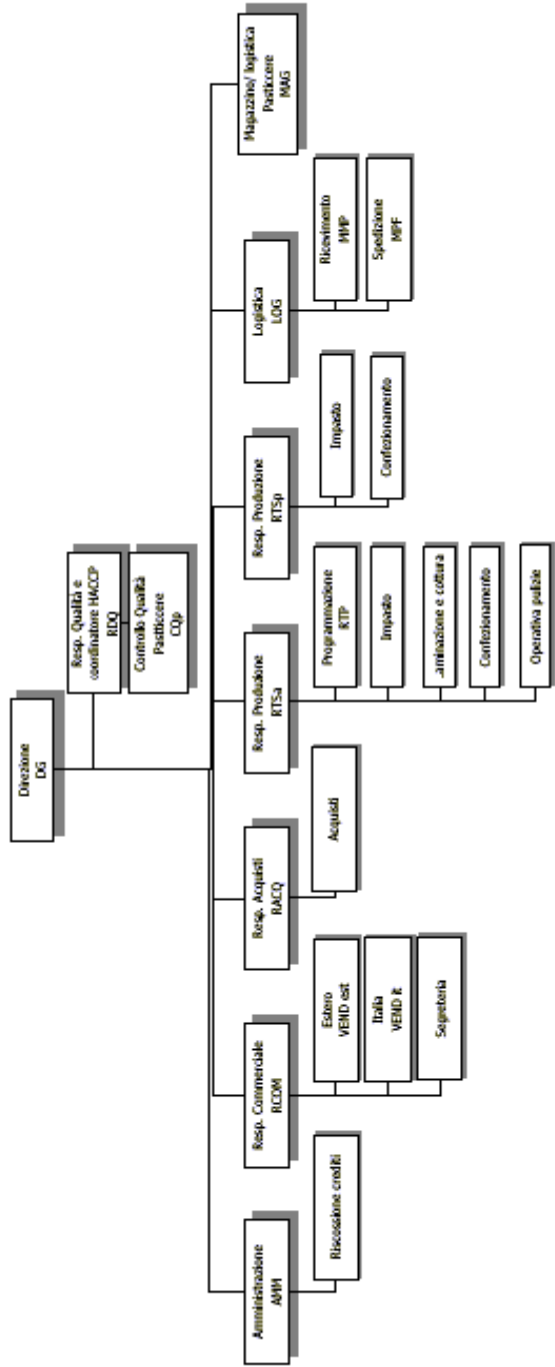
L'azienda ha il sistema di certificazione di qualità **ISO 9001**. Esso è stato definito e realizzato per il raggiungimento di standard qualitativi richiesti da eventuali normative mantenendo per quanto possibile le caratteristiche di flessibilità della produzione e tempestività di fornitura ed ulteriori elementi caratterizzanti il posizionamento dell'azienda nel mercato.

Questa certificazione consente la corrispondenza del sistema aziendale a un

modello nel quale si possa arrivare a garantire la soddisfazione del cliente attraverso la prevenzione di tutte le potenziali cause di inefficienza.

ORGANIGRAMMA AZIENDALE

ORGANIGRAMMA FUNZIONALE GENERALE
siti di Asolo e Rignano sull'Arno



DATA dc-16

Capitolo 2

Serie Storiche economiche e previsioni

2.1. Introduzione alle serie storiche economiche

Una serie storica $X_t, t=1,2,3,\dots,n$ è una successione ordinata di numeri che misura un certo fenomeno x_t esaminato con riguardo alla sua evoluzione rispetto alla variabile tempo che indicherò con lettera t .

Poiché si tratta di una funzione di t osservata ad istanti discreti, la rappresentazione grafica più ovvia è quella cartesiana, e consiste nel disegnare i punti $(t,x_t), t=1,2,\dots,n$ come illustrato nella figura 2.1 per una serie delle vendite di un prodotto.

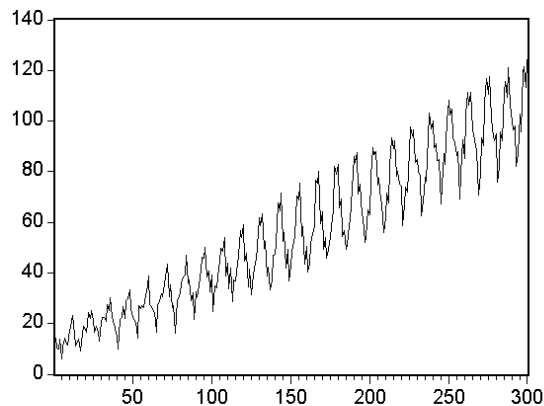


fig. 2.1 : grafico della serie di vendite mensili di un prodotto.

Risulta utile per la modellazione pensare x_t come una parte finita di una realizzazione de un processo stocastico X , dove per processo stocastico si intende una famiglia di variabili casuali descritte da un parametro $t \in T$.

Quindi una serie storica è una parte finita di una singola realizzazione del processo stocastico generatore dei dati; questo sarebbe noto se potessimo

conoscere le funzioni di densità congiunte e quelle di tutti i possibili raggruppamenti delle variabili casuali che lo compongono. Se ad ogni punto o in un determinato intervallo di tempo si verifica un solo fenomeno, la serie viene detta univariata; altrimenti con più fenomeni si ha una serie storica a più variabili detta multivariata.

Le serie storiche sono classificate in base alla natura dell'oggetto di studio:

Di carattere economico: come ad esempio il mercato dei titoli; se invece siamo nel campo aziendale si prendono in considerazione le vendite, i bilanci, le esportazioni,...ecc.

Di carattere demografico: così vengono analizzati i vari processi che caratterizzano l'aspetto vitale di un essere umano: nascita morte, etc.

Generate da un controllo processuale: cioè si considerano le variazioni che si sono avute in base alle prestazioni offerte da un processo produttivo. Se le osservazioni sono di gran lunga lontane dal livello obiettivo che si è prefisso (target), c'è la necessità di riportare il processo sotto controllo.

Generate da processi binari: se la serie può essere considerata come un processo binario ovvero che può assumere solo due valori: lo zero e l'uno.

Generate da un processo di punti: quando si hanno successioni di eventi ad intervalli casuali. Nel nostro caso la serie può essere definita univariata ed è del tipo stocastico perché i dati vengono studiati in funzione del tempo ovvero mensilmente ed è di carattere economico.

2.2 Obiettivi dell'analisi delle serie storiche

Distinguiamo 5 fasi del processo di analisi di una serie storica:

1. La descrizione dell'andamento del fenomeno, fase importante e preliminare per individuare il meccanismo generatore (modello) della serie, solitamente realizzato attraverso un modello grafico.
2. Il filtraggio (detto anche stima dei parametri) delle componenti non osservabili della serie stessa.

3. Il controllo(o verifica del modello) che tiene sotto sorveglianza un processo produttivo.Va identificata la natura del fenomeno rappresentato dalla sequenza di osservazioni a disposizione.
4. La previsione dei valori futuri.
5. Il controllo ottimale del sistema.

Quando si parla di “modello generatore di una serie storica” si intende un modello matematico che è in grado di generare i dati della serie che si sta analizzando. Tuttavia questa operazione è tutt’altro che semplice: i valori di una serie storica sono influenzati da una molteplicità di fattori che rende praticamente impossibile costruire un modello che li consideri tutti.

Ad esempio, il consumo nazionale di un certo prodotto è influenzato da dinamiche della popolazione, del reddito, dei prezzi, della produzione, dalle condizioni atmosferiche e da un complesso di altri fattori che non è possibile individuare.

L’analisi delle serie storiche si differenzia a seconda dell’obiettivo :

- Fare previsioni
- Descrivere la serie

Quando si fanno delle previsioni bisogna stabilire l’orizzonte temporale ossia se si intendono prevedere valori futuri della serie nel breve, medio, lungo periodo.

Per fare previsioni si cerca di individuare il modello generatore della serie storica (ossia una funzione matematica che permette di descrivere l’andamento della serie) e, grazie a tale modello, si fanno previsioni con il metodo della estrapolazione.

L’analisi descrittiva mira a scorporare tra loro le diverse componenti della serie.

Abbiamo principalmente due diversi approcci all’analisi delle serie storiche:

- L’approccio Classico
- L’approccio moderno

2.3.1 Approccio classico: le componenti di una serie storica

Nell'ambito dell'approccio classico all'analisi delle serie storiche, si suppone che

il fenomeno studiato risulti dall'aggregazione delle seguenti quattro componenti che verranno successivamente analizzate:

- Trend.
- Ciclo.
- Stagionalità.
- Componente casuale o errore

Trend (componente tendenziale): rappresenta l'andamento riferito nel lungo periodo della serie; esso cambia nel corso del tempo senza tuttavia presentare dei cicli prevedibili a priori. Per quanto riguarda nell'analisi del trend va sottolineato che non esistono tecniche sempre valide per evidenziarlo, tuttavia se è monotono crescente o decrescente l'analisi risulta facilitata; solitamente viene rappresentato con una funzione matematica di tipo semplice.

Molto spesso già un'osservazione visiva della serie permette di diagnosticare la presenza di un trend, ma nel caso in cui non si è certi di poter fare affidamento su semplici impressioni, può essere utile analizzare le funzioni di autocorrelazione.

Sovente è necessario rimuovere la componente di trend. Per far questo esistono varie metodologie, tra tutte la più usata risulta essere quella delle differenze successive. Tale approccio risulta molto valido se utilizziamo un modello di tipo **ARIMA** in seguito analizzato. In generale, possiamo affermare che risulta utile "ridurre" la presenza di fattori di "disturbo" che possono "nascondere" la componente di trend.

Componente ciclica: rappresenta le oscillazioni attorno al trend dovute alle espansioni e contrazioni congiunturali dell'attività economica. Nello studio dei fenomeni economici si osserva l'alternarsi di fasi di prosperità e di depressione che, nel loro insieme, costituiscono quello che si definisce ciclo economico.

Generalmente il ciclo è meno nitido delle altre componenti in quanto spesso si confonde col trend e più volte viene ignorato come elemento autonomo ed è incorporato nel trend stesso, ottenendo così una componente mista che va sotto il nome di ciclo-trend.

L'analisi delle tendenze è molto importante in quanto la tendenza è la "vita" stessa del prodotto, inserita in un determinato contesto di mercato e in relazione alle azioni di marketing dell'azienda e dei *competitors*. Della componente ciclica è importante capire l'evoluzione futura e prevedere, con la necessaria tempestività, i cambiamenti di tasso crescita/diminuzione della domanda.

Stagionalità: la componente stagionale è costituita da movimenti del fenomeno, nel corso dell'anno, che tendono a ripetersi in modo quasi analogo nello stesso periodo di tempo in anni successivi. Per dare una definizione più precisa, si può affermare che essa rende conto della fluttuazione del fenomeno, nel corso dell'anno, dovuta a fattori sociali e climatici e si riproduce in maniera più o meno permanente da un anno all'altro.

Questi eventi sono difficilmente controllati, ancor più, è difficile modificarne gli eventuali effetti negativi, perché si tratta di cause in genere esterne al sistema economico. Sono quindi dovuti a diversi fattori che non possono essere perfettamente prevedibili.

Ogni gruppo di cause può più o meno influenzare i vari settori produttivi; la difficoltà consiste nel quantificare il valore di questa influenza. Ci sono vari modi di trattare la componente stagionale, ma tutti principalmente mirano ad eliminare gli effetti della stagionalità per trovare le "leggi" che realmente governano una serie economica. La stagionalità infatti rappresenta un elemento di disturbo poiché può offuscare l'effettivo movimento ciclico della serie.

Errore: è la componente irregolare che raccoglie tutte le variazioni dovute alle cause più particolari ricorrenti, che non compaiono nelle componenti precedenti. La componente casuale fa sì che tutte le serie storiche che non siano di tipo deterministico presentano delle irregolarità di segno positivo o negativo, prodotte da un comportamento di tipo aleatorio, ma che in genere si annullano tra di loro ossia tendono a compensarsi. Per questo motivo l'errore è una variabile casuale con media nulla, varianza costante e autocorrelazione pari a zero.

Le quattro componenti, rispettivamente indicate con:

T = trend

C = ciclo

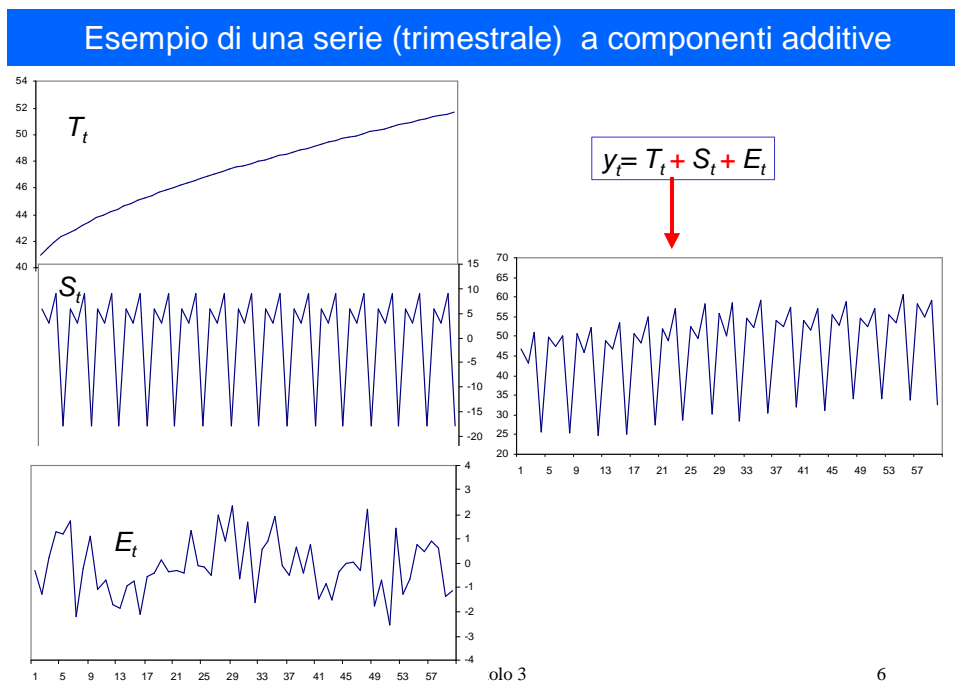
S = stagionalità

E = residuo

sono generalmente combinate in due tipi di modelli di aggregazione:

- MODELLO ADDITIVO:

$$X_t = T_t + S_t + C_t + E_t$$

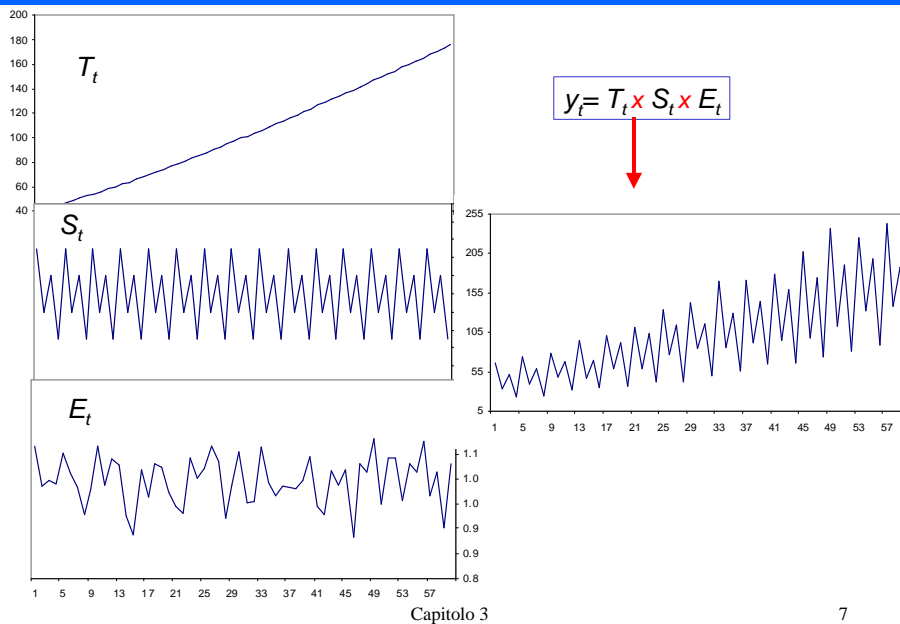


E' il modello più semplice e più utilizzato; la serie è la somma dei valori assoluti associati alle varie componenti; l'entità di ciascuna delle componenti non influenza le altre.

- MODELLO MOLTIPLICATIVO:

$$X_t = T_t S_t C_t E_t$$

Esempio di una serie (trimestrale) a componenti moltiplicative



7

E' un modello meno usato del precedente; il trend è espresso in termini assoluti, mentre le altre componenti sono espresse come fattore di proporzionalità, (esempio, vendite di un grande magazzino: la componente casuale è proporzionale al volume delle vendite; nei mesi estivi le vendite raddoppiano rispetto al periodo invernale;...).

Il modello moltiplicativo può essere ricondotto ad un modello additivo tramite una trasformazione logaritmica:

$$\log X_t = \log T_t + \log S_t + \log C_t + \log E_t$$

- MODELLO MISTO

In alternativa ai due modelli precedenti, si può considerare un modello intermedio: i **modelli misti**. Questi ultimi sono caratterizzati da relazioni tra componenti sia additive che moltiplicative.

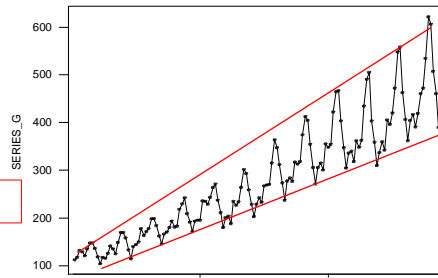
$$Y_t = (T_t + C_t) S_t \varepsilon_t$$

I modelli misti sono innumerevoli, ognuno di essi può soddisfare esigenze particolari compatibili con il fenomeno di studio.

Che serie è?

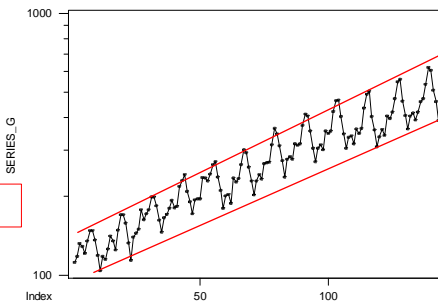
Dati originari (mensili)
Passeggeri aerei 1949-60

↓
E' una serie moltiplicativa



**Trasformazione
logaritmica**

↓
Diventa una serie additiva



2.3.2 Analisi CLASSICA delle diverse componenti delle serie storiche.

L'analisi viene in genere condotta con il metodo dei residui: si procede eliminando dalla serie le diverse componenti. Individuata una componente della serie, si procede depurando la serie da quella componente (per differenza, se il modello è additivo, dividendo se il modello è moltiplicativo).

Si procede poi eliminando via via le varie componenti. L'eliminazione di una o più componenti ha anche lo scopo di far risaltare una componente della serie.

Si parla, allora, di filtraggio della serie storica (s.s.) da detta componente, che poi viene individuata come differenza tra la s.s. originaria e la s.s. depurata.

Individuazione del trend.

Detrendizzazione - (si rende la s.s. stazionaria)

Per individuare il trend si possono utilizzare diversi metodi:

> metodo dei minimi quadrati

> metodo empirico (perequazioni grafiche)

- > metodo della media mobile
- > metodo delle medie parziali
- > metodo delle differenze prime (o seconde, ecc.)
- > metodo dei minimi quadrati.

Si tratta di individuare una particolare funzione matematica (la retta, la parabola, l'esponenziale, l'iperbole, la logistica q, ecc) e di individuare con il metodo dei minimi quadrati i parametri della funzione scelta.

Esempio: supponendo di ipotizzare un trend lineare, la funzione sarà:

$$T_t = a + bt$$

Dove : se $t=0 \rightarrow a=T_0$ per cui ci dice qual è il valore del trend al tempo zero.

Considerando: $T_{t+1} - T_t = b \rightarrow$ il parametro b indica la variazione (assoluta) del trend nell'unità di tempo. I parametri a e b vengono individuati con il metodo dei minimi quadrati. Una volta individuati i parametri vengono calcolati i valori teorici del trend.

- > metodo empirico (perequazioni grafiche)

Si basa sulle osservazione grafica della serie; è un metodo troppo soggettivo.

- > metodo della media mobile (m.m.)

Usando medie mobili appropriate vengono eliminate le componenti cicliche, stagionali e casuali: così facendo si evidenzia il trend. Esse hanno la caratteristica di ridurre l'intensità della variazione di un insieme di dati e vengono quindi usate per lisciare la serie.

Svantaggi: usando le m.m. si perdono dati all'inizio e alla fine della serie; inoltre le m.m. sono influenzate dai valori estremi.

- > metodo delle medie parziali

Consiste nel separare i dati in due parti (preferibilmente uguali) e nel fare la media dei dati di ciascuna parte. Si ottengono così due punti sul grafico della serie e si traccia poi la retta passante per quei due punti. I valori del trend sono quindi i valori ottenuti su questa retta. Tale metodo è applicabile solo se il trend è

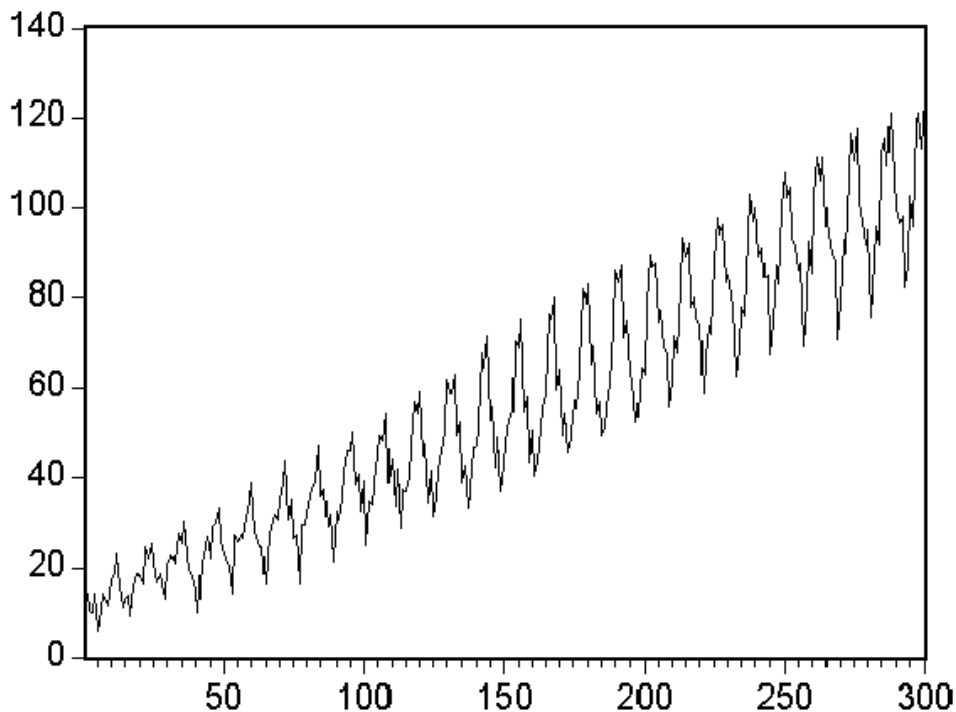
lineare e può essere esteso ai casi in cui i dati possono essere divisi in più parti in ciascuna delle quali il trend è lineare

2.4 Approccio moderno

Processi stazionari

Solamente per introdurre il concetto di stazionarietà, parlando di serie storiche, esse si dicono stazionarie se non ci sono cambiamenti sistematici nella media (non c'è trend) e, in varianza, se sono state rimosse le variazioni periodiche. Molte delle teorie probabilistiche delle serie storiche si basano sull'ipotesi di stazionarietà e affinché possano essere applicate, l'analisi delle serie storiche richiede che le serie non stazionarie siano trasformate in serie stazionarie.

Questi due grafici sottostanti (figure 2.2 e 2.3) mostrano la differenza, anche solo graficamente, di una serie stazionaria da una serie non stazionaria



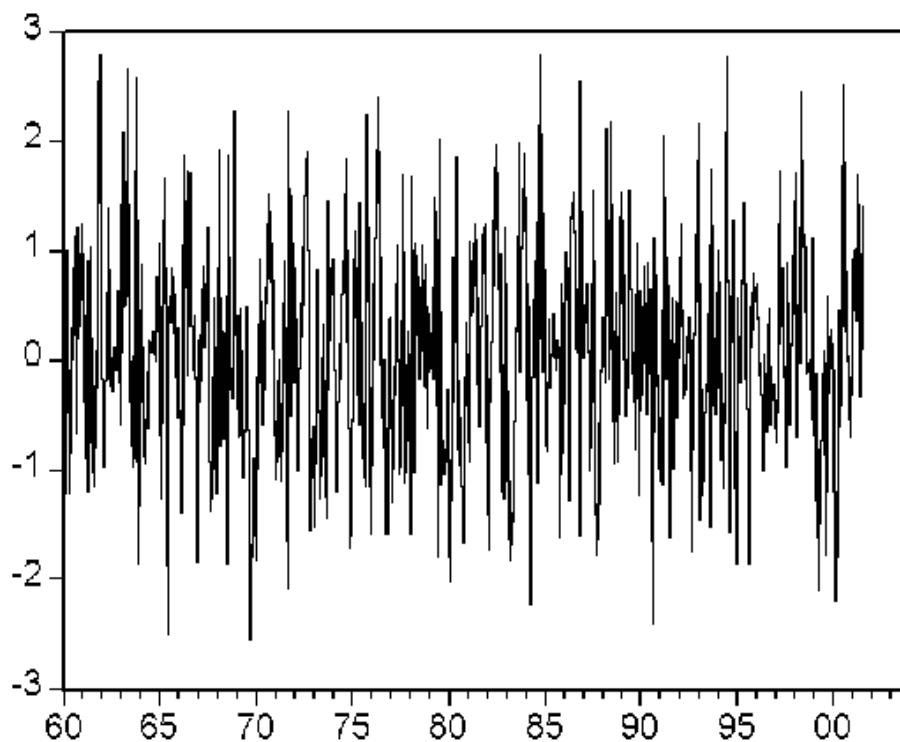


Illustrazione Ifig.2.3: serie stazionaria

Processi stocastici lineari

I principali processi stocastici stazionari sono:

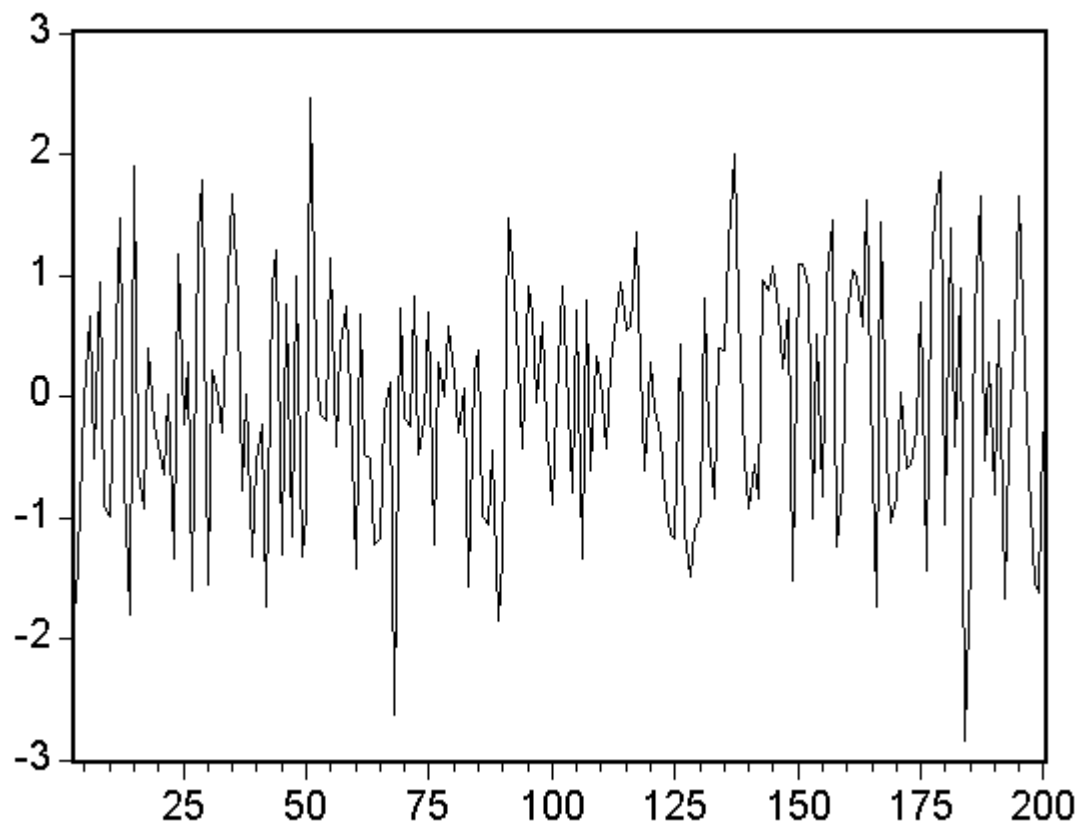
1) **WHITE NOISE (WN)**

Questo processo rappresenta la base di molti altri modelli. Le sue caratteristiche principali sono quelle di avere media uguale a zero, varianza costante, e assenza di autocorrelazione:

$$E(\varepsilon_t)=0 \quad \text{Var}(\varepsilon_t)= \quad \text{Cov}(\varepsilon_t, \varepsilon_{t-\tau})=0 \quad \text{per } \tau \neq 0$$

Se ε_t è distribuito “normalmente”, il processo è noto anche come un processo *white noise* di tipo Gaussiano. Chiaramente non è possibile effettuare previsioni relative ad un fenomeno generato da un processo di questo tipo.

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$



Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
█	█	1	0.001	0.001	0.0002	0.988
█	█	2	-0.062	-0.062	0.7875	0.675
█	█	3	0.011	0.011	0.8129	0.846
█	█	4	0.044	0.040	1.2086	0.877
█	█	5	-0.087	-0.086	2.7451	0.739
█	█	6	-0.009	-0.003	2.7600	0.838
█	█	7	0.042	0.032	3.1332	0.872
█	█	8	0.013	0.012	3.1675	0.923
█	█	9	-0.142	-0.132	7.3696	0.599
█	█	10	0.018	0.014	7.4378	0.684
█	█	11	-0.070	-0.092	8.4729	0.670
█	█	12	-0.090	-0.084	10.200	0.598
█	█	13	-0.060	-0.059	10.972	0.613
█	█	14	0.049	0.015	11.486	0.647
█	█	15	-0.059	-0.065	12.246	0.660
█	█	16	0.066	0.076	13.207	0.658
█	█	17	0.054	0.040	13.847	0.678
█	█	18	0.013	-0.005	13.885	0.737
█	█	19	0.070	0.101	14.957	0.725
█	█	20	0.013	-0.016	14.993	0.777
█	█	21	-0.051	-0.060	15.573	0.793
█	█	22	-0.006	-0.016	15.582	0.836
█	█	23	0.091	0.080	17.464	0.786

2) PROCESSI A MEDIA MOBILE MA(q)

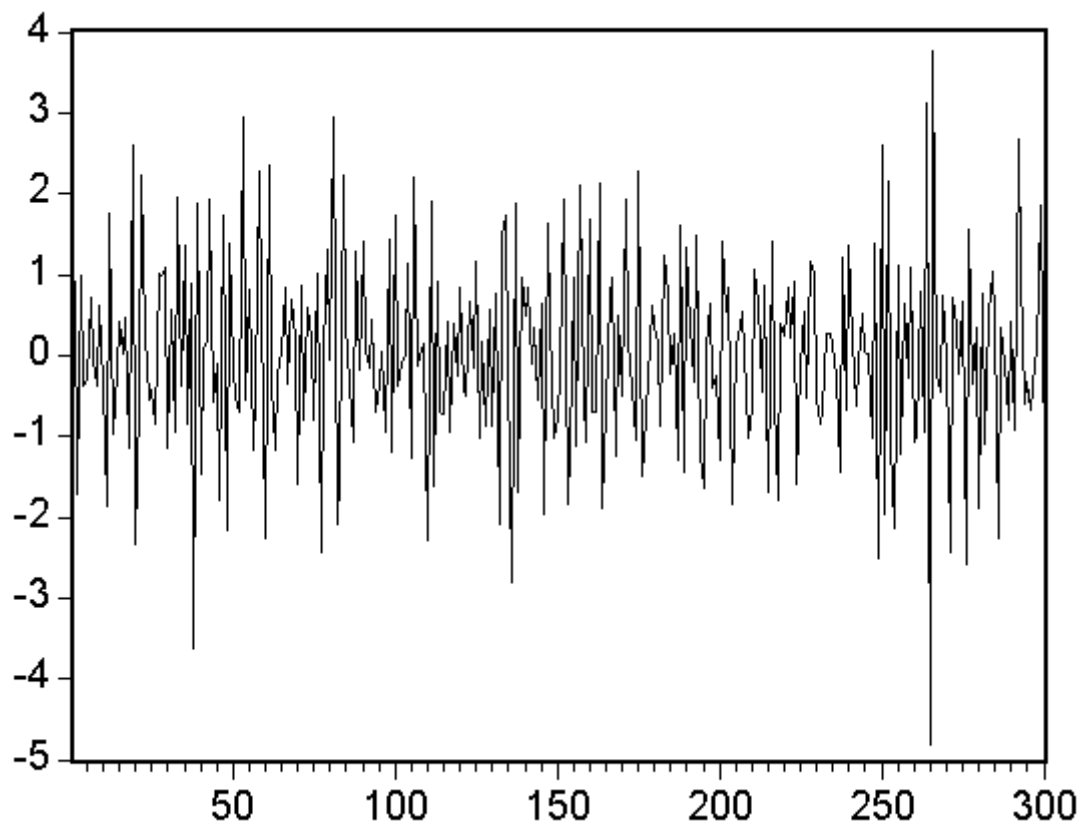
Questo tipo di modello da una descrizione di una serie storica assumendo che il valore osservato della serie nel generico istante t sia una combinazione lineare dei valori di una componente casuale rilevati nello stesso istante di tempo t e in q istanti immediatamente precedenti:

$$X_t = \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q}$$

Il corrispondente processo stocastico generatore è detto "processo di medie mobili di ordine q " e viene indicato con **MA(q)**:

$$X_t - \mu = \Theta(B)z_t$$

è sempre stazionario e la sua caratteristica è che l'autocorrelazione parziale (PACF) decresce a zero secondo una misura di esponenziali e oscillazioni (pseudo) periodiche la cui ampiezza tende a zero. Grafico della serie:



Disegno 1fig2.4 grafico di una serie a media mobile

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.497	-0.497	74.712	0.000
		2	0.096	-0.200	77.519	0.000
		3	-0.098	-0.193	80.451	0.000
		4	0.013	-0.165	80.506	0.000
		5	0.011	-0.102	80.545	0.000
		6	0.060	0.017	81.638	0.000
		7	-0.093	-0.072	84.304	0.000
		8	0.052	-0.036	85.149	0.000
		9	-0.017	-0.010	85.244	0.000
		10	0.046	0.044	85.899	0.000
		11	-0.051	-0.009	86.707	0.000
		12	0.049	0.038	87.463	0.000
		13	-0.123	-0.102	92.204	0.000
		14	0.098	-0.041	95.272	0.000
		15	-0.002	0.022	95.273	0.000
		16	-0.008	-0.003	95.293	0.000
		17	-0.098	-0.145	98.365	0.000
		18	0.151	0.042	105.67	0.000
		19	-0.085	0.024	107.99	0.000
		20	0.058	0.011	109.06	0.000
		21	-0.096	-0.073	112.07	0.000
		22	0.070	0.005	113.67	0.000
		23	0.006	0.059	113.68	0.000

Disegno 2 correlogramma di una serie ma(1)

3) PROCESSI AUTOREGRESSIVI AR(p)

Questo modello assume che il generico valore osservato al tempo (t, x_t) , sia funzione lineare di un certo numero p di termini immediatamente precedenti e di una componente casuale ϵ_t :

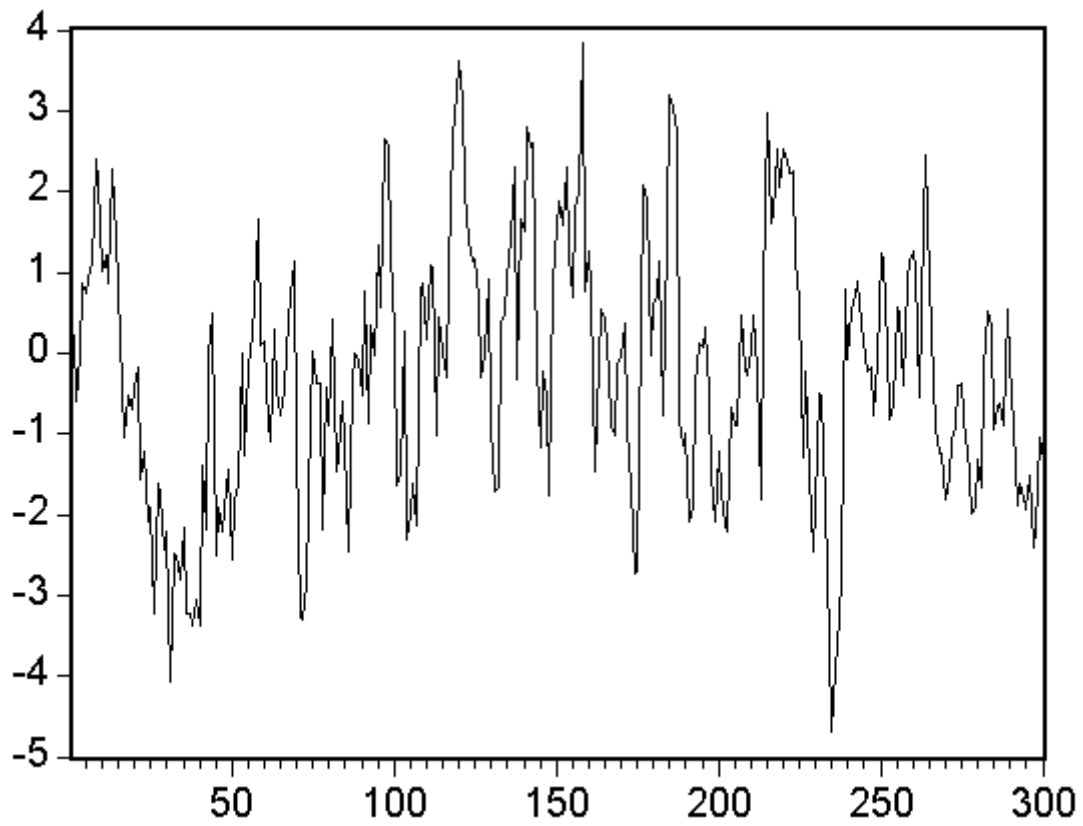
$$X_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + \epsilon_t$$

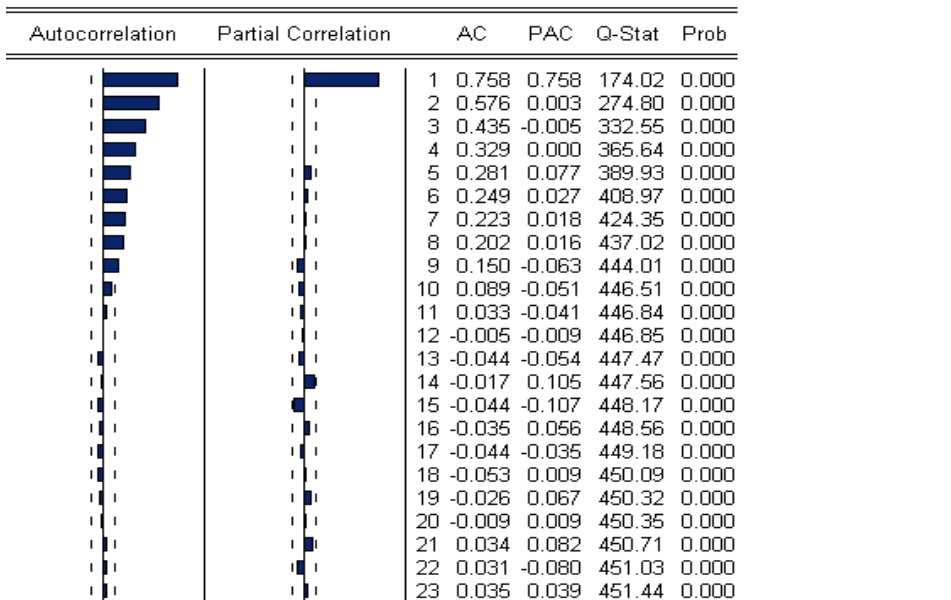
Il suo corrispondente processo stocastico generatore è detto "autoregressivo di ordine p " (**AR(p)**):

$$\Phi(B)(x_t - \mu) = z_t$$

è sempre invertibile e la sua caratteristica è che l'autocorrelazione (ACF)

decrese a zero secondo una mistura di esponenziali e oscillazioni (pseudo) periodiche la cui ampiezza tende a zero. Grafico della serie:





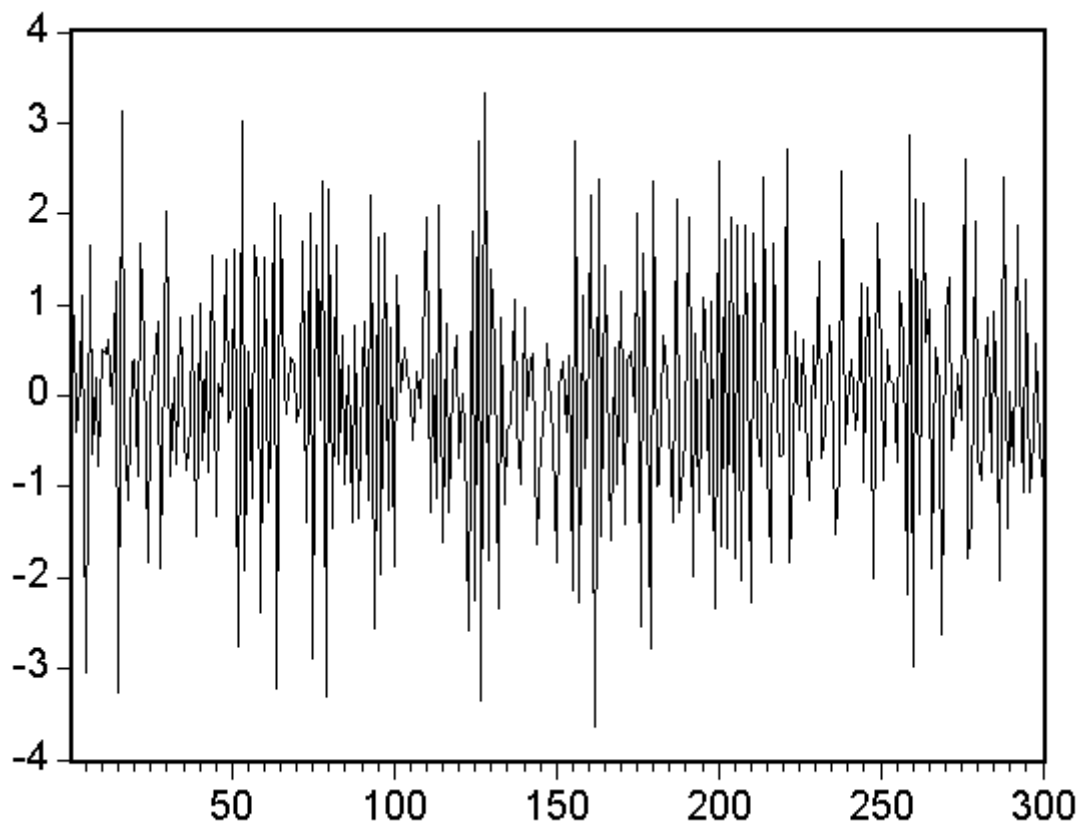
Disegno 3 correlogrammo serie ar(1)

PROCESSI AUTOREGRESSIVI A MEDIA MOBILE ARMA (p,q).

La combinazione dei due metodi, descritti poc'anzi, consiste nell'assumere che x_t dipenda sia dalle x in p tempi precedenti, sia dalle ε allo stesso istante e in q tempi precedenti. Si descrive quindi un modello "autoregressivo a media mobile(p,q)":

$$\Phi(B)(X_t - \mu) = \Theta(B)Z_t$$

E' sia stazionario che invertibile e la sua caratteristica è che l'autocorrelazione parziale (PACF) si comporta come l'autocorrelazione parziale di un processo a media mobile, cioè decresce esponenzialmente a zero. L'interesse di unire un modello $AR(p)$ e un modello $MA(q)$ in un modello **ARMA(p,q)** è nata dall'esigenza di descrivere una serie storica con pochi parametri, cosa non semplice con un solo processo autoregressivo oppure a medie mobili.



Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.663	-0.663	133.15	0.000
		2 0.270	-0.302	155.31	0.000
		3 -0.122	-0.175	159.85	0.000
		4 0.083	-0.047	161.94	0.000
		5 -0.032	0.028	162.26	0.000
		6 -0.022	-0.028	162.41	0.000
		7 0.057	0.032	163.41	0.000
		8 -0.109	-0.098	167.09	0.000
		9 0.098	-0.073	170.06	0.000
		10 -0.002	0.079	170.06	0.000
		11 -0.058	0.012	171.11	0.000
		12 0.062	0.033	172.32	0.000
		13 0.008	0.125	172.34	0.000
		14 -0.057	0.015	173.36	0.000
		15 0.080	0.073	175.40	0.000
		16 -0.132	-0.106	180.97	0.000
		17 0.139	-0.036	187.13	0.000
		18 -0.108	-0.014	190.91	0.000
		19 0.134	0.116	196.73	0.000
		20 -0.178	-0.029	207.00	0.000
		21 0.213	0.138	221.79	0.000
		22 -0.192	-0.016	233.79	0.000
		23 0.134	-0.005	239.63	0.000

Processi non stazionari

La classe dei modelli *ARMA* viene ampliata introducendo i modelli *ARIMA* e le

opportune varianti che meglio rappresentano le dinamiche delle situazioni reali: non stazionarietà del livello, componenti stagionali, trend polinomiali,.....,etc.

Con *ARIMA* (autoregressivi di media mobile integrati) si definisce una serie storica non stazionaria che dopo una serie di differenziazioni d , è stata ricondotta ad una serie stazionaria che può essere identificata come *MA*, *AR* o *ARMA*.

Dato quindi un intero non negativo d , il processo iniziale x_t è un processo **ARIMA** (p,d,q) se il processo

$$\Phi(B)(1-B)^d\}x_t = \Theta(B)z_t + \Theta_0$$

è un processo **ARMA** (p,q) stazionario

$$\Phi(B)(x_t - \mu) = \Theta(B)z_t$$

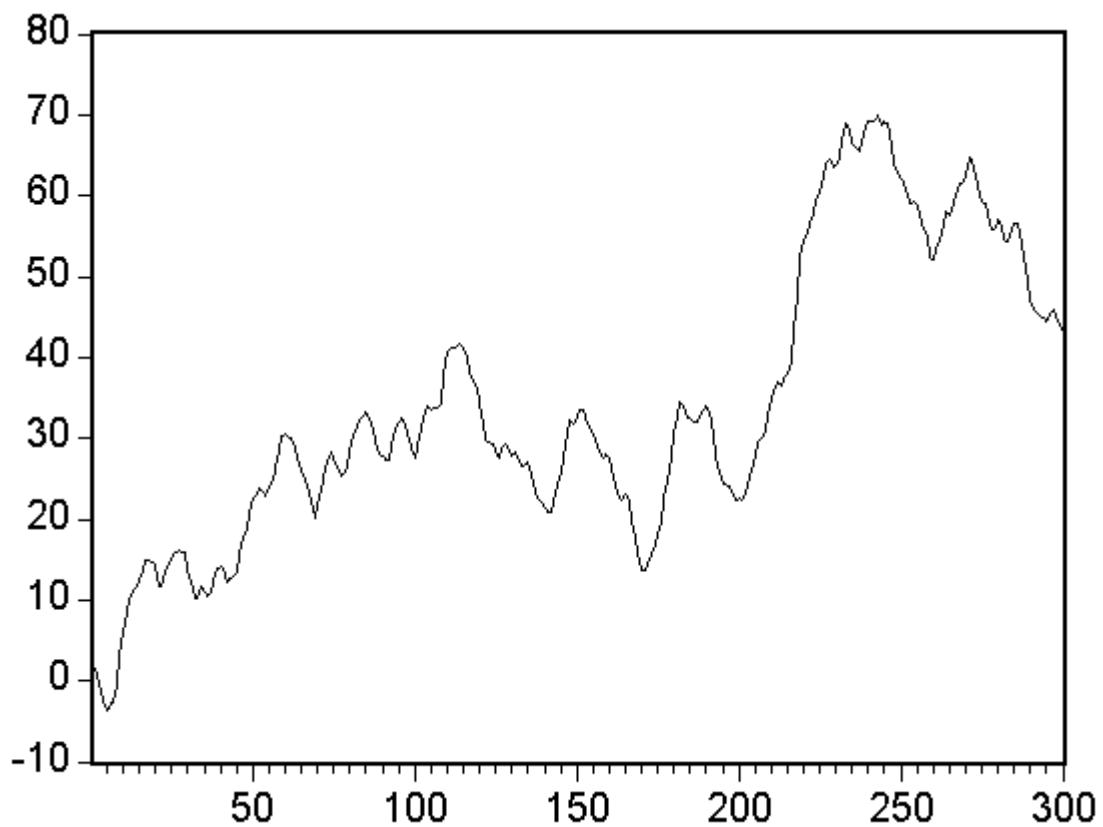
con $\Phi(B)$ e $\Theta(B)$ rispettivamente l'operatore stazionario **AR**

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

e l'operatore invertibile **MA**

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

La maggior parte delle serie che si incontrano nella pratica possono essere ricondotte ad un processo **ARIMA** (p,d,q) con p , d e q non superiori a 2.



Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
██████████	██████████	1	0.991	0.991	297.33	0.000
██████████	██████████	2	0.976	-0.274	587.00	0.000
██████████	██████████	3	0.958	-0.120	867.16	0.000
██████████	██████████	4	0.939	-0.053	1136.8	0.000
██████████	██████████	5	0.918	-0.018	1395.3	0.000
██████████	██████████	6	0.897	0.030	1643.1	0.000
██████████	██████████	7	0.875	-0.033	1880.0	0.000
██████████	██████████	8	0.854	0.017	2106.5	0.000
██████████	██████████	9	0.834	0.020	2323.1	0.000
██████████	██████████	10	0.814	-0.001	2530.4	0.000
██████████	██████████	11	0.796	0.022	2728.9	0.000
██████████	██████████	12	0.778	0.008	2919.2	0.000
██████████	██████████	13	0.760	-0.028	3101.6	0.000
██████████	██████████	14	0.742	-0.025	3276.1	0.000
██████████	██████████	15	0.725	0.014	3443.2	0.000
██████████	██████████	16	0.708	0.024	3603.2	0.000
██████████	██████████	17	0.693	0.030	3756.8	0.000
██████████	██████████	18	0.678	0.017	3904.6	0.000
██████████	██████████	19	0.664	0.010	4046.9	0.000
██████████	██████████	20	0.652	0.016	4184.3	0.000
██████████	██████████	21	0.639	-0.046	4316.8	0.000
██████████	██████████	22	0.626	0.000	4444.5	0.000
██████████	██████████	23	0.614	0.032	4568.0	0.000

Nel 1970 **Box e Jenkins** hanno presentato nel loro libro un metodo generale sviluppando un appropriato modello di serie storica ARIMA. Nel caso di serie storiche stagionali si parla di modelli **SARIMA o ARIMA stagionali**, i quali permettono di dare una natura stocastica anche alla stagionalità.

I modelli ARIMA stagionali sono stati introdotti da Box e Jenkins e sono dei modelli ARIMA(p,d,q) che, a loro volta, possiedono dei residui modellabili secondo una struttura ARIMA(P,D,Q)^s con operatori polinomiali costruiti in funzione di B^s e potenze successive.

Il metodo di Box e Jenkins trae origine dallo studio dei filtri lineari. Se si immagina di alimentare il filtro con una serie storica corrispondente al *white noise*, la risposta sarà identica alla risposta di frequenza del filtro. Il metodo Box-Jenkins ha lo scopo di identificare questo filtro e di generare la serie storica, in modo diretto dal segnale di *white noise*. Tale metodo permette di ottenere una lettura diretta del processo ARIMA che genera il processo stocastico stazionario in esame. La costruzione di un modello ARIMA è possibile perché la corrispondenza fra funzione di autocovarianza e parametri del processo è biunivoca e la stima delle autocovarianze risulta essere consistente, efficiente e sufficiente.

Questo prerequisito teorico non consentirebbe la costruzione di uno specifico modello ARIMA sulla base di un numero finito e limitato di osservazioni in quanto la classe ARIMA a sua volta consta di infiniti elementi, ciascun sottoinsieme dei quali è specificato dagli ordini (p,d,q). Per questo è necessaria e opportuna una procedura operativa fatta di decisioni coordinate allo scopo di specificare il modello, stimare i parametri e verificarne la significatività statistica.

La procedura proposta da Box e Jenkins consiste in tre tappe: ***identificazione, stima, verifica*** che sono state nel tempo arricchite in numero e in contenuti introducendo nuovi strumenti statistici.

- **Analisi preliminari.**

Le analisi preliminari consistono nelle trasformazioni iniziali da apportare a una serie storica perchè possa essere considerata come una parte finita di una realizzazione di un processo Gaussiano stazionario.

- Identificazione del modello.

E' il momento della specificazione degli ordini (p,d,q) del modello ARIMA e consente di individuare il numero e la natura dei parametri da stimare. Dei valori "ragionevoli" per p e q si possono individuare esaminando il correlogramma e l'ACF parziale della serie differenziata, dove c termini si sono persi differenziando.

In tutte le analisi dei dati e in particolar modo, in analisi di serie storiche, ci possono essere più modelli che rappresentano bene una stessa serie di dati.

Certe volte è semplice fare la scelta migliore altre volte risulta più difficile.

Per questo motivo sono stati introdotti numerosi criteri per poter confrontare i modelli trovati e poter scegliere il migliore. Il criterio **AIC** (*Akaike's information criterion*) serve a valutare la qualità di un modello che si è scelto per adattare i dati a disposizione è stato introdotto da **Akaike** (1973,1974) ed è definito come:

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

in cui $k=p+q+2$ è il numero dei parametri del modello,

$$\hat{\delta} = (\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q, \hat{\sigma}_\epsilon)'$$

è il vettore (k x 1) contenente i parametri stimati e $L(\hat{\delta})$

$$L(\hat{\delta})$$

è la funzione di verosimiglianza calcolata in $\hat{\delta}$ ed in ipotesi di gaussianità del *white noise*.

La quantità $\frac{2k}{n}$ rappresenta una sorta di penalità legata al numero di parametri del modello, che diminuisce al crescere di n, diventando del tutto trascurabile per valori elevati di n. Da questo deriva una certa tendenza del criterio AIC alla

sovraparametrizzazione.

Per ovviare a questo inconveniente si può considerare un secondo criterio, detto **SC (Schwarz Criterion, Schwarz 1978)**:

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

I valori prescelti di p e q sono quelli che minimizzano AIC o SC.

- Stima dei parametri.

Questa è la fase dedicata alla stima efficiente, consistente dei parametri di un modello ARIMA, specificato negli stadi precedenti. I due metodi di stima più usati sono il metodo della massima verosimiglianza e il metodo dei minimi quadrati non lineari.

In ipotesi di gaussianità del *white noise* il metodo della massima verosimiglianza fornisce stimatori con migliori proprietà statistiche.

Il metodo dei minimi quadrati non lineari ha invece il pregio di non dipendere da specifiche ipotesi distributive.

- Verifica del modello

In questa fase si cerca di controllare la qualità statistica del risultato ottenuto dagli algoritmi di stima, sia con test specifici sui parametri, che mediante il controllo complessivo del modello e dei residui da esso stimati. In questa fase decido se accettare o rifiutare il modello. Inoltre, in caso di rifiuto, ottengo informazioni sulla procedura da seguire nel ripercorrere il ciclo di "identificazione – stima – verifica".

Quando si è individuato un modello ARIMA soddisfacente si possono calcolare le previsioni. Disponendo di dati fino al tempo N, le previsioni coinvolgeranno sia le osservazioni sia i residui fino al tempo N compreso.

2.5 PREVISIONE

Quando si fanno delle previsioni in campo economico bisogna stabilire l'orizzonte temporale ossia se si intendono prevedere valori futuri delle serie nel breve, nel medio e nel lungo periodo.

Il **breve termine** si suddivide in due parti:

- A brevissimo termine (fino ad un mese): sono previsioni JIT (Just In Time). La previsione è condizionata dalla situazione presente che si ritiene senz'altro continuata nell'immediato futuro. In genere basta traslare la previsione di un solo periodo.
- A breve termine (fino ad un anno): hanno un ruolo di primo piano i fattori accidentali o erratici: scioperi, leggi speciali, catastrofi naturali, etc. oppure stagionali. La previsione è legata all'inerzia intrinseca in ogni evoluzione dinamica e che impedisce repentini cambiamenti. In questo caso le previsioni si estendono per più periodi ma non troppo.

A **medio termine** (da 1 a 5-10 anni): il ruolo centrale è svolto dalle fasi di espansione e contrazione che coinvolgono più settori di attività.

A **lungo termine** (dai 10 anni in poi): dominano i fattori demografici, le modifiche strutturali del sistema economico, l'evoluzione tecnologica, il sistema dei valori, etc. La previsione non può limitarsi alla mera estensione del passato ma deve comprendere varie alternative anche accompagnate da considerazioni qualitative (scenari).

In molte analisi economiche l'interesse si concentra su un periodo tra i 12 e i 24 mesi. Tale orizzonte è abbastanza ampio da dare spazio ad elementi decisionali, ma ancora breve perché si possa sentire l'influenza del passato più recente.

Per fare previsioni si cercherà di individuare quello che è chiamato il "modello generatore della serie storica" ossia una funzione matematica che permette di descrivere l'andamento della serie.

La previsione sarà realizzata con il metodo dell'estrapolazione.

Il modello generatore di una serie storica = modello matematico che è in grado di generare i dati della serie che si sta analizzando. Tuttavia quest'operazione è molto difficile: i valori di una s.s. sono influenzati da una molteplicità di fattori che rende praticamente impossibile costruire un modello che li consideri tutti.

Come detto, l'analisi delle s.s. si differenzia a seconda dell'obiettivo:

....fare previsioni;

....descrivere la serie.

>> Per fare previsioni si cerca di individuare il modello generatore della serie e grazie a tale modello si fanno previsioni.

>> L'analisi descrittiva mira a scorporare tra loro le diverse componenti della serie.

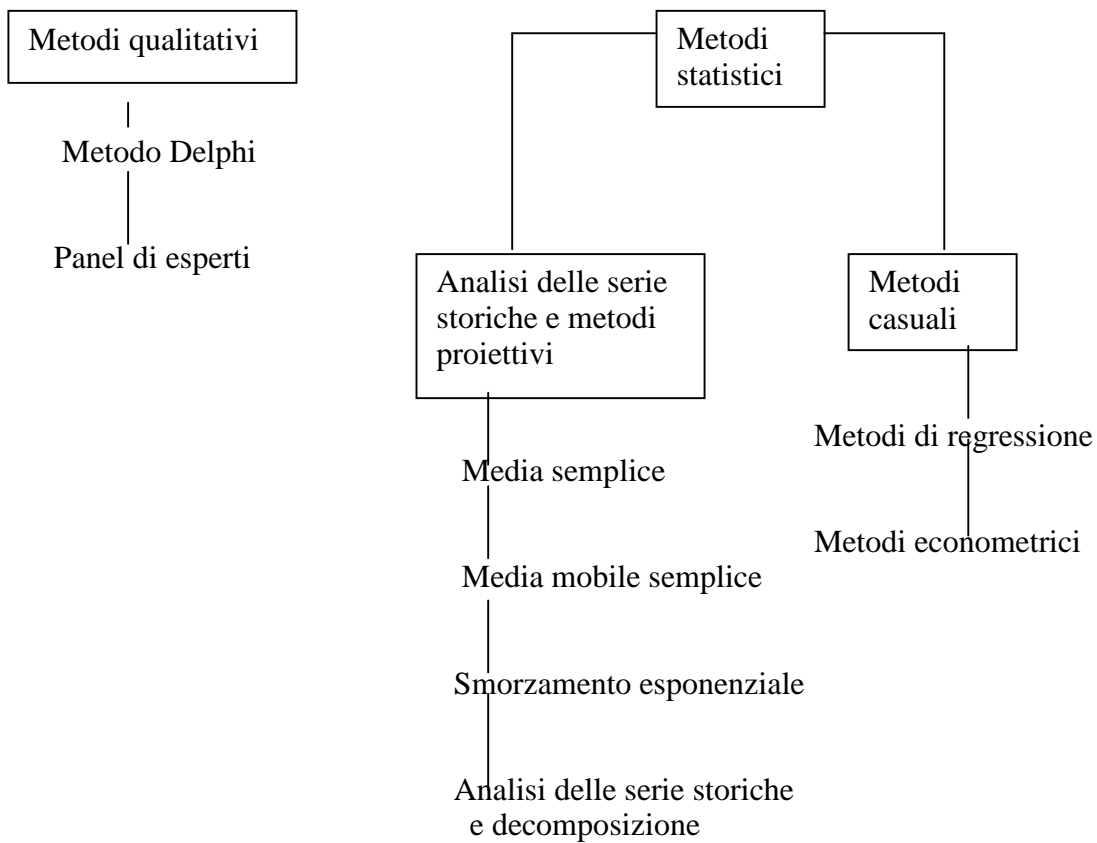
I metodi di previsione di tipo classico si fondano essenzialmente sull'idea seguente: se esiste un modello generatore dei dati di una serie, questo può essere posto in luce eliminando l'effetto che il caso ha sulla serie.

Si tratta quindi di individuare dei metodi che consentono di eliminare l'effetto della componente casuale della serie. Tali metodi sono detti di “*smoothing*”, ossia di lisciamento della serie. Sono metodi che non mirano a scorporare le varie componenti della serie, ma solo ad eliminare i disturbi casuali in modo da poter proiettare nel futuro vicino il modello sottostante così evidenziato.

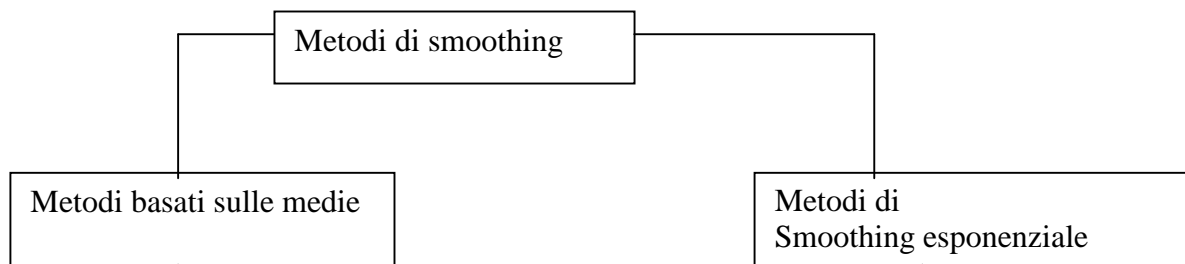
Come si può vedere dagli schemi seguenti ci sono vari approcci alle previsioni.

Metodi di previsione_schema²

² (Cfr. R. CASTAGNA, *Corso di gestione Aziendale, Modulo 13-22/04/2002, Facoltà di Ingegneria, Cremona*; MARKLAND SWIEGARD, *Quantitative Methods, applications to managerial decision making, Wiley International Editino, 1984*).



Metodi di previsione di tipo classico³

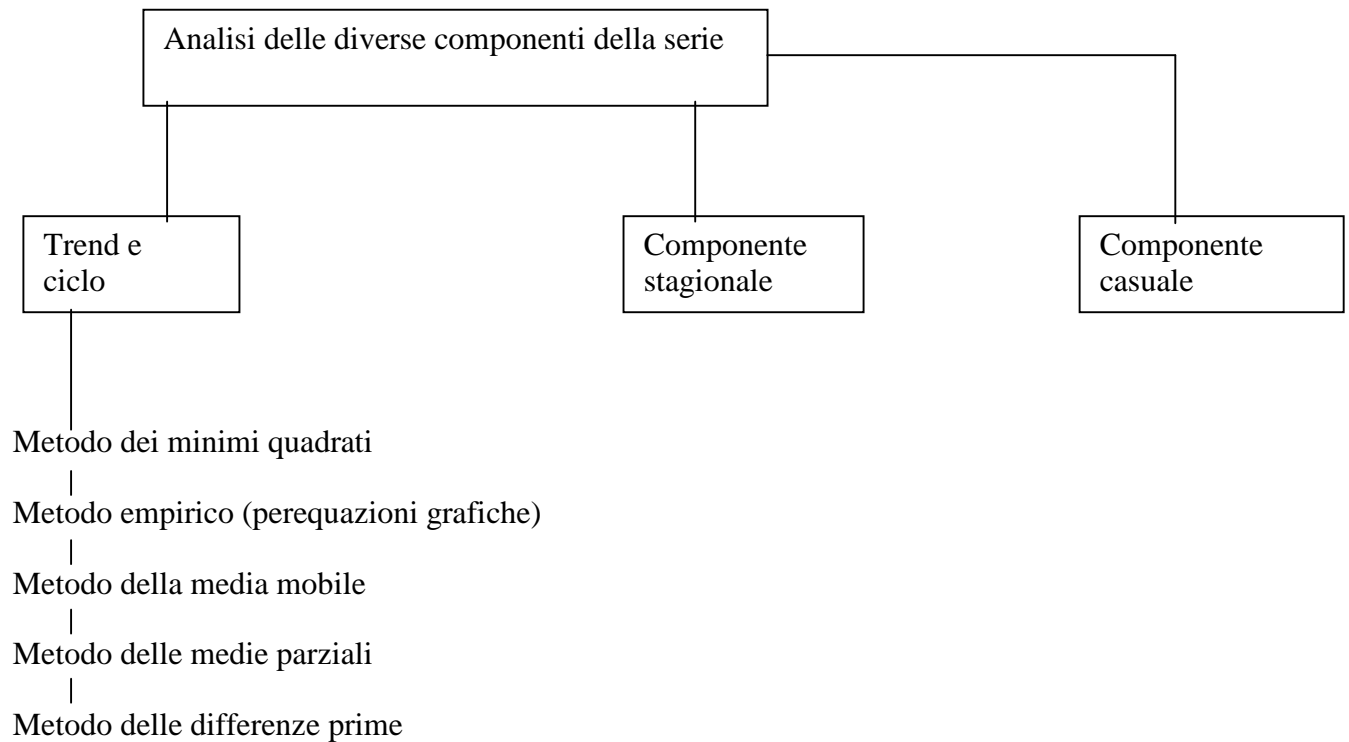


³ (Cfr. R. CASTAGNA, Corso di gestione Aziendale, Modulo 13-22/04/2002, Facoltà di Ingegneria, Cremona; MARKLAND SWIEGARD, Quantitative Methods, applications to managerial decision making, Wiley International Editino, 1984)..

Media aritmetica
|
Media mobile semplice
|
Media mobile doppia (lineare)

smoothing esponenziale semplice
|
metodo di Brown
|
metodi di Holt e Winters

Metodi di previsione di tipo classico: incorporare le componenti della serie



Metodi di previsione statistica: introduzione

Come anticipato prima, la previsione (o forecast) si distingue in base a tre elementi:

1) Tempo

- Breve periodo
- Medio periodo
- Lungo periodo

2) Obiettivo

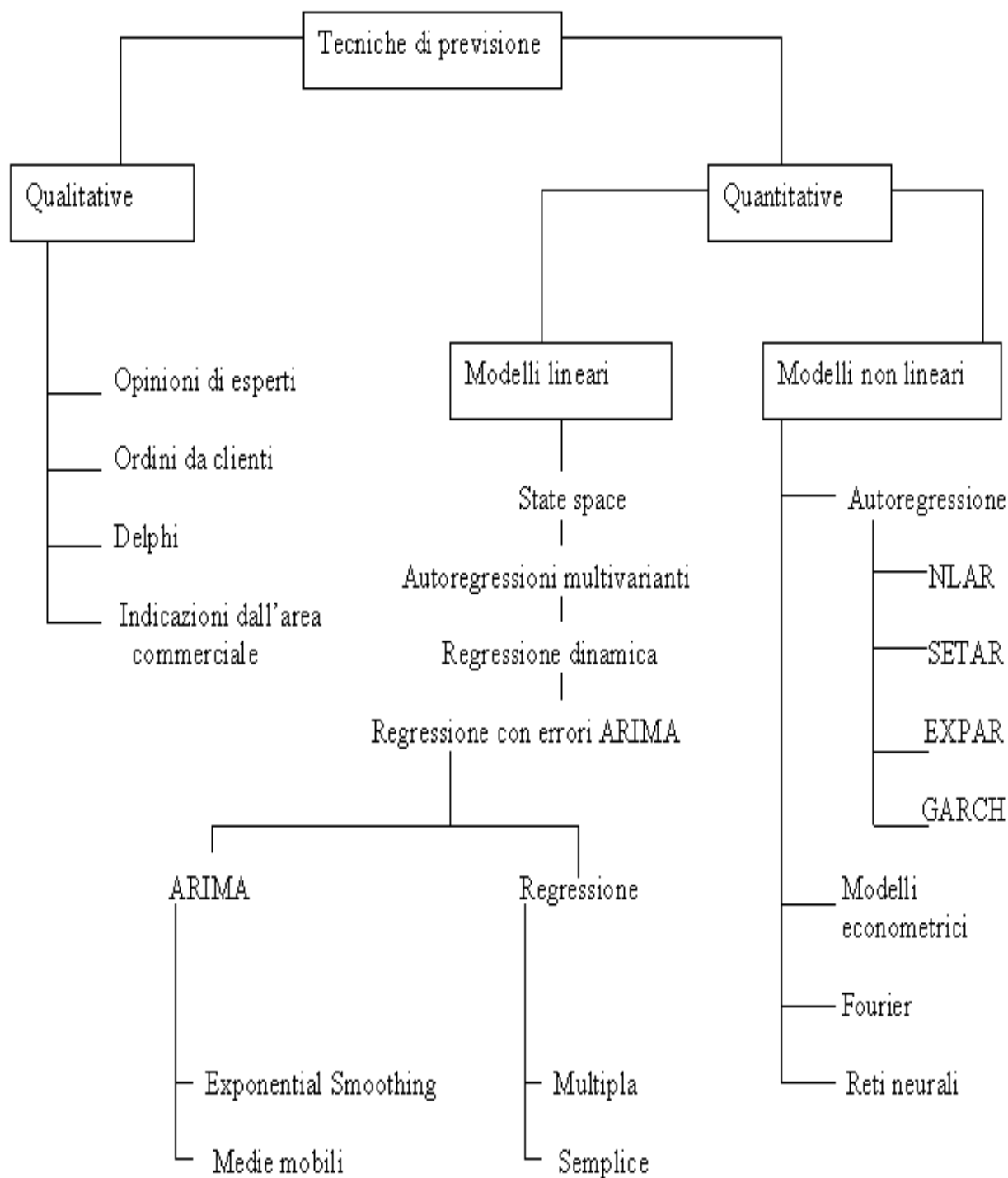
- Strumentale
- Tendenziale
- Condizionale
- Normativo

3) Metodo impiegato (una delle possibili suddivisioni) può essere

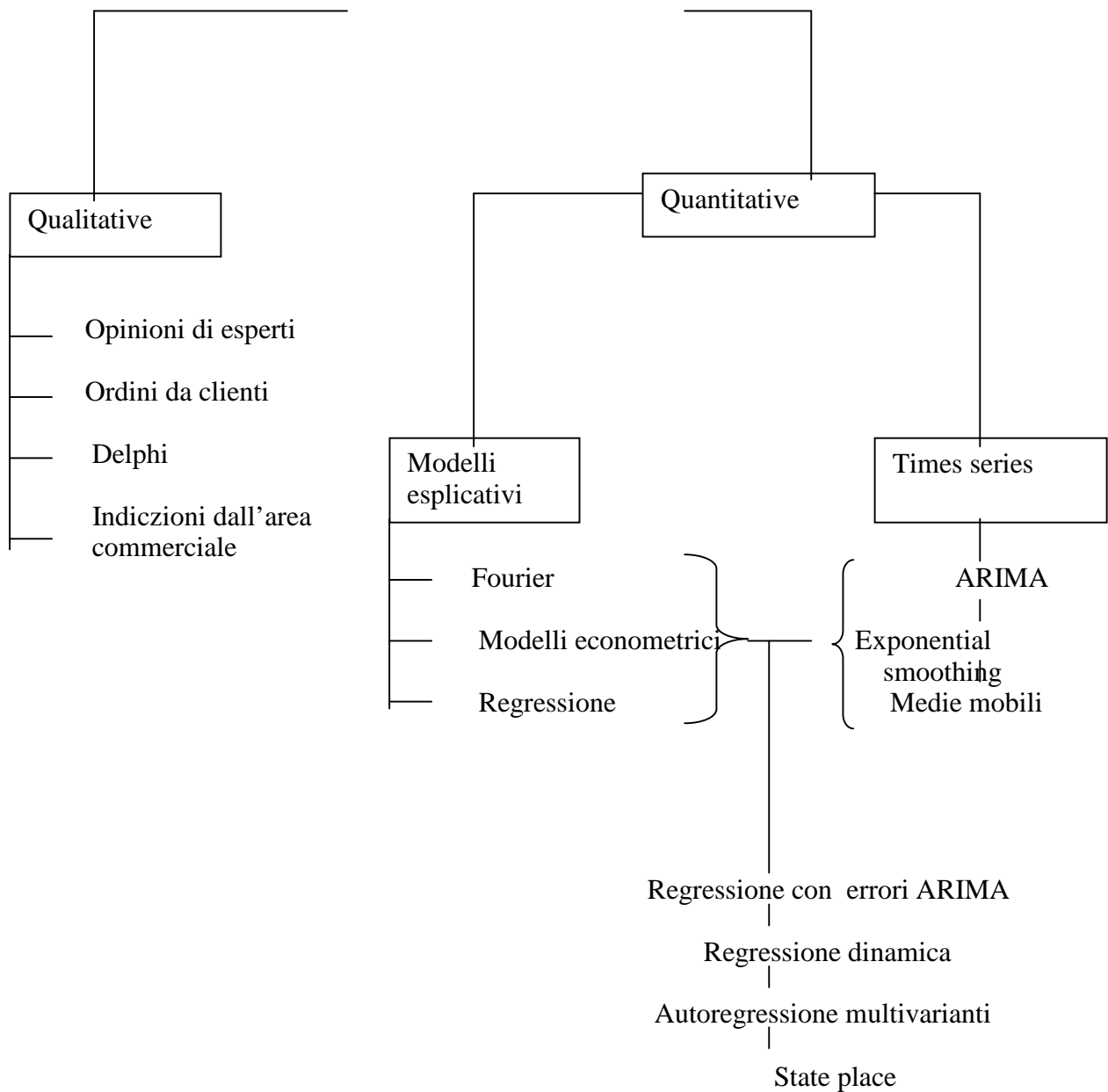
- Delle serie storiche;
- Le regressioni e modelli econometrici.

Le **procedure di previsione** sono innumerevoli ma possono essere classificate in tre gruppi principali:

- metodi univariati;
- metodi multivariati;
- metodi di giudizio o soggettivi



MOSCA R., (1998 – 1999), *Dispense al Corso di Economia Aziendale 1 e 2*, DIPEM, Pubblicazioni Interno



MOSCA R., (1998 – 1999), *Dispense al Corso di Economia Aziendale 1 e 2*, DIPEM, Pubblicazioni Interno

Modelli qualitativi.

I modelli qualitativi si basano sull'esperienza di manager/esperti nella determinazione della domanda di mercato. Per formulare previsioni soggettive

vengono impiegati dati qualitativi e informazioni su eventi e sul contesto ambientale. All'aumentare dell'arco temporale di previsione, aumenta il livello di aggregazione dell'analisi e diminuisce la significatività dei dati storici che riflettono l'andamento passato della domanda. Per questi motivi le tecniche impiegate diventano via via più qualitative e si svincolano dall'utilizzo dei dati storici.

Metodo Delphi: depura il giudizio degli esperti da componenti soggettive, pregiudizi e predisposizioni personali e ha le seguenti caratteristiche:

- a) Anonimato. Evita condizionamenti causati dal cambio di idea o dallo status/influenza dei membri più importanti.
- b) Interazione e feedback controllato. I giudizi e le previsioni degli esperti sono elaborate e filtrate da un moderatore che ad esempio opera in modo da focalizzare l'attenzione sul problema originario.
- c) Sintesi statistica delle risposte. Il metodo Delphi si propone di sfruttare il giudizio di esperti rendendolo il più oggettivo possibile.

Il metodo Delphi consiste nella somministrazione di questionari ad un gruppo di esperti (*panel*); tali questionari, oltre a richiedere giudizi e pareri, riportano il grado di consenso espresso dai membri sugli argomenti emersi. Nella versione originale, il metodo è composto da 4 cicli. I parametri più importanti per determinare l'attrattività del metodo Delphi sono:

- precisione ossia il grado di consenso ottenuto dal panel;
- affidabilità ossia il grado di stabilità dei risultati al variare degli esperti scelti (ottenimento di previsioni simili con gruppi di esperti diversi).

La precisione di una previsione ottenuta con il metodo Delphi viene misurata come la distanza interquantilica tra le risposte degli esperti.

I problemi del metodo Delphi sono riassunti brevemente in:

1) La definizione di esperto.

- La loro conoscenza degli eventi passati è rilevante per il futuro?
- Il parere degli esperti è sempre migliore dei pareri di individui mediamente informati?

2) Relazioni di conoscenza personale con gli esperti da parte del moderatore oppure la scelta di esperti all'interno dello stesso ambiente (scientifico o

industriale).

- 3) Il consenso non è garanzia di elevata precisione. Il metodo Delphi è strutturato in modo tale da favorire il consenso tra gli esperti. Un elevato consenso può scaturire anche da comportamenti conformisti da parte degli esperti.
- 4) Scarsa possibilità di verificare e misurare l'impegno e lo sforzo profusi da parte degli esperti.

MODELLI QUANTITATIVI

Nelle moderne organizzazioni si riscontra una forte tendenza a guardare verso il futuro, di conseguenza i pianificatori sono interessati ai tempi, alla magnitudo e agli effetti di eventi futuri che influenzeranno le loro attività. Il “*forecasting*” è l'attività di previsione, proiezione e stima di occorrenze di eventi futuri incerti. Poiché il futuro è raramente noto, gli strumenti di *forecasting* sono decisamente importanti e spesso assolutamente necessari per l'esito positivo delle attività industriali.

L'obiettivo di tali strumenti è quello di sfruttare al meglio le informazioni passate e presenti per guidare le attività delle organizzazioni verso i target prestabiliti.

Nelle organizzazioni industriali l'attività di *forecasting* è spesso usata per stimare la domanda futura di beni e/o servizi in quanto la maggior parte di tali organizzazioni non sono nella posizione di poter aspettare di aver ricevuto degli ordini per procedere con l'attività di :

- pianificazione e programmazione della produzione;
- definizione dei corretti livelli di manodopera;
- acquisizione materie prime e items.

Le organizzazioni industriali più avanzate anticipano la domanda futura dei loro prodotti e trasferiscono queste informazioni in fattori input per soddisfare tale domanda.

Affinché un'attività di business sia in grado di sopravvivere è necessario soddisfare le richieste del cliente in modo tanto rapido quanto i *competitors*.

I fattori che influenzano la domanda di beni e servizi sono molteplici e i più importanti possono essere così riassunti:

- 1) condizioni generali di business e stato dell'economia;
- 2) azioni e reazioni dei competitors.

A seconda della realtà in oggetto i tipi di previsione possono essere assai diversi e si basano su diverse unità di misura: unità fisiche, ore macchina, pesi e/o volumi, costo dei beni prodotti, ore di manodopera diretta, numero di clienti, etc.

Un modello previsionale rappresenta una metodologia che permette di :

- descrivere un fenomeno,
- prevedere un evento,
- controllare il sistema.

In ambito industriale, in particolare nel settore manifatturiero, l'utilizzo di questi modelli trova largo uso nell'ambito della gestione delle scorte e della programmazione della produzione.

La previsione può essere considerata quindi come il link tra il sistema esterno, non controllabile, e il sistema interno dell'organizzazione le cui attività sono governabili.

Il metodo qualitativo si basa su esperienze empiriche che variano molto in base allo specifico modello di *business*. Esso è meno complesso ma meno attendibile.

Il metodo quantitativo è un metodo oggettivo e prevede la possibilità di misurare l'accuratezza della previsione.

Esso è più accurato e preciso anche se risulta più oneroso.

Nei modelli esplicativi la variabile di uscita Y è funzione di una o più variabili dipendenti X_i .

L'obiettivo è quello di scoprire la forma della relazione analitica f (in maniera

esplicita o implicita) al fine di poterle utilizzare per formulare la previsione

$$Y=f(X_1,X_2,\dots,X_n)$$

Nelle serie storiche la variabile da prevedere dipende dalle caratteristiche della serie stessa e l'obiettivo è quello di scoprire le caratteristiche della serie storica ed utilizzarle per fare una estrapolazione del futuro.

L'estrapolazione dei dati è un procedimento che cerca di ricavare i valori prossimi di una grandezza mediante lo studio dei valori che si sono via via presentati. In generale abbiamo tre diverse tipologie di studio:

- Analisi Serie Storica (*Times Series Analysis*);
- Smorzamento Esponenziale (*Exponential Smoothing*)
- Reti neurali Artificiali(ANN).

Tutti i metodi si basano sulla significatività degli eventi passati verso il futuro prossimo, ma ovviamente tale assunto non è sempre vero.

Il metodo dell'analisi della serie storica cerca di identificare le componenti base (trend,ciclo e stagionalità)filtrando i valori precedentemente acquisiti.

Il metodo dello smorzamento esponenziale si basa invece sul concetto di media mobile pesando in maniera opportuna i dati più recenti.

Le reti neurali artificiali derivano la loro capacità previsionale cercando di approssimare la funzione incognita $C_t=\phi(C_{t-1},C_{t-2},\dots,C_{t-k})$ secondo il teorema di Kolmogorov.

Nei modelli esplicativi talvolta può risultare molto difficile individuare la relazione f tra variabili dipendenti e indipendenti. Però le serie storiche hanno come punto di forza la capacità di interpretare il fenomeno in assenza di una formulazione esplicativa. Come punto debole però le serie storiche non contemplano la possibilità di introdurre variabili che potenzialmente possono influenzare il fenomeno oggetto delle nostre "attenzioni".

In un'ottica previsionale non sempre i dati hanno lo stesso peso.Spesso infatti i dati più recenti hanno un contenuto informativo più significativo rispetto a quelli più distanti nel tempo.

Esistono dei modelli previsionali che tengono conto di questo fatto.

Analisi di tipo statico⁴

A) Last Period Demand (LPD)

La previsione del periodo successivo coincide con il livello accaduto nel periodo precedente.

$$Y_t = Y_{t-1}$$

Dove:

Y_t = domanda prevista per il periodo t

Y_{t-1} = domanda attuale del periodo t-1

Esso ha un'utilità pressoché nulla.

B) Studio delle curve di probabilità(p(x);F(x))

Il massimo contenuto di informazioni ottenibile è del tipo:

esiste una probabilità xx% che il consumo sia pari a C, o ad un valore ad esso inferiore(F(x)), ovvero che esso sia compreso tra due valori noti. Tale metodologia è poco utile.

C) Media Aritmetica Semplice.

La previsione per il periodo successivo coincide con la media dei livelli di domanda passati, cioè la previsione è formulata utilizzando tutti i valori passati.

$$P_{t+1} = \frac{1}{t} \sum_{i=1}^t y_i$$

Questa metodologia fornisce solamente un valore medio della domanda nel periodo considerato più l'eventuale S.Q.M.

⁴ ROMEO CASTAGNA, Op. Cit..

Analisi di tipo dinamico

Metodi di Averaging

Metodi di *Averaging* sono modelli “time series” così definiti:

$$Y = \sum_{n=0}^{\infty} \alpha_n y_{t-n}$$

E il loro obiettivo è quello di utilizzare le informazioni passate per sviluppare una previsione sul futuro. Si cerca attraverso opportune relazioni analitiche di stimare i coefficienti α_n . In generale i metodi di *averaging* non garantiscono elevate performance in presenza di trend o stagionalità.

Medie mobili semplici di ordine K

La media mobile di ordine k calcola la previsione sulla base di k osservazioni precedenti. La previsione MA(k) dipende solo da k valori passati:

$$P_{t+1} = \frac{\sum_{i=1}^k X_{t-k+i}}{K}$$

- P_{t+1} : previsione della domanda per il periodo t+1;
- X_i : Domanda effettiva nel periodo i (la sommatoria è estesa agli ultimi K periodi).

Ogni previsione è semplicemente un aggiustamento della previsione precedente.

Bisogna determinare il numero di periodi K da prendere in considerazione.

$$\sigma_k = \sqrt{\frac{\sum_{j=K+1}^t (P_j - X_j)^2}{t - (k + 1)}}$$

- P_j : previsione della domanda per il periodo j
- X_j : domanda effettiva nel periodo j

La scelta di K determina il grado di smorzamento del modello ossia se k assume un valore alto le previsioni sono maggiormente distribuite con oscillazioni ampie

mentre con K basso avremo previsioni con possibilità di variazioni repentine. Rispetto alla media aritmetica semplice ,la media mobile MA(k) presenta i seguenti vantaggi:

- si riferisce solo agli ultimi k periodi della serie
- non modifica il numero di osservazioni di riferimento nel tempo.

Verrà scelto l'ordine K della media mobile in modo da minimizzare il valore dello scarto quadratico medio associato.

Procedura:

- 1) Calcolo di σ per i diversi k;
- 2) Scelta del K con minore σ_k ;
- 3) Definizione della previsione P_{t+1} solo per il k definito al punto 2;
- 4) Determinazione della banda di confidenza della previsione:

$$P_{t+1} \pm t_{\frac{\alpha}{2}, t-(k+1)} \sigma_k \cdot$$

Smoothing Techniques

Le *smoothing techniques* sono metodi previsionali basati su tecniche di lisciamento o smorzamento e consentono, rimuovendo le fluttuazioni puntuali, di evidenziare il trend secolare attraverso il quale effettuare poi le previsioni per il futuro in modo semplice.

Medie mobili ponderate

$$P_{t+1} = \sum_{i=1}^k X_{t-k+i} W_i$$

$$\sum_i^k W_i = 1$$

$-W_i$: peso associato alla domanda effettiva nel periodo i (normalmente $W_i < W_{i+1}$)

Bisogna determinare sia il numero di periodi k sia i k valori dei pesi W_i

Exponential smoothing (ES)

Lo svantaggio delle medie mobili ponderate è rappresentato dal dover determinare i k valori dei pesi W_i .

Il modello dell'Exponential Smoothing permette di associare pesi decrescenti, in modo esponenziale, ai valori della domanda meno recenti.

$$P_{t+1} = P_t + a(X_t - P_t)$$

con $0 < a < 1$ e P_t : previsione della domanda per il periodo t effettuata all'istante $t-1$.

Si analizza per quale valore di a si ha la minimizzazione della deviazione standard della previsione (serie storiche).

Procedura

1) Si assume $P_1 = X_1$

2) Per ogni a e per ogni $j=2, t$ si calcola $P_j = aX_{j-1} + (1-a)P_{j-1}$

3) Calcolo σ per i diversi a

4) Scelta del valore a con σ_a minore ossia $\sigma_a = \sqrt{\frac{\sum_{j=2}^t (P_j - X_j)^2}{t-2}}$

5) Definizione della previsione P_{t+1} solo per il valore di a definito al punto 4

6) Determinazione della banda di confidenza della previsione $P_{t+1} \pm t \frac{\alpha}{2, t-2} \sigma_a$

Exponential Smoothing doppio (HES)

Exponential Smoothing doppio (HES) prevede l'utilizzo di due parametri a e b .

La previsione si ottiene come:

$$P_{t,t+\tau} = S_t + \tau G_t$$

$$S_t = aX_t + (1-a)(S_{t-1} + G_{t-1}) \quad 0 < a < 1$$

$$G_t = b(S_t - S_{t-1}) + (1-b)G_{t-1} \quad 0 < b < 1$$

S_t è una stima del livello della serie al tempo t (intercetta)

G_t è una stima della pendenza della serie al tempo t

P_t è pari alla somma dell'intercetta con un termine che fornisce da indicatore dell'inclinazione della serie.

Con $\tau \neq 1$ le previsioni risultano essere delle rette inclinate di un valore pari a G_t passanti, all'istante t , per il valore S_t . Il processo di inizializzazione dei parametri richiede due stime: S_1 e G_1 . Essi possono essere stimati:

a partire dalla serie storica ponendo $S_1 = X_1$ e $G_1 = X_2 - X_1$

oppure utilizzando un modello di regressione.

La scelta dei parametri a e b può essere realizzata attraverso un algoritmo di ottimizzazione non lineare oppure tramite un algoritmo che esplora tutte le combinazioni di a e b scegliendo poi quella che minimizza l'errore.

Rispetto ai modelli MA ed ES, HES modella il trend di una serie storica, anche se non è possibile utilizzare HES con serie storiche affette da fenomeni di stagionalità.

Exponential Smoothing Triplo (TES)

Exponential Smoothing Triplo (TES) si applica a serie storiche stagionali dotate di trend.

La previsione può essere effettuata in due modi differenti a seconda che si basino

sul modello:

moltiplicativo

$$P_{t,t+\tau} = (S_t + \tau G_t) C_{t-s} + \tau$$

oppure sul modello **additivo**.

$$P_{t,t+\tau} = S_t + \tau G_t + C_{t-s+\tau}$$

Con S_t stima del livello della serie al tempo t

G_t stima della pendenza della serie al tempo t

C_{t-s-t} stima della componente stagionale dipendente dalla lunghezza di una stagione (S) da definire a priori

Le stime sono calcolate in funzione dei parametri a b c .

Nel caso del modello moltiplicativo i valori della componente orizzontale S_t , di trend G_t e stagionale

C_{t-s-t} sono calcolati attraverso le seguenti relazioni:

$$P_{t,t+\tau} = (S_t + \tau G_t) C_{t-s} + \tau$$

$$S_t = a \frac{X_t}{C_{t-s}} + (1-a)(S_{t-1} + G_{t-1})$$

$$G_t = b(S_t - S_{t-1}) + (1-b)G_{t-1}$$

$$C_t = c \frac{X_t}{S_t} + (1-c)C_{t-s}$$

Come tutti i modelli di Exponential Smoothing, l'Esponential Smoothing triplo prevede che venga stimato il valore iniziale delle componenti all'inizio dell'algoritmo. Il metodo più semplice si basa sul calcolo dei valori di una stagione completa (cioè s periodi).

Intercetta all'istante s
$$S_s = \frac{1}{s} \sum_{t=1}^s X_t$$

Trend all'istante s
$$G_s = \frac{1}{s} \sum_{t=1}^s \frac{X_{t+s} - X_t}{s}$$

I-esimo indice stagionale
$$C_i = \frac{X_i}{S_s}$$

Nel caso di un modello additivo, l'unica differenza consiste nel calcolo della componente stagionale.

$$P_{t,t+\tau} = S_t + \tau G_t + C_{t-s} + \tau$$

$$S_t = a(X_t - C_{t-s}) + (1-a)(S_{t-1} + G_{t-1})$$

$$G_t = b(S_t - S_{t-1}) + (1-b)G_{t-1}$$

$$C_t = c(X_t - S_t) + (1-c)C_{t-s}$$

L'inizializzazione del modello è analoga a quella del modello moltiplicativo, escluso il calcolo della componente stagionale $C_i = X_i - S_s$.

La scelta dei parametri ottimi può essere fatta minimizzando MSE.

La scelta più conveniente per trovare i coefficienti ottimi è quella di applicare un algoritmo di ottimizzazione lineare.

Possiamo usare poi la :

- regressione lineare;
- regressione non lineare;
- regressione multipla.

Su questi metodi non mi soffermo in questa sede .

Modelli dinamici e modelli statici

Un modello dinamico di decisione implica –come nel caso di flussi di risorse- la presenza di trasferimenti di risorse ma questa volta l'evento dinamico ,cioè che conferisce una dimensione temporale al sistema ,è una decisione presa a causa di un evento.

La complessità del sistema economico e la natura particolare dell'attività economica sono tali da far ritenere che il processo che genera le osservazioni sulle s.s.e. sia un processo stocastico multidimensionale e dinamico,cioè un processo in cui il presente è influenzato ,almeno in parte dal passato.Se si è concordi con ciò ,il modello lineare classico analizzato nei paragrafi precedenti,può rappresentare una buona modellazione del processo generatore delle variabili osservabili solo se sono verificate determinate condizioni di indipendenza condizionale tra la variabile

dipendente y_t ed il passato suo e/o delle variabili esplicative.

In econometria esistono almeno due ragioni che giustificano il ricorso alla specificazione dinamica dei modelli. Un modello dinamico è caratterizzato principalmente dal fatto che la variabile dipendente è funzione di un insieme di variabili esplicative osservate in periodi diversi del tempo, producendo relazioni non solo istantanee tra variabili, ma anche differite nel tempo.

Esempio di modello dinamico $c_t = \alpha + \beta y_{t-1}$ oppure $c_t = \alpha + \beta_0 y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_k y_{t-k}$ con c e y variabili e con α e β parametri.

Il modello di regressione lineare statico (intervengono solo le variabili *correnti* cioè associate allo stesso tempo t , esempio $c = \alpha + \beta(y - v)$ oppure $c = \alpha + \beta(y)$ con c e y e v variabili e α e β parametri) si ottiene eliminando l'effetto del passato ed avendo una dipendenza pura ed istantanea tra variabili. Limite del modello statico è la mancanza di una qualsiasi forma di dinamica al punto da renderli poco verosimili e poco utili per lo studio dei fenomeni economici. Quindi un modello statico si rivela spesso insoddisfacente a causa della limitata capacità esplicativa che riesce ad esprimere per l'analisi dei fenomeni economici reali. Le sue capacità previsive non sono soddisfacenti poiché non contenendo informazione sul passato della variabile y_t .⁵ Esso spesso produce errori elevati di previsione.

Indici per valutare la qualità delle previsioni⁶

Una volta conosciuto l'andamento effettivo della serie, sorge naturale l'esigenza di valutare l'attendibilità delle previsioni formulate in precedenza. Solitamente ci si prefigge di rispondere a queste domande:

- Date due o più serie di previsioni, quale ha fornito i risultati migliori?
- In che misura una serie di previsioni può considerarsi soddisfacente?

Non esiste un metodo di previsione migliore in senso universale, ma bensì ogni metodo ha i suoi vantaggi e si adatta a specifiche esigenze. Inoltre la misura

⁵ LORENZO FARINA, Modelli Dinamici per l'Ingegneria Gestionale, AA 2005/06, Facoltà di Ingegneria, Uniroma1

⁶ DI FONZO TOMMASO, Materiali per il corso di Statistica economica, AA 2001/02, Facoltà di Scienze Statistiche, Università degli Studi di Padova.

della qualità della previsione è data, di solito, dal confronto con la realtà, ossia dalla distanza tra previsione e la realizzazione effettiva.

Nell'analisi congiunturale (cioè che contempla un periodo con limite superiore fino ai 12-18 mesi) l'obiettivo primario fondamentale è la previsione dei punti di svolta ciclici dell'economia.

Valutare la bontà delle previsioni in questo caso comporta il tenere un considerazione i punti di svolta previsti ed i punti di svolta effettivi, focalizzando l'attenzione esclusivamente sul messaggio che esse offrono circa la dinamica del fenomeno (in crescita, in diminuzione o stabile) rispetto al periodo precedente.

Indichiamo con p_t il valore previsto al tempo t , valore disponibile con un certo anticipo rispetto al tempo t ; e con r_t indichiamo il valore realizzato al tempo t conosciuto di solito con un certo ritardo.

Previsione dei punti di svolta

Come si è detto la validità di una serie di valori previsti viene giudicata in primo luogo sulla base dei punti di svolta. I punti di svolta riscontrati nella serie di valori previsti p_t vengono confrontati per ogni tempo t con la situazione effettivamente accertata che è solitamente costituita dai punti di svolta della corrispondente serie di valori realizzati r_t .

valori previsti p_t	valori realizzati r_t		
	p.s	no p.s	Totale
p.s	n_{11}	n_{12}	$n_{1.}$
no p.s	n_{21}	n_{22}	$n_{2.}$
totale	$n_{.1}$	$n_{.2}$	n

Con n_{11} e n_{22} =punto di svolta in cui le previsioni si sono rivelate esatte.

n_{12} =errore di prima specie ossia punto di svolta previsto e non realizzato.

n_{21} =errore di seconda specie ossia punto di svolta non previsto ma verificatosi.

Definiamo due indici relativi per descrivere i due errori:

- Indice relativo degli errori di prima specie (rapporto tra il numero di errori di prima specie e il numero totale di punti di svolta previsti): $E_1 = \frac{n_{12}}{n_{.1}}$
 - Indice relativo degli errori di seconda specie (rapporto tra il numero di errori di seconda specie e il numero totale di punti di svolta previsti): $E_2 = \frac{n_{21}}{n_{.1}}$
- Essi valgono zero quando non si è commesso nessun errore mentre valgono uno quando nessuno dei punti di svolta previsti corrisponde ai punti di svolta effettivi

Valutazione delle performance⁷

Valutare le performance di un particolare modello significa misurare quanto è accurato un particolare modello previsionale dato un set di valori.

L'errore puntuale commesso si calcola come differenza tra osservazione e previsione.

$$E_t = X_t - P_{t-\tau,t}$$

Avendo n valori di e_t è possibile definire alcune grandezze statistiche per l'errore.

Si definiscono :

- Errore Medio $ME = \frac{1}{n} \sum_{t=1}^n e_t$
- Errore Medio Assoluto $MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$
- Errore Medio Quadratico $MSE = \frac{1}{n} \sum_{t=1}^n e_t^2$
- Errore Medio Assoluto Percentuale $MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{X_t - P_t}{X_t} \right|$

Ogni grandezza rappresenta una precisa misura di accuratezza:

⁷ DI FONZO, Op. Cit.

ME esprime la presenza o meno di una tendenza a sovrastimare o sottostimare la previsione.

MAE esprime lo scostamento medio della previsione dal dato reale.

MSE indica la somma quadratica degli errori.

MAPE è uno scostamento medio percentuale che tiene conto del valore dell'osservazione.

Sia che vengano adottate tecniche di tipo statico che tecniche di tipo dinamico il criterio di scelta della bontà previsionale di un modello consiste sempre nel minimizzare l'errore di previsione.

Indico con σ l'errore globale di previsione con σ così definito:

$$\sigma = \sqrt{\sum_{i=0}^n (P_i - X_i)^2}$$

Esso rappresenta l'errore globalmente commesso dal modello usato.

La procedura standard per valutare la performance prevede sei step:

- 1) Dividere la serie storica in due parti: - set di training;
- set di testing.

Talvolta alcune tecniche richiedono di utilizzare tre sottoinsiemi della serie storica: training, validation e testing.

- 2) Utilizzare i dati relativi al set di training per formulare le previsioni;
- 3) Memorizzare gli scostamenti tra le previsioni e i dati del set di prova;
- 4) Aggiornare il set di training con il nuovo dato reale;
- 5) Iterare gli step 2-3-4 finché il set di inizializzazione non contiene tutti i valori della serie storica;
- 6) Applicare una tecnica statistica per analizzare gli scostamenti memorizzati.

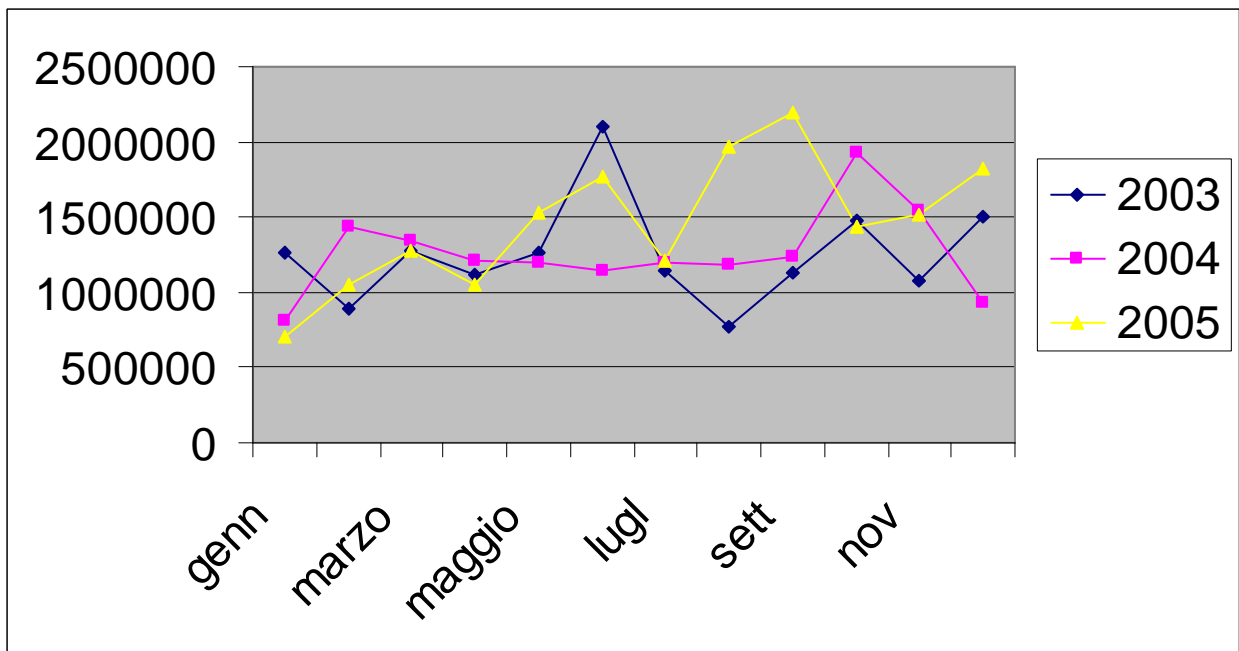
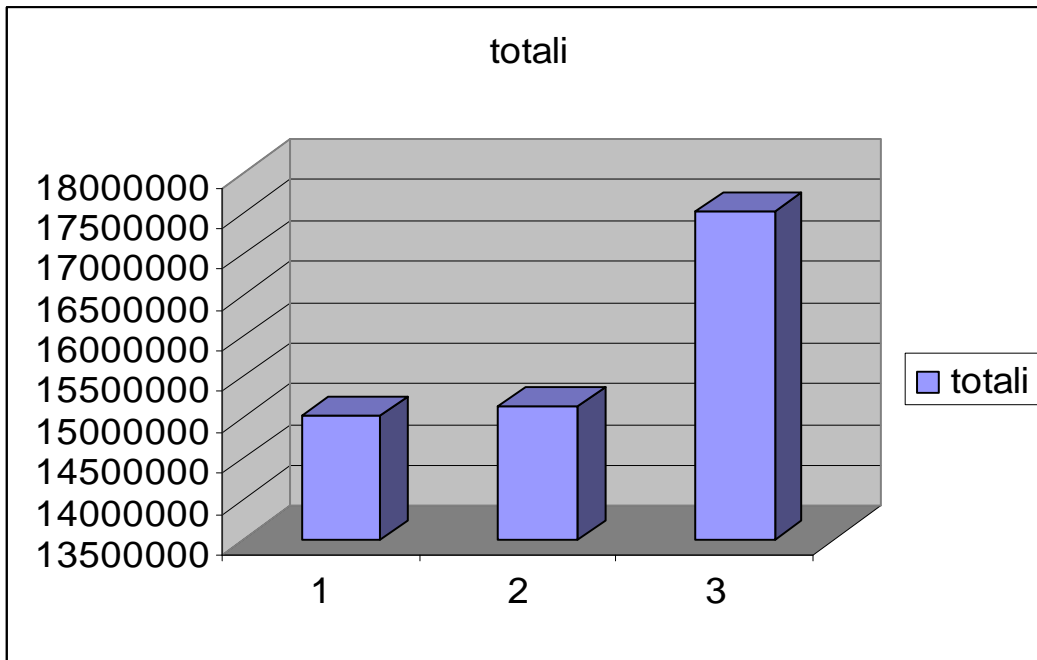
E' importante quindi valutare lo scostamento tra la previsione ed il dato futuro, essendo il dato futuro non ancora utilizzato dal modello per formulare la previsione stessa.

Capitolo 3

3.1 Breve panoramica sull'azienda

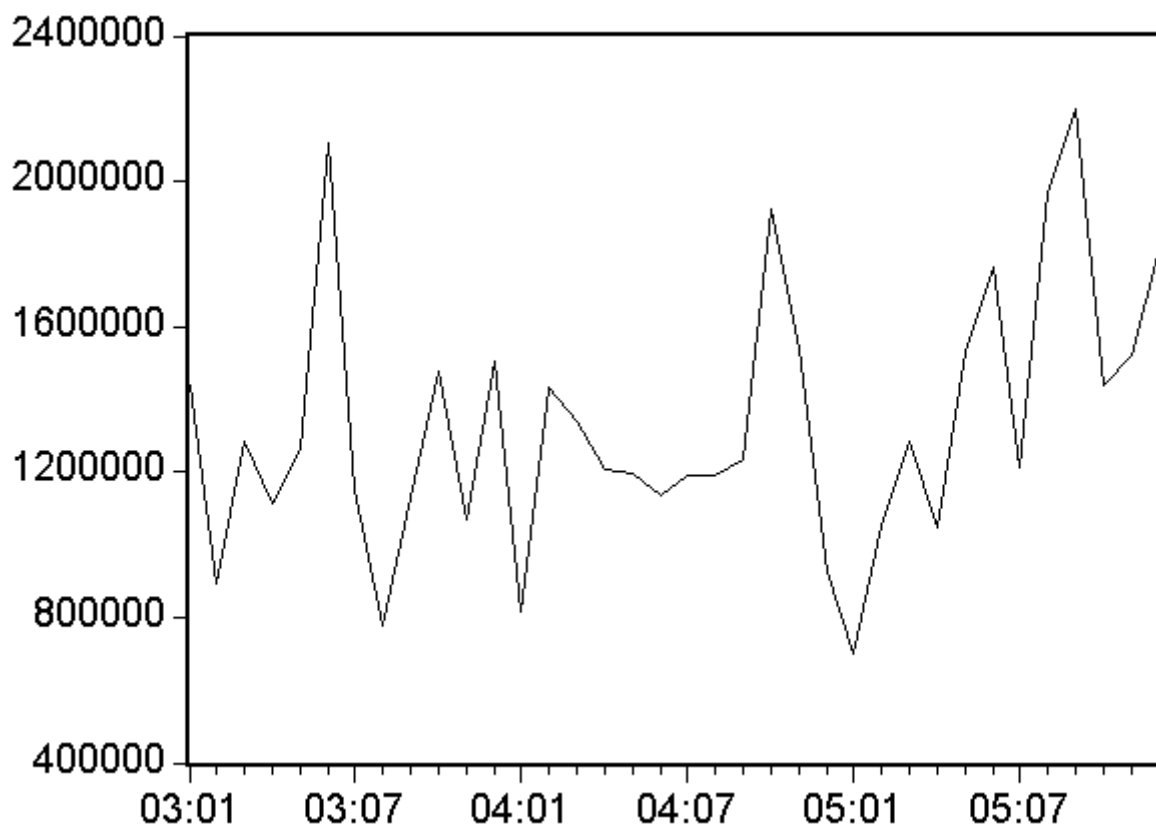
L'azienda *Asolo Dolce s.a.s* gentilmente ha messo a disposizione il suo database con i dati relativi alle vendite mensili con partenza dal gennaio 2003 fino al dicembre 2005

	duemilatre	duemilaquattro	duemilacinque
gennaio	1263170,67	813992,21	702696,26
febbraio	892014,56	1432598,65	1050326,64
marzo	1282003,61	1339991,77	1282003,61
aprile	1111600,41	1207732,05	1045065,8
maggio	1263170,67	1196800,84	1529632,4
giugno	2104643,57	1138048,26	1765067,65
luglio	1149598,31	1191826,52	1212040,67
agosto	776429,26	1189112,96	1966026,4
settembre	1131196,53	1233592,69	2197204,18
ottobre	1476108,88	1925137,56	1436302,23
novembre	1070788,51	1537001,58	1522369,17
dicembre	1507054,03	928400,95	1821096,9
totali	15027779,01	15134236,04	17529831,91

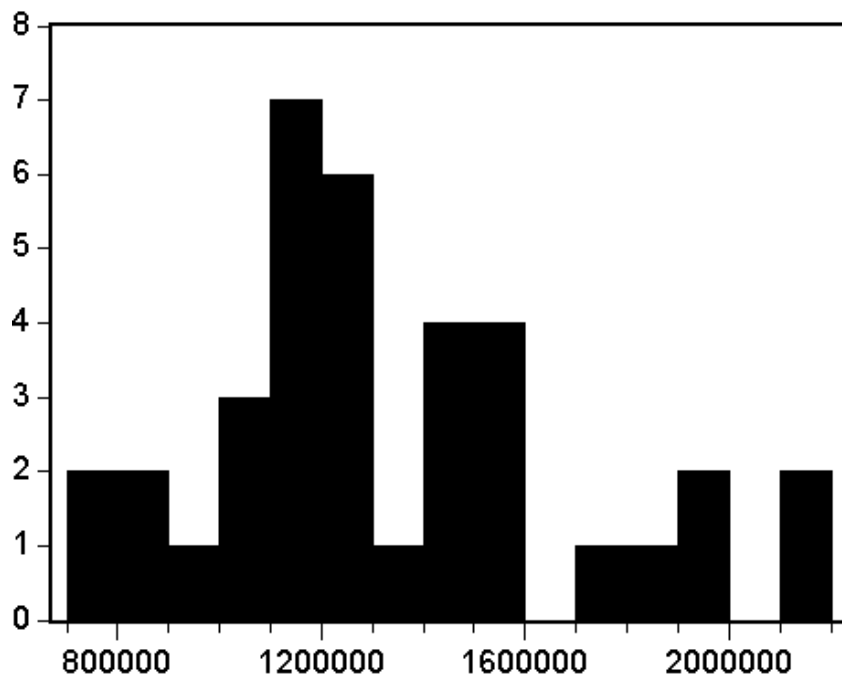


Con Eviews analizzo i dati in sequenza temporale e ottengo : il grafico

della serie mensile



L'istogramma con principali dati statistici mi dà questo risultato:



Series: Y	
Sample 2003:01 2005:12	
Observations 36	
Mean	1329617.
Median	1248381.
Maximum	2197204.
Minimum	702696.0
Std. Dev.	360699.2
Skewness	0.628744
Kurtosis	2.993897
Jarque-Bera	2.371967
Probability	0.305446

Correlogramma della serie presa in esame:

File Edit Objects View Procs Quick Options Window Help
View Procs Objects Print Name Freeze Sample Genr Sheet Stats Ident Line Bar

Correlogram of Y

Date: 02/28/06 Time: 20:09
Sample: 2003:01 2005:12
Included observations: 36

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.211	0.211	1.7398	0.187
		2 -0.042	-0.091	1.8109	0.404
		3 0.146	0.185	2.6955	0.441
		4 0.006	-0.081	2.6971	0.610
		5 -0.050	-0.005	2.8070	0.730
		6 0.068	0.054	3.0154	0.807
		7 -0.085	-0.124	3.3548	0.850
		8 -0.075	0.002	3.6277	0.889
		9 -0.057	-0.092	3.7947	0.924
		10 0.042	0.117	3.8855	0.952
		11 0.075	0.043	4.1967	0.964
		12 -0.036	-0.056	4.2687	0.978
		13 0.001	0.034	4.2688	0.988
		14 -0.070	-0.142	4.5739	0.991
		15 -0.067	0.024	4.8658	0.993
		16 0.096	0.074	5.4934	0.993

Dall'analisi del correlogramma vedo che la serie è un **wn**. Ossia le osservazioni componenti la serie sono tra loro incorrelate.

Dalla teoria sappiamo che per prevedere il futuro il futuro è necessario utilizzare l'informazione del passato. In una partita di testa o croce, la memoria non gioca alcun ruolo; il presente è indipendente dal passato. Non vi è memoria. La memoria di ciò che osserviamo è la parte utile di tale osservazione, parte che permette di tradurre la dipendenza probabilistica tra presente e passato.

L'indipendenza è difficile da ammettere. Tutti sanno che i giocatori dei casinò che hanno visto uscire il rosso dieci volte di seguito, pensano che all'undicesima partita, ci sono maggiori probabilità di vedere uscire il nero. Osservare un milione di partite non apporterebbe nessuna informazione supplementare quanto al risultato della partita successiva.

La memoria è di lunghezza L se la conoscenza di un passato più lontano di L non apporta più informazione di quanto non faccia il passato di lunghezza L .

I fenomeni senza alcuna memoria in matematica sono designati con il nome di “**rumore bianco**”. La probabilità è tale che ciò che succede ad ogni istante è indipendente sia dal passato che dal futuro. Ogni fenomeno che si svolge nel corso del tempo senza memoria si chiama rumore. Quando costruisco un modello ciò che non è possibile spiegare resta rappresentato da un rumore.

Definisco “innovazione” tutto ciò che avviene (in un mese, in questo caso, trattando di serie mensili) ed è un elemento indipendente dal passato. Tali innovazioni sono indipendenti e costituiscono una sequenza senza memoria. Hanno per ipotesi tutte le stesse caratteristiche probabilistiche ,con media nulla.

Possiamo quindi dire che questo è” rumore di innovazione” che riunirà tutto ciò che può succedere in un mese e non è tenuto in considerazione dal modello. Ad esempio conflitti sociali, avvenimenti monetari inaspettati o eventi meteorologici eccezionali. Considerare tale innovazione aleatoria ci permette di riassumere tutto quello che non si sa o che non si vuole introdurre nel modello ma che ci permette facilmente di creare una previsione.

Prevedere si riassume nel ”prolungare il funzionamento del sistema dal momento presente ad un dato momento futuro”, supponendo che il fattore di innovazione sia nullo. L'innovazione non ha memoria, ed è indifferente al passato, quindi, la migliore previsione dell'innovazione è “nessuna innovazione”.

3.2 IL MERCATO DELL'AZIENDA

L'instabilità geo-politica continua a penalizzare i mercati mondiali e le economie sia a livello macro-economico sia nel nostro piccolo mondo delle famiglie che compongono la nostra società. Tuttavia il sedimentarsi del mercato obbliga l'azienda a proporsi in ambiti più differenziati. Da qui nasce la ricerca continua di nuovi mercati di sbocco e di nuove prodotti con cui presentarsi al consumatore.

3.3 RETE DISTRIBUTIVA DELL'AZIENDA

Come esposto, la crisi economico la crisi economica continua a condizionare gli acquisti delle persone, ci pone l'obbligo di concentrarci su obiettivi precisi. La scelta di determinati mercati implica una riorganizzazione delle risorse umane, della logistica ed un servizio al cliente molto preciso. La vendita al dettaglio sta subendo una crisi che ormai dura da diverso tempo. Vi è in atto una metamorfosi della rete di vendita e la GDO sta avendo sempre maggiore incidenza nelle scelte del consumatore finale. I costi di esercizio elevato e i ridotti spazi espositivi, impongono al dettagliante criteri di selezione dei fornitori. Per contro alle aziende crea solo una notevole frammentazione di fatturato con conseguente aumento dei costi aziendali; unicamente per soddisfare quantitativi limitati e logisticamente distribuiti in tutto il territorio nazionale. Risulta per tanto doveroso proporsi in modo più incisivo nella GDO.

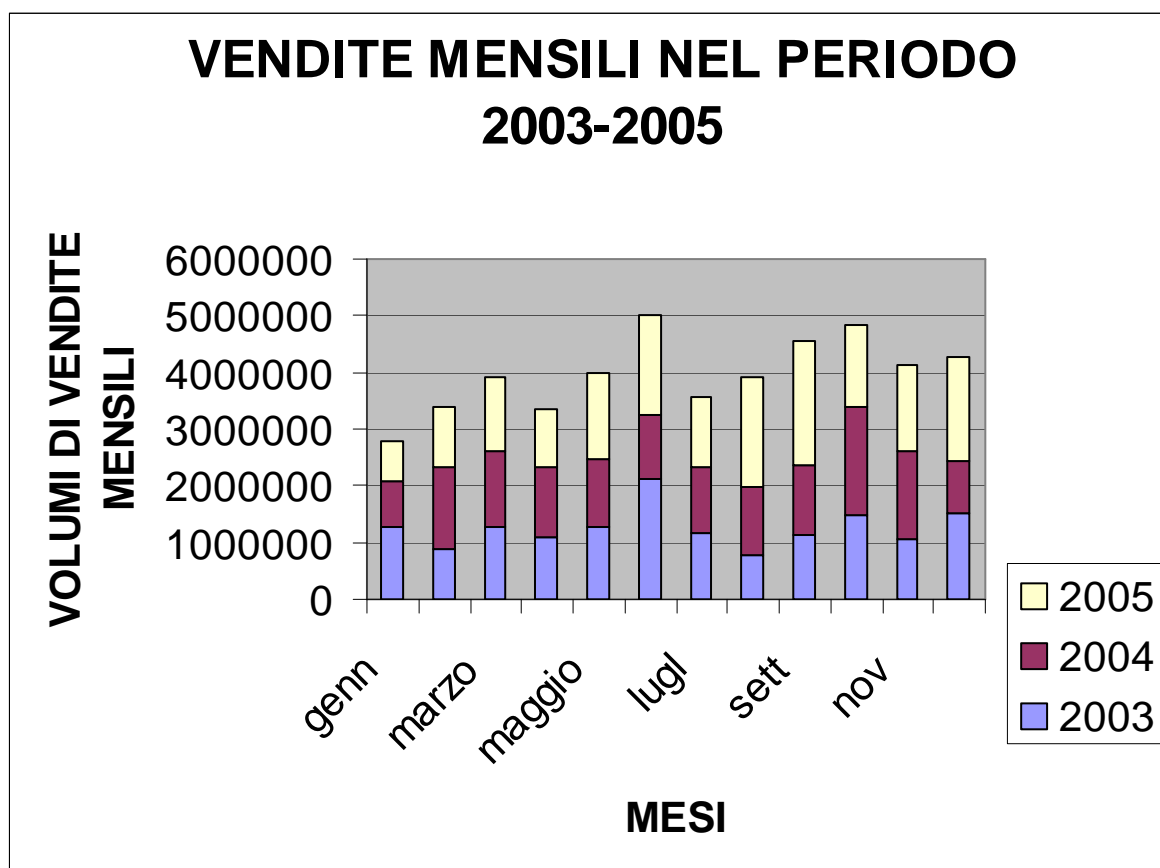
Inoltre l'azienda è dotata di una buona rete di vendita a livello nazionale composta da validi agenti.

Il loro compito non si limita solo alla fase di vendita vera e propria ma anche ad affiancare la clientela nell'assistenza post-vendita. I titolari spesso e volentieri prendono le valigie di campionario e partono per trattare di persona nuovi accordi commerciali soffermandosi fuori sede anche più giorni.

Capitolo 4

4.1 ANALISI DEI DATI DELLE QUANTITA' VENDUTE DAL 2003 AL 2005

In questo capitolo cercherò di fare un'analisi delle vendite che sono state fatte dal 2003 al 2005. L'andamento delle quantità vendute di questa azienda è molto simile all'andamento generale dell'economia italiana che soprattutto dalla nascita della nuova moneta europea si trova di fronte ad una grande crisi economica in tutti i settori.



La suddivisione del volume di vendite per articolo non mi risulta utile in quanto il numero di prodotti è notevole e risulterebbe troppo dispersivo ai nostri fini. Sarebbe stato interessante soffermarci solo sugli articoli che danno una maggiore redditività all'azienda ma i dati non sono disponibili. Non sempre il prodotto che viene più

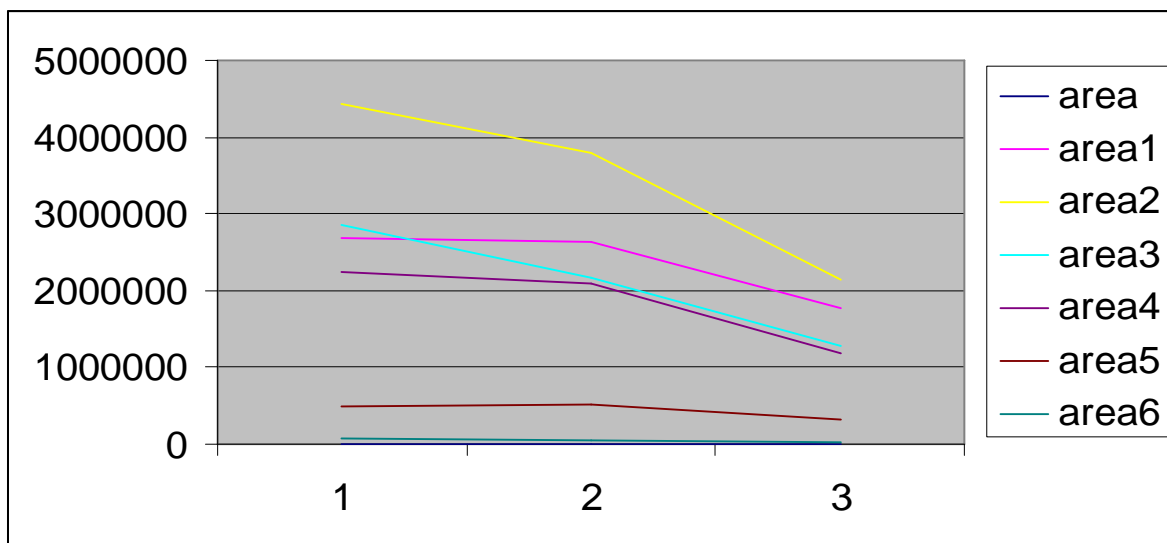
venduto corrisponde a quello che effettivamente dà una maggiore resa all'azienda in termini economici.

L'analisi dei dati per clienti si è anche questa rivelata poco sintetica in quanto mediamente, ho calcolato, l'azienda effettua circa duemila "operazioni" mensili con circa duecento clienti diversi.

Il database fornisce cioè "classifiche" "riepilogative di circa 25 pagine al mese nelle quali l'ordine con cui sono elencati i dati è quello del codice cliente. Stando così le cose, scoprire tra i vari clienti, coloro che spendono di più e analizzare a quale canale distributivo appartengono è un compito immane.

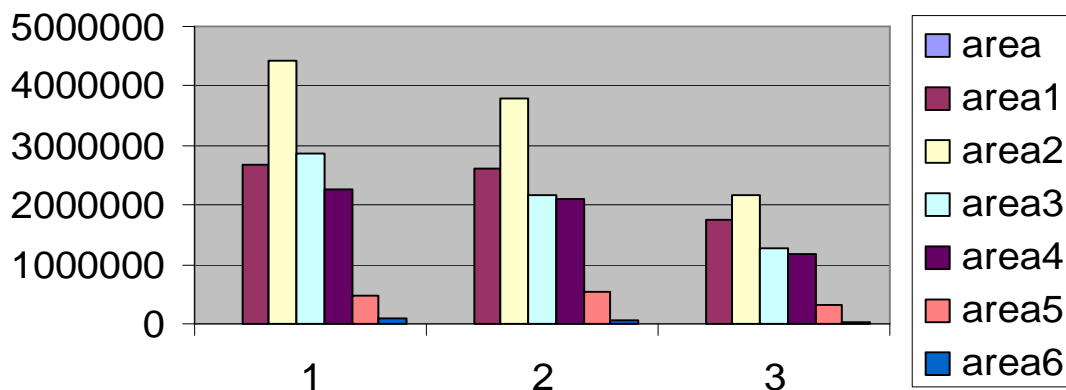
Dati delle vendite per area geografica (area Nielsen)

Area 1=Nord-Ovest **Area2**=Nord-Est **Area3**=Centro Italia **Area4**=Sud-Italia
Area5=Sicilia **Area6**=Sardegna



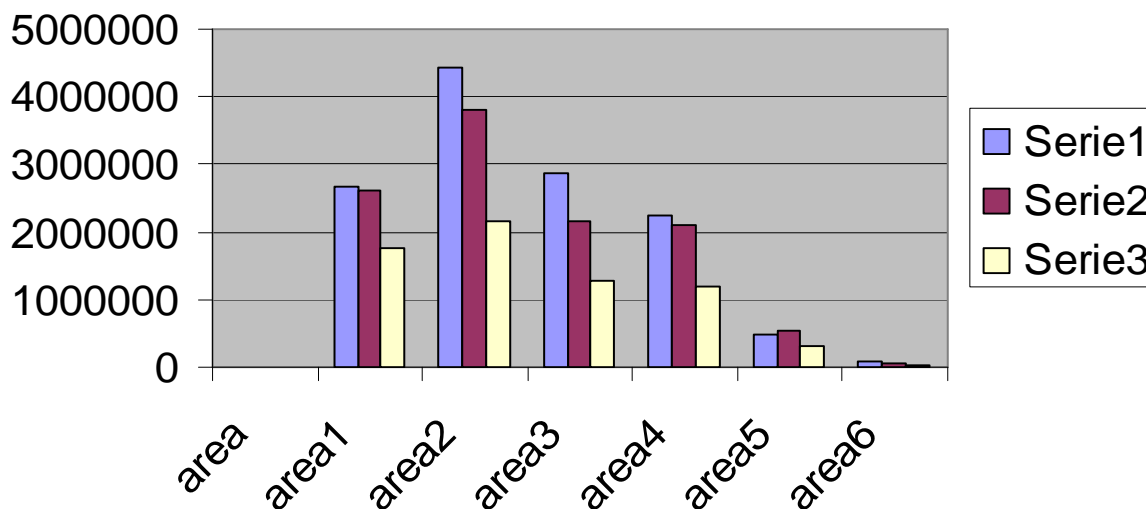
oppure li posso visualizzare così:

vendite annuali per area geografica



dove l'anno 2003=1 2004=2 2005=3

vendite per area nel periodo 2003-2004-2005



4.2 APPROCCIO CLASSICO

Ossia le componenti “interessanti” di una serie sono deterministiche (in funzione del tempo)

Calcolo la previsione mediante la regressione lineare.

1) regressione

faccio la regressione lineare in funzione del tempo con questi coefficienti.

intercetta $\alpha=1120377$

pendenza $\beta=11310,29$

Ossia applico la formula

$y=1120377+11310,29*x$ con $x=1,2,\dots,36$ e con i residui $e_t \approx WN(0, \sigma^2)$.

e con Excel ottengo questi risultati :

OUTPUT RIEPILOGO

	Coefficienti	Errore standard	Stat t	Valore di significatività	Inferiore 95%	Superiore 95%	Inferiore 95,0%	Superiore 95,0%
Intercetta	1120377	117580,6	9,52858	3,96E-11	881424,1	1359329	881424,1	1359329
Variabile X								
1	11310,29	5541,779	2,040913	0,049084	48,04868	22572,53	48,04868	22572,53

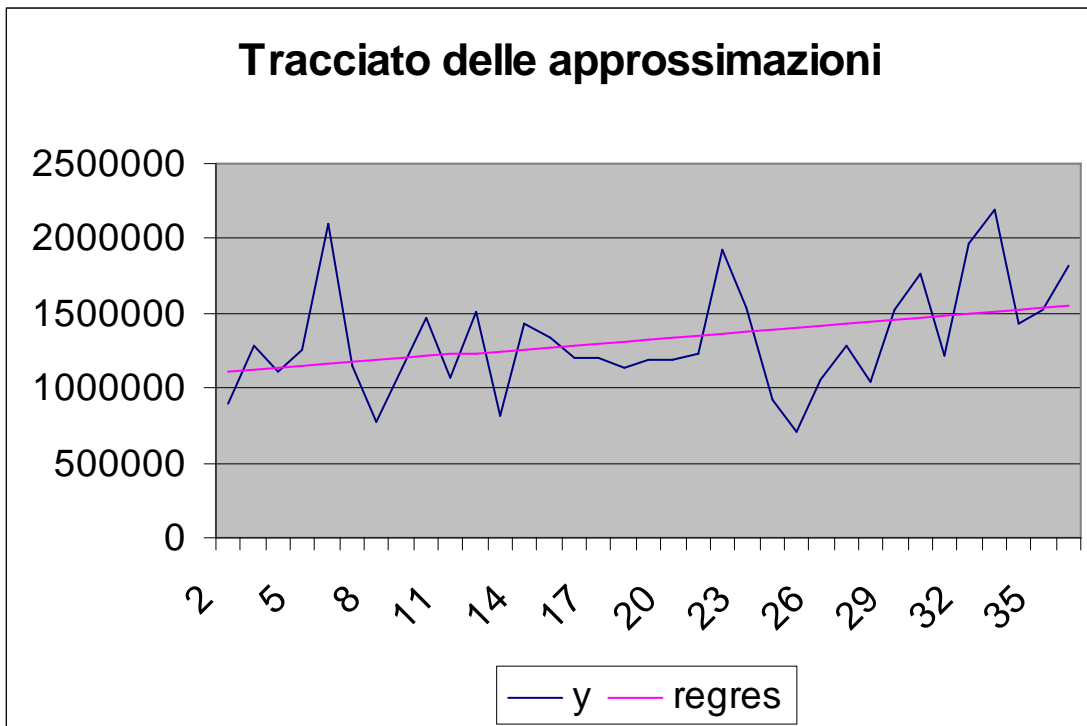
OUTPUT RIEPILOGO

Statistica della regressione	
R multiplo	0,330362
R al quadrato	0,109139
R al quadrato corretto	0,082937
Errore standard	345417,8
Osservazioni	36

ANALISI VARIANZA

	gdl	SQ	MQ	F	Significatività F
Regressione	1	4,97E+11	4,97E+11	4,165328	0,049084
Residuo	34	4,06E+12	1,19E+11		
Totale	35	4,55E+12			

Questa sotto è la loro rappresentazione grafica dove : y =serie osservata , **regres**=serie ottenuta con la regressione lineare.



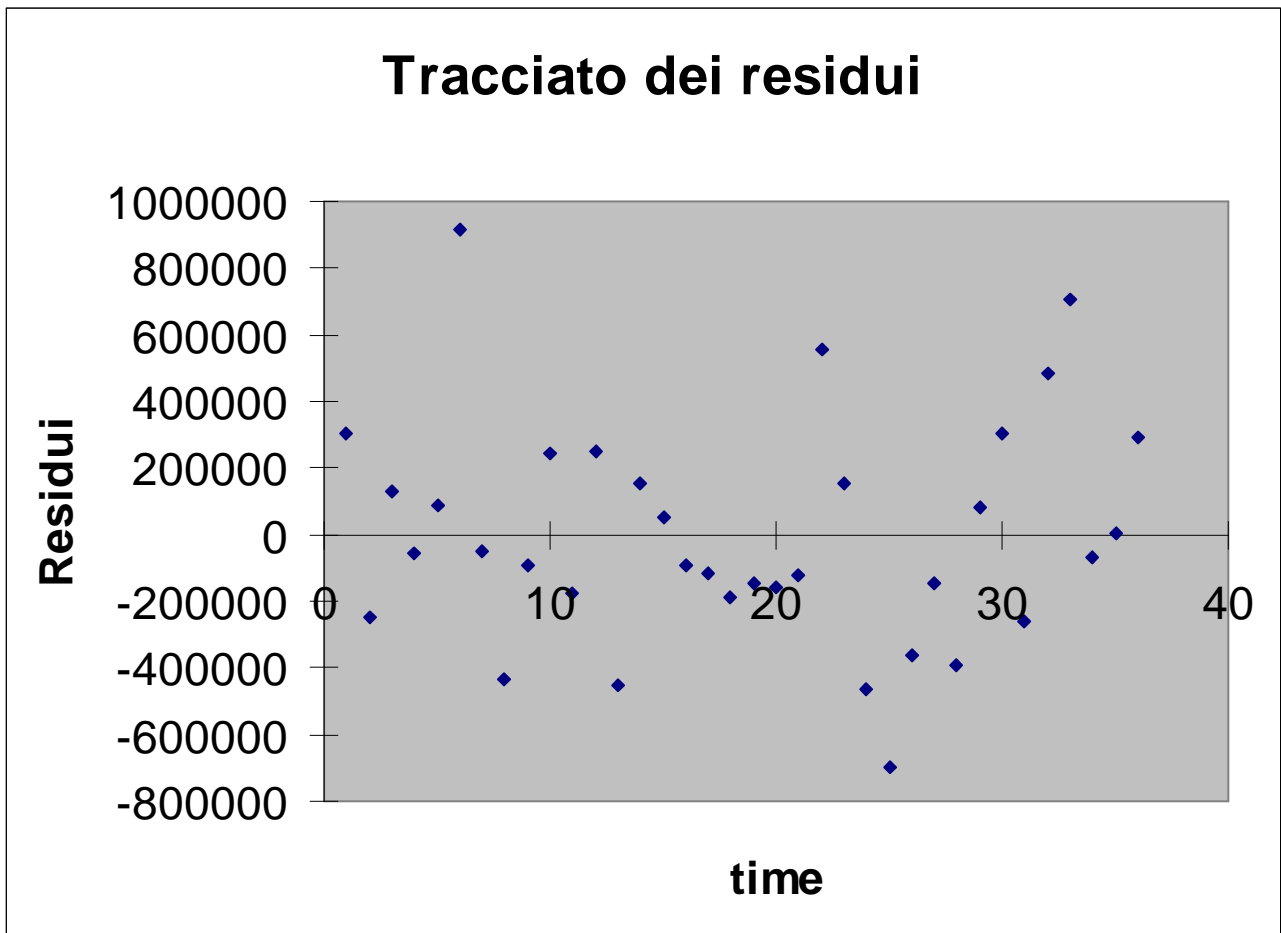
questo sotto è il grafico che rappresenta l'andamento dei residui ottenuti.

I residui sono dati dalla seguente formula:

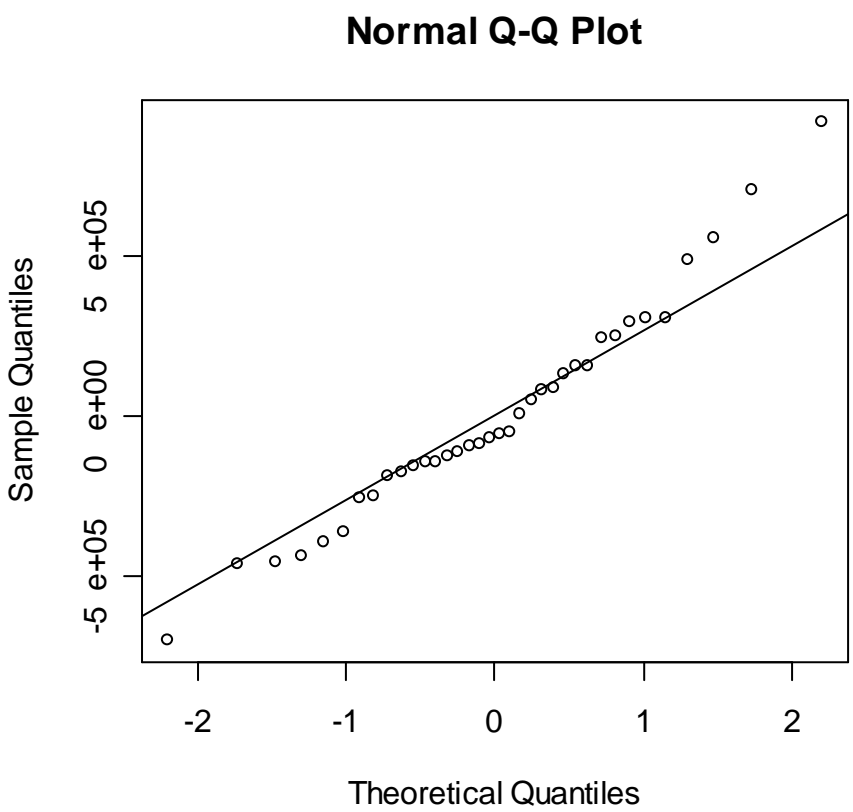
$$\text{residui} = y - \hat{y}$$

Noto che i residui oscillano attorno allo zero la loro distribuzione è abbastanza soddisfacente.

Per sicurezza ho analizzato i residui con R dove il plot e il qqnorm confermano questa ipotesi di non violazione dell'omoschedaticità dei residui.



Ho analizzato i residui con R dove il qqnorm confermano questa ipotesi di non violazione dell'omoschedaticità dei residui.



Da questa formula:

$$y=1120377+11310,29*x$$

posso ottenere

$$\hat{y}_{37}=1120377+11310,29*37 =1538857,286$$

$$\hat{y}_{38}=1120377+11310,29*38= 1538857,286$$

$$\hat{y}_{39}=1120377+11310,29*39=1561477,866$$

$$\hat{y}_{40}=1120377+11310,29*40=1561477,866$$

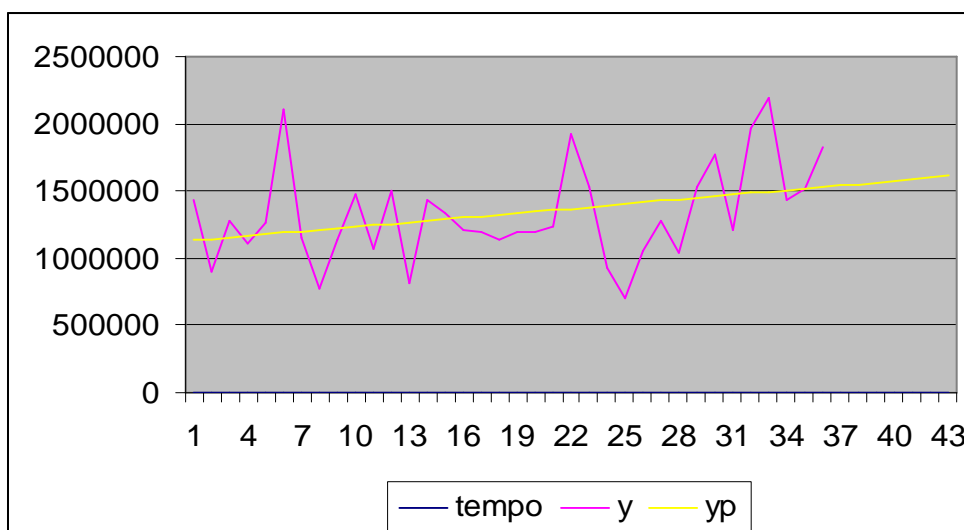
cioè riesco anche ad ottenere delle previsioni

$$T=37 \ y_regr=1538857,286 \ t=38 \ y_reg= 1538857,286 \ t=39$$

$$y_reg=1561477,866$$

$$T=40 \ y_reg=1561477,866$$

Il cui grafico è riportato qui sotto.



Possiamo affermare che il trend è previsto in rialzo.

Naturalmente più le previsioni si allontanano dall'odierno ossia più mi proietto nel futuro e maggiormente esiste la possibilità che il margine di errore tra il previsto e il realizzato aumenti. Quindi ritengo affidabile la previsione di gennaio in misura maggiore di quella di febbraio.

2) lisciamiento esponenziale

METODO STAGIONALE MOLTIPLICATIVO

L'ipotesi di base è che la previsione in prossimità di n possa essere fornita dalla retta di regressione:

$$[\bar{y}_n + T_n(t - n)]S_n$$

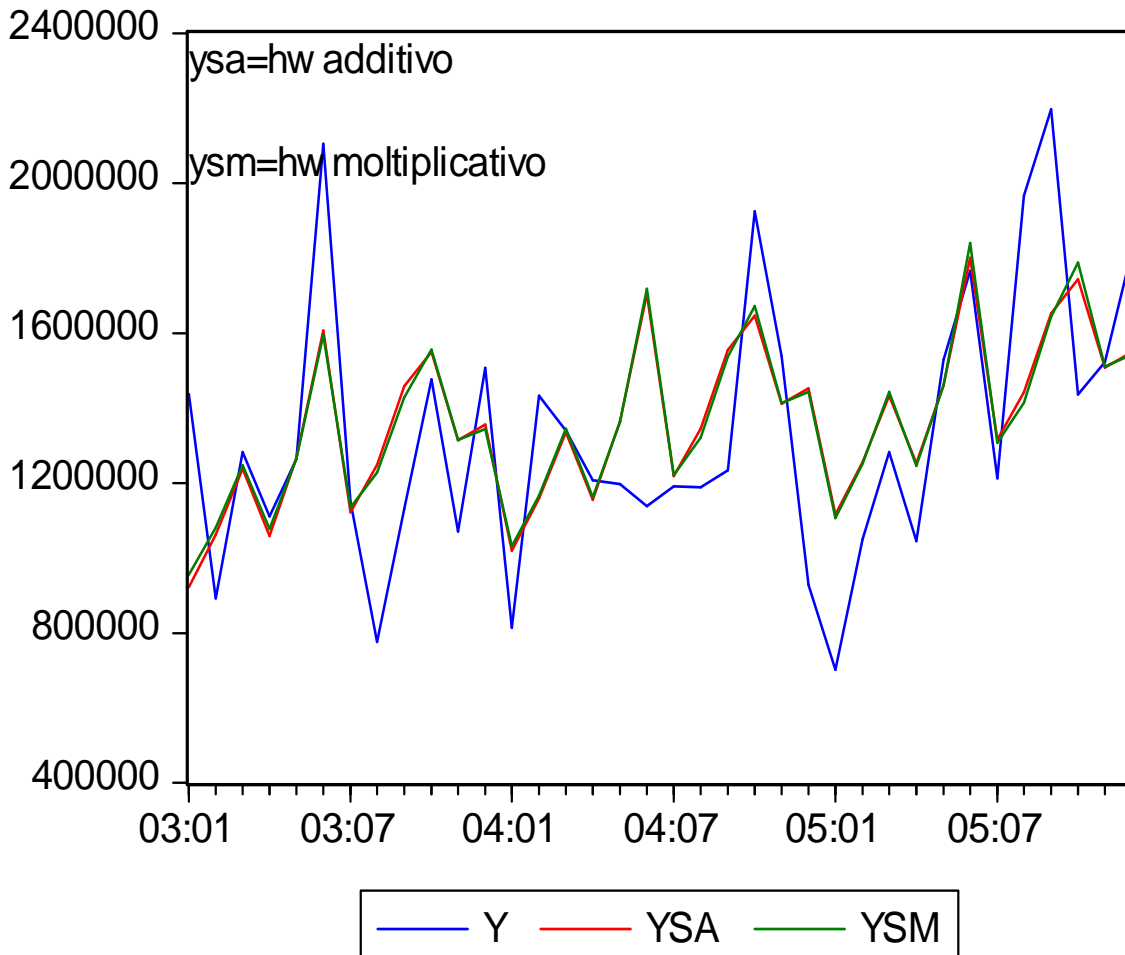
Le stime di \bar{y}_n , T_n e S_n sono date dalle seguenti formule di aggiornamento :

$$\begin{aligned}\bar{y}_n &= \alpha(\bar{y}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha) \frac{y_n}{\hat{S}_{n-s}}, 0 < \alpha < 1 \\ \hat{T}_n &= \beta \hat{T}_{n-1} + (1 - \beta)(\bar{y}_n - \bar{y}_{n-1}), 0 < \beta < 1 \\ \hat{S}_n &= \gamma \hat{S}_{n-s} + (1 - \gamma) \frac{y_n}{y_n}, 0 < \gamma < 1\end{aligned}$$

La previsione è data da:

$$\begin{aligned}F_{n,k} &= [\bar{y}_n + \hat{T}_n k] \hat{S}_{n+k-s}, 1 \leq k \leq s \\ F_{n,k} &= [\bar{y}_n + \hat{T}_n k] \hat{S}_{n+k-2s}, 1 \leq k \leq 2s\end{aligned}$$

Confronto graficamente il metodo di Holt-Winters stagionale moltiplicativo e additivo così da poter già a prima vista avere un'idea di quale tra i due possa essere il metodo migliore.



Da questa osservazione grafica preliminare non noto eccessive differenze tra i due metodi.

Ho usato il metodo hw-additivo (chiamato **ysa**) e il metodo hw-moltiplicativo (definito con **ysm**).

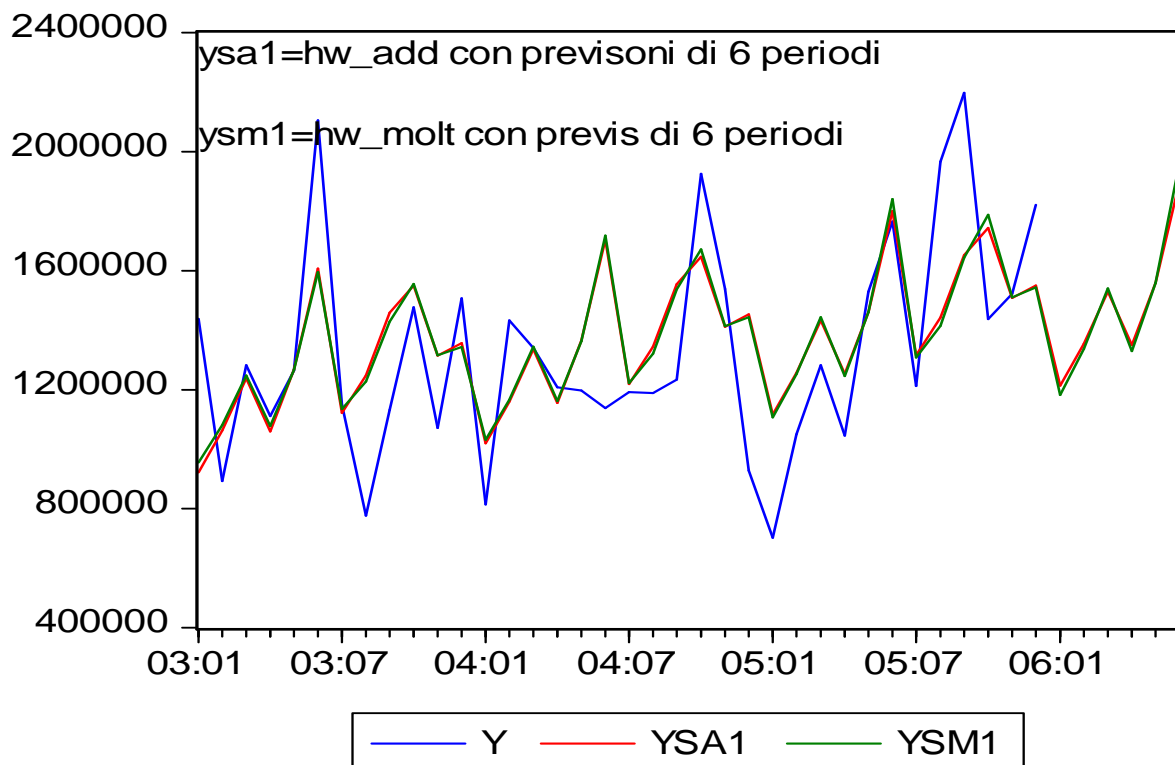
Il valore di alpha beta e gamma ho lasciato che fosse **Eviews** (4.1) a sceglierli.

Di seguito presento le principali statistiche delle tre serie.

Sample: 2003:01 2005:12

	Y	YSA	YSM
Mean	1329617.	1363833.	1365826.
Median	1248381.	1350409.	1344385.
Maximum	2197204.	1800455.	1839820.
Minimum	702696.0	921975.6	955262.8
Std. Dev.	360699.2	215873.3	216713.4
Skewness	0.628744	0.068394	0.278621
Kurtosis	2.993897	2.341036	2.469392
Jarque-Bera	2.371967	0.679418	0.888093
Probability	0.305446	0.711978	0.641436
Sum	47866209	49097970	49169724
Sum Sq. Dev.	4.55E+12	1.63E+12	1.64E+12

Ora modifico range e sample al fine di ottenere delle previsioni (portandolo a giugno del 2006).



Indico con y=serie osservata, ysa1=serie ysa a cui aggiungo 6 mesi di previsione e

con ysm1=serie ysm a cui aggiungo 6 mesi di previsione.

Queste sono le principali statistiche ottenute dalle serie “estese” (cioè che includono anche le previsioni):

Date: 06/04/06
Time: 18:21
Sample: 2003:01 2006:06

	Y	YSA1	YSM1
Mean	1329617.	1363833.	1365826.
Median	1248381.	1350409.	1344385.
Maximum	2197204.	1800455.	1839820.
Minimum	702696.0	921975.6	955262.8
Std. Dev.	360699.2	215873.3	216713.4
Skewness	0.628744	0.068394	0.278621
Kurtosis	2.993897	2.341036	2.469392
Jarque-Bera Probability	2.371967 0.305446	0.679418 0.711978	0.888093 0.641436
Sum	47866209	49097970	49169724
Sum Sq. Dev.	4.55E+12	1.63E+12	1.64E+12
Observations	36	36	36

Riporto l'mse per ciascuno dei due metodi così da giustificare la scelta di uno o dell'altro :

msem= 286501.0

msea= 286973.6

Scelgo il metodo di Holt-Winters stagionale moltiplicativo in quanto minimizza l'errore quadratico medio.

Indici dell'errore medio di previsione

Volendo giudicare l'accuratezza complessiva delle previsioni nell'intervallo dei tempi da 1 a 36 ,è ragionevole proporre una funzione monotona decrescente che assume valore zero se e solo se gli errori sono tutti nulli. Risponde a tali requisiti la media potenziata di ordine s degli errori di previsione presi in valore assoluto.

$$I_s = \left(\frac{1}{n} \sum |e_t|^s \right)^{\frac{1}{s}}$$

In particolare per $s=1$ si ottiene l'errore medio assoluto di previsione

$$EMA = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Questo indice viene spesso detto errore medio di previsione

$$EM = \frac{1}{n} \sum_{t=1}^n e_t$$

Dalla teoria apprendo che se i due indici sono abbastanza vicini (in valore assoluto) vi è sistematicità negli errori di previsione caratterizzati prevalentemente dallo stesso segno. Se invece EM è di entità ridotta rispetto a EMA gli errori di previsione non presentano sistematicità particolari.

Con R mi calcolo i due valori precedenti

$$EMA = 224730.7$$

$$EM = -36208.74$$

Devo concludere che non vi è sistematicità negli errori.

Ulteriori conferme le posso avere calcolando la media quadratica degli errori di previsione ponendo $s=2$

$$\text{ossia } Is = \left(\frac{1}{n} \sum_{t=1}^n |e_t|^{\frac{1}{2}} \right)$$

$$(\text{mse additivo})^2 = 82353822220.9$$

$$(\text{mse moltiplicativo})^2 = 82082836902.7$$

A questo punto con R scompongo il quadrato della media quadratica degli errori. La scomposizione la effettuo solamente sulla serie lisciata con HW moltiplicativo in quanto, come già detto, il suo MSE risulta inferiore all' MSE ottenuto dalla serie lisciata con HW additivo.

$$\text{Ossia decompongo } MSE^2 = ES + EV + EC \quad \text{con } ES + EV + EC = 1$$

I tre addendi (ES,EV,EC) rappresentano le quote parti dell'errore complessivo di previsione ,misurato tramite MSE dovute rispettivamente:

- a diversa media dei valori previsti e realizzati
- a diversa variabilità dei valori previsti e realizzati
- a imperfetta correlazione lineare fra i valori previsti e quelli realizzati.

Definisco:

$$ES = \frac{(\bar{p} - \bar{r})^2}{MSE^2}$$

$$EV = \frac{(\sigma_p - \sigma_r)^2}{MSE^2}$$

$$EC = \frac{(1 - \rho_{pr})}{MSE^2}$$

con $ES+EV+EC=1$

La frazione di errore dovuta a diversa media, è particolarmente grave poiché indica che non si è stati in grado di prevedere la tendenza del fenomeno. Nel caso esaminato risulta pari a $ES=0.01597255$ quindi non grave.

Ho calcolato l'errore nelle variabilità $EV = 0.2525729$.La frazione di errore imputabile a diversa variabilità segnala che non si sono previste esattamente le oscillazioni dei valori effettivi attorno alla linea di trend.

La quota parte di errore dovuta all'imperfetta correlazione , $EC = 0.7595696$, appare di minore gravità rispetto alle altre poiché essa può venire attribuita a fattori accidentali che si manifestano in maniera differente nei dati previsti e nei dati realizzati.

Volendo giudicare l'accuratezza complessiva delle previsioni nell'intervallo dei tempi da 1 a 36 ,è ragionevole proporre una funzione monotona decrescente che assume valore zero se e solo se gli errori sono tutti nulli.Risponde a tali requisiti la media potenziata di ordine s degli errori di previsione presi in valore assoluto.

$$I_s = \left(\frac{1}{n} \sum |e_t|^s \right)^{\frac{1}{s}}$$

In particolare per $s=1$ si ottiene l'*errore medio assoluto di previsione*

$$EMA = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Questo indice viene spesso detto errore medio di previsione

$$EM = \frac{1}{n} \sum_{t=1}^n e_t$$

Dalla teoria apprendo che se i due indici sono abbastanza vicini (in valore assoluto) vi è sistematicità negli errori di previsione caratterizzati prevalentemente dallo stesso segno. Se invece EM è di entità ridotta rispetto a EMA gli errori di previsione non presentano sistematicità particolari.

Con R mi calcolo i due valori precedenti

$$EMA = 224730.7$$

$$EM = -36208.74$$

Devo concludere che non vi è sistematicità negli errori.

Previsione dei punti di svolta

E' possibile giudicare la validità di una serie di valori previsti anche dai suoi punti di svolta.

Come già detto nella teoria per *l'indice relativo degli errori di prima specie* intendiamo il rapporto tra il numero di errori di prima specie (ossia :punto di svolta previsto e non realizzato) e il numero totale di punti di svolta previsti.

$$E_1 = \frac{n_{12}}{n_1} = 0.35$$

Per indice relativo degli errori di seconda specie intendiamo il rapporto tra il numero

di errori di seconda specie (cioè: punto di svolta non previsto ma verificatosi) e il numero totale di punti di svolta effettivi

$$E_2 = \frac{n_{21}}{n} = 0.18$$

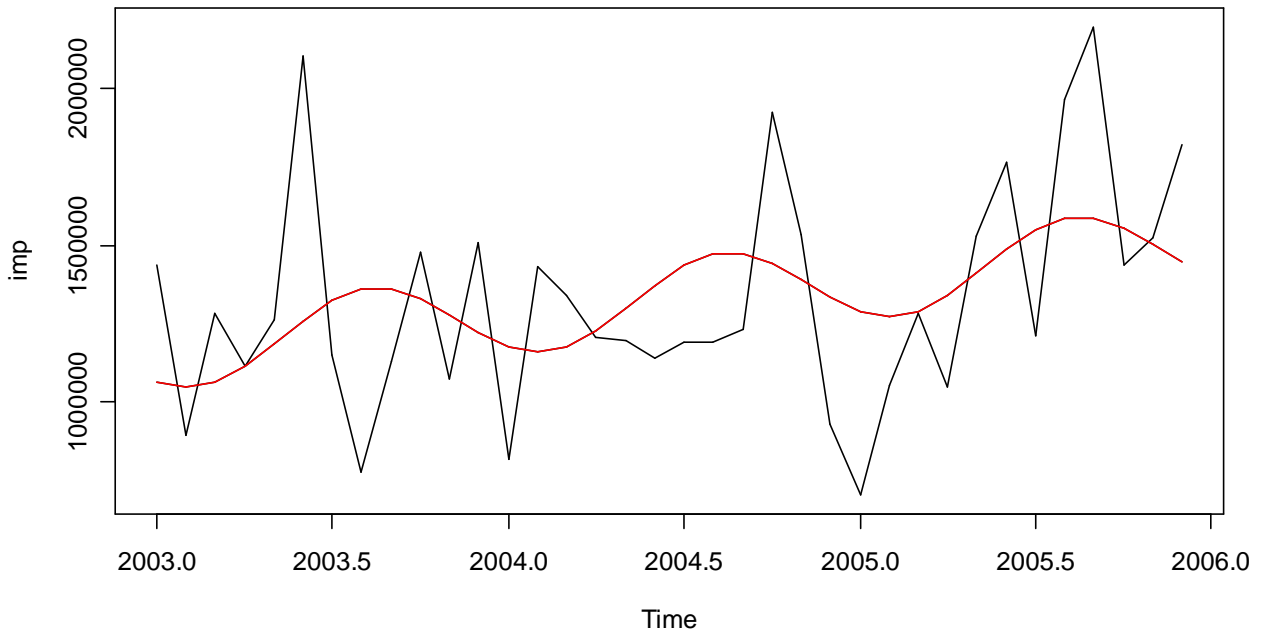
Questi due indici, il cui valore varia nell'intervallo $[0,1]$, non è eccessivo e quindi posso concludere che la previsione è abbastanza buona.

Uso di funzioni Trigonometriche

Un altro metodo per stimare il trend e la stagionalità può essere effettuato mediante la regressione con utilizzo della funzione $\text{lm}()$. Il modello da stimare è il seguente ed esprime il trend in funzione del tempo e la stagionalità come somma di funzioni trigonometriche :

$$Y_t = f(t) + \alpha \cos\left(\frac{2\pi}{12}\right) + \beta \sin\left(\frac{2\pi}{12}\right) + \varepsilon_t$$

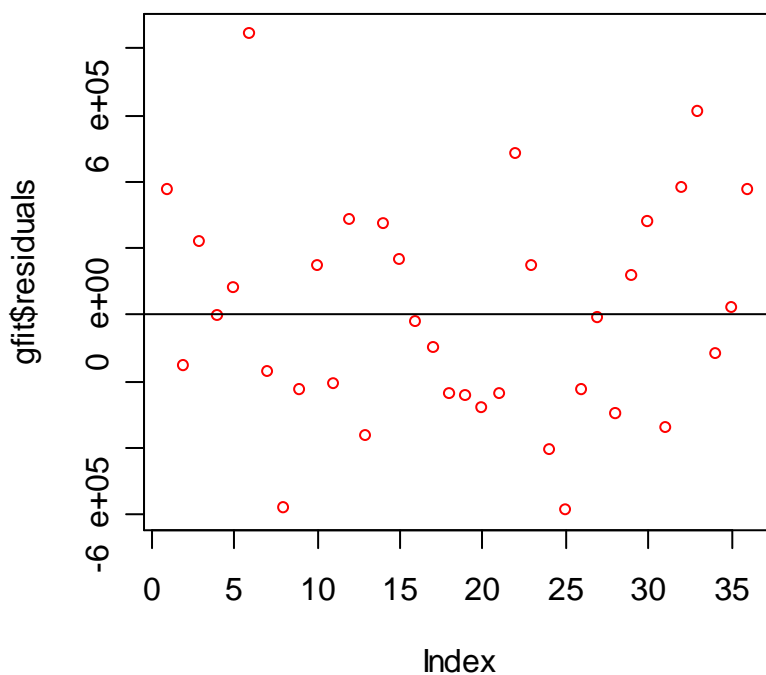
Stima del trend e stagionalità con funzioni sin e cos



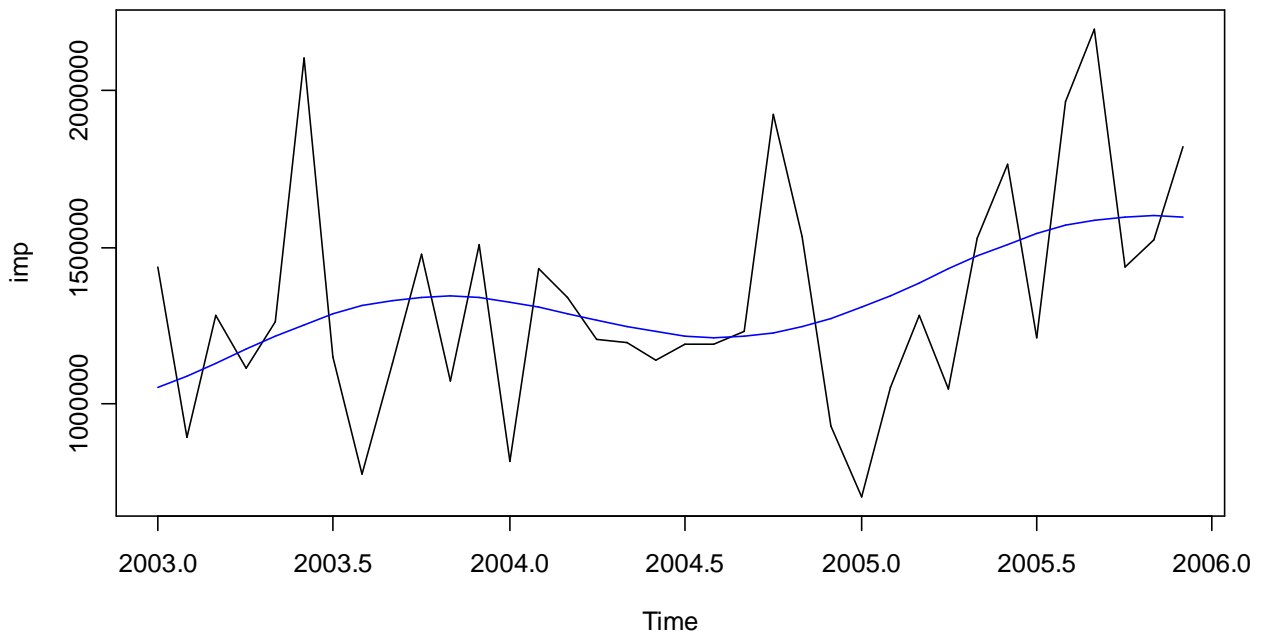
Dal grafico risalta subito che questa non è una buona stima anche se possiamo dire viene evidenziato che il trend è crescente nel tempo.

Mediante la regressione : $\text{imp.vec} \sim t + \cos.t + \sin.t$

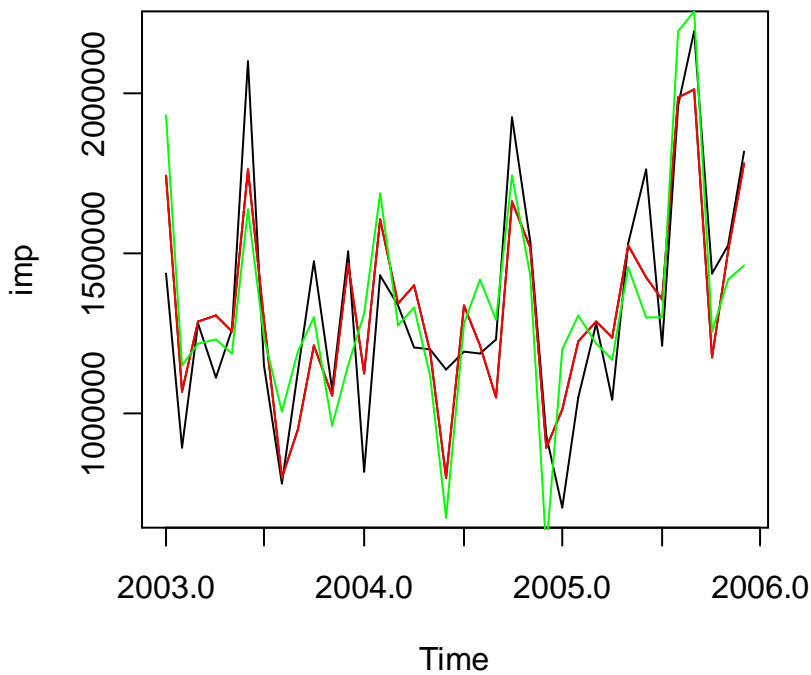
L' R^2 è molto basso (0.1717) e indice di un cattivo adattamento .



anche apportando modifiche non ho buoni risultati con l' R^2

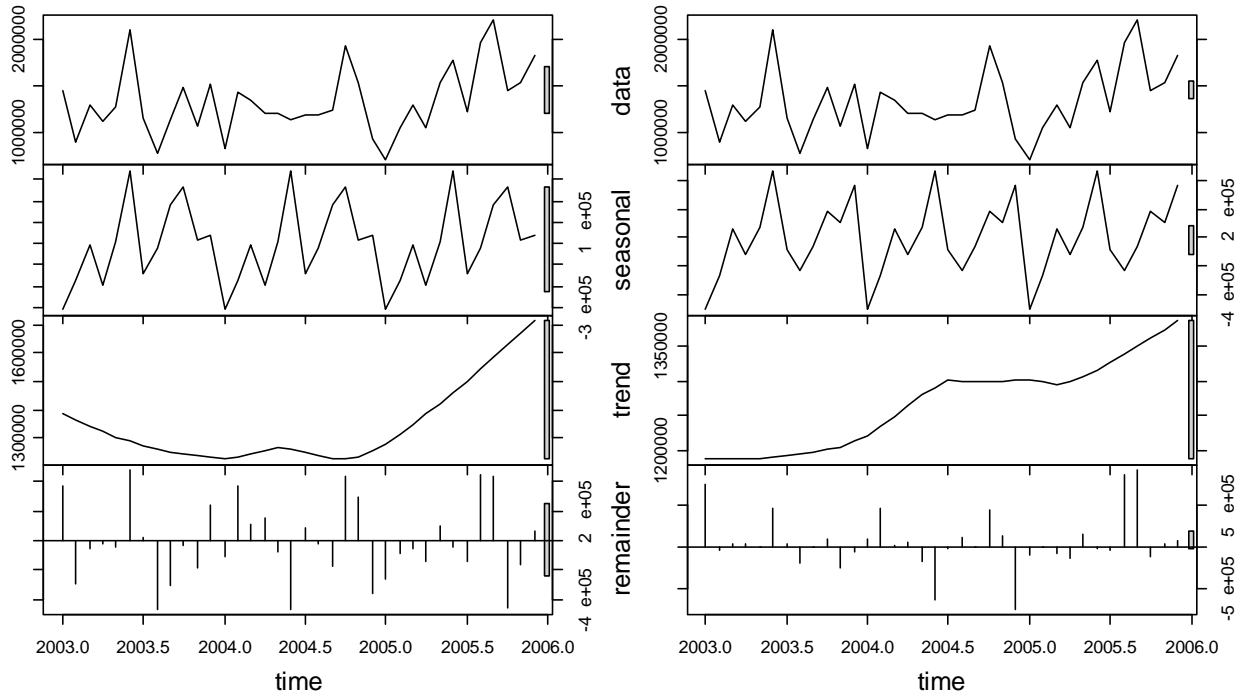


Possiamo osservare in questo grafico un confronto tra la serie originale e la serie destagionalizzata con la stima robusta (tratteggiata in verde) e la serie destagionalizzata con la stima non robusta (tratteggiata con la linea rossa)

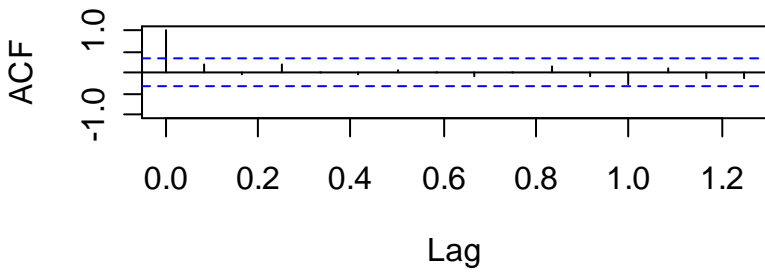


Sotto vediamo i plot relativi

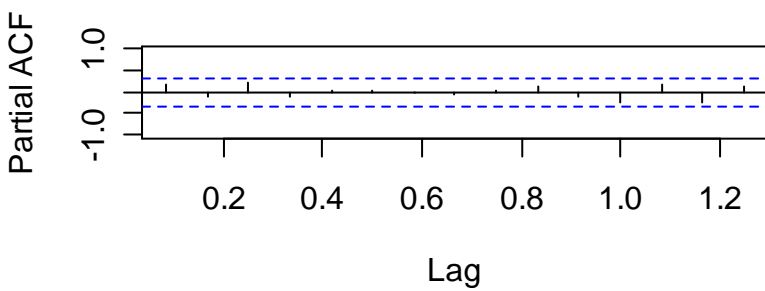
st(imp, s.w = "per", robust = FALSE / TRUE)



Series dimp.stl

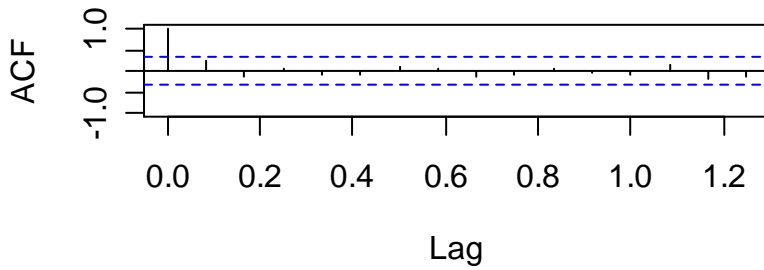


Series dimp.stl

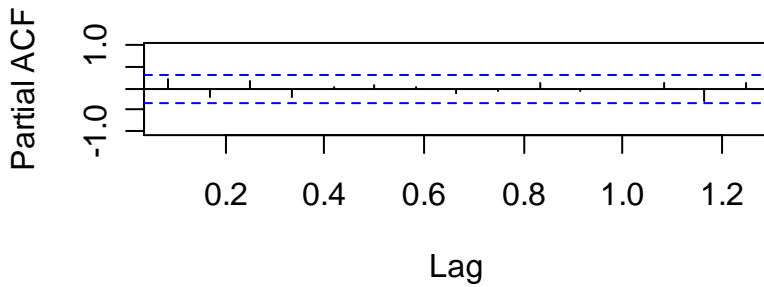


Questi sono i correlogrammi della serie destagionalizzata con la componente stagionale gaussiana

Series dimp.rstl



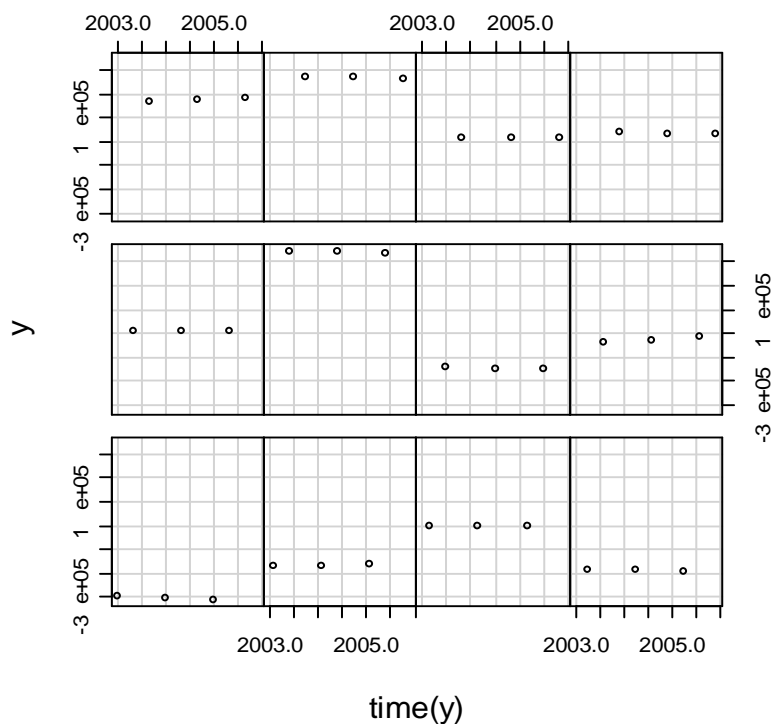
Series dimp.rstl



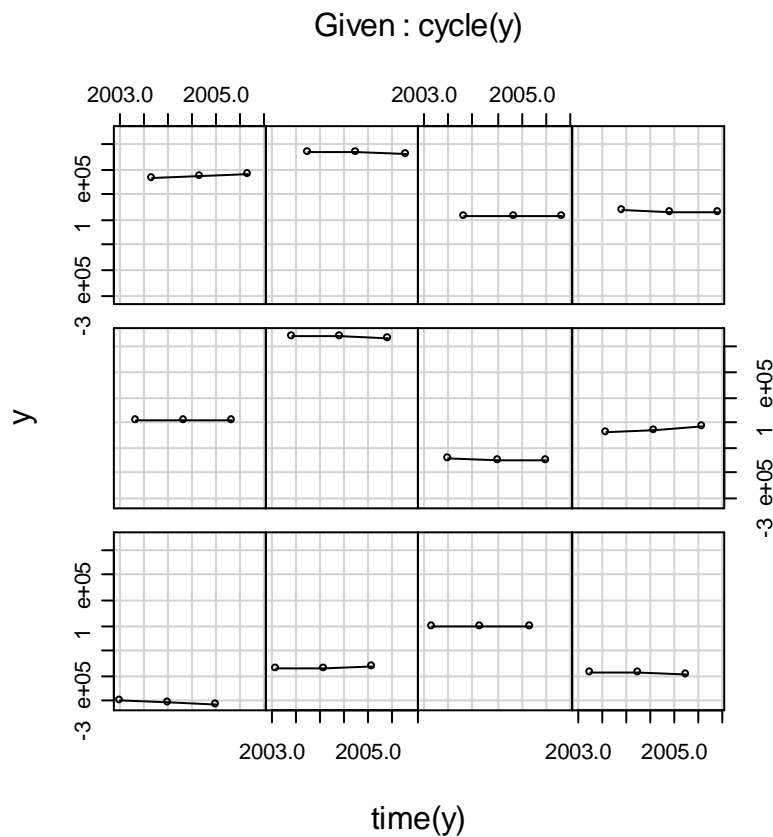
questi sopra sono correlogrammi della serie destagionalizzata con la serie stimata con la stima robusta.

E' necessario e opportuno verificare che la stagionalità nel corso degli anni non presenti qualche trend. Nel caso succedesse si parla allora di trend-stagionalità.

Given : cycle(y)



Dal grafico emerge che la stagionalità non è costante da un anno all'altro ma presenta un trend all'incirca lineare per quasi tutti i mesi dell'anno. In tale circostanza è necessario stimare questo trend –stagionalità attraverso la regressione lineare (detto modello di stagionalità variabile). Se la stagionalità oscilla intorno ad un dato valore costante, ed è probabile che tali oscillazioni siano di natura casuale, una stima della stagionalità di quel mese può essere ottenuta con la media delle componenti stagionali che quel mese ha presentato nel corso dei vari anni (detto modello di stagionalità costante)



Come si vede il trend della stagionalità può essere ben stimato con una retta, e una stima della componente stagionale per i prossimi dodici mesi del 2006 riassunta in questa tabella

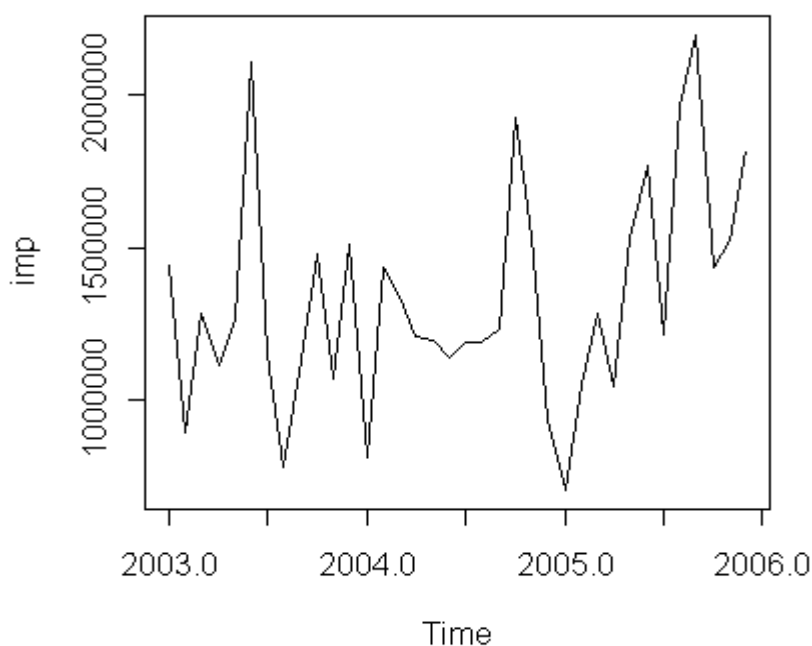
	Jan	Feb	Mar	Apr	May	Jun
2006	-319660.741	-163648.185	-2612.508	-192892.327	14012.729	328382.891
	Jul	Aug	Sep	Oct	Nov	Dec
2006	-149666.518	-5431.823	193340.435	257516.771	14587.281	30535.825

4.3 APPROCCIO MODERNO

L'approccio moderno, come già detto, ha la componente deterministica che è stata rimossa.

Ossia con l'approccio moderno ho rimosso trend, ciclo e stagionalità.

Questo è un diagramma della serie mensile del fatturato di Asolo Dolce dea 2003 al 2005 fatta con R 1.9.1



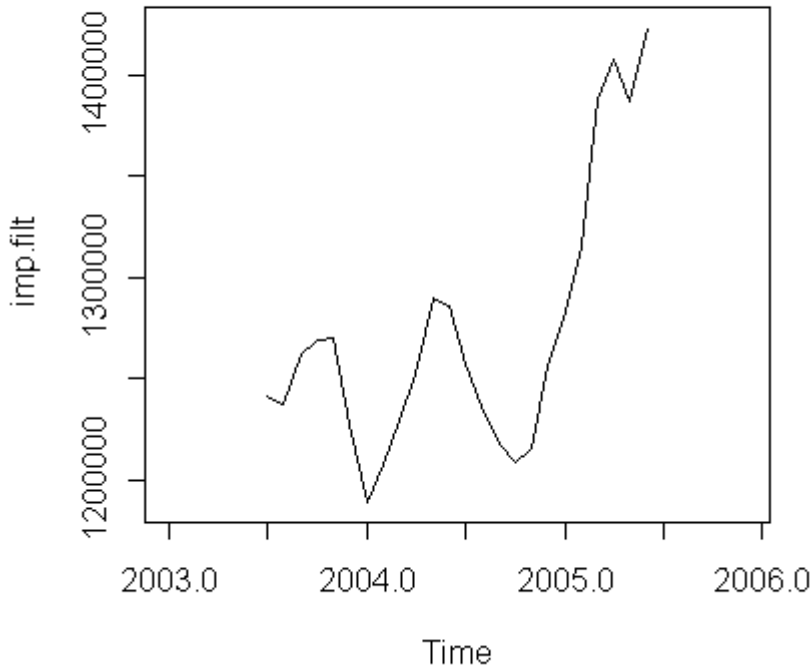
Provo a rimuovere la parte deterministica.

Medie Mobili

R mette a disposizione la libreria **ast** con le due funzioni “filter” e “sfilter”.

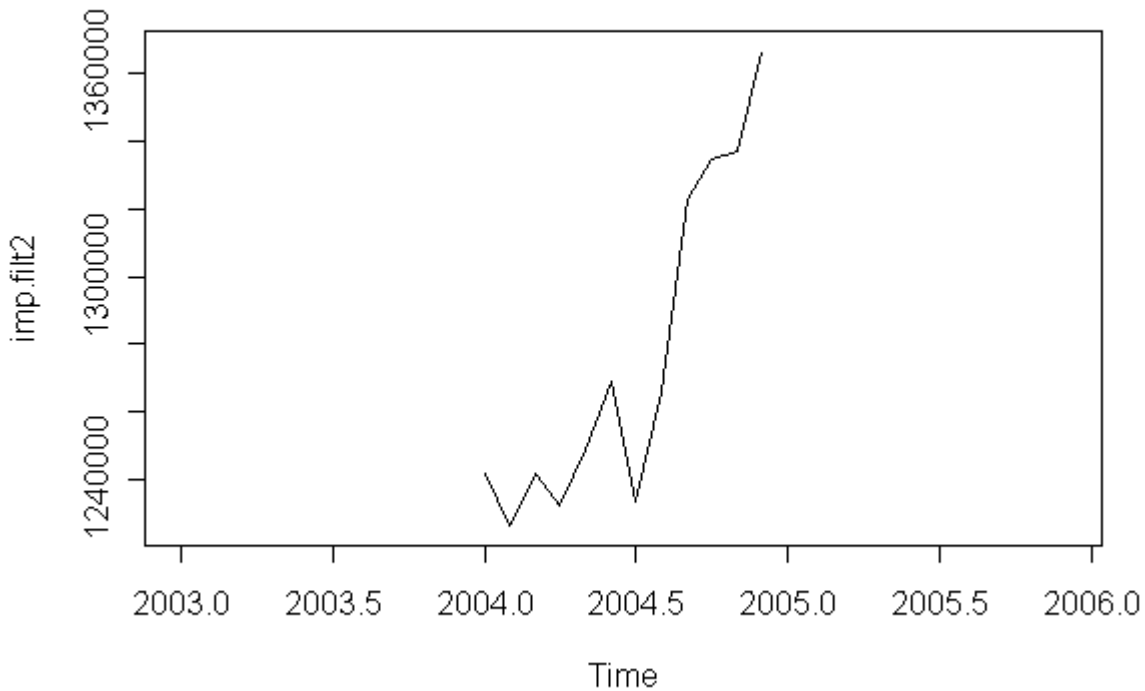
La funzione **sfilter** applica una media mobile ponderata a 13 termini per eliminare la stagionalità e mettere in risalto solo la componente di fondo.

Trend stimato con la media mobile a 13 term



la funzione **filter** è una m.m. con pesi di valore “1/25” ripetuti 25 volte (essendo 25 l’ordine della m.m.).Quindi è una m.m. semplice (ossia i pesi sono sempre gli stessi, non variano a seconda del dato preso in esame)

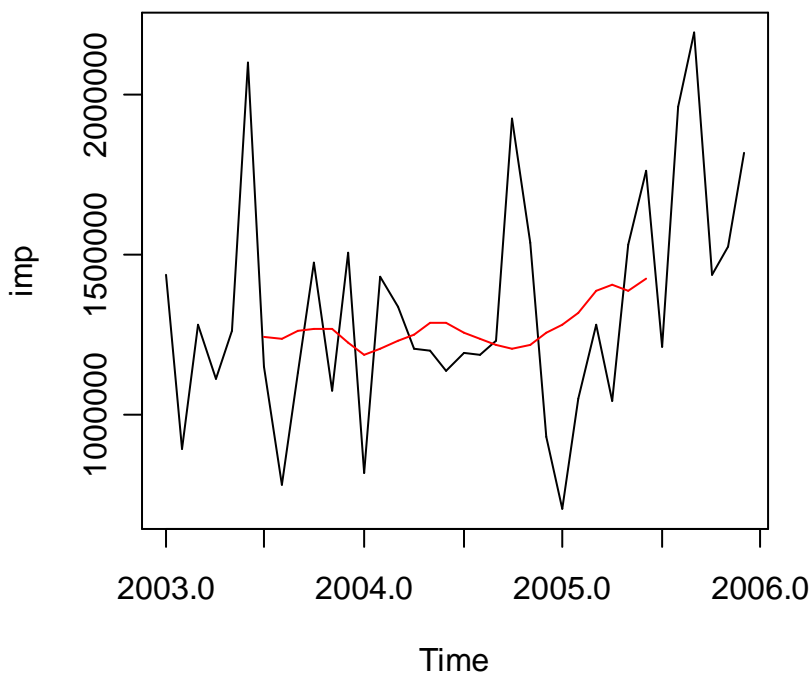
Trend stimato con media mobile semplice-funzione 'filter'



Per eliminare la componente deterministica devo scegliere una media mobile che non mi elimini troppe mensilità ,dato che già ne ho pochine a disposizione.

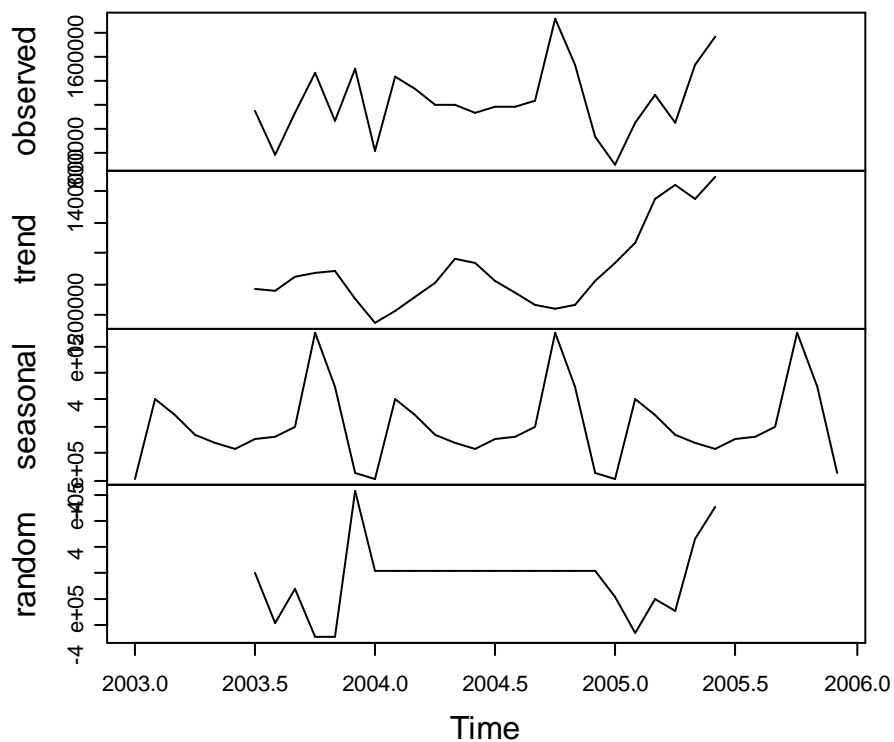
Provo la funzione **decompose**.La funzione in oggetto utilizza per decomporre la serie le medie mobili (che io abbrevierò in “m.m.”).Qui ne usa una di ordine tredici.

Vediamo il grafico della decomposizione appena attuata.



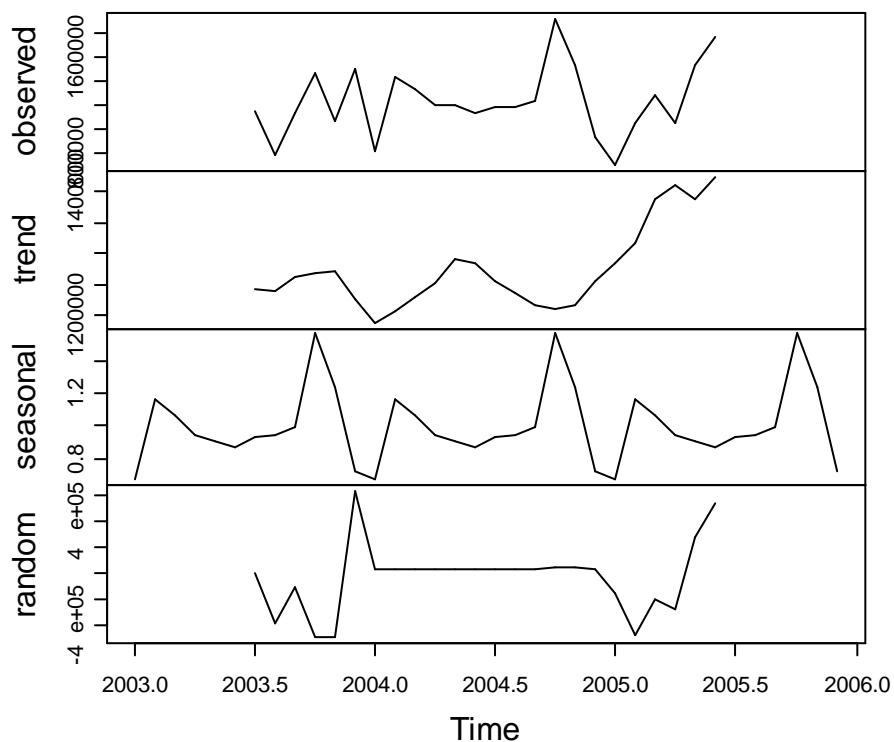
Come si può notare dalla linea rossa (trend stimato) ho perso un numero notevole di osservazioni.

Decomposition of additive time series



Ora provo decomporre la serie con il metodo moltiplicativo.

Decomposition of multiplicative time series



Io proseguo con i dati ottenuti con il metodo additivo.

Si è supposto che la componente erratica fosse distribuita normalmente con media

pari a zero, varianza costante e che non vi fosse autocorrelazione. Tali ipotesi alla base del modello vanno verificate con opportuni test statistici detti test di specificazione del modello. Il venir meno a questa ipotesi potrebbe inficiare la validità del modello adottato e far propendere per altri modelli più complessi oppure di intervenire sulla serie con delle trasformazioni atte, ad esempio, a stabilizzare la varianza oppure ad eliminare l'autocorrelazione.

Voglio verificare il valore della media degli errori ottenuti con la funzione decompose. Effettuo il test t: $\text{test.t} = -0.681008$

e calcolo il p-value $\text{pt}(\text{test.t}, n-1, \text{lower.tail}=\text{F}) = 0.7486671$

oppure considero il valore soglia del test per un livello di significatività al 99% :

$$\text{qt}(.99, n-1) = 2.499867$$

Questo risultato ci consente di concludere che la media degli errori non è significativamente diversa da zero.

Prima di procedere a verificare che i residui si distribuiscano secondo una variabile aleatoria normale, andiamo ad esaminare i residui standardizzati al fine di individuare i valori anomali. E' preferibile operare sui residui standardizzati per avere a che fare con numeri puri.

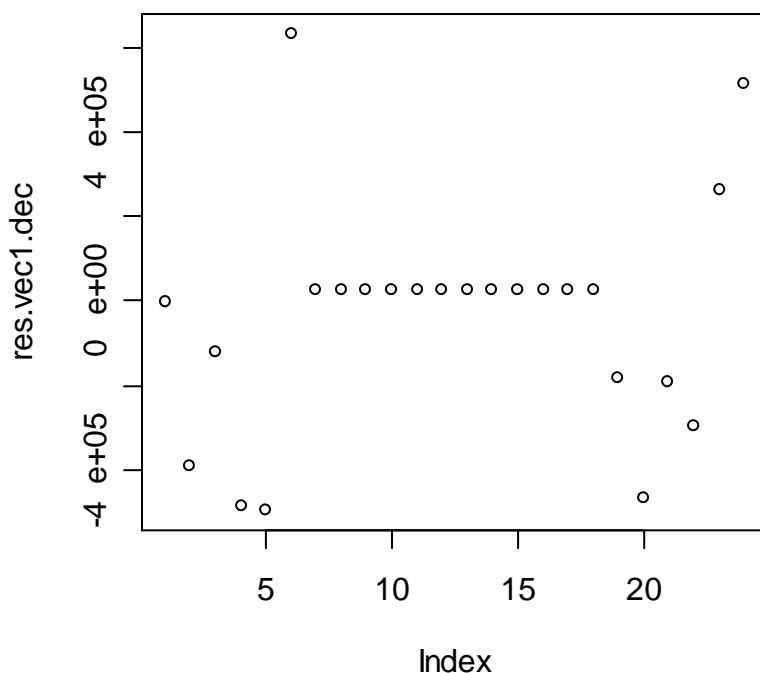
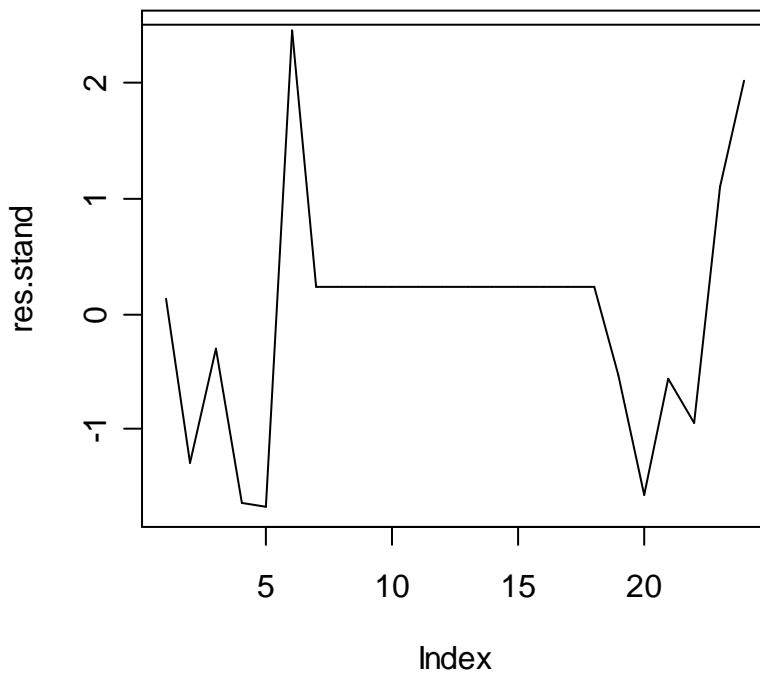


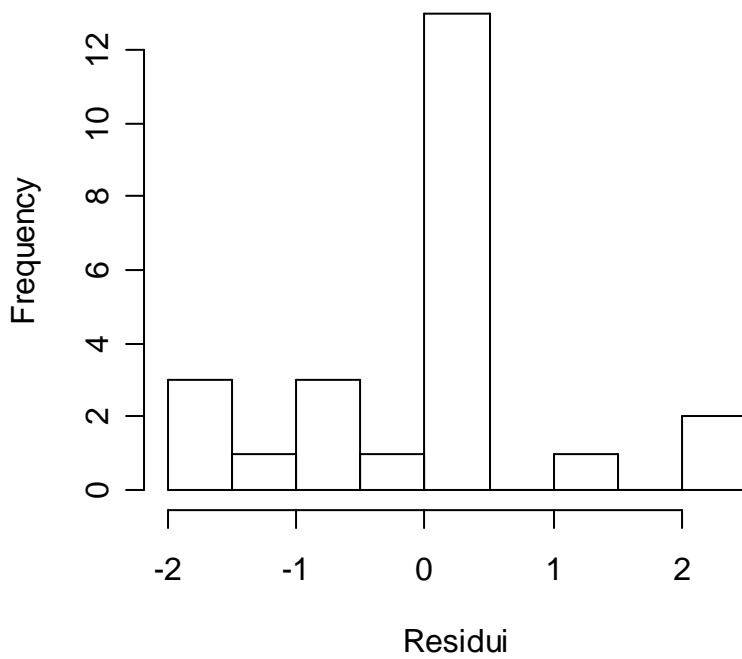
diagramma dei residui standardizzati



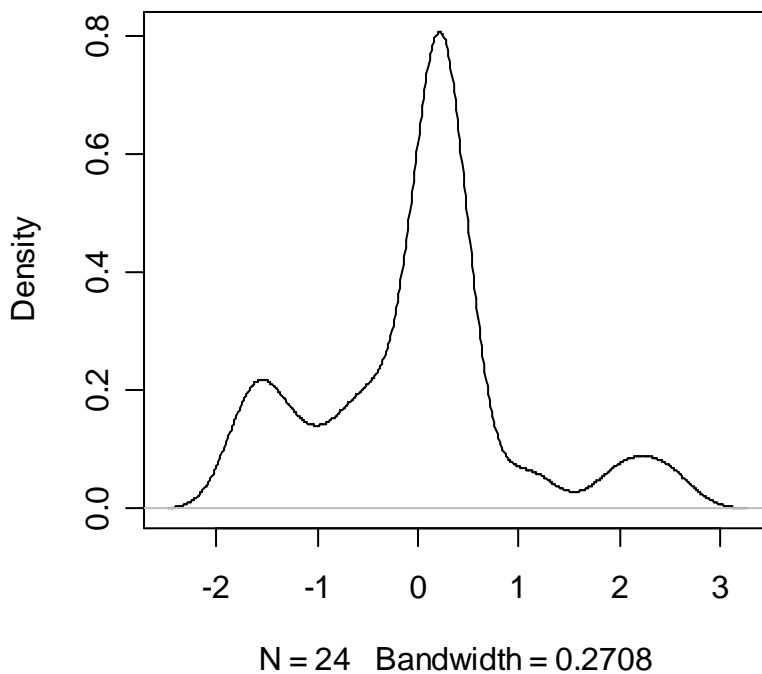
Dal diagramma emerge che nessuna osservazione risulta trovarsi al di là della banda di confidenza del 99% (banda compresa tra il -2.5 e il +2.5). Dal grafico non sembra che la distribuzione degli errori sia una normale.

Un modo abbastanza semplice ed intuitivo per verificare la normalità della distribuzione degli errori è quello di ricorrere all'aiuto di un istogramma

Distribuzione dei residui:istogramma

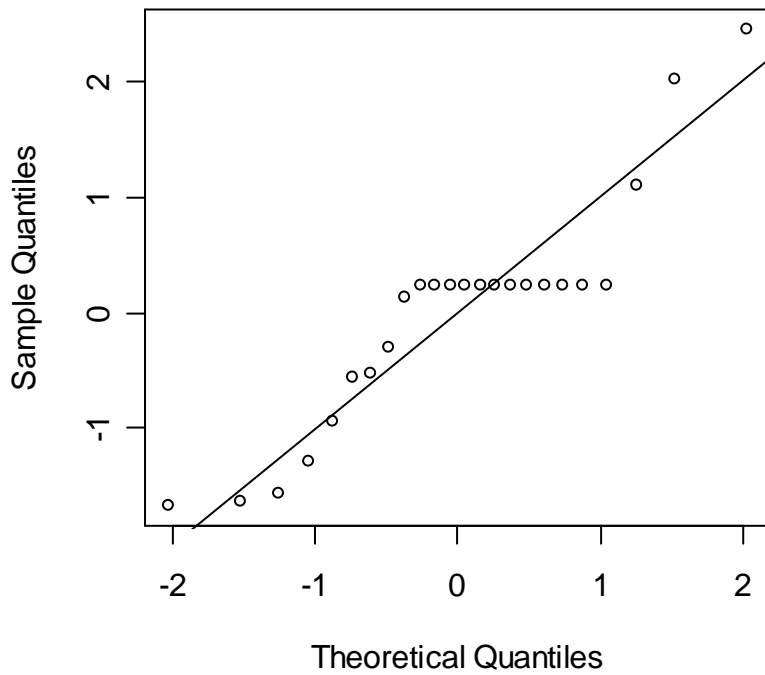


Distribuzione dei residui:lisciamento



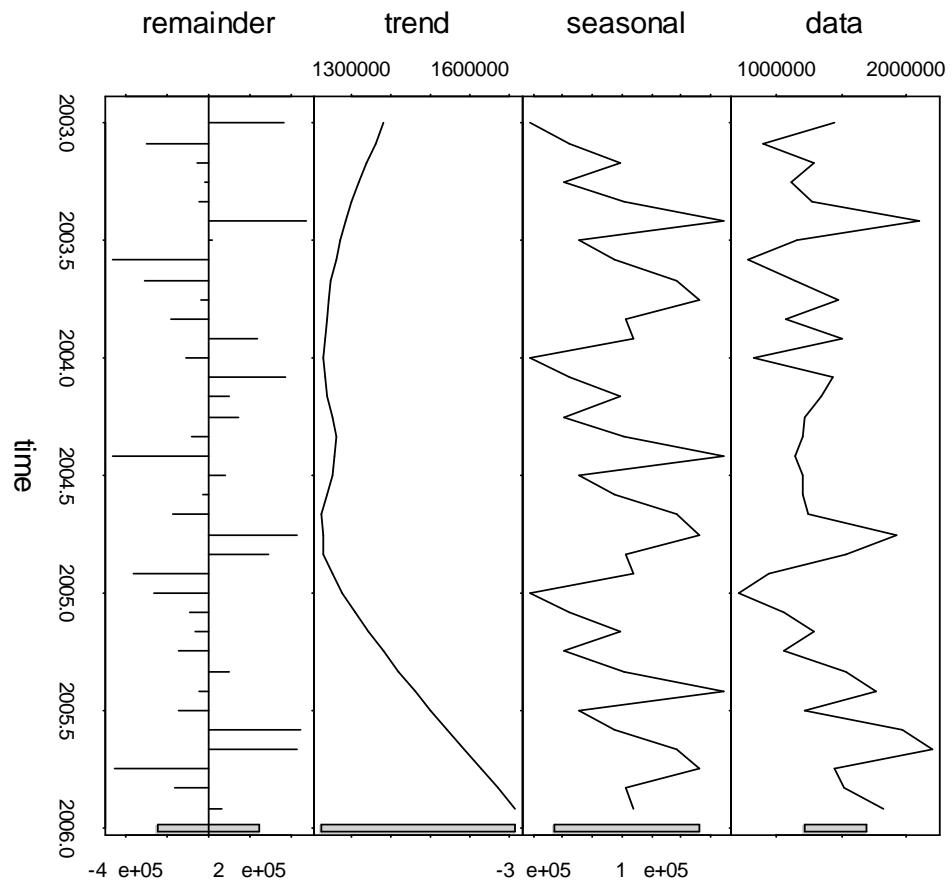
.....e di un QQ-plot (per verificare la normalità della distribuzione degli errori)

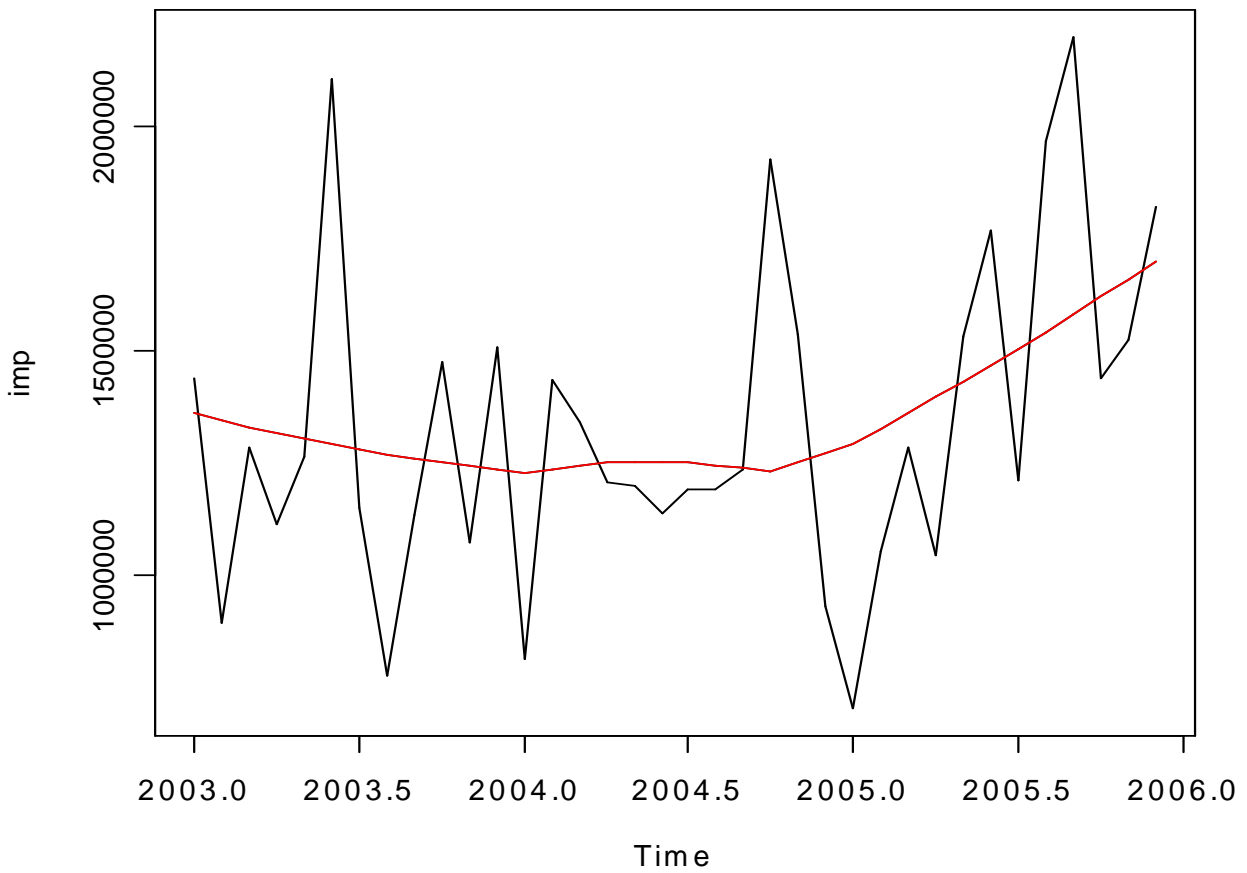
Normal Q-Q Plot



Tutti e tre i grafici ci danno indicazioni nettamente contrarie alla distribuzione normale dei residui.

La funzione di decomposizione stl usa **“loess”** per decomporre la serie in trend, ciclo e stagionalità.





Con questo grafico esaminiamo la serie osservata indicata con la linea nera e il trend stimato con la linea rossa.

Ora mi trovo le principali statistiche rispettivamente della stagionalità, del trend e dei residui.

rispettivamente della stagionalità:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-316700.0	-151800.0	3782.0	599.3	65890.0	347800.0

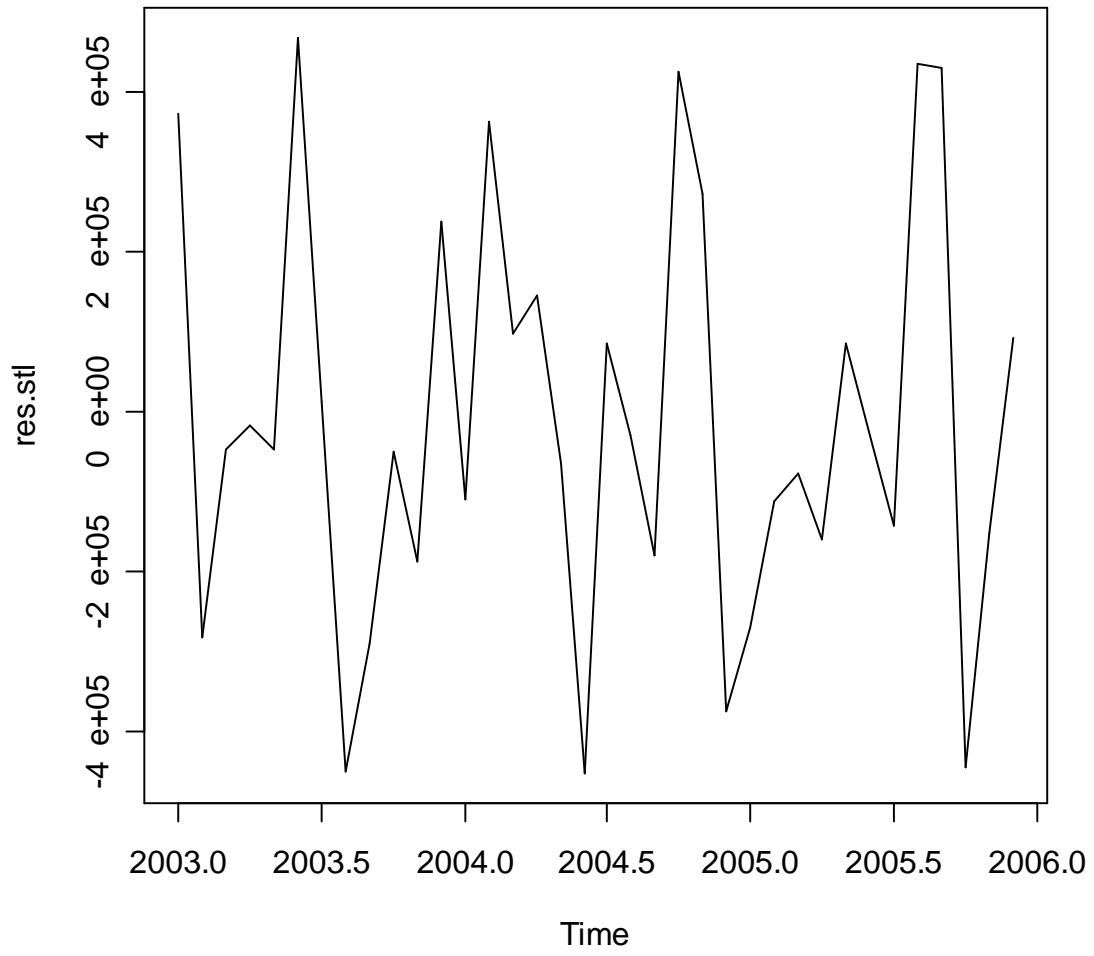
del trend:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1227000	1250000	1284000	1341000	1369000	1699000

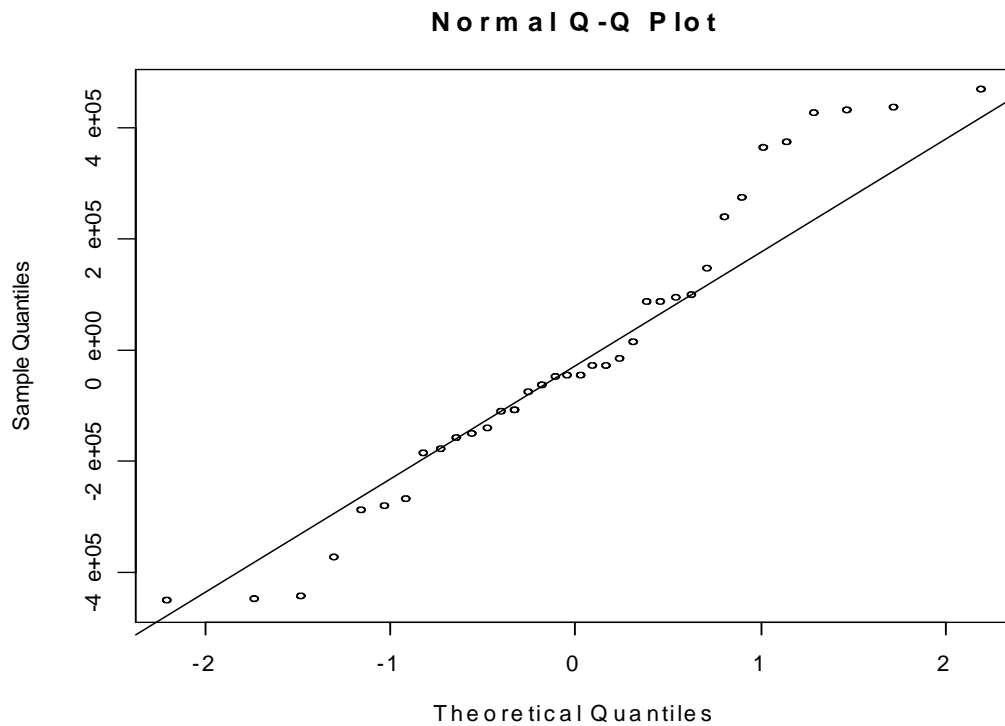
e dei residui:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-451300	-164400	-46740	-12200	110100	466600

e osservo il grafico dei residui.



Vediamo che i residui hanno una distribuzione abbastanza normale:



Verifico che la media dei residui non sia significamene diversa da zero.Effettuo il test **t** :

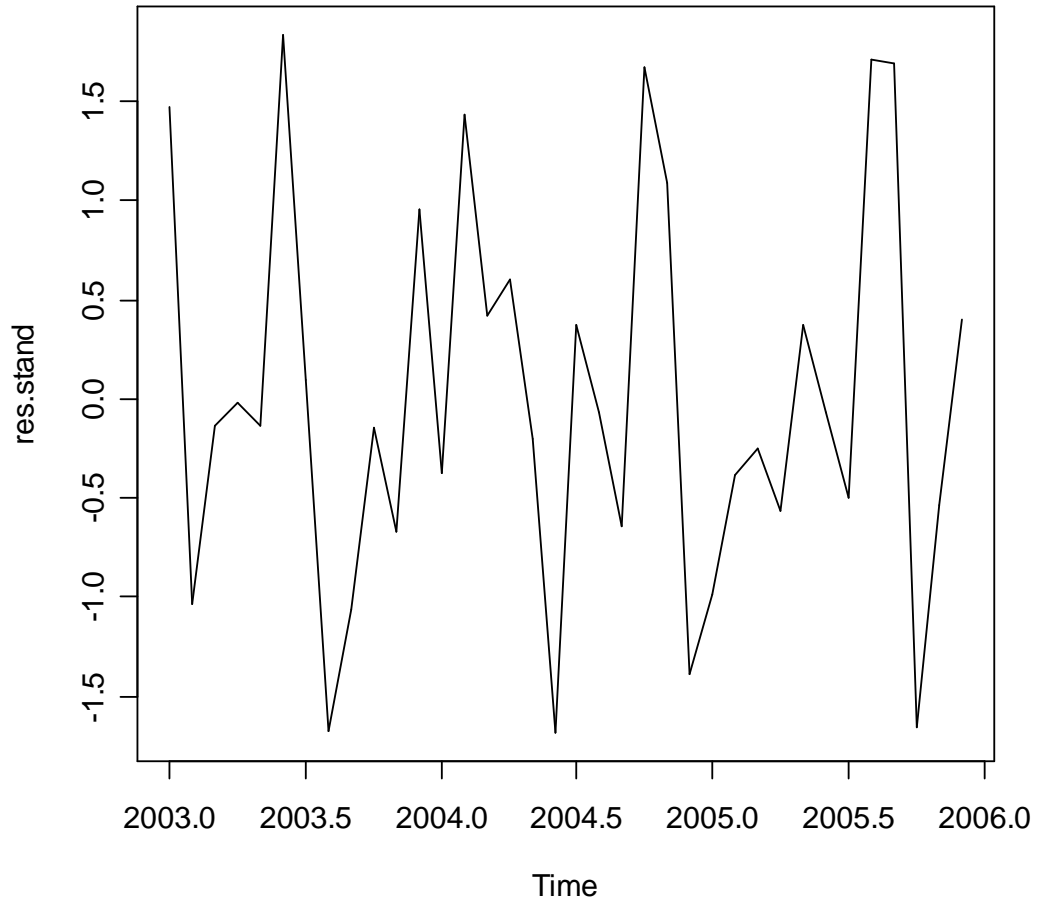
test.t=-0.2768752 il cui p-value e pari a 0.6082477

oppure considero il valore soglia del test per un livello di significatività pari al 99%:

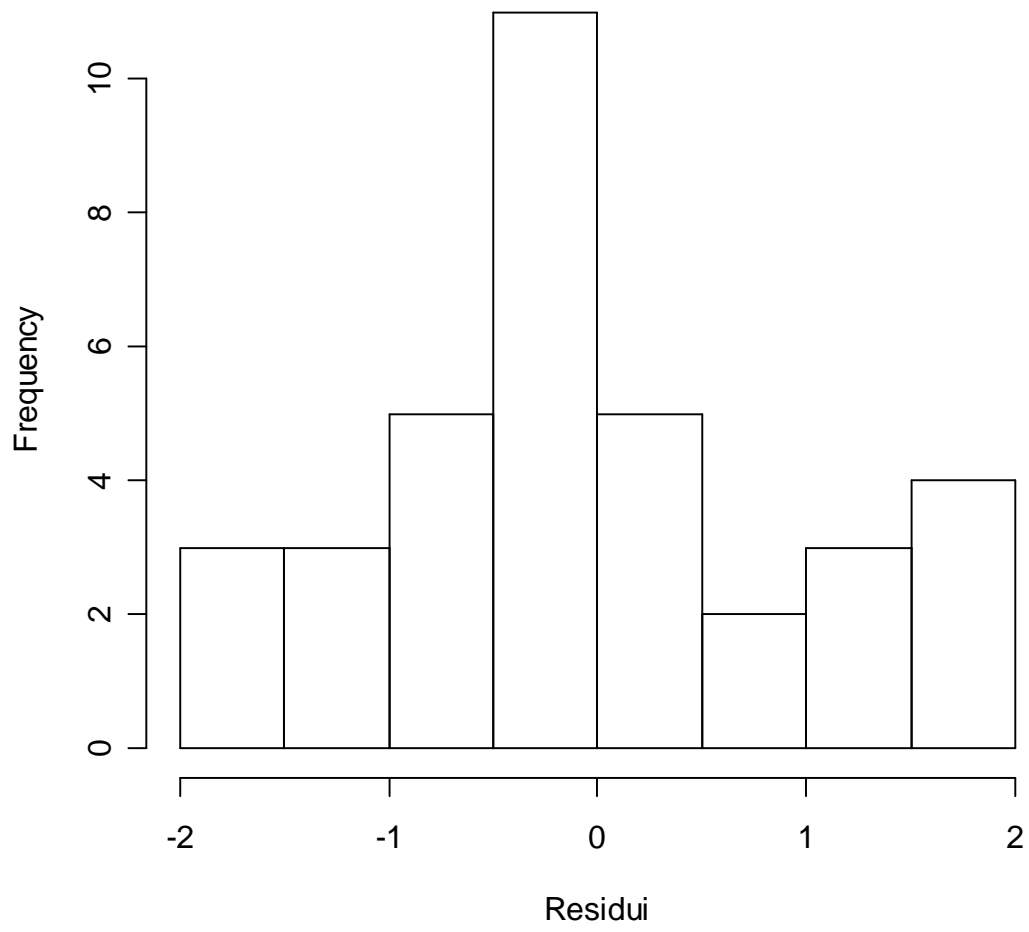
qt(.99,n-1)= 2.437723

Voglio verificare che i residui si distribuiscano secondo una variabile aleatoria normale.E' preferibile operare sui residui standardizzati e ne traccio il grafico.

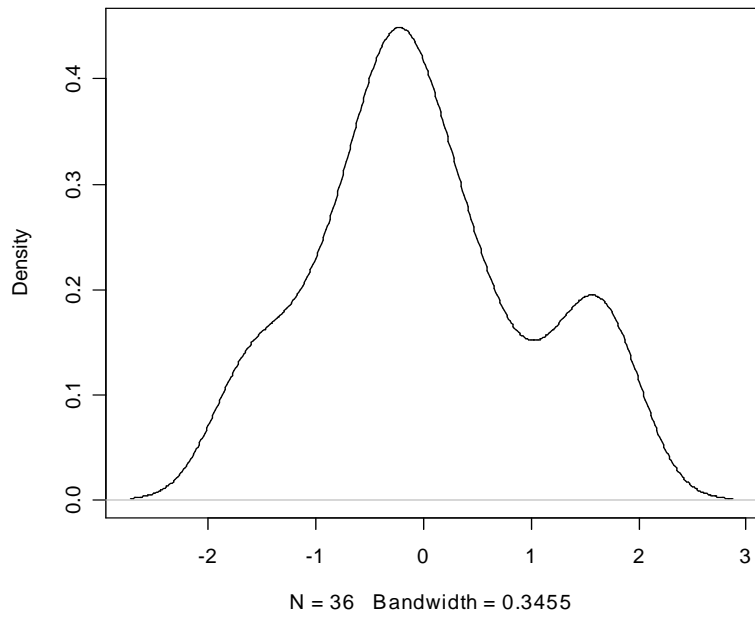
Diagramma dei residui standardizzati



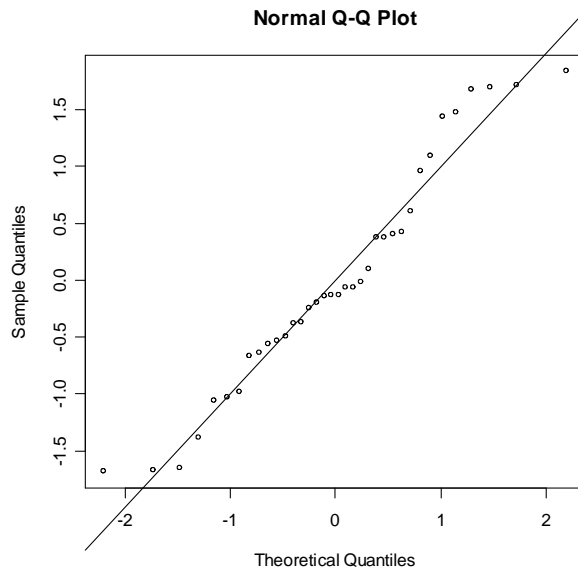
Distribuzione dei residui:istogramma



Distribuzione dei residui:lisciamento



Normal Q-Q Plot



Per avere un risultato statisticamente più affidabile bisogna effettuare un test di normalità.

Effettuo il Shapiro test e ottengo: $W = 0.9536$, con $p\text{-value} = 0.1359$

Il p-value è decisamente rispetto ai livelli di significatività a cui di solito si fa riferimento: ciò ci fa propendere per l'ipotesi nulla ovvero per la normalità della distribuzione degli errori.

Provo ad aumentare il numero di parametri equivqlenti nella stima e ottengo : la

media dei residui $\bar{e} = \frac{\sum e_t}{n} = -11141.46$, la varianza campionaria corretta

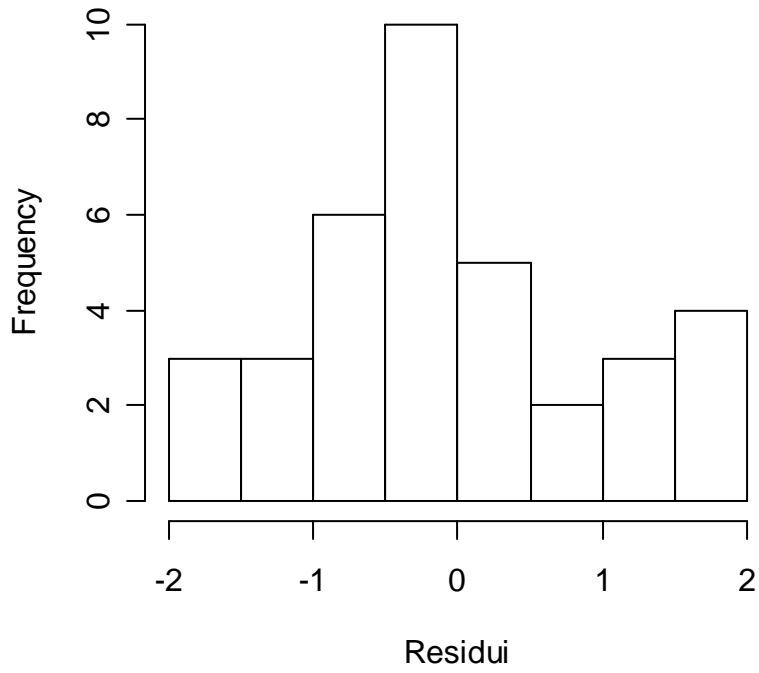
$$s^2 = \frac{\sum (e_t - \bar{e})^2}{n-1} = 71852442452$$

e lo scarto quadratico medio $s = 268053.1$

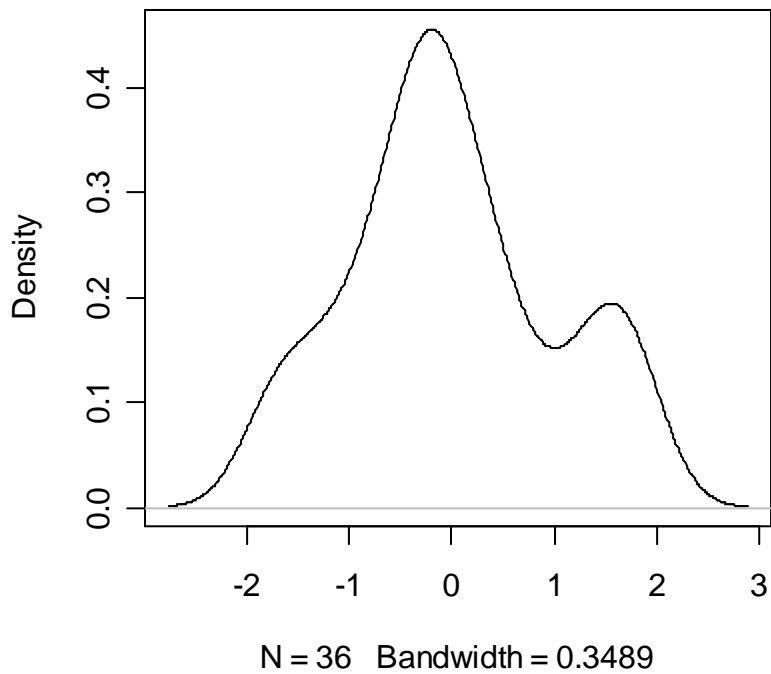
Determino il valore del test $t = \frac{\bar{e}}{\frac{s}{\sqrt{n}}} = 0.2493864$ con p-value=0.5977401 e

$qt(.99, n-1) = 2.437723$

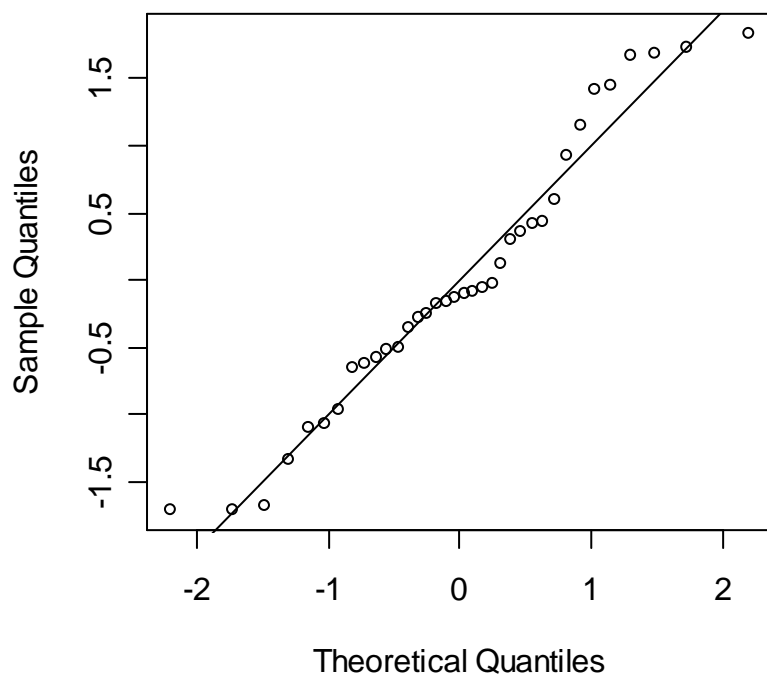
Distribuzione dei residui:istogramma



Distribuzione dei residui:lisciamento



Normal Q-Q Plot



Come prima effetto il test di Shapiro per verificarne la normalità e ottengo: $W = 0.9555$,

con $p\text{-value} = 0.1558$

Il $p\text{-value}$ è elevato rispetto ai livelli di significatività a cui di solito si fa riferimento.

Questo fa propendere per l'ipotesi nulla ossia la normalità della distribuzione degli errori.

Ottingo un risultato migliore del precedente (quello che aveva "s.window" pari a 11), per cui scelgo questo ultimo (quello con "s.window=27")

Effettuo il test di Jarque-Bera che si basa sulla misura della asimmetria e della curtosi della distribuzione ottenendo: $X\text{-squared} = 0.947$, $df = 2$, $p\text{-value} = 0.6228$

Anche in questa circostanza il $p\text{-value}$ è sufficientemente elevato per impedirci di rifiutare l'ipotesi nulla ossia rifiutare la normalità della distribuzione dei residui.

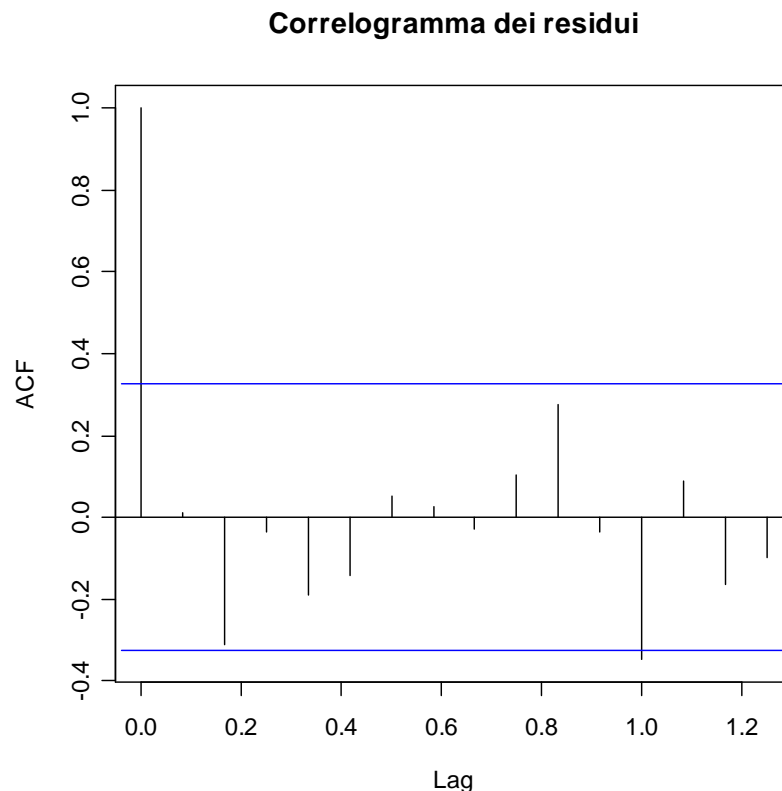
Si può avere un fenomeno di autocorrelazione temporale per cui ogni valore è influenzato da quello precedente e determina in parte rilevante quello successivo. Esistono vari test per saggiare la presenza di una correlazione seriale dei residui di una serie storica ma in questo lavoro farò riferimento solo ai test di Box-Pierce, Ljung-Box, Durbin-Watson.

Un modo abbastanza semplice per vedere se una serie presenta autocorrelazione è

quello di tracciarne il correlogramma .

In caso di assenza di autocorrelazione la distribuzione asintotica della stima del coefficiente di autocorrelazione è di tipo normale ed avremo una banda di confidenza del tipo come riportato nella sezione teorica.

Valori esterni a tale intervallo indicano la presenza di autocorrelazione significativa.



Verifico che nella serie di residui standardizzati non vi sia correlazione seriale prima con il test di Ljung-Box $X\text{-squared} = 18.0376$, $df = 12$, $p\text{-value} = 0.1146$ e poi con quello di Box-Pierce che mi dà $X\text{-squared} = 13.2014$, $df = 12$, $p\text{-value} = 0.3546$

In entrambi i casi il risultato del test non ci consente di rifiutare l'ipotesi nulla e quindi concludo sostenendo che i residui della serie storica non sono autocorrelati.

Un'ulteriore modalità per verificare l'esistenza di autocorrelazione tra i residui è il test di

Durbin-Watson che ha come ipotesi nulla l'assenza di autocorrelazione tra i residui:
 $\text{res. stand} \sim 1$

con statistica pari a $DW = 1.9157$ e $p\text{-value} = 0.7986$

oppure facendo regredire i valori osservati della serie sul trend e sulla stagionalità, senza considerare l'intercetta ed effettuare il test di Durbin Watson sui residui di tale regressione:

$\text{imp} \sim -1 + \text{trend} + \text{stagionalità}$ ottenendo $DW = 1.9171$, $p\text{-value} = 0.7802$

Entrambe le verifiche hanno come ipotesi alternativa :l'autocorrelazione è diversa da zero.

In ambo i casi esaminati si conferma l'assenza di autocorrelazione dei residui.

Concludendo posso affermare che la componente erratica della serie storica risulta soddisfare tutti i

requisiti richiesti: normalità , omoschedaticità e assenza di autocorrelazione seriale.

Essa può essere assimilata ad un processo stocastico detto White noise gaussiano.

Correlogramma della serie dei residui

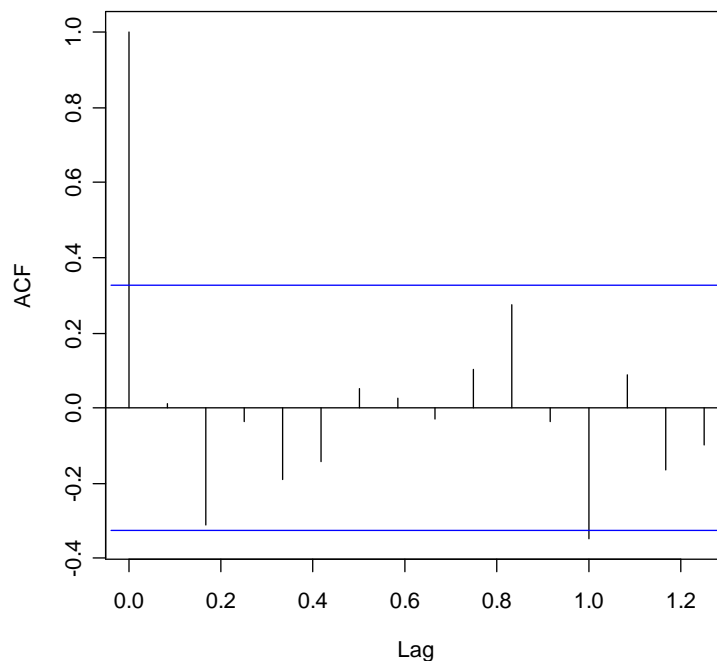


Grafico delle autocovarianze

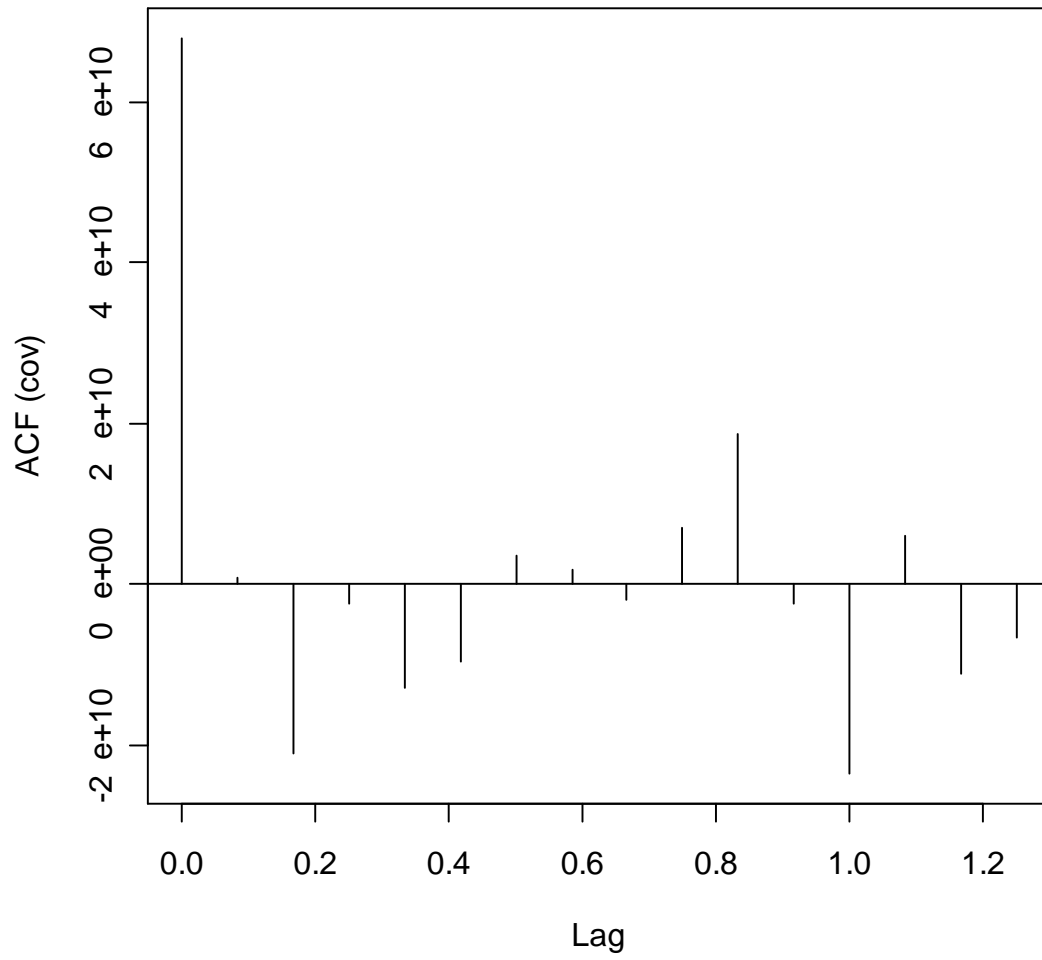
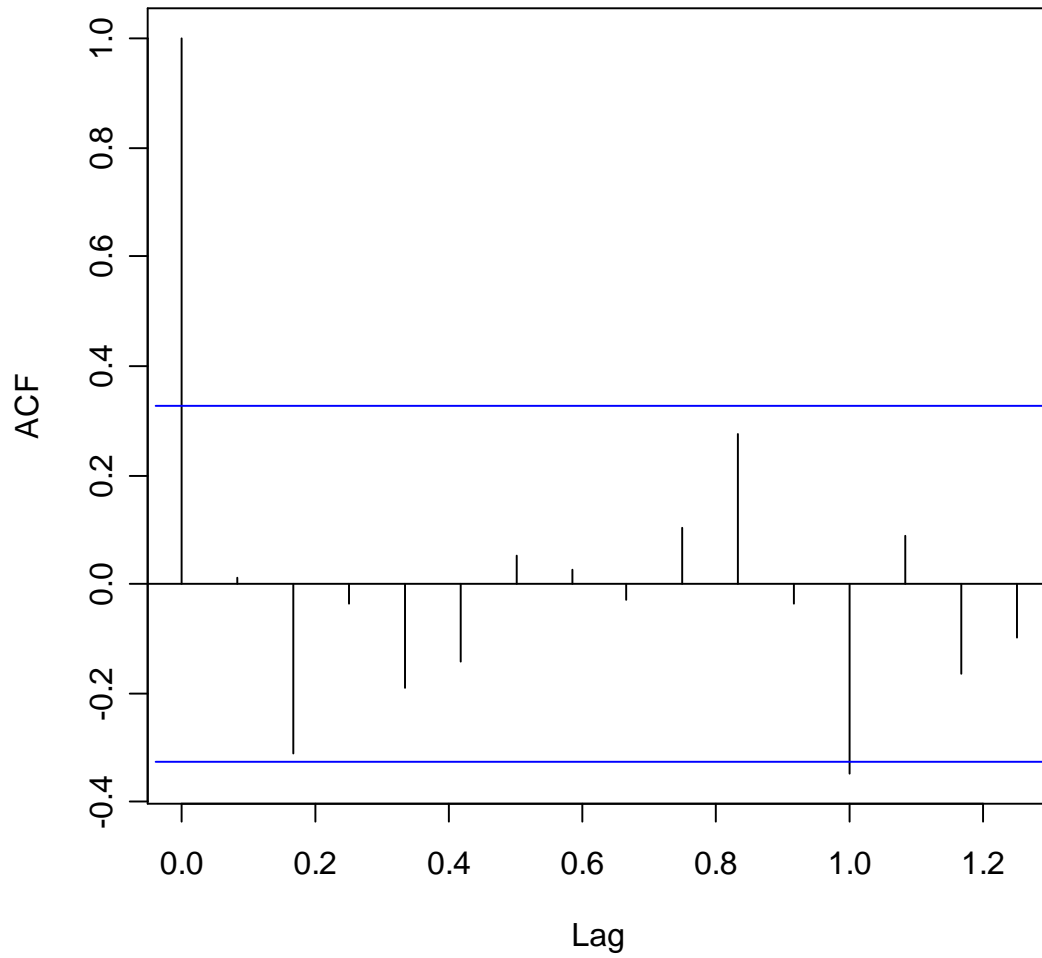
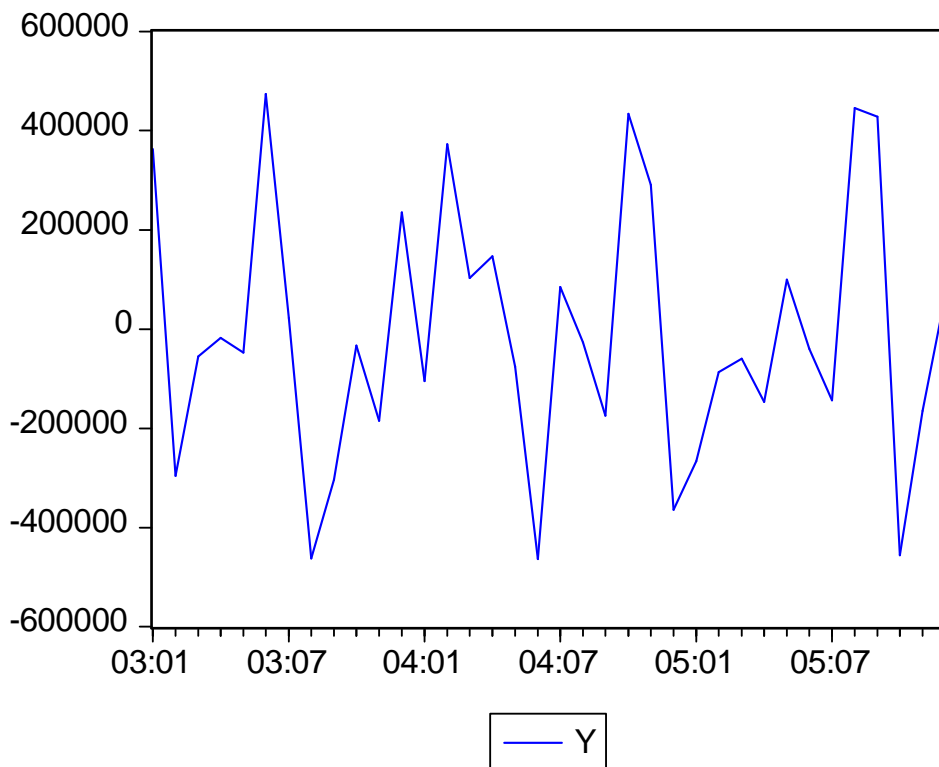


Grafico delle correlazioni parziali

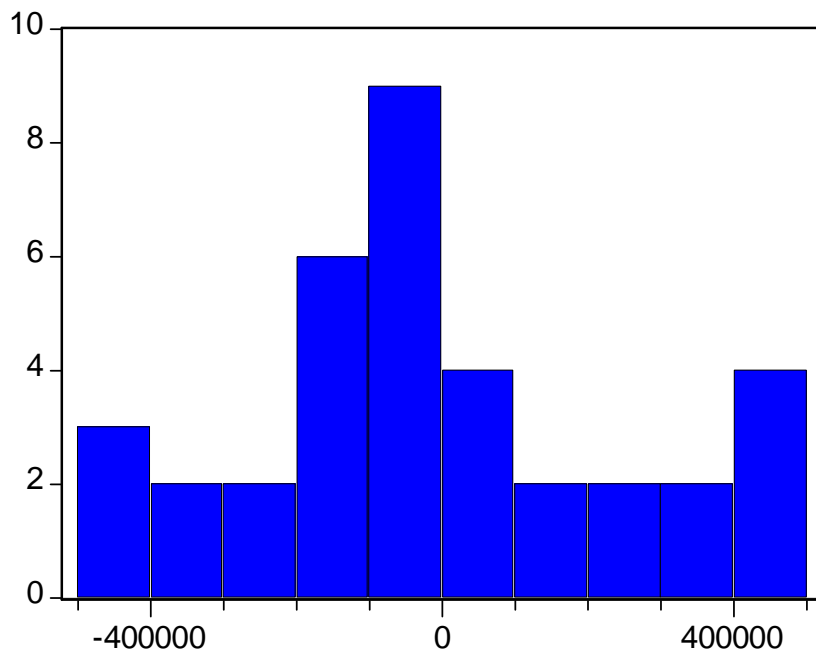


Per una diversa visualizzazione del mio operato importo la serie dei residui (res.stl) in Eviews.

Ottingo il grafico della serie a cui ho eliminato, con stl, il trend e la stagionalità.



Inoltre ottengo anche questa tabella che mi riporta riunite l'istogramma e altre statistiche fondamentali:



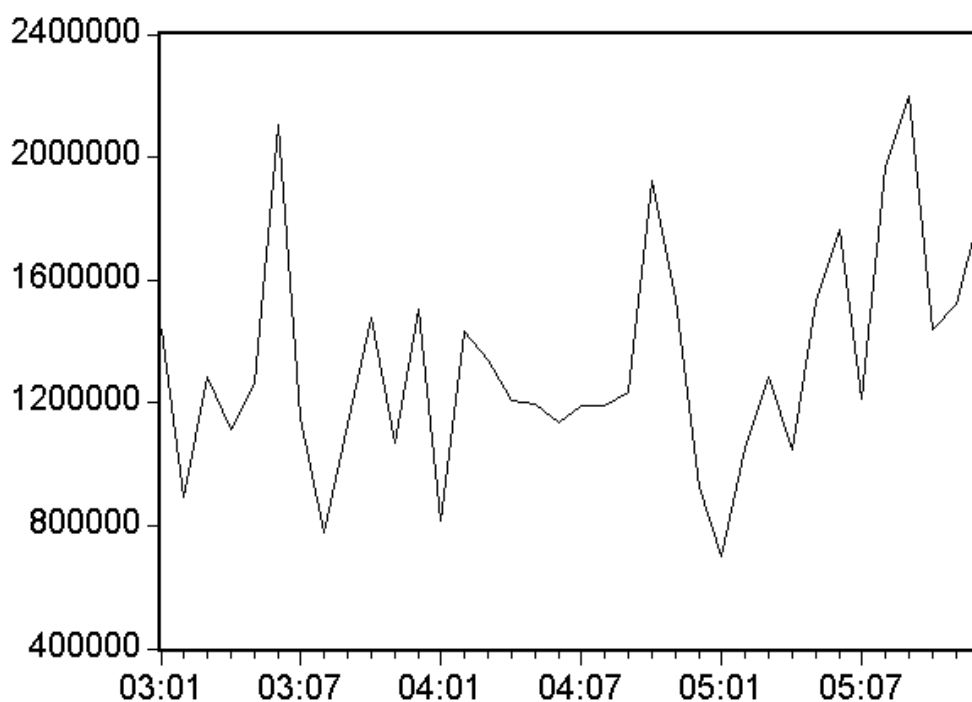
Series: Y	
Sample 2003:01 2005:12	
Observations 36	
Mean	-11141.47
Median	-43168.73
Maximum	473837.2
Minimum	-463668.3
Std. Dev.	264303.9
Skewness	0.222104
Kurtosis	2.341193
Jarque-Bera	0.947020
Probability	0.622812

Il correlogramma della serie è questo:

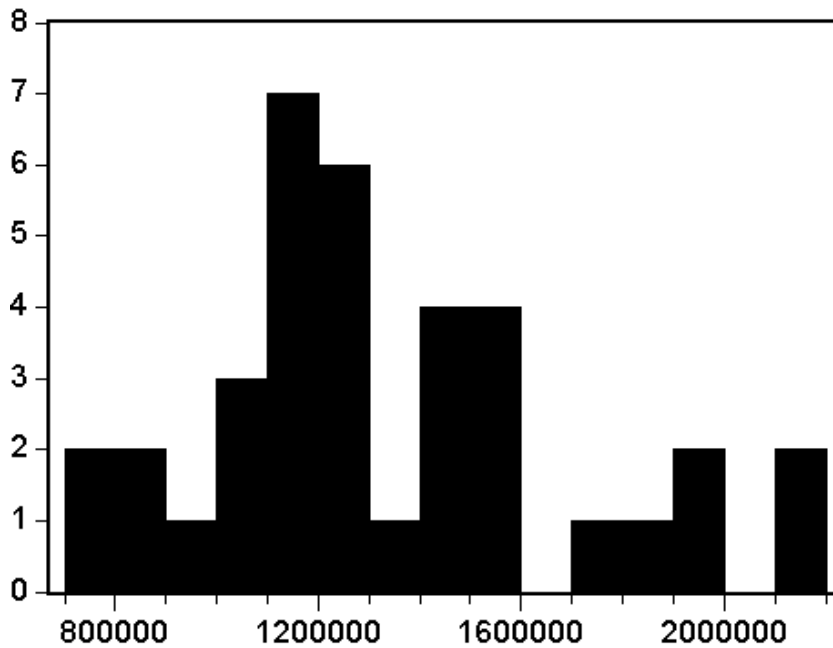
Date: 06/13/06 Time: 14:12
 Sample: 2003:01 2005:12
 Included observations: 36

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	0.012	0.012	0.0059	0.939
.** .	.** .	2	-0.310	-0.310	3.8780	0.144
. .	. .	3	-0.036	-0.030	3.9309	0.269
.** .	.** .	4	-0.191	-0.317	5.4864	0.241
.* .	.** .	5	-0.143	-0.201	6.3911	0.270
. .	.* .	6	0.053	-0.174	6.5197	0.368
. .	.* .	7	0.025	-0.184	6.5489	0.477
. .	.** .	8	-0.028	-0.237	6.5865	0.582
. *	.* .	9	0.104	-0.126	7.1386	0.623
. **	. *	10	0.274	0.152	11.093	0.350
. .	. .	11	-0.037	-0.040	11.170	0.429
*** .	.** .	12	-0.347	-0.298	18.038	0.115
. *	. *	13	0.090	0.087	18.521	0.139
. *	*** .	14	-0.162	-0.388	20.160	0.125
. *	. *	15	-0.100	-0.158	20.811	0.143
. **	. *	16	0.301	-0.123	26.998	0.042

Con Eviews ora lavoro direttamente sulla serie osservata e analizzo i dati in sequenza temporale ottenendo :
 il grafico della serie mensile



e l'istogramma con principali dati statistici mi dà questo risultato:



Series: Y	
Sample 2003:01 2005:12	
Observations 36	
Mean	1329617.
Median	1248381.
Maximum	2197204.
Minimum	702696.0
Std. Dev.	360699.2
Skewness	0.628744
Kurtosis	2.993897
Jarque-Bera	2.371967
Probability	0.305446

Correlogramma della serie presa in esame:

File Edit Objects View Procs Quick Options Window Help
View Procs Objects Print Name Freeze Sample Genr Sheet Stats Ident Line Bar

Correlogram of Y

Date: 02/28/06 Time: 20:09
Sample: 2003:01 2005:12
Included observations: 36

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.211	0.211	1.7398	0.187
		2	-0.042	-0.091	1.8109	0.404
		3	0.146	0.185	2.6955	0.441
		4	0.006	-0.081	2.6971	0.610
		5	-0.050	-0.005	2.8070	0.730
		6	0.068	0.054	3.0154	0.807
		7	-0.085	-0.124	3.3548	0.850
		8	-0.075	0.002	3.6277	0.889
		9	-0.057	-0.092	3.7947	0.924
		10	0.042	0.117	3.8855	0.952
		11	0.075	0.043	4.1967	0.964
		12	-0.036	-0.056	4.2687	0.978
		13	0.001	0.034	4.2688	0.988
		14	-0.070	-0.142	4.5739	0.991
		15	-0.067	0.024	4.8658	0.993
		16	0.096	0.074	5.4934	0.993

Dall'analisi del correlogramma vedo che la serie è un **wn**. Ossia le osservazioni componenti la serie sono tra loro incorrelate. Devo concludere che la migliore "soluzione" sia stata quella di estrarre le componenti con la funzione stl.

Ho ottenuto una grande quantità di informazioni che altrimenti Eviews avrebbe tenuto "nascoste", pur arrivando con entrambi i metodi alla stessa conclusione.

Dalla teoria sappiamo che per prevedere il futuro il futuro è necessario utilizzare l'informazione del passato. In una partita di testa o croce, la memoria non gioca alcun ruolo; il presente è indipendente dal passato. Non vi è memoria. La memoria di ciò che osserviamo è la parte utile di tale osservazione, parte che permette di tradurre la dipendenza probabilistica tra presente e passato.

L'indipendenza è difficile da ammettere. Tutti sanno che i giocatori dei casinò che hanno visto uscire il rosso dieci volte di seguito, pensano che all'undicesima partita, ci sono maggiori probabilità di vedere uscire il nero. Osservare un milione di partite non apporterebbe nessuna informazione supplementare quanto al risultato della partita successiva.

La memoria è di lunghezza L se la conoscenza di un passato più lontano di L non apporta più informazione di quanto non faccia il passato di lunghezza L .

I fenomeni senza alcuna memoria in matematica sono designati con il nome di “**rumore bianco**”. La probabilità è tale che ciò che succede ad ogni istante è indipendente sia dal passato che dal futuro. Ogni fenomeno che si svolge nel corso del tempo senza memoria si chiama rumore. Quando costruisco un modello ciò che non è possibile spiegare resta rappresentato da un rumore.

Definisco “innovazione” tutto ciò che avviene (in un mese, in questo caso, trattando di serie mensili) ed è un elemento indipendente dal passato. Tali innovazioni sono indipendenti e costituiscono una sequenza senza memoria. Hanno per ipotesi tutte le stesse caratteristiche probabilistiche, con media nulla.

Possiamo quindi dire che questo è “rumore di innovazione” che riunirà tutto ciò che può succedere in un mese e non è tenuto in considerazione dal modello. Ad esempio conflitti sociali, avvenimenti monetari inaspettati o eventi meteorologici eccezionali. Considerare tale innovazione aleatoria ci permette di riassumere tutto quello che non si sa o che non si vuole introdurre nel modello ma che ci permette facilmente di creare una previsione.

Prevedere si riassume nel “prolungare il funzionamento del sistema dal momento presente ad un dato momento futuro”, supponendo che il fattore di innovazione sia nullo. L'innovazione non ha memoria, ed è indifferente al passato, quindi la migliore previsione dell'innovazione è “nessuna innovazione”.

CONCLUSIONI

Da quanto emerso alla fine del lavoro si può affermare che, sia con l'approccio classico sia con l'approccio moderno nella pratica aziendale non ho ottenuto buoni risultati. Nell'approccio classico con entrambe i metodi presi in esame il trend è in rialzo, e sostanzialmente positivo. Nell'approccio moderno il risultato dell'analisi della serie ottenuta dalla rimozione della parte deterministica non è utilizzabile a fini previsivi. Praticamente questi risultati tradotti in "gergo aziendale" possono essere espressi come segue.

<<Il peso assunto dall'export nel fatturato globale è sempre più importante, sia per volumi e sia perché il mercato interno è in crisi. Anzi possiamo dire che la sua esistenza (dell'export) è diventata di vitale importanza per l'azienda date le condizioni di mercato interno. Si può pensare che almeno nell'est europeo l'azienda consoliderà e incrementerà la sua presenza su quei mercati anche nel prossimo futuro. Ferme restando naturalmente le condizioni sociali, politiche ed economiche che hanno permesso la nascita e lo sviluppo di tali mercati per Asolo Dolce >>. Per l'azienda Asolo Dolce ai fini pratici, questo lavoro può non sembrare di grossa utilità. Io invece spero che questo lavoro possa aver messo in luce che esistono diverse angolazioni da cui esaminare una determinata azienda e che ogni problema per essere affrontato richiede il giusto approccio mentale.

BIBLIOGRAFIA

Didier Dacunha-Castelle.*La scienza del caso*;1998.Dedalo Edizioni.

Di Fonzo-Lisi.*Complementi di statistica economica*;2000.Cleup Edizioni

Sergio Zani,*Modelli di serie storiche per analisi economiche*;1983.Editore Giuffrè.

Piccolo D.,*Analisi moderna delle serie storiche*;Angeli edizioni.

Di Fonzo,*Statistica e analisi di mercato*;Il mulino edizioni

D.Besanko,D.Dranove,M.Shanley,"Economia dell'industria e strategia d'impresa.Utet libreria.2001

Cippoletta I.(1992),*Congiuntura economica e previsione*,Bologna,Il Mulino

Theil H.(1965) *Economic forecast and policy*,Amsterdam,North-Holland.

Battaglia F.(2004),*Metodi di previsione statistica*. Dip. di Statistica,Probabilità e Statistiche Applicate,Università La Sapienza,Roma.

Tassinari F. *Lezioni di Statistica aziendale*. Università di Bologna.

Mazzali A. *Analisi previsionale*. Università degli studi di Brescia.

Mazzali A. *Analisi e previsione delle Serie Storiche*. Università degli studi di Brescia.

Revetria R. *Impianti industriali e Studi di fattibilità. Modulo 1*. Università degli studi di Genova.

Piccolo D. (1990). *Introduzione all'analisi delle serie storiche*. NIS, Roma

Alvaro G. (1995). *Contabilità nazionale e statistica economica*. Cacucci editore.

Mosca R. (1998-1999). *Dispense al Corso di Economia Aziendale 1 e 2*. DIPEM Pubblicazioni Interno.

Castagna R., *Corso di Gestione Aziendale, Modulo 13*, AA 2001/02, Facoltà di Ingegneria Cremona

Markland, Swiegard, *Quantitative Methods, applications to managerial decision making*, Wiley International Edition (1984)

Di Fonzo T. *Materiale per il corso di Statistica economica*, AA2000/01. Facoltà di Scienze Statistiche, Padova.

LORENZO FARINA, *Modelli Dinamici per l'Ingegneria Gestionale*, AA 2005/06, Facoltà di Ingegneria, Uniroma1