

3. Metodi di scomposizione

3.1 Introduzione

Molti metodi di previsione si basano sul fatto che, se esiste un *pattern* sistematico, questo possa essere individuato e separato da eventuali oscillazioni accidentali, mediante metodi di perequazione o smussamento (lisciamento, *smoothing*) dei dati della serie storica. L'effetto dello smussamento è quello di eliminare disturbi casuali cosicché, una volta individuato il *pattern*, questo possa essere proiettato nel futuro per produrre la previsione.

I metodi di scomposizione tendono, di norma, a individuare due *pattern*: il *trend-ciclo* e la *stagionalità*. Il *trend-ciclo* può essere scomposto, a sua volta, in componente di fondo (*trend*) e oscillazioni congiunturali (*ciclo*)^(*).

In questo capitolo sarà illustrato il metodo classico di scomposizione. Come si vedrà, tale procedimento è più che altro uno strumento di analisi della serie storica, e necessita pertanto di alcuni perfezionamenti per poter produrre la previsione.

I metodi di scomposizione hanno costituito il primo approccio di analisi delle serie storiche. Il *metodo classico* risale agli anni 20 e costituisce ancora oggi la base per i metodi più frequentemente usati. Attualmente il metodo di scomposizione più diffuso è quello denominato *Census II* che viene usato per numerose serie economiche sia in ambito privato sia in ambito pubblico (Makridakis, Wheelwright, Hyndman, 19XX).

E' importante precisare che, nel seguito, se non viene specificato diversamente, useremo il simbolo y_t , $t=1, \dots, n$ per indicare la serie storica, eventualmente aggiustata per le variazioni dei prezzi, per le oscillazioni di calendario, ecc..

3.2 Il modello di scomposizione

Il modello matematico ipotizzato nel metodo classico di scomposizione è:

$$(3.1) \quad y_t = f(S_t, T_t, E_t)$$

dove y_t è il dato riferito al periodo t

S_t è la componente stagionale al periodo t

T_t è la componente *trend-ciclo* al periodo t

E_t è la componente irregolare al tempo t .

^(*) In queste note non affronteremo il problema della stima del *ciclo*.

S_t e T_t sono le componenti sistematiche e cioè componenti per le quali è ipotizzabile un andamento regolare (che risponde a certe 'leggi') rispetto al tempo. La possibilità di operare previsioni, risiede proprio nella individuazione di queste componenti, assumendo che la legge che ha regolato il loro andamento nel passato continui a valere anche nel futuro.

La forma di $f()$ dipende dall'approccio seguito. Una forma molto comune è la seguente:

$$(3.2) \quad y_t = S_t + T_t + E_t$$

che viene definita *modello additivo*. La somma $(S_t + T_t)$ costituisce la parte sistematica del modello.

Un'altra forma altrettanto frequente è il *modello moltiplicativo*:

$$(3.3) \quad y_t = S_t \times T_t \times E_t$$

Il prodotto $(S_t \times T_t)$ costituisce la parte sistematica del modello.

Infine, il *modello misto* ma in alcuni casi detto anch'esso modello moltiplicativo è il seguente, in cui la parte sistematica ha componenti moltiplicative mentre quella erratica è additiva:

$$(3.4) \quad y_t = S_t \times T_t + E_t$$

Il modello con componenti sistematiche additive è appropriato quando l'ampiezza dell'oscillazione stagionale non varia col livello della serie. Se invece la fluttuazione stagionale aumenta (o diminuisce) proporzionalmente con l'aumento (diminuzione) del livello della serie, allora è più adeguato un modello con componenti sistematiche moltiplicative. Molte serie economiche esibiscono fluttuazioni stagionali che crescono all'aumentare del livello della serie; per tale motivo, in ambito economico, i modelli con componenti sistematiche moltiplicative trova più larga applicazione.

Nel modello additivo, le componenti S_t, T_t, E_t sono espresse nella stessa unità di misura di y_t ; nel modello moltiplicativo, solo T_t (per convenzione) viene espresso nell'unità di misura di y_t ; E_t e S_t sono numeri puri. Nel modello additivo l'errore può assumere valori positivi o negativi; 0 è il valore *neutrale*, nel senso che non influenza la serie. Nel modello moltiplicativo l'errore può assumere solo valori non negativi e ha 1 come valore *neutrale*. Nel modello misto (v. espressione 3.4) solo T_t e E_t sono espresse nella stessa unità di misura del fenomeno y_t mentre S_t è un numero puro che oscilla intorno a 1 assumendo valori non negativi. Sempre nel modello misto (v. espressione 3.4), l'errore può assumere valori positivi o negativi; 0 è il valore *neutrale*, nel senso che non influenza la serie.

La differenza fra modello moltiplicativo e modello misto risiede nella posizione di E_t . Infatti, nel modello moltiplicativo l'effetto dell'errore dipende anche dall'ordine di grandezza della componente sistematica. Ad esempio, a parità di $E_t=0.8$ se la componente sistematica è $T_t S_t=100$, $y_t=180$; se invece $T_t S_t=10$, $y_t=18$. Nel modello misto, invece, l'effetto di E_t non risente della grandezza della componente sistematica. Ad esempio, a parità di $E_t=10$, se la componente sistematica è $T_t S_t=100$, $y_t=110$; se invece $T_t S_t=10$, $y_t=20$.

Si noti che, col modello moltiplicativo, potrebbe essere utile ricorrere alla trasformazione della serie. Poiché la funzione logaritmica trasforma una espressione moltiplicativa in una additiva si ha:

$$\ln y_t = \ln(S_t \times T_t \times E_t) \iff \ln y_t = \ln S_t + \ln T_t + \ln E_t$$

Quindi, invece del modello moltiplicativo sui dati originari della serie, si potrebbe applicare il modello additivo sulle trasformate logaritmiche.

Nel seguito, analizzeremo i modelli di scomposizione di tipo additivo e misto.

3.3 La media mobile

La media mobile è un semplice metodo che smussa (liscia, perequa) la serie storica. Tale procedura è basilare nei metodi di scomposizione. Se la serie è composta solo da *trend* e dalla componente residua, la media mobile elimina gli effetti dei disturbi. Se nella serie originaria è presente anche il fenomeno stagionale di periodo k , allora una media mobile di ampiezza k è in grado di eliminare *anche* la stagionalità. Nei due casi, la media mobile si propone di isolare il *trend-ciclo*.

Vediamo un esempio di serie che presenta *trend* e stagionalità (Tab. 3.1). Si tratta delle vendite mensili di shampoo (litri) venduti in tre anni. Il *time plot* della serie (Fig. 3.6) evidenzia la presenza di un marcato *trend* crescente e di disturbi di un certo rilievo; non si individuano ciclicità periodiche tipiche della stagionalità. La Fig. 3.6 riporta l'evoluzione delle medie mobili a 3 termini (MM3) e a 7 termini (MM7); come si vede, la media mobile elimina una certa quota di oscillazioni perturbatrici.

La media mobile a tre termini ci dà una stima del *trend* T_2 del mese di Febbraio 1999, mediante la media aritmetica dei dati di Gennaio, Febbraio, Marzo 1999:

$$T_2 = (y_1 + y_2 + y_3) / 3$$

Generalizzando, la media mobile a tre termini *centrata su* t è:

$$(3.5) \quad T_t = (y_{t-1} + y_t + y_{t+1}) / 3, \quad t = 2, \dots, n-1$$

Si notino, nella Tab. 3.1, i valori della media mobile a tre termini: non c'è stima del *trend* per i tempi $t=1$ e $t=n$ perché mancano le osservazioni al tempo 0 e al tempo $n+1$.

Come si sarà capito, questa procedura è denominata *media mobile* perché ogni successiva media viene calcolata eliminando il valore 'più vecchio' e inserendone un nuovo. La media mobile è un *metodo di adattamento locale* in quanto produce una serie di valori smussati ognuno in corrispondenza del punto di osservazione t , in base ai $k/2$ nel caso di k pari o $(k-1)/2$, nel caso di k dispari, valori prossimi *localmente* al tempo t .

Tab. 3.1 – Esempio di serie con trend e disturbi casuali

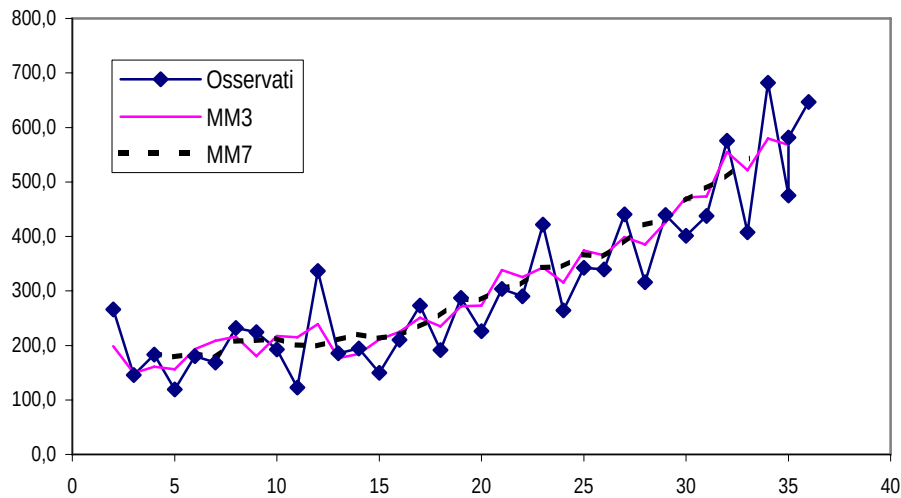
Mese	y_t	MM3	MM5	MM7
1	266,0			
2	145,9	198,3		
3	183,1	149,4	178,9	
4	119,3	160,9	159,4	185,0
5	180,3	156,0	176,6	179,1
6	168,5	193,5	184,9	185,8
7	231,8	208,3	199,6	177,2
8	224,5	216,4	188,1	208,2
9	192,8	180,1	221,7	209,0
10	122,9	217,4	212,5	212,7
11	336,5	215,1	206,5	200,9
12	185,9	238,9	197,8	198,9
1	194,3	176,6	215,3	210,4
2	149,5	184,6	202,6	220,1
3	210,1	211,0	203,7	213,1
4	273,3	224,9	222,3	218,8
5	191,4	250,6	237,6	234,4
6	287,0	234,8	256,3	254,5
7	226,0	272,2	259,6	284,7
8	303,6	273,2	305,6	283,4
9	289,9	338,4	301,1	305,0
10	421,6	325,3	324,4	312,5
11	264,5	342,8	331,6	343,1
12	342,3	315,5	361,7	344,9
1	339,7	374,1	340,6	366,2
2	440,4	365,3	375,5	363,3
3	315,9	398,5	387,3	388,0
4	439,3	385,5	406,9	421,4
5	401,3	426,0	433,9	431,1
6	437,4	471,4	452,2	465,6
7	575,5	473,5	500,8	488,3
8	407,6	555,0	515,6	508,6
9	682,0	521,6	544,3	543,7
10	475,3	579,5	558,6	
11	581,3	567,8		
12	646,9			

Fonte: Makridakis, Wheelwright, Hyndman (19XX)

Il numero di termini coinvolti nella media mobile influenza il risultato della perequazione. All'aumentare dei termini, la spezzata che unisce i punti individuati dalle medie mobili si fa più smussata. Nella Fig. 3.1 compare l'andamento della media mobile a 7 termini, che è calcolata come:

$$T_t = (y_{t-3} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + y_{t+3}) / 7, \quad t = 4, \dots, n-3$$

Fig. 3.1 – Valori osservati e medie mobili (MM) per i dati di Tab. 3.1



Come si può facilmente verificare, una media mobile a k termini, con k dispari, fa perdere $(k-1)/2$ termini all'inizio e altrettanti termini alla fine della serie (Tab. 3.1). La perdita dei primi termini ha poca importanza; al contrario la perdita degli termini più recenti ha conseguenze rilevanti ai fini della operazione di previsione. Una possibile soluzione consiste nell'effettuare, agli estremi, delle medie mobili con un numero inferiore di termini. Ad esempio, nel caso di media mobile a tre termini si può calcolare T_1 come $T_1 = (y_1 + y_2)/2$ e T_n come $T_n = (y_{n-1} + y_n)/2$.

Le medie mobili fin qui introdotte hanno un numero dispari di termini e perciò risultano automaticamente *centrate* su un punto t di osservazione. Tali medie sono dette *semplici* poiché tutti i termini della media hanno associato lo stesso peso.

Supponiamo che si voglia calcolare una media mobile con numero pari di termini. Ad esempio, posto $k=4$, sui dati di Tab. 3.1 si ha:

$$T' = (y_1 + y_2 + y_3 + y_4)/4 = (266,0 + 145,9 + 183,1 + 119,3)/4$$

$$T'' = (y_2 + y_3 + y_4 + y_5)/4 = (145,9 + 183,1 + 119,3 + 180,3)/4$$

La prima media sarebbe centrata fra il secondo e il terzo termine; la seconda media centrata fra il terzo e il quarto. Per risolvere la questione della *centratura*, si effettua una media mobile a 2 termini sulle due successive medie mobili a termini pari. Con questo procedimento la media aritmetica delle due medie mobili a 4 termini sopra calcolate, viene ad essere centrata nel punto $t=3$. Quindi:

$$T_3 = (T' + T'')/2$$

Sostituendo a T' e T'' le espressioni precedenti, la formula di T_3 diventa:

$$(3.6) \quad T_3 = (y_1 + 2y_2 + 2y_3 + 2y_4 + y_5)/8$$

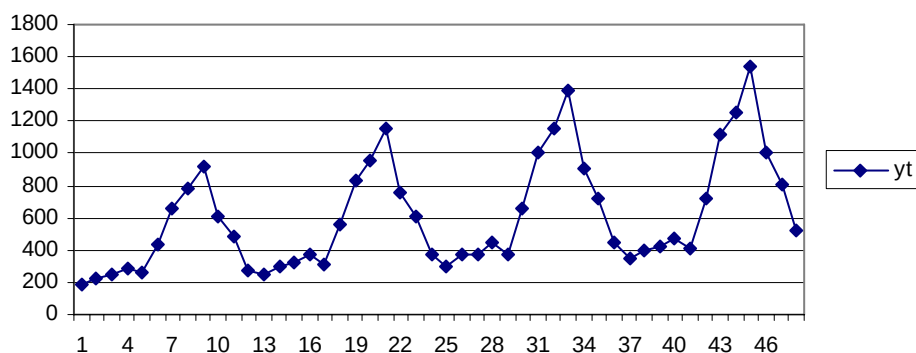
che è una media *ponderata*: i termini centrali hanno peso 2, i termini estremi peso 1; il denominatore è, ovviamente, la somma dei pesi. Essa è detta *media mobile centrata a k termini* (k pari).

Medie mobili con numero pari di termini sono usate per eliminare l'oscillazione stagionale. Su dati mensili si userà $k=12$; $k=4$ su dati trimestrali e $k=2$ su dati semestrali. Ovviamente, con k pari, si perdono $k/2$ termini all'inizio e alla fine della serie.

3.4 Scomposizione classica: il modello additivo

Nella esposizione delle fasi per la scomposizione, ci riferiamo ai dati mensili sulle vendite di bottiglie QQQ, relativi a 4 anni (Fig. 3.2 e Tab. 3.2).

Tab. 3.2 – Scomposizione modello additivo (vendite bibita QQQ)



Supponiamo che il modello di scomposizione sia additivo:

$$y_t = S_t + T_t + E_t$$

La scomposizione classica viene condotta svolgendo le fasi seguenti.

1. Calcolo del trend-ciclo di prima approssimazione. Si tratta di una fase strumentale che non produce una stima definitiva della componente *trend-ciclo*. Il *trend-ciclo di prima approssimazione* viene calcolato con una media mobile centrata a 12 termini. Indichiamo con MM_t il valore di detta media, dove $t=7, \dots, n-6$ a causa della perdita di dati all'inizio e al termine della serie.

2. Calcolo della componente $(S+E)_t$: serie della stagionalità mista e errore. Anche questa è una fase strumentale. La serie $(S+E)_t$ è calcolata sottraendo dalla serie originale, la grandezza MM_t :

$$(S+E)_t = y_t - MM_t$$

Si noti che abbiamo usato la notazione $(S+E)_t$ per indicare che, una volta sottratta la media mobile, non è possibile distinguere la parte stagionale dalla parte disturbo.

3. Stima della componente stagionale. Dalla componente $(SE)_t$ si elimina il disturbo e si perviene alla stima di S_t . Nell'approccio classico si ipotizza che l'oscillazione stagionale sia costante da anno in anno, per cui, con dati mensili, si assume che $S_t = S_{t+12} = S_{t+24} = \dots$

Si parla di *modello di stagionalità costante*¹. Il coefficiente di stagionalità S_m per il mese m ($m=1, \dots, 12$) viene calcolato effettuando la media aritmetica dei termini $(S+E)_t$ dove $t=m, m+12, m+24, \dots$. In altre parole la stima della stagionalità per gennaio è data dalla media aritmetica dei valori $(S+E)_t$ riferiti a gennaio. Il risultato di questa operazione produce 12 coefficienti di stagionalità $\hat{S}_m$, $m=1, \dots, 12$ (dove m indica il mese), che si ripetono per ogni anno. Trattandosi di un modello additivo, i valori $\hat{S}_m$ devono verificare la seguente proprietà:

$$\sum_{m=1}^{12} \hat{S}_m = 0$$

Infatti, per definizione di stagionalità nel modello additivo, le oscillazioni stagionali esauriscono il loro effetto all'interno dell'anno (ovvero, la media dei coefficienti di stagionalità deve essere uguale al valore *neutrale* che, nel modello additivo, è 0). Se i coefficienti di stagionalità calcolati mediante la media sommano ad un valore $w \neq 0$, si riaggiustano tali coefficienti sottraendo da ciascuno il valore $w/12$.

4. Derivazione della serie destagionalizzata D_t . Il dato destagionalizzato D_t è calcolato nel modello additivo come:

$$D_t = y_t - \hat{S}_t$$

La serie D_t contiene dunque il *pattern* del *ciclo-trend* e l'effetto del disturbo. Essa è pertanto utile per lo studio del *ciclo-trend*.

5. Stima del ciclo-trend. La stima $\hat{T}_t$ del *ciclo-trend* può essere ottenuta mediante una media mobile a 3 termini sui dati D_t .

6. Stima dell'intera componente sistematica della serie. Mediante le stime della stagionalità e del *trend-ciclo* si ottiene la stima $\hat{y}_t$, che contiene solo il *pattern* sistematico della serie, dove:

$$\hat{y}_t = \hat{T}_t + \hat{S}_t$$

7. Calcolo del residuo del modello. Il residuo del modello Res_t è, infine:

$$Res_t = y_t - \hat{y}_t$$

La Tab. 3.2 riporta i risultati delle fasi sopra elencate, per la serie delle vendite di bottiglie della bibita QQQ.

¹ Si può condurre un'analisi grafica per controllare la validità dell'ipotesi di stagionalità costante. Si costruisce, per ognuno dei 12 mesi, un *plot* di $(SE)_t$ in ordinata *versus* gli anni (in ascissa). Se i punti individuati si trovano collocati parallelamente all'asse delle ascisse, allora il modello di stagionalità costante è adeguato.

Tab. 3.2 – Scomposizione modello additivo (vendite bibita QQQ)

Anno	Mese	t	yt	MMt	Stima (EST)	Stima St (media)	Stima St	Dt	Stima Tt	Stima yt	Rest
1999	1	1	189				-289,5	478,5	473,0	183,5	5,5
	2	2	229				-238,5	467,5	480,7	242,2	-13,2
	3	3	249				-247,2	496,2	485,1	237,9	11,1
	4	4	289				-202,6	491,6	508,1	305,5	-16,5
	5	5	260				-276,5	536,5	493,1	216,6	43,4
	6	6	431				-20,2	451,2	454,7	434,5	-3,5
	7	7	660	450,1	209,9		283,7	376,3	395,4	679,1	-19,1
	8	8	777	455,2	321,8		418,3	358,7	340,8	759,1	17,9
	9	9	915	460,9	454,1		627,5	287,5	358,1	985,6	-70,6
	10	10	613	467,2	145,8		184,9	428,1	396,8	581,6	31,4
	11	11	485	472,8	12,2		10,3	474,7	476,7	486,9	-1,9
	12	12	277	480,2	-203,2		-250,2	527,2	511,8	261,6	15,4
2000	1	13	244	492,5	-248,5	-279,1	-289,5	533,5	531,7	242,2	1,8
	2	14	296	507,3	-211,3	-228,1	-238,5	534,5	544,7	306,2	-10,2
	3	15	319	524,8	-205,8	-236,8	-247,2	566,2	557,8	310,6	8,4
	4	16	370	540,8	-170,8	-192,2	-202,6	572,6	576,1	373,5	-3,5
	5	17	313	551,9	-238,9	-266,1	-276,5	589,5	579,4	303,0	10,0
	6	18	556	560,9	-4,9	-9,8	-20,2	576,2	571,0	550,8	5,2
	7	19	831	567,1	263,9	294,1	283,7	547,3	555,1	838,8	-7,8
	8	20	960	572,8	387,3	428,7	418,3	541,7	537,8	956,1	3,9
	9	21	1152	578,4	573,6	637,9	627,5	524,5	546,8	1174,3	-22,3
	10	22	759	583,7	175,3	195,3	184,9	574,1	565,1	750,0	9,0
	11	23	607	589,3	17,7	20,6	10,3	596,7	597,4	607,6	-0,6
	12	24	371	596,2	-225,2	-239,8	-250,2	621,2	601,8	351,6	19,4
2001	1	25	298	607,7	-309,7		-289,5	587,5	608,4	318,9	-20,9
	2	26	378	623,0	-245,0		-238,5	616,5	608,1	369,6	8,4
	3	27	373	640,8	-267,8		-247,2	620,2	627,4	380,2	-7,2
	4	28	443	656,7	-213,7		-202,6	645,6	638,8	436,1	6,9
	5	29	374	667,3	-293,3		-276,5	650,5	658,8	382,3	-8,3
	6	30	660	674,7	-14,7		-20,2	680,2	683,7	663,5	-3,5
	7	31	1004	679,8	324,3		283,7	720,3	711,7	995,4	8,6
	8	32	1153	682,8	470,2		418,3	734,7	738,5	1156,8	-3,8
	9	33	1388	685,7	702,3		627,5	760,5	738,1	1365,6	22,4
	10	34	904	688,8	215,2		184,9	719,1	728,1	913,0	-9,0
	11	35	715	691,4	23,6		10,3	704,7	705,0	715,3	-0,3
	12	36	441	695,4	-254,4		-250,2	691,2	678,5	428,3	12,7
2002	1	37	350	702,8	-352,8		-289,5	639,51	656,4	366,89	-16,9
	2	38	400	711,6	-311,6		-238,5	638,51	648,4	409,9	-9,9
	3	39	420	722,0	-302,0		-247,2	667,2	659,4	412,2	7,8
	4	40	470	732,7	-262,7		-202,6	672,62	675,4	472,8	-2,8
	5	41	410	741,0	-331,0		-276,5	686,47	699,8	423,3	-13,3
	6	42	720	748,3	-28,3		-20,2	740,18	754,3	734,1	-14,1
	7	43	1120				283,7	836,31	802,7	1086,4	33,6
	8	44	1250				418,3	831,68	860,1	1278,5	-28,5
	9	45	1540				627,5	912,45	856,1	1483,6	56,4
	10	46	1009				184,9	824,14	845,4	1030,3	-21,3
	11	47	810				10,3	799,74	798,0	808,3	1,7
	12	48	520				-250,2	770,18	785,0	534,8	-14,8
Media						10,4	0,0				

Come si vede, la media mobile a 12 termini centrata fa perdere 6 dati all'inizio e 6 alla fine della serie. I dati sono recuperati attraverso la stima della stagionalità mediante l'ipotesi di modello stagionale costante. Si può notare, infatti, che la sequenza dei coefficienti di stagionalità si ripete nei tre anni. Il segno e l'entità del coefficiente di stagionalità esprime l'importanza e la direzione dell'effetto: quando il segno del coefficiente è negativo significa che la stagionalità provoca una contrazione (rispetto al livello che avrebbe il *trend-ciclo* senza l'effetto dell'oscillazione stagionale); il segno è

positivo, quando la stagionalità amplifica il fenomeno. Coefficienti di stagionalità positivi ed elevati si presentano nei mesi più caldi.

Si precisa che, per la stima della stagionalità, abbiamo utilizzato solo i dati dei due anni centrali che hanno i dati $(ES)_t$ completi (cioè per tutti i 12 mesi). Da notare, inoltre, che è stato necessario riaggiustare le stime della stagionalità in modo che queste sommassero a zero.

Il *trend-ciclo* è stato stimato con una media mobile a tre termini dei dati destagionalizzati, con eccezione dei due termini estremi (v. valori in grassetto) per i quali è stata impiegata una media a due termini.

3.5 Scomposizione classica: il modello misto

Utilizzando ancora i dati mensili sulle vendite di bibita QQQ, deriviamo la scomposizione mediante il modello misto:

$$y_t = S_t \times T_t + E_t$$

Le fasi della scomposizione sono elencate qui di seguito.

1. Calcolo del trend-ciclo di prima approssimazione. Esso viene calcolato con una media mobile centrata a 12 termini. Indichiamo con MM_t il valore di detta media, $t=7, \dots, n-6$ (stesso procedimento del modello additivo).

2. Calcolo della componente $(SE)_t$: serie della stagionalità mista e errore. La serie $(SE)_t$, composta da stagionalità ed errore, è calcolata dividendo la serie y_t per MM_t :

$$(SE)_t = y_t / MM_t$$

3. Stima della componente stagionale. Dalla serie $(SE)_t$ si elimina il disturbo e si perviene alla stima di S_t . Si ipotizza, anche qui, che l'oscillazione stagionale sia costante di anno in anno per cui, con dati mensili, $S_t = S_{t+12} = S_{t+24} = \dots$. il coefficiente di stagionalità S_m per il mese m ($m=1, \dots, 12$) viene calcolato effettuando la media aritmetica dei termini $(SE)_t$ dove $t=m, m+12, m+24, \dots$. Ancora, la stima della stagionalità per gennaio è data dalla media aritmetica dei valori $(SE)_t$ riferiti a gennaio. Il risultato sarà dato da 12 coefficienti di stagionalità che si ripetono per ogni anno. Si ricava quindi la stima $\hat{S}_k$ $k=1, \dots, 12$ dove k indica il mese. Nel caso del modello misto, i valori $\hat{S}_k^*$ devono verificare la seguente proprietà:

$$\frac{1}{12} \sum_{k=1}^{12} \hat{S}_k = 1$$

Infatti, per definizione di stagionalità, le oscillazioni stagionali esauriscono il loro effetto all'interno dell'anno (ovvero, la media dei coefficienti di stagionalità deve essere uguale al valore *neutrale* che, nel modello moltiplicativo, è 1). Se i coefficienti di stagionalità calcolati mediante la media hanno media $w \neq 1$, si riaggiustano tali coefficienti dividendo ciascuno per il valore w .

4. *Derivazione della serie destagionalizzata D_t* Il dato destagionalizzato D_t si ricava come:

$$D_t = y_t / \hat{S}_t$$

Come si vede, a questo punto è necessario ritornare alla serie originaria, operando la trasformazione esponenziale.

Questa grandezza contiene il pattern del *ciclo-trend* e l'effetto del disturbo. Essa è utile per il successivo studio del *ciclo-trend*.

5. *Stima del ciclo-trend.* La stima del *ciclo-trend* $\hat{T}_t$ può essere ottenuta mediante una media mobile a 3 termini sui dati D_t .

6. *Stima dell'intera componente sistematica della serie.* Mediante le stime della stagionalità e del *trend-ciclo*, si ricava la stima $\hat{y}_t$ che contiene solo il *pattern* sistematico della serie, dove:

$$\hat{y}_t = \hat{T}_t \times \hat{S}_t$$

7. *Calcolo del residuo del modello.* Si ricava, infine, il residuo del modello Res_t come:

$$Res_t = y_t - \hat{y}_t$$

La Tab. 3.3 riporta i risultati delle fasi sopra elencate, per la serie delle vendite di bottiglie della bibita QQQ.

Come si vede, la media mobile a 12 termini centrata è la stessa usata nel caso del modello additivo (v. Tab. 3.2). La media mobile fa perdere 6 dati all'inizio e 6 dati alla fine del periodo di analisi. I dati sono recuperati attraverso la stima della stagionalità mediante l'ipotesi di modello stagionale costante. Si può notare, infatti, che la sequenza dei coefficienti di stagionalità si ripete nei tre anni. L'entità del coefficiente di stagionalità esprime l'importanza e la direzione dell'effetto: quando il coefficiente è minore di 1 significa che la stagionalità provoca una contrazione (rispetto al livello che avrebbe il *trend-ciclo* senza l'effetto dell'oscillazione stagionale); quando il coefficiente è maggiore di 1, la stagionalità amplifica il fenomeno. Coefficienti di stagionalità maggiori di 1 si presentano nei mesi più caldi.

Si precisa che, per la stima della stagionalità, abbiamo utilizzato solo i dati dei due anni centrali che hanno i valori $(ES)_t$ completi (cioè per tutti i 12 mesi). Da notare, inoltre, che è stato necessario riaggiustare le stime della stagionalità in modo che queste avessero media uguale a 1.

Il *trend-ciclo* è stato stimato con una media mobile a tre termini dei dati destagionalizzati, con eccezione dei due termini estremi (v. valori in grassetto) per i quali è stata impiegata una media a due termini.

Tab. 3.3 – Scomposizione con modello misto (vendite bibita QQQ)

Anno	Mese	t	yt	MMt	Stima (Est)	Stima St (media)	Stima St	Dt	Stima Tt	Stima yt	Rest
1999	1	1	189				0,491	384,7	385,4	189,3	-0,3
	2	2	229				0,593	386,0	396,9	235,4	-6,4
	3	3	249				0,593	419,9	410,9	243,7	5,3
	4	4	289				0,677	426,8	436,4	295,5	-6,5
	5	5	260				0,562	462,6	442,8	248,9	11,1
	6	6	431				0,982	439,1	450,6	442,3	-11,3
	7	7	660	450,1	1,47		1,466	450,1	450,8	661,1	-1,1
	8	8	777	455,2	1,71		1,677	463,4	456,9	766,1	10,9
	9	9	915	460,9	1,99		2,001	457,2	463,8	928,2	-13,2
	10	10	613	467,2	1,31		1,302	470,8	466,5	607,4	5,6
	11	11	485	472,8	1,03		1,029	471,5	461,5	474,8	10,2
	12	12	277	480,2	0,58		0,626	442,4	470,2	294,4	-17,4
2000	1	13	244	492,5	0,50	0,493	0,491	496,7	479,3	235,5	8,5
	2	14	296	507,3	0,58	0,595	0,593	499,0	511,2	303,2	-7,2
	3	15	319	524,8	0,61	0,595	0,593	537,9	527,8	313,0	6,0
	4	16	370	540,8	0,68	0,679	0,677	546,4	547,1	370,5	-0,5
	5	17	313	551,9	0,57	0,564	0,562	557,0	556,6	312,8	0,2
	6	18	556	560,9	0,99	0,985	0,982	566,5	563,4	553,0	3,0
	7	19	831	567,1	1,47	1,471	1,466	566,7	568,5	833,7	-2,7
	8	20	960	572,8	1,68	1,682	1,677	572,5	571,6	958,5	1,5
	9	21	1152	578,4	1,99	2,008	2,001	575,6	577,0	1154,8	-2,8
	10	22	759	583,7	1,30	1,306	1,302	582,9	582,8	758,9	0,1
	11	23	607	589,3	1,03	1,032	1,029	590,0	588,5	605,4	1,6
	12	24	371	596,2	0,62	0,628	0,626	592,5	596,4	373,4	-2,4
2001	1	25	298	607,7	0,49		0,491	606,6	612,1	300,7	-2,7
	2	26	378	623,0	0,61		0,593	637,2	624,3	370,3	7,7
	3	27	373	640,8	0,58		0,593	629,0	640,1	379,6	-6,6
	4	28	443	656,7	0,67		0,677	654,2	649,5	439,9	3,1
	5	29	374	667,3	0,56		0,562	665,5	664,0	373,2	0,8
	6	30	660	674,7	0,98		0,982	672,4	674,2	661,7	-1,7
	7	31	1004	679,8	1,48		1,466	684,7	681,5	999,4	4,6
	8	32	1153	682,8	1,69		1,677	687,6	688,6	1154,7	-1,7
	9	33	1388	685,7	2,02		2,001	693,5	691,8	1384,5	3,5
	10	34	904	688,8	1,31		1,302	694,2	694,3	904,0	0,0
	11	35	715	691,4	1,03		1,029	695,0	697,8	717,9	-2,9
	12	36	441	695,4	0,63		0,626	704,3	703,9	440,8	0,2
2002	1	37	350	702,8	0,50		0,491	712,4	697,0	342,4	7,6
	2	38	400	711,6	0,56		0,593	674,3	698,3	414,2	-14,2
	3	39	420	722,0	0,58		0,593	708,2	692,2	410,5	9,5
	4	40	470	732,7	0,64		0,677	694,0	710,6	481,2	-11,2
	5	41	410	741,0	0,55		0,562	729,6	719,0	404,1	5,9
	6	42	720	748,3	0,96		0,982	733,5	742,3	728,6	-8,6
	7	43	1120				1,466	763,8	747,6	1096,3	23,7
	8	44	1250				1,677	745,4	759,6	1273,7	-23,7
	9	45	1540				2,001	769,5	763,3	1527,6	12,4
	10	46	1009				1,302	774,9	777,2	1012,1	-3,1
	11	47	810				1,029	787,4	797,6	820,5	-10,5
	12	48	520				0,626	830,4	808,9	506,5	13,5
Media						1,003	1,000				

3.6 Valutazione della scomposizione ottenuta

Abbiamo già accennato (par. 2.4) che la valutazione dell'adattamento ottenuto mediante il modello di scomposizione può essere condotta mediante indici quali MSE, MAE, MAPE, riferiti alla serie storica disponibile. Oltre al calcolo di tali indici, è buona norma condurre anche delle analisi grafiche dei residui Res_t . L'idea che sta alla base di questi controlli è la seguente: se la scomposizione è buona allora il residuo non dovrebbe evidenziare oscillazioni sistematiche di nessun tipo ma dovrebbe essere assimilabile ad un disturbo casuale (come è la natura di E_t nel modello di scomposizione ipotizzato).

I grafici che ci possono interessare sono i seguenti.

1. *Andamento di Res_t rispetto al tempo t .* L'ideale è che non si presentino oscillazioni sistematiche, come avviene in Fig. 3.3. Situazione dubbia è quella di Fig. 3.4 (la scomposizione ottenuta è più efficace per periodi più remoti). La Fig. 3.5, infine, evidenzia che non siamo stati in grado di individuare un andamento ciclico (o comunque curvilineo)

2. *Normal probability plot di Res_t .* L'idea che giustifica questo tipo di analisi è la seguente: se siamo stati in grado di scomporre bene la serie e abbiamo colto in modo adeguato la componente sistematica, il residuo dovrebbe comportarsi approssimativamente come un errore puramente accidentale di tipo normale.

3. *Andamento dei dati destagionalizzati rispetto al tempo t .* I dati destagionalizzati, in quanto composti solo da trend-ciclo e errore, non devono mostrare i picchi tipici della stagionalità.

Fig. 3.3 – Esempio di residui privi di andamenti sistematici

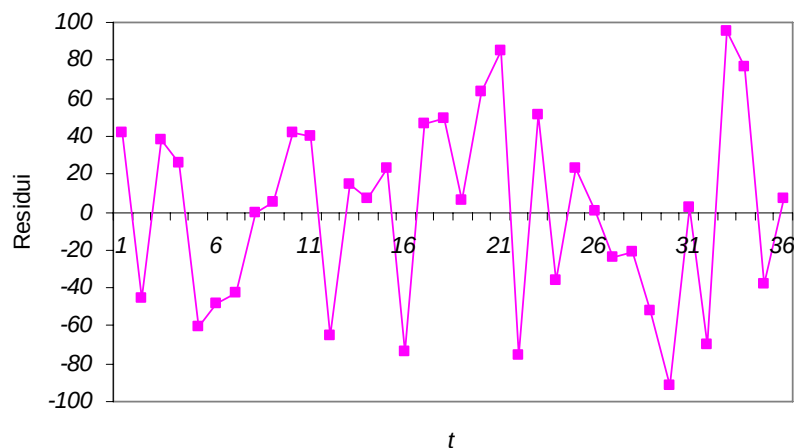


Fig. 3.4 – Esempio di residui con variabilità più alta per periodi recenti

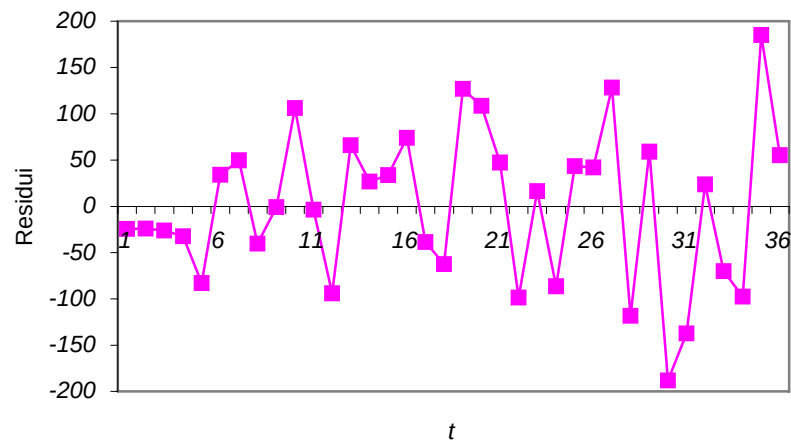
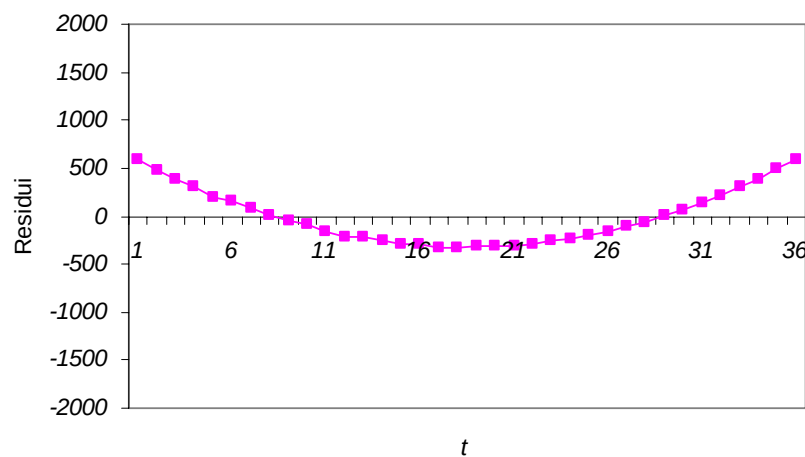


Fig. 3.5 – Esempio di residui con andamento curvilineo



Applichiamo queste regole ai risultati della scomposizione ottenuta col modello additivo e con quello misto per la serie relativa alla bibita QQQ.

Il calcolo del MAPE, 2.78% per il modello additivo, 1.38% per quello misto, mostra la superiorità del secondo. Ciò viene ulteriormente confermato dal *time plot* dei residui riportato in Fig. 3.6. I residui del modello additivo mostrano sono più ampi nel periodo iniziale e finale della serie. Ciò viene chiarito esaminando i dati destagionalizzati (Fig. 3.7). Infatti, il modello additivo tende a destagionalizzare peggio soprattutto agli estremi della serie: destagionalizzato *troppo* all'inizio della serie e *poco* alla fine della serie. Ciò è tipico con serie di tipo moltiplicativo o misto.

Fig. 3.6 – Residui dei modelli additivo e misto (Tabb. 3.2, 3.3)

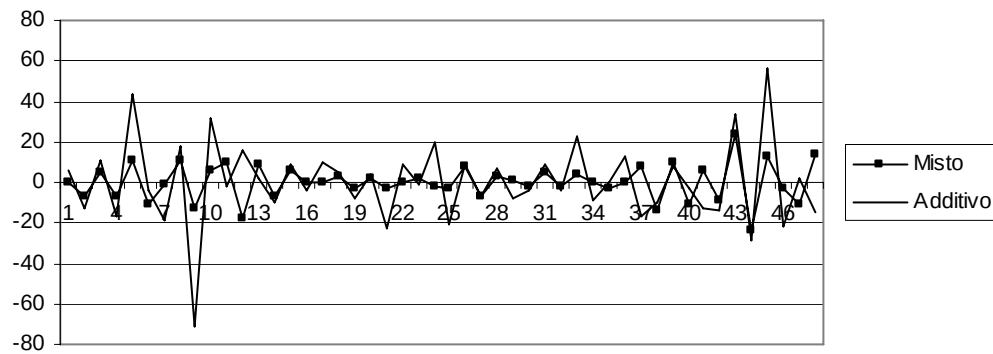
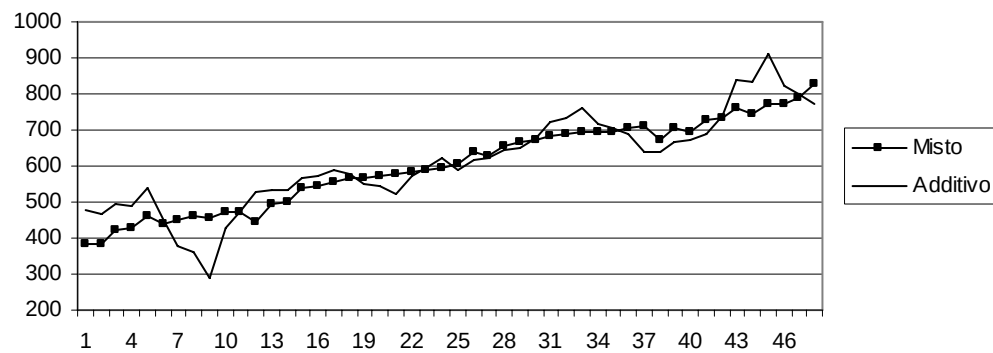
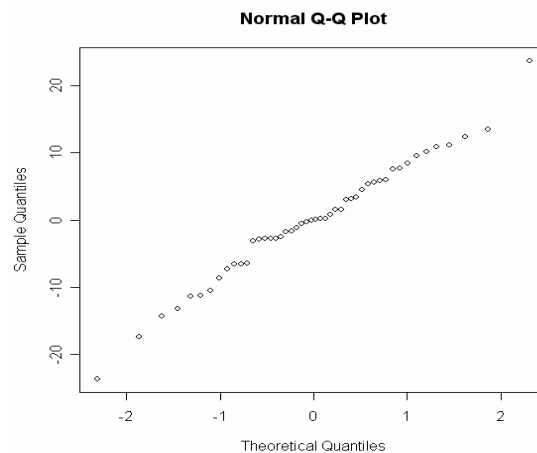


Fig. 3.7 – Serie destagionalizzata (Tabb. 3.2, 3.3)



Vediamo, infine, il *normal probability plot* dei residui del modello misto, avendo preferito questo al modello additivo.

Fig. 3.8 – Normal plot dei residui da modello misto (Tab. 3.3)



Come si è visto, la prima stima ottenuta nella fase di scomposizione è quella della stagionalità (infatti la media mobile iniziale dà solo una stima di prima approssimazione del ciclo-trend, dando luogo alla perdita di dati). Pertanto, prima di procedere alla stima del ciclo-trend, è preferibile controllare subito l'andamento dei dati destagionalizzati.

3.7 La previsione

A conclusione di questo capitolo, vogliamo soffermarci a descrivere come si eseguono le previsioni una volta che si è condotta un'analisi di scomposizione.

Il metodo classico di scomposizione propone uno studio della serie storica e solo con ulteriori strumenti analitici è possibile arrivare a produrre delle previsioni. Ad ogni modo, è necessario ottenere previsioni delle singole componenti sistematiche (*trend*, *ciclo*, *stagionalità*) per poi ricomporre il dato futuro della serie: come prodotto dei valori delle singole componenti sistematiche nel caso del modello misto o moltiplicativo, come somma nel caso del modello additivo.

Non sempre questo procedimento produce risultati soddisfacenti. In generale, il *trend-ciclo* è la componente più difficile da prevedere soprattutto a causa dell'oscillazione ciclica essere più agevole se il ciclo è assente oppure ha intensità trascurabile rispetto all'ordine di grandezza del *trend*. In tale circostanza, il *trend* può essere validamente stimato mediante una funzione analitica del tempo $f(t)$. La previsione del trend $\hat{T}_{n+1}$ per il tempo futuro $t=n+1$ (dove n è la lunghezza della serie disponibile) viene ottenuta estrapolando la retta, per cui $\hat{T}_{n+1}=f(n+1)$.

La componente stagionale è relativamente semplice da prevedere nell'ipotesi di stagionalità costante. Infatti, l'effetto stimato della stagionalità del mese (trimestre, semestre, ecc.) m ($m=1, \dots, 12$) è $\hat{S}_m$, che è costante negli anni. Quindi, supponendo che il tempo $n+1$ sia riferito al mese di gennaio, la previsione dell'effetto stagionale è $\hat{S}_1$.

La previsione F_{n+1} per il fenomeno in esame, si ottiene ricomponendo le previsioni per le due componenti sistematiche. Supponendo che il tempo $n+1$ sia riferito al mese di gennaio, si ottiene

$$\begin{aligned} F_{n+1} &= \hat{T}_{n+1} + \hat{S}_1 && \text{per il modello additivo} \\ F_{n+1} &= \hat{T}_{n+1} \times \hat{S}_1 && \text{per il modello misto o moltiplicativo} \end{aligned}$$

Nel capitolo seguente affronteremo la stima del *trend* mediante adattamento di una forma analitica e deriveremo le previsioni per le vendite della bibita QQQ.

Anche se per la previsione sono stati sviluppati metodi più sofisticati e più efficaci, i procedimenti di scomposizione restano un valido strumento per comprendere le caratteristiche evolutive passate della serie.

4. Studio del trend mediante forma analitica

4.1 Introduzione

L'evoluzione di lungo periodo di una serie storica è denominata *trend*. Nell'economia, ad esempio, il *trend* è determinato dal lento sviluppo delle tecnologie, dei fenomeni demografici e sociali, ecc.

L'esistenza di una evoluzione di lungo periodo può essere evidenziata dall'andamento dei dati destagionalizzati risultanti da un'analisi di scomposizione, oppure dalla serie di dati annuali (che, come si sa, sono privi della stagionalità).

Nel capitolo precedente abbiamo illustrato la stima del *trend* mediante le medie mobili; tale procedimento è denominato *adattamento locale del trend* o *stima locale*. In questo capitolo affronteremo la stima del *trend* mediante specificazione e stima di una funzione analitica del tempo t . Questo procedimento è denominato *analisi globale* poiché la funzione stimata definisce come una sorta di legge di evoluzione del *trend* rispetto al tempo t . Varie forme funzionali sono utilizzate per rappresentare il *trend*. Quelle che vedremo in queste note sono: la forma lineare, la quadratica e l'esponenziale.

4.2 Forme analitiche per rappresentare il trend

Ipotizziamo che $y_t = T_t + e_t$, dove y_t è una serie priva di stagionalità (ovvero può essere il dato annuale o quello destagionalizzato) e e_t la componente di disturbo.

La forma lineare nel tempo t è:

$$(4.1) \quad T_t = \beta_0 + \beta_1 t \quad t=1, \dots, n$$

dove β_0 è l'intercetta e β_1 è la pendenza della retta. Se $\beta_1 > 0$ il *trend* è crescente rispetto al tempo; se $\beta_1 < 0$, il *trend* è decrescente; se $\beta_1 = 0$ esiste un *pattern* orizzontale. In genere, t è rappresentato dalla serie di numeri interi $1, 2, \dots, n$.

Avendo a disposizione una serie priva di stagionalità, un modo per capire se la forma lineare è appropriata per rappresentare il trend consiste nel verificare se le differenze successive dei valori della serie sono approssimativamente costanti rispetto a t . Ciò scaturisce dal fatto che, se vale la (4.1), si ha:

$$(4.2) \quad \Delta_t = T_t - T_{t-1} = \beta_0 + \beta_1 t - \beta_0 - \beta_1(t-1) = \beta_1 \quad \text{costante in } t.$$

Talvolta il *trend* può esibire andamenti non lineari rispetto al tempo. Allora può essere appropriata la forma quadratica:

$$(4.3) \quad T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad t=1, \dots, n$$

La (4.2) può assumere una varietà di forme a seconda del segno dei coefficienti β_1 e β_2 ; da notare, ad esempio, che quando $\beta_2 = 0$ si ottiene una forma lineare in t .

Un modo per verificare l'adeguatezza di un modello quadratico consiste nell'effettuare le seconde differenze fra termini successivi della serie (cioè le differenze successive della serie Δ_t definita in (4.2)). Se i valori di tali differenze appaiono

approssimativamente costanti rispetto a t , può essere valido un modello quadratico. Ciò scaturisce dal fatto che, se vale la (4.3), allora $\Delta_{t+1}-\Delta_t$ è costante in t . Infatti, se vale la (4.3), si ha:

$$\Delta_{t+1}-\Delta_t=2\beta_2 \quad \text{costante in } t$$

dove:

$$\Delta_t = T_t - T_{t-1} = \beta_0 + \beta_1 t + \beta_2 t^2 - \beta_0 - \beta_1(t-1) - \beta_2(t-1)^2$$

$$\Delta_{t+1} = T_{t+1} - T_t = \beta_0 + \beta_1(t+1) + \beta_2(t+1)^2 - \beta_0 - \beta_1 t - \beta_2 t^2$$

Un'altra forma spesso usata è quella esponenziale:

$$(4.4) \quad T_t = \beta_0 \exp(\beta_1 t) \quad t=1, \dots, n$$

Se vale la (4.4), si verifica la seguente espressione:

$$\frac{T_t}{T_{t-1}} = \frac{\beta_0 \exp(\beta_1 t)}{\beta_0 \exp(\beta_1(t-1))} = \exp(\beta_1) \quad \text{costante in } t$$

Un modo per verificare l'adeguatezza di un modello esponenziale consiste pertanto nell'effettuare i rapporti fra termini successivi della serie e verificare se questi sono approssimativamente costanti rispetto a t . Alternativamente, possiamo effettuare la trasformazione logaritmica della serie e verificare se le differenze successive sono approssimativamente costanti. Infatti, se operiamo la trasformazione logaritmica, la (4.4.) diventa una forma lineare. Infatti:

$$\ln T_t = \ln \beta_0 + \beta_1 t \quad t=1, \dots, n$$

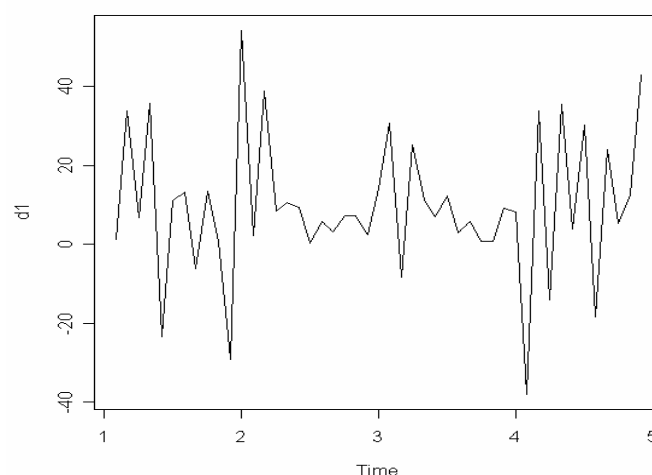
per cui verificare l'appropriatezza di una forma esponenziale per la serie equivale a verificare l'appropriatezza di una forma lineare per la trasformata logaritmica.

4.3 Stima del trend

Una volta che è stata scelta una forma analitica per rappresentare il trend, è necessario passare alla sua stima a partire da dati di osservazione. E' importante ribadire che i dati sui quali viene stimato il trend devono essere privi di andamento stagionale e ciclico. In altre parole, l'unica componente sistematica presente nei dati deve essere quella tendenziale di lungo periodo. In assenza di significative oscillazioni cicliche, i dati più idonei all'analisi del trend sono: i valori destagionalizzati oppure la serie di dati annuali.

A scopi di esemplificazione, viene condotta un'analisi del trend sui dati relativi alle vendite di bibita QQQ. Supponiamo quindi di avere già derivato le stime della stagionalità e di aver verificato, dal grafico di Fig. 3.7, che il modello moltiplicativo è quello migliore. Sui dati destagionalizzati mediante il modello moltiplicativo verifichiamo la validità di un modello lineare mediante le differenze successive (Fig. 4.1). Come si vede, il *time plot* delle differenze prime evidenzia un *pattern* stazionario e quindi la forma lineare può essere appropriata.

Tab. 4.1 – Time plot delle differenze prime dei dati destagionalizzati (mod. moltip.)



Mediante il metodo dei minimi quadrati ordinari, la funzione analitica che rappresenta il trend è in tal caso: 392.988 8.539

$$(4.5) \quad \hat{T}_t = 393 + 8.54t$$

che registra un indice di determinazione lineare R^2 pari a 0,976.

Nello studio del trend mediante funzione analitica, viene usato il metodo dei minimi quadrati come in una consueta analisi di regressione. Tuttavia non si procede alla verifica delle ipotesi sui coefficienti dal momento che la stima del modello ha lo scopo di *descrivere* l'andamento del trend rispetto al tempo e non quello di valutare l'effetto di t sul trend, nel senso causale o esplicativo.

4.4 Stima della componente sistematica e previsione

Mediante la stima della stagionalità ottenuta per il modello moltiplicativo (v. Tab. 3.3) e la stima del *trend* mediante la funzione (4.5), è possibile proporre una stima dell'intera componente sistematica del modello moltiplicativo che è:

$$\hat{y}_t = \hat{T}_t \times \hat{S}_t = (393 + 8.54t) \times \hat{S}_t$$

I residui della scomposizione sono:

$$Res_t = y_t - \hat{y}_t$$

mediante i quali si può calcolare ad esempio il MAPE. In questo caso, il MAPE è pari a 2.47%, valore più alto di quello ottenuto mediante l'impiego della media mobile a 3 termini per la stima del *trend* (v. paragrafo 3.6).

Questo tipo di scomposizione, che utilizza una forma analitica per rappresentare il *trend*, ci consente in modo estremamente semplice di proporre delle previsioni per i periodi futuri. E' necessario, però, ipotizzare che il modello stimato sui dati passati rimanga valido anche nel futuro. Per non rischiare troppo, conviene limitarsi ad effettuare delle previsioni nel breve periodo.

A titolo esemplificativo, proponiamo qui le previsioni delle vendite di bibita QQQ, per i primi tre mesi del 2003: gennaio, febbraio e marzo ai quali corrisponde $t=49, 50, 51$ e coefficienti di stagionalità pari a 0,491, 0,593, 0,593. Le previsioni che indichiamo come F_{49}, F_{50}, F_{51} sono:

$$F_{49}=(393+8,54 \times 49) \times 0,491= 398,4$$

$$F_{50}=(393+8,54 \times 50) \times 0,593= 486.3$$

$$F_{51}=(393+8,54 \times 51) \times 0,593= 491.3$$

Come si vede le ipotesi di validità nel futuro del modello riguardano:

- la validità, anche nel futuro, del modello moltiplicativo;
- la validità, anche nel futuro, della forma lineare per rappresentare il trend e la stabilità dei valori dei relativi parametri;
- la validità, anche nel futuro, del modello di stagionalità costante e la stabilità dei coefficienti stagionali per i tre mesi futuri, per i quali si costruisce la previsione.