

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..

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

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 .

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

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

Un modello additivo 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 moltiplicativo. Molte serie economiche esibiscono fluttuazioni stagionali che crescono all'aumentare del livello della serie; per tale motivo, in ambito economico, il modello moltiplicativo 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*.

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:

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

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

3.3 Rappresentazioni grafiche negli approcci di scomposizione

La Fig. 3.1 mostra i principali grafici derivanti da una analisi di scomposizione mediante un modello additivo. I risultati fanno riferimento ai dati di Tab. 2.4 (vendite di bottiglie della bibita QQQ).

Richiamiamo l'attenzione soprattutto sui seguenti grafici:

- (i) il *time plot* della serie *destagionalizzata* ovvero depurata dalla stagionalità (*seasonally adjusted data*), riportato in Fig. 3.1;
- (ii) il *time plot* della componente stagionale (Fig. 3.2);
- (iii) la serie storica dei residui (Fig. 3.3);
- (iv) il *time plot* della serie originale e della serie stimata (Fig. 3.5).

L'andamento dei dati destagionalizzati dovrebbe mostrare una linea abbastanza liscia, priva di quelle oscillazioni regolari e marcate tipiche della stagionalità della serie. La Fig. 3.1 mostra che la destagionalizzazione mediante il modello additivo non è soddisfacente: si tende a destagionalizzare 'troppo' nel primo anno (infatti, si può notare una inversione delle punte per il mese di giugno e di settembre); si destagionalizza 'poco' nell'ultimo anno (si nota, infatti, che rimane un picco molto elevato). Il motivo risiede nel fatto che il modello additivo assume implicitamente che il *range* di variazione delle oscillazioni stagionali all'interno dell'anno rimanga costante mentre, nel caso in esame, tende ad aumentare dal 1999 al 2001 (Fig. 2.5). Pertanto la stima della stagionalità, rappresentata nella Fig. 3.2, non è da ritenersi valida.

Un'altra verifica sul modello di scomposizione adottato ci proviene dal grafico dei residui. Se la scomposizione è valida allora i residui devono presentare un andamento accidentale rispetto al tempo. Dalla Fig. 3.3 si nota invece la presenza di una certa ciclicità degli stessi: i residui sono più vicini a zero nella parte centrale della serie mentre sono maggiori (in valore assoluto) alle estremità. L'andamento dei residui può essere apprezzato anche dalla Fig. 3.4, in cui si riporta la serie osservata e quella stimata, nell'ipotesi di aver stimato il trend mediante una funzione lineare di t (su questo punto di vedrà più avanti il procedimento).

Nel seguito, quando verrà illustrato in dettaglio il procedimento di scomposizione della serie, vedremo anche i risultati del modello moltiplicativo.

Fig. 3.1 – Dati destagionalizzati con modello additivo (dati di Tab. 2.4)

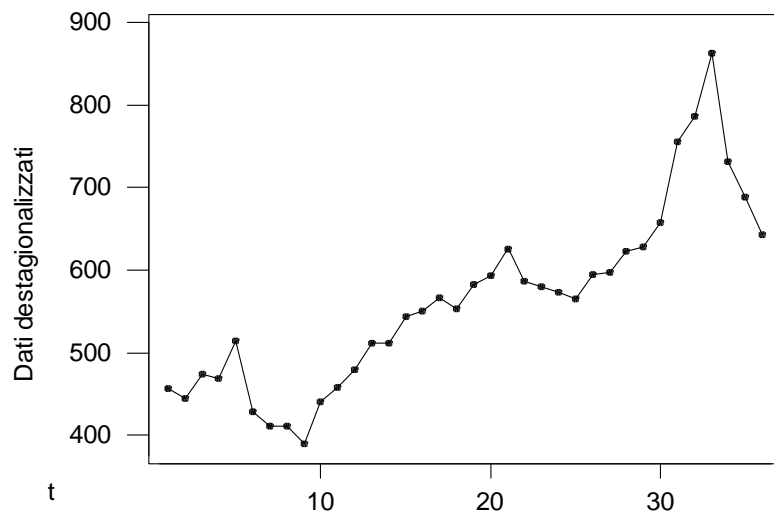


Fig. 3.2 – Stagionalità stimata con modello additivo (dati di Tab. 2.4)

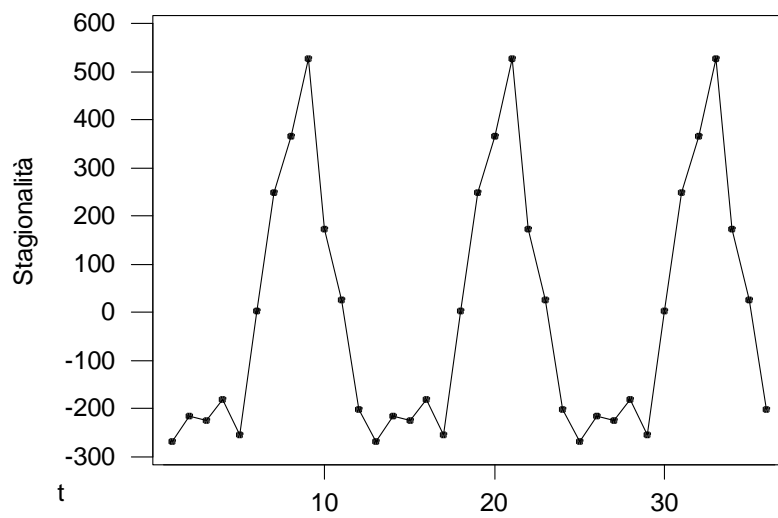


Fig. 3.3 – Residui del modello additivo (dati di Tab. 2.4)

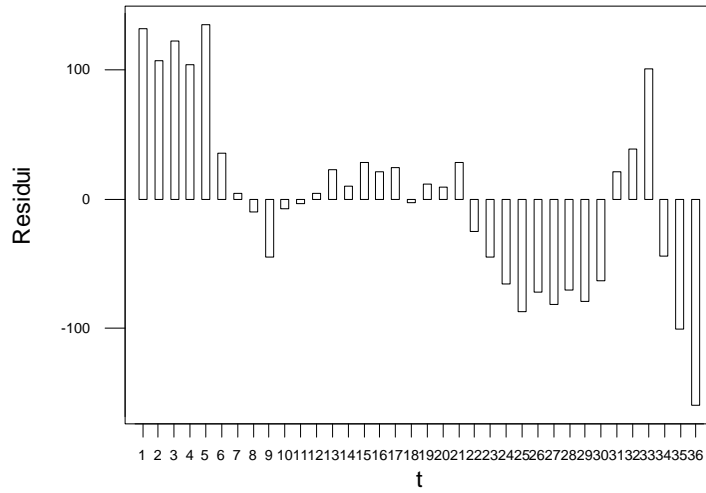
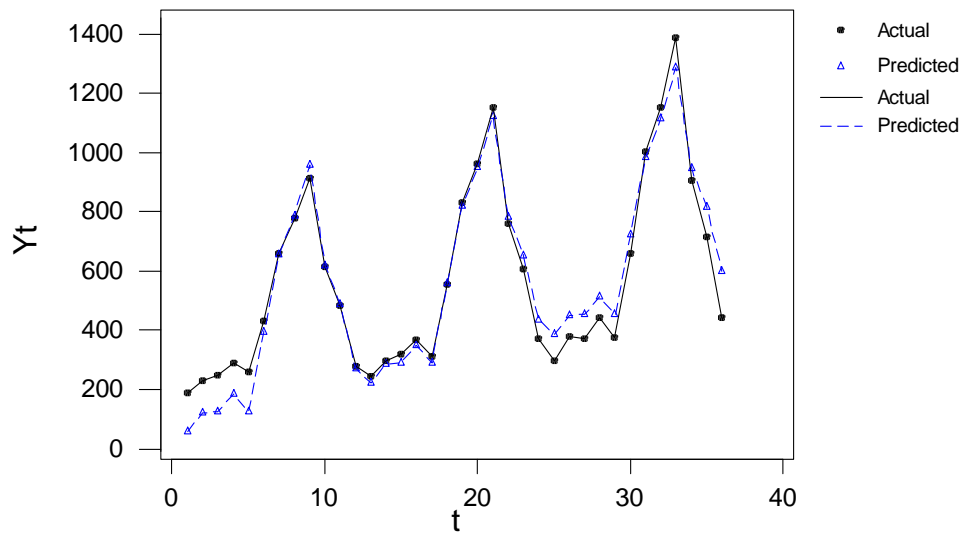


Fig. 3.4 – Dati stimati e osservati (serie di Tab. 2.4; modello additivo)



A conclusione di questo paragrafo, vogliamo puntualizzare il significato di *serie destagionalizzata*. Nel caso di un modello additivo, il dato destagionalizzato D_t è derivato come:

$$(3.1) \quad D_t = y_t - S_t = y_t + E_t$$

mentre nel modello moltiplicativo:

$$(3.2) \quad D_t = y_t / S_t = y_t \times E_t$$

una volta che è stata stimata la componente stagionale S_t .

3.4 La media mobile

La media mobile è un semplice metodo che smussa (liscia, peregua) 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 p , allora una media mobile di ampiezza p 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:

$$(3.3) \quad T_t = (y_{t-1} + y_t + y_{t+1}) / 3, \quad t = 2, \dots, n-1$$

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 crea una serie di valori smussati di lunghezza pari alla serie originaria, ognuno in corrispondenza del punto di osservazione t .

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

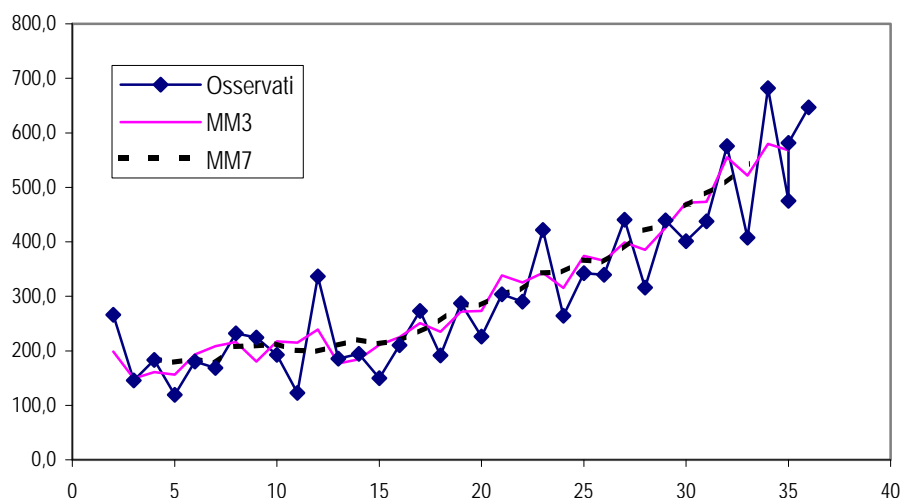
Mese	y_t	MM3	MM5	MM7
1	266,0			
2	145,9	198,3		
3	183,1	149,4	178,9	
4	119,3	160,9	159,4	185,0
5	180,3	156,0	176,6	179,1
6	168,5	193,5	184,9	185,8
7	231,8	208,3	199,6	177,2
8	224,5	216,4	188,1	208,2
9	192,8	180,1	221,7	209,0
10	122,9	217,4	212,5	212,7
11	336,5	215,1	206,5	200,9
12	185,9	238,9	197,8	198,9
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. 3.6 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. 3.5 – 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:

$$(3.4) \quad T_3 = (y_1 + 2y_2 + 2y_3 + 2y_4 + y_5)/8$$

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.

3.5 Scomposizione classica: il modello additivo

Riferendoci ai dati mensili sulle vendite di bottiglie 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 $(SE)_t$: serie della stagionalità mista e errore. Anche questa è una fase strumentale. La serie $(SE)_t$ è calcolata sottraendo dalla serie originale, la grandezza MM_t :

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

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, $S_t = S_{t+12} = S_{t+24} = \dots$. 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 $(SE)_t$, dove $t=m, m+12, m+24, \dots$. In altre parole la stima della stagionalità per gennaio è data dalla media aritmetica dei valori (SE) 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. 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.

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

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

e $\hat{S}_t = \hat{S}_m$ se t si riferisce al mese m . 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* è 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 \hat{E}_t è, infine:

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

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 – Risultati del modello additivo (vendite bibita QQQ)

ANNO MESE	t	y_t	Stima S_t	Stima D_t	Stima T_t	Stima y_t	\hat{E}_t
1999	1	189	-267,0	456,0	450,5	184	5,50
	2	229	-216,0	445,0	450,5	235	-5,50
	3	249	-224,7	473,7	458,2	234	15,46
	4	289	-180,1	469,1	462,6	282	6,51
	5	260	-253,9	513,9	485,6	232	28,38
	6	431	2,3	428,7	470,6	473	-41,91
	7	660	249,0	411,0	451,2	700	-40,22
	8	777	366,7	410,3	416,7	783	-6,31
	9	915	526,0	389,0	403,4	929	-14,42
	10	613	172,7	440,3	413,2	586	27,09
	11	485	27,1	457,9	429,1	456	28,82
	12	277	-202,1	479,1	459,1	257	19,96
2000	1	244	-267,0	511,0	482,7	216	28,34
	2	296	-216,0	512,0	500,7	285	11,31
	3	319	-224,7	543,7	522,2	298	21,46
	4	370	-180,1	550,1	535,3	355	14,84
	5	313	-253,9	566,9	553,6	300	13,38
	6	556	2,3	553,7	556,9	559	-3,24
	7	831	249,0	582,0	567,5	817	14,44
	8	960	366,7	593,3	576,3	943	17,02
	9	1152	526,0	626,0	600,4	1126	25,58
	10	759	172,7	586,3	601,9	775	-15,58
	11	607	27,1	579,9	597,4	625	-17,51
	12	371	-202,1	573,1	579,8	378	-6,71
2001	1	298	-267,0	565,0	572,7	306	-7,66
	2	378	-216,0	594,0	577,3	361	16,65
	3	373	-224,7	597,7	585,6	361	12,13
	4	443	-180,1	623,1	604,9	425	18,17
	5	374	-253,9	627,9	616,2	362	11,71
	6	660	2,3	657,7	636,2	639	21,42
	7	1004	249,0	755,0	680,2	929	74,78
	8	1153	366,7	786,3	733,0	1100	53,35
	9	1388	526,0	862,0	801,1	1327	60,92
	10	904	172,7	731,3	793,2	966	-61,91
	11	715	27,1	687,9	760,4	788	-72,51
	12	441	-202,1	643,1	665,5	463	-22,43

Come si vede, si perdono 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.

Il *trend-ciclo* è stato stimato con una media 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.

In base al procedimento descritto nella Tab. 3.2 si ottiene un MAPE pari al 2,8%. Tuttavia (v. Fig. 2.1, che si riferisce a questo stesso procedimento, ma si veda anche la Fig. 3.6), il modello additivo non realizza una soddisfacente destagionalizzazione dei dati.

3.6 Scomposizione classica: il modello moltiplicativo

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

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

La scomposizione classica viene condotta svolgendo le fasi seguenti.

1. Calcolo del trend-ciclo di prima approssimazione. Come accade per il modello moltiplicativo, 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}_m $m=1, \dots, 12$ dove m indica il mese. I valori \hat{S}_m devono verificare la seguente proprietà:

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

Infatti, per definizione di stagionalità, le oscillazioni stagionali esauriscono il loro effetto all'interno dell'anno.

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

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

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 è 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 \hat{E}_t come:

$$\hat{E}_t = y_t / \hat{y}_t$$

Tuttavia, per consentire un confronto con l'adattamento del modello additivo, ai fini del calcolo degli indici MAPE, MAE, ecc., conviene utilizzare i residui calcolati nel modo consueto:

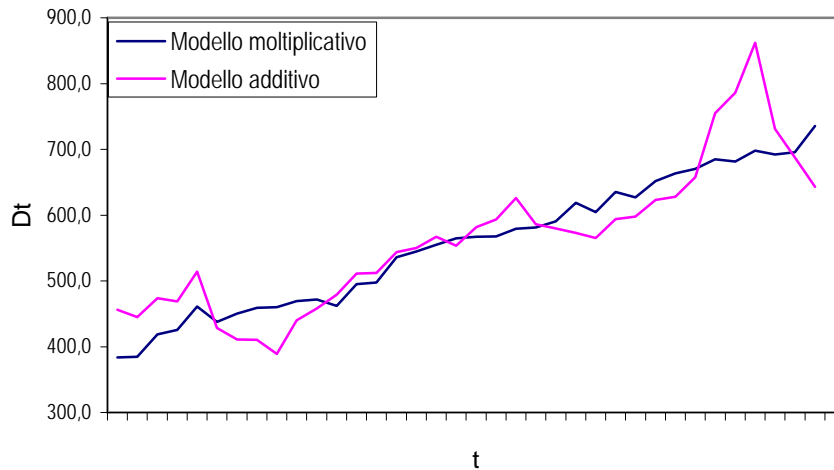
$$Res_t = y_t - \hat{y}_t$$

I dettagli della scomposizione mediante il modello moltiplicativo sono riportati in Tab. 3.3. Vale la pena notare che il modello moltiplicativo realizza un miglioramento del MAPE che è uguale a 1,8% e, soprattutto, una più che soddisfacente destagionalizzazione dei dati. Ciò non ci sorprende in quanto, già l'analisi preliminare della serie aveva messo in luce la maggiore adeguatezza della forma moltiplicativa.

Tab. 3.3 – Risultati del modello moltiplicativo (vendite bibita QQQ)

ANNO	MESE	t	y_t	Stima S_m	Stima Dt	Stima T_t	Stima y_t	Res_t
1999	1	1	189	0,493	383	384	189	-0,326
	2	2	229	0,595	385	384	229	0,394
	3	3	249	0,595	419	396	235	13,642
	4	4	289	0,679	425	410	278	10,744
	5	5	260	0,564	461	435	245	14,735
	6	6	431	0,985	438	441	435	-3,666
	7	7	660	1,466	450	450	659	0,828
	8	8	777	1,692	459	449	760	17,348
	9	9	915	1,988	460	457	908	7,108
	10	10	613	1,306	469	463	605	8,315
	11	11	485	1,028	472	467	480	4,853
	12	12	277	0,600	462	468	280	-3,424
2000	1	13	244	0,493	495	476	235	9,247
	2	14	296	0,595	497	485	289	7,475
	3	15	319	0,595	536	510	303	15,852
	4	16	370	0,679	545	526	357	12,603
	5	17	313	0,564	555	545	307	5,550
	6	18	556	0,985	565	555	546	9,677
	7	19	831	1,466	567	562	824	6,875
	8	20	960	1,692	568	566	958	2,001
	9	21	1152	1,988	579	571	1136	16,081
	10	22	759	1,306	581	576	752	6,650
	11	23	607	1,028	591	584	600	7,046
	12	24	371	0,600	619	597	358	13,183
2001	1	25	298	0,493	605	605	298	-0,009
	2	26	378	0,595	635	620	369	9,306
	3	27	373	0,595	627	622	370	2,798
	4	28	443	0,679	652	638	433	9,511
	5	29	374	0,564	663	647	365	8,966
	6	30	660	0,985	670	662	652	8,227
	7	31	1004	1,466	685	673	986	17,742
	8	32	1153	1,692	682	679	1148	4,574
	9	33	1388	1,988	698	688	1368	19,564
	10	34	904	1,306	692	691	902	1,963
	11	35	715	1,028	696	695	715	0,341
	12	36	441	0,600	736	716	429	11,979

Fig. 3.6 – Dati destagionalizzati (vendite di bibita QQQ)



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 \hat{E}_t . L'idea che sta alla base di questi controlli è la seguente: se la scomposizione è valida allora il residuo non dovrebbe evidenziare oscillazioni sistematiche di nessun tipo e il suo *line plot* dovrebbe oscillare intorno al valore neutrale (0 per il residuo del modello additivo, 1 per il residuo del modello moltiplicativo), in modo accidentale. Vediamo il caso del modello additivo dove il residuo è:

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

I grafici che ci possono interessare sono i seguenti.

1. *Andamento di \hat{E}_t rispetto al tempo t .* L'ideale è che non si presentino oscillazioni sistematiche, come avviene in Fig. 3.7. Situazione dubbia è quella di Fig. 3.8 (la scomposizione ottenuta è più valida per periodi più remoti). La Fig. 3.9, infine, evidenzia che non siamo stati in grado di individuare un andamento ciclico (o comunque curvilineo)
2. *Correlogramma di \hat{E}_t .* L'ideale è quello di non avere valori elevati dell'autocorrelazione tuttavia, difficilmente da un'analisi di scomposizione, si ottengono residui ottimali.

3. Normal probability plot di \hat{E}_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 accidentale.

Nel caso di modello moltiplicativo, resta in massima parte valido quanto affermato nei punti 1 e 2. Non ha senso, invece, costruire un *normal probability plot* perché il residuo moltiplicativo \hat{E}_t può assumere solo valori non negativi mentre una v.c. normale assume valori sia negativi, sia positivi. E' più corretto, se mai, costruire il *normal probability plot* delle trasformate logaritmiche dei residui.

La Fig. 3.10 riporta il *line plot* dei residui del modello additivo, per i dati relativi alle vendite della bibita QQQ. Si nota un risultato un po' insoddisfacente per il periodo iniziale e finale; sappiamo che ciò è dovuto al fatto che il modello additivo tende a destagionalizzare male agli estremi della serie quando questa è meglio rappresentata da un modello moltiplicativo.

Fig. 3.8 – Esempio di residui privi di andamenti sistematici

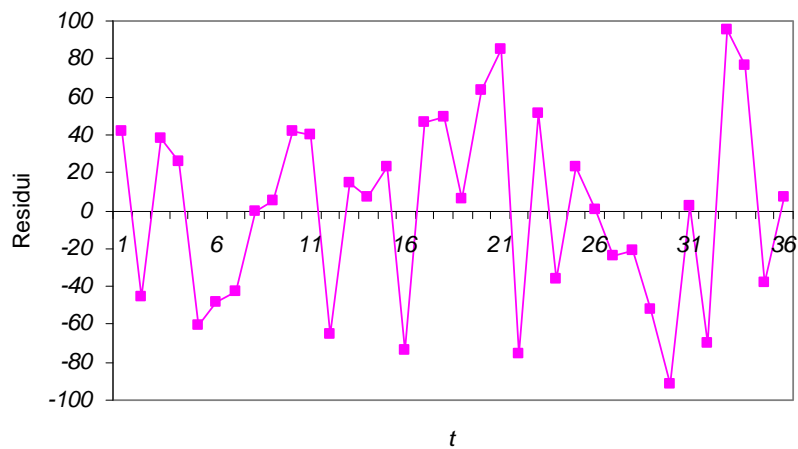


Fig. 3.9 – Esempio di residui con variabilità più alta per periodi recenti

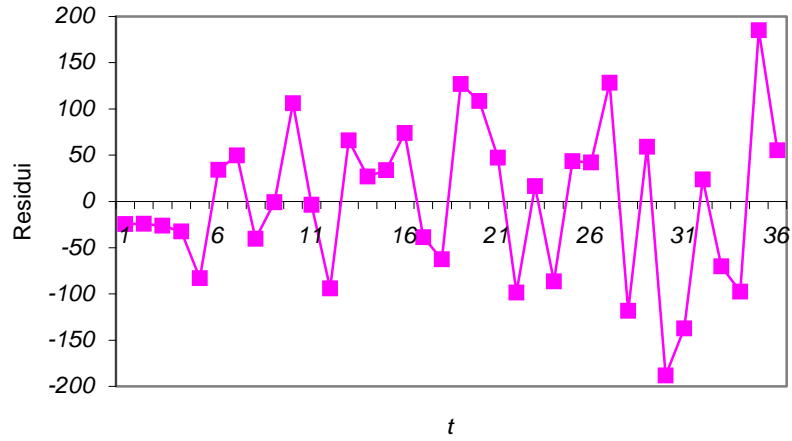


Fig. 3.10 – Esempio di residui con andamento curvilineo

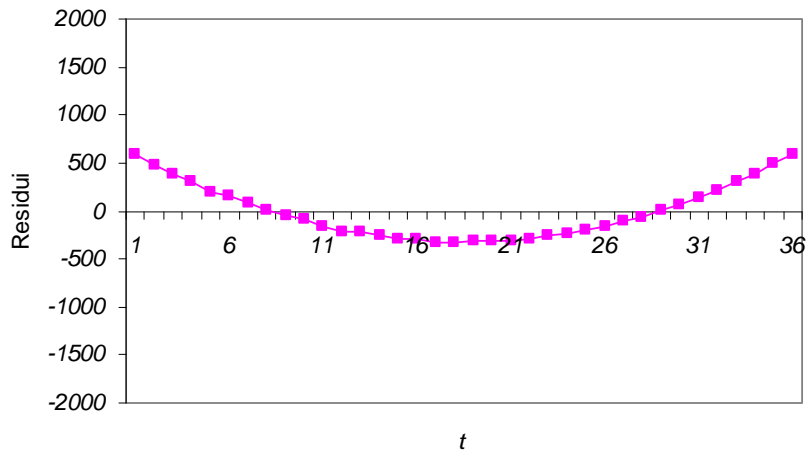
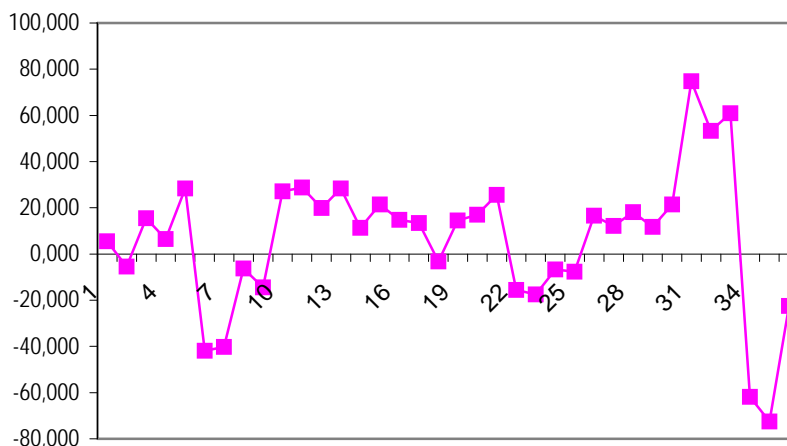


Fig. 3.11 – Residui del modello additivo di Tab. 3.2



3.7 Alcune osservazioni ulteriori sui metodi di scomposizione

A conclusione di questo capitolo, vogliamo soffermarci a discutere alcune tematiche riguardanti i metodi di scomposizione e precisamente:

- 1) l'impiego della trasformazione logaritmica per il modello moltiplicativo;
- 2) come eseguono la scomposizione procedure contenute nei pacchetti statistici Statistica (www.statsoft.com) e Minitab (www.minitab.com);
- 3) come si eseguono le previsioni una volta che si è condotta un'analisi di scomposizione.

Abbiamo già accennato al fatto che si può passare da un modello moltiplicativo rispetto alla serie originarie y_t ($t=1, \dots, n$) ad uno additivo rispetto alla serie $\ln y_t$. Quindi, volendo adottare per y_t il modello moltiplicativo, si potrebbe decidere di eseguire la scomposizione con modello additivo sui logaritmi poiché si ha:

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

E' bene precisare, tuttavia, che la scomposizione di y_t mediante modello moltiplicativo conduce a risultati diversi alla scomposizione di $\ln y_t$ mediante modello additivo (come descritto nel paragrafo 3.4). Per capire quanto appena affermato, vediamo nel dettaglio l'espressione della media mobile nei due approcci. Ad esempio, la media mobile a $2k+1$ termini (dispari) centrata sul tempo t è, sui dati originari:

$$\frac{1}{2k+1} \sum_{i=t-k}^{t+k} y_i = \frac{1}{2k+1} \sum_{i=t-k}^{t+k} T_i \times S_i \times E_i$$

Invece, se si usano dati trasformati:

$$(3.5) \quad \frac{1}{2k+1} \sum_{i=t-k}^{t+k} \ln y_i = \frac{1}{2k+1} \sum_{i=t-k}^{t+k} (\ln T_i + \ln S_i + \ln E_i)$$

Quest'ultima espressione equivale al logaritmo della media geometrica dei valori originali:

$$(3.6) \quad \frac{1}{2k+1} \sum_{i=t-k}^{t+k} (\ln T_i + \ln S_i + \ln E_i) = \ln \left(\prod_i T_i \times S_i \times E_i \right)^{1/(2k+1)}$$

Dal confronto fra (3.5) e (3.6) si capisce che i due approcci conducono, in generale, a risultati diversi.

Vediamo, a questo punto, le procedure di scomposizione classica delle serie storiche, contenute nei pacchetti statistici, Statistica e Minitab.

Il pacchetto Statistica, nella procedura *Time Series>Decomposition CensusI*, esegue i passi esattamente come da noi descritti.

Nel pacchetto Minitab, la procedura *Time series>Decomposition*, esegue le fasi dell'analisi secondo un ordine diverso da quello da noi presentato. Come prima fase, Minitab stima il *trend* mediante interpolazione di una retta rispetto al tempo t :

$$\hat{T}_t = a + bt$$

Successivamente determina il dato detrendizzato D_t che è:

$$D_t = (S + E)_t = y_t - \hat{T}_t \quad \text{per il modello additivo}$$

$$D_t = (S \times E)_t = y_t / \hat{T}_t \quad \text{per il modello moltiplicativo}$$

Vengono poi calcolati i coefficienti di stagionalità come da noi descritto, ipotizzando il modello di stagionalità costante. Si nota che la procedura Minitab non tiene conto dell'eventuale presenza del ciclo, in quanto stima il solo *trend* con una retta (la retta non ha andamenti oscillatori!).

Si può, tuttavia, risolvere questo problema lavorando con Minitab in modo più flessibile e cioè ricorrendo a diverse procedure. In particolare, la sequenza di fasi indicata per condurre l'analisi di scomposizione può essere realizzata in Minitab nel seguente modo:

- i) mediante la procedura *Time series>Decomposition* adattare un modello senza *trend* e memorizzare i dati destagionalizzati;
- ii) mediante la procedura *Time series>Moving average* procedere alla stima del ciclo-trend.

Veniamo infine al tema della previsione. Ci sono stati molti tentativi di sviluppare previsioni basate direttamente sul risultato di una analisi di scomposizione mediante medie mobili. Generalmente si procede alla

previsione della singola componente (*trend*, *ciclo*, ecc.) per poi ricomporre il dato futuro della serie (come prodotto dei valori delle singole componenti nel caso del modello 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 che non ha andamenti regolari periodici. Il procedimento può avere successo 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 \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 .

Per la componente irregolare viene adottato, come previsione, il valore neutrale, che è pari a 0 per il modello additivo, a 1 per il modello moltiplicativo.

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

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

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 numerosi metodi più sofisticati e più efficaci, i procedimenti di scomposizione restano un valido strumento per comprendere le caratteristiche evolutive passate della serie. La scomposizione costituisce la fase preliminare per decidere sul metodo di previsione da adottare.

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 (anch'essi 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 dipendenza del *trend* dal 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 qui rappresenta o il dato annuale o quello destagionalizzato e e_t la componente di disturbo.

La forma lineare in 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; se $\beta_1 < 0$, il *trend* è decrescente; se $\beta_1 = 0$ esiste un *pattern* orizzontale. Un modo per capire se il trend lineare è appropriato per rappresentare la nostra serie, consiste nel verificare se le differenze successive della serie (destagionalizzata o annuale) sono approssimativamente costanti rispetto a t . Ciò scaturisce dal fatto che, se vale la (4.1), si ha che:

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

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 .

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. I dati sui quali viene stimato il trend dovrebbero 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. In particolare, dalla Fig. 3.6, si nota che i dati destagionalizzati mediante il modello moltiplicativo, possono essere rappresentati mediante una funzione lineare nel tempo. Anziché procedere alla media mobile a 3 termini, decidiamo quindi di adottare la stima di una funzione analitica, lineare in t .

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

$$(4.5) \quad \hat{T}_t = 380,3 + 9,505t$$

che registra un indice di determinazione lineare R^2 pari a 0,985.

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 (v. Tab. 4.1):

$$\hat{y}_t = \hat{T}_t \times \hat{S}_t = (380,3 + 9,505t) \times \hat{S}_t$$

A questo punto è possibile pervenire ad una determinazione di E_t , componente casuale del modello, mediante la grandezza \hat{E}_t dove:

$$\hat{E}_t = \frac{y_t}{\hat{y}_t} = \frac{y_t}{\hat{T}_t \hat{S}_t}$$

Tuttavia, per proporre un confronto con altri metodi di scomposizione e con altri modelli, si usa frequentemente il residuo consueto:

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

mediante il quale si può calcolare ad esempio il MAPE. In questo caso, il MAPE è pari a 1,93%, valore leggermente più alto di quello ottenuto mediante l'impiego della media mobile a 3 termini per la stima del *trend* (v. paragrafo 3.6).

Tab. 4.1 – Risultati del modello moltiplicativo (vendite bibita QQQ).
Trend stimato mediante funzione lineare del tempo

ANNO	MESE	t	y_t	Stima S_m	Stima Dt	Stima T_t	Stima y_t	Res_t
1999	1	1	189	0,493	383	390	192	-3,13
	2	2	229	0,595	385	399	238	-8,64
	3	3	249	0,595	419	409	243	5,77
	4	4	289	0,679	425	418	284	4,79
	5	5	260	0,564	461	428	241	18,79
	6	6	431	0,985	438	437	431	0,34
	7	7	660	1,466	450	447	649	10,57
	8	8	777	1,692	459	456	772	5,09
	9	9	915	1,988	460	466	926	-11,29
	10	10	613	1,306	469	475	621	-7,89
	11	11	485	1,028	472	485	498	-13,40
	12	12	277	0,600	462	494	296	-19,40
2000	1	13	244	0,493	495	504	248	-4,34
	2	14	296	0,595	497	513	306	-9,53
	3	15	319	0,595	536	523	311	7,91
	4	16	370	0,679	545	532	362	8,30
	5	17	313	0,564	555	542	306	7,48
	6	18	556	0,985	565	551	543	13,02
	7	19	831	1,466	567	561	815	15,80
	8	20	960	1,692	568	570	965	-4,84
	9	21	1152	1,988	579	580	1153	-1,09
	10	22	759	1,306	581	589	770	-10,87
	11	23	607	1,028	591	599	616	-8,65
	12	24	371	0,600	619	608	365	6,21
2001	1	25	298	0,493	605	618	305	-6,56
	2	26	378	0,595	635	627	373	4,59
	3	27	373	0,595	627	637	379	-5,95
	4	28	443	0,679	652	646	439	3,80
	5	29	374	0,564	663	656	370	4,17
	6	30	660	0,985	670	665	655	4,70
	7	31	1004	1,466	685	675	981	23,02
	8	32	1153	1,692	682	684	1158	-4,78
	9	33	1388	1,988	698	694	1380	8,11
	10	34	904	1,306	692	703	919	-14,86
	11	35	715	1,028	696	713	733	-17,89
	12	36	441	0,600	736	722	433	7,82

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 2002: gennaio, febbraio e marzo ai quali corrisponde $t=37, 38, 39$ e coefficienti di stagionalità pari a 0,493, 0,595, 0,595. Le previsioni che indichiamo come F_{37}, F_{38}, F_{39} sono:

$$F_{37}=(380,3+9,505 \times 37) \times 0,493= 361$$

$$F_{38}=(380,3+9,505 \times 38) \times 0,595= 441$$

$$F_{39}=(380,3+9,505 \times 39) \times 0,595= 447$$

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

- la stabilità dei valori dei parametri della forma analitica usata per rappresentare il *trend*;
- la stabilità dei coefficienti stagionali per i tre mesi futuri, per i quali si costruisce la previsione.