CAPITOLO SECONDO

ANALISI DELLE SERIE STORICHE

SOMMARIO: 1. Introduzione alle serie storiche. - 2. Analisi classica delle serie storiche. - 3. Analisi moderna delle serie storiche. 4. Procedura TRAMO-SEATS dell'ISTAT. - Esercizi svolti e commentati.

1. INTRODUZIONE ALLE SERIE STORICHE

Molti fenomeni presentano andamenti nel tempo caratterizzati da una certa regolarità o costanza strettamente legate alla posizione occupata dall'osservazione nella sequenza di dati osservati.

L'**analisi delle serie storiche**, o **cronologiche**, o **temporali**, è la metodologia statistica che si occupa dello studio di tali fenomeni i quali sono rappresentabili, appunto, tramite serie storiche.

In termini rigorosi, una **serie storica**, x_t, è una successione di osservazioni ordinate logicamente secondo un indice temporale *t* che definisce l'ordinamento dei dati e tale che $t \in T$. L'ordine degli elementi di una serie storica non è casuale, anzi i dati sono ordinati in modo naturale dal valore assunto dall'indice temporale. Fondamentale, nell'analisi delle serie storiche, è la scelta dei periodi di osservazione: tutti devono essere contraddistinti da omogeneità di caratteristiche.

La periodicità delle osservazioni è, inoltre, variabile; generalmente le serie storiche sono mensili, trimestrali o annuali. Nell'analisi economica, tuttavia, si incontrano spesso serie settimanali, giornaliere o orarie (si consideri lo studio del traffico telefonico o del traffico stradale), o al contrario serie biennali, decennali (si considerino i censimenti della popolazione).

Numerosi sono gli esempi di serie storiche, le tipologie che sovente si incontrano nella pratica sono le seguenti:

— **serie storiche economiche** fornite dagli uffici di statistica, tra queste rientrano:
  — i numeri indici dei prezzi all'ingrosso, al consumo etc.;
  — i dati sull'occupazione;
  — i dati desunti dallo schema dei conti della contabilità nazionale;

— **serie storiche demografiche** tra cui rientrano:
  — il numero di nascite e/o i corrispondenti tassi di natalità;
  — il numero di matrimoni e/o i corrispondenti tassi di nuzialità;
  — il numero di morti e/o i corrispondenti tassi di mortalità;

— **serie storiche fisiche** tra cui rientrano:
  — il livello di precipitazioni nevose mensili;
  — la temperatura minima e massima settimanale;

— **serie storiche binarie** in cui le osservazioni possono assumere due soli valori come:
  — variazioni positive/negative del mercato azionario;
  — astensione/partecipazione alle elezioni politiche.

Una serie storica si dice:
— **univariata** o **multivariata** a seconda che si riferisca ad uno solo o a più fenomeni;
— **a parametro discreto** (o **discreta**) o **a parametro continuo** (o **continua**) a seconda che T sia, rispettivamente, discreto o continuo, pur potendo essere il fenomeno (o i fenomeni) osservato, rispettivamente, continuo o discreto.

Le serie storiche di cui ci occuperemo in questo capitolo sono univariate, discrete e le cui osservazioni si riferiscono ad intervalli temporali tra loro equidistanti.

Sostanzialmente si distinguono due tipi di approcci di analisi statistica delle serie storiche, il primo concerne l'analisi tradizionale o **classica**, di cui ci occuperemo nel paragrafo seguente, mentre il secondo concerne l'analisi **moderna** o stocastica delle stesse di cui ci occuperemo nel paragrafo terzo.

Un altro approccio alle serie storiche è di tipo frequenziale o spettrale, si parla in tal caso di **analisi spettrale**, in cui lo studio di una serie di osservazioni è condotto esaminando le sue proprietà in funzione della frequenza che caratterizza il suo comportamento periodico. In questo tipo di analisi si intende una serie storica come la risultante della somma di infinite serie periodiche di periodo, ampiezza e fasi varie, e che sono scomponibili in serie di Fourier.

2. ANALISI CLASSICA DELLE SERIE STORICHE

L'**analisi classica** delle serie storiche, che si adatta bene alle stesse solo se sono regolari, concepisce la **serie storica** come la **somma di varie componenti**, e si basa sulla scomposizione della stessa fino alla completa specificazione delle sue componenti attraverso opportuni metodi statistici.

Sia x_t la serie storica osservata, per l'approccio considerato, la serie è esprimibile in funzione di una **componente sistematica** di tipo deterministico, esprimente la **legge di evoluzione temporale** del fenomeno, e di una **componente stocastica** (ε_t) avente funzioni compensative per le discrepanze esistenti tra il modello e la realtà, e includente l'effetto indistinto di fattori omessi o che, secondo il risultato della condotta umana, risultano imprevedibili.

La componente sistematica di una serie storica è del tipo:

$$f(t) = f(T_t, C_t, S_t)$$

infatti, essa, a sua volta, è costituita da:
— una **componente di fondo** o **trend**, T_t, che indica la tendenza di fondo che caratterizza l'andamento di un fenomeno in un certo periodo di tempo. Il concomitante effetto di molteplici fattori fa sì che il fenomeno assuma un andamento regolare, sovente crescente o decrescente, nel lungo periodo;
— una **componente congiunturale** o **ciclo**, C_t, che indica il susseguirsi di fasi ascendenti e discendenti intorno al trend in un periodo di tempo, necessariamente maggiore di un anno; si tratta delle successioni di fasi del ciclo economico: prosperità, crescita, depressione, ripresa;

— una **componente stagionale** o **stagionalità**, S_t, solo eventuale se le osservazioni sono relative a periodi inferiori all'anno, ed esprime l'alternanza legata al ciclo solare annuale. Essa, in sostanza, deriva dal movimento della Terra attorno al Sole che genera la ripetizione annuale del clima, delle usanze, dei consumi e degli adempimenti legali ed amministrativi.

2.1 Modelli decompositivi utilizzati nell'analisi classica

Dopo aver individuato le componenti di una serie storica si deve individuare il modello che sia in grado di metterle in relazione. In generale, il modello di una serie storica è del tipo:

$$x_t = \phi(T_t, C_t, S_t, \varepsilon_t)$$

Esistono tre tipologie di modelli decompositivi: il *modello additivo*, il *modello moltiplicativo* e il *modello misto*.

Modello additivo

Il **modello additivo** che assume l'ipotesi che le **componenti**, espresse tutte nella medesima unità di misura di x_t, siano **indipendenti tra loro**, per cui non possono influenzarsi vicendevolmente; in tal caso l'espressione analitica del modello è la seguente:

$$x_t = T_t + C_t + S_t + \varepsilon_t \tag{2.1}$$

Modello moltiplicativo

Il **modello moltiplicativo** che assume l'ipotesi che solo il **trend** sia espresso nella stessa unità di misura di x_t, mentre le altre componenti si configurino come numeri puri, sotto forma di indici, e che siano legate al trend da una relazione di proporzionalità, per cui esiste **dipendenza tra le componenti**, ipotesi questa che consente un utilizzo frequente di tale modello nella pratica; in tal caso il modello assume la seguente espressione analitica:

$$x_t = T_t C_t S_t \varepsilon_t \tag{2.2}$$

L'espressione appena ottenuta può essere linearizzata, riconducendola alla corrispondente del modello additivo, considerando la trasformata logaritmica:

$$\log x_t = \log T_t + \log C_t + \log S_t + \log \varepsilon_t$$

Modello misto

Il **modello misto** può assumere due espressioni analitiche:

— se la componente stagionale è di tipo additivo:

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

— se la componente stagionale sia di tipo moltiplicativo:

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

2.2 Stima del trend-ciclo

Le variazioni cicliche sono state oggetto di numerose teorie economiche, tuttavia, nell'analisi dell'andamento dei principali fenomeni concernenti la situazione economica di sistema ciò che è interessante studiare, dal punto di vista statistico, sono le interrelazioni esistenti tra fenomeni di lungo periodo, e che, in quanto tali esercitano effetti sul trend di una serie storica, e fenomeni che hanno carattere congiunturale, e che, in quanto tali, esercitano effetti sul ciclo.

Per questo motivo è opportuna la stima della componente costituita dalla commistione delle due citate, il **trend-ciclo** (indicato con il simbolo $(TC)_t$), rappresentante la tendenza di fondo della serie.

Nella stima del trend-ciclo bisogna operare una distinzione tra il caso in cui nei dati sia presente una componente stagionale dal caso in cui questa sia assente.

1. Stima del trend-ciclo in presenza di stagionalità

In questo caso, la stima del trend-ciclo implica la **depurazione della componente stagionale**.

A questo punto occorre, per l'approccio decompositivo:

A) individuare il trend-ciclo calcolando medie mobili centrate con un numero di termini in funzione della cadenza delle osservazioni;
B) verificare l'esistenza della componente stagionale ed eliminarla in due fasi attraverso:
   I. il calcolo dei rapporti o coefficienti lordi di stagionalità (comprensivi di componente stagionale e di componente accidentale);
   II. il calcolo dei coefficienti netti di stagionalità.
C) eventualmente, perequare il trend-ciclo attraverso adattamento di una funzione polinomiale alla serie.

A) Individuazione del trend-ciclo

In generale, la **media mobile** è la media aritmetica di *k* osservazioni consecutive. Nell'analisi delle serie storiche le medie mobili consentono di sostituire ai dati osservati la media aritmetica di valori contigui; tale media costituisce, appunto, una stima del trend-ciclo.

Si ipotizzi che la serie storica sia generata da un modello moltiplicativo del tipo (2.1).

La prima operazione per l'utilizzo della media mobile riguarda la definizione di un **intervallo di perequazione**, cioè il dominio o **span** della media; tale intervallo è in grado di stabilire la maggiore o minore reattività della media alle variazioni proprie della serie storica.

In questa analisi, le medie mobili sono di 12, 6, 4, 7 o 24 termini a seconda che le osservazioni siano, rispettivamente, mensili, bimensili, trimestrali, giornaliere o orarie.

Se la media mobile è centrata di ordine *k* si distingue a seconda se il numero di osservazioni è:

— dispari, la media mobile è centrata automaticamente;
— pari, la media mobile è centrata applicando la sua formula su $k+1$ dati, con pesi, rispettivamente, da 1, 2, 2, …, 2, 2, 1, ed aventi per somma 2*k*.

La media mobile centrata relativa al mese *i*-esimo e all'anno *j*-esimo, stima del **trend-ciclo di prima approssimazione** è espressa nel modo seguente:

$$m_{ij} = (TC)_{ij}$$

Tale procedura è soggetta al limite, peraltro notevole, che risultano persi, rispetto alla serie d'origine, i dati relativi all'inizio e alla fine della serie.

Esempio

La tabella seguente riporta le vendite (in migliaia di litri) trimestrali in un supermercato di una bevanda analcolica:

Anni	Trimestri			
	Primo	**Secondo**	**Terzo**	**Quarto**
2002	170	300	610	120
2003	250	410	790	190
2004	290	460	890	250
2005	450	550	1.100	270
2006	320	600	1.260	280

Tabella 1

qual è la tendenza di fondo delle vendite?

Una stima del trend-ciclo delle vendite è ottenuta calcolando le medie mobili che, poiché si è in presenza di dati trimestrali, con quattro osservazioni ogni anno, sono medie mobili di ordine quattro, ottenute raggruppando i dati di quattro in quattro.

Indicando con t = 1, 2, …, 20, rispettivamente, il primo, secondo, …, ventesimo trimestre avremo:

Date t	1	2	3	4	5	6	....
x_t	170	300	610	120	250	410	....

$$\underbrace{170\quad 300\quad 610\quad 120}_{m(4)_1}$$

$$\underbrace{300\quad 610\quad 120\quad 250}_{m(4)_2}$$

$$\underbrace{610\quad 120\quad 250\quad 410}_{m(4)_3}$$

— a $t = \dfrac{1+2+3+4}{4} = 2{,}5$, si associa $m(4)_{2,5} = \dfrac{170+300+610+120}{4} = 300$

— a $t = \dfrac{2+3+4+5}{4} = 3{,}5$, si associa $m(4)_{3,5} = \dfrac{300+610+120+250}{4} = 320$

etc.

t	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5
$m(4)_t$	300	320	347,5	392,5	410	420	432,5	457,5	472,5

t	11,5	12,5	13,5	14,5	15,5	16,5	17,5	18,5
$m(4)_t$	512,5	535	587,5	592,5	560	572,5	612,5	615

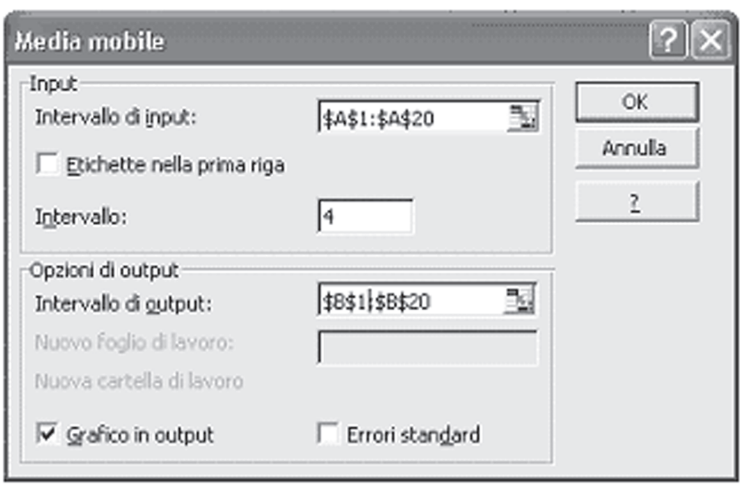
Diversi problemi di interesse statistico possono essere risolti tramite l'uso di fogli elettronici, come Microsoft Excel, grazie alle numerose funzioni definite nella categoria «Statistiche».

In questo volume ci riferiremo più volte a esse nello svolgere gli esempi. Talvolta, anche se il problema da risolvere non appartiene a tale categoria, imposteremo un foglio elettronico per rendere meccanici i calcoli più laboriosi.

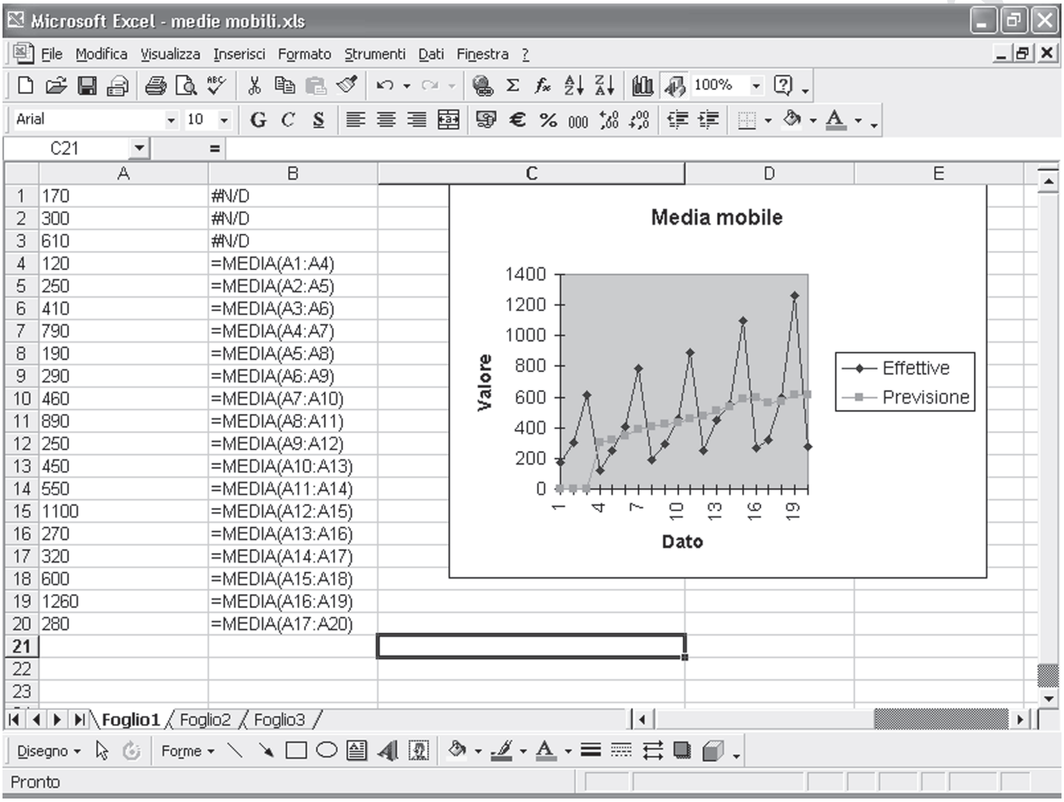
A fini puramente didattici, nel foglio, invece dei risultati figureranno le funzioni. Tale visualizzazione si ottiene aprendo il menu «Strumenti» scegliendo «Opzioni», quindi, in «Visualizza» selezionando l'opzione «Formule».

Excel mette a disposizione uno strumento di analisi in grado di porre in evidenza la tendenza di fondo di una serie di dati. Lo strumento in questione è **Media mobile**.
A parte i componenti aggiunti con l'installazione predefinita in Excel, è possibile che sul vostro computer non sia installata questa funzionalità, per cui occorre avviare Excel, selezionare il menu Strumenti e scegliere Componenti aggiuntivi.
Facendo riferimento alla serie riportata in tabella 1, inseriamo in un foglio i dati nelle celle da A1 ad A20, selezioniamo il menu **Strumenti**, quindi **Analisi dati**. Nella finestra di dialogo **Strumenti di analisi** selezioniamo **Media mobile**.

Completiamo la finestra di dialogo come illustrato di seguito:



Il foglio di calcolo con le medie mobili e il grafico in output è il seguente:



	A	B
1	170	#N/D
2	300	#N/D
3	610	#N/D
4	120	=MEDIA(A1:A4)
5	250	=MEDIA(A2:A5)
6	410	=MEDIA(A3:A6)
7	790	=MEDIA(A4:A7)
8	190	=MEDIA(A5:A8)
9	290	=MEDIA(A6:A9)
10	460	=MEDIA(A7:A10)
11	890	=MEDIA(A8:A11)
12	250	=MEDIA(A9:A12)
13	450	=MEDIA(A10:A13)
14	550	=MEDIA(A11:A14)
15	1100	=MEDIA(A12:A15)
16	270	=MEDIA(A13:A16)
17	320	=MEDIA(A14:A17)
18	600	=MEDIA(A15:A18)
19	1260	=MEDIA(A16:A19)
20	280	=MEDIA(A17:A20)

È evidente che la curva delle medie mobili è alquanto regolare per cui si può sostanzialmente affermare che la serie presenta una tendenza di fondo di tipo lineare crescente.

B) Destagionalizzazione della serie storica

L'eliminazione della componente stagionale da una serie storica è una questione di fondamentale importanza in quanto la disponibilità di dati privi di tale componente costituisce non solo un fabbisogno informativo diffuso da parte di utilizzatori poco esperti ma, addirittura, è sanzionata da regolamenti comunitari che esortano gli Istituti nazionali di statistica dei paesi membri a produrre serie storiche depurate della componente stagionale, si pensi all'importanza assunta dal fatto che dati relativi agli stessi mesi ma a diversi anni e depurati di tale componente possono essere confrontati tra loro. La procedura attraverso cui si rimuovono le oscillazioni che, in una serie storica, avvengono nell'arco di un anno e che sono attribuibili alla componente stagionale, è detta, appunto, **destagionalizzazione**.

Supponendo un modello di tipo moltiplicativo il rapporto tra il valore osservato per ciascun dato della serie e la stima del trend-ciclo costituisce il **rapporto lordo di stagionalità**:

$$x_{ij}^{*} = \frac{(TC)_{ij} S_{ij} \varepsilon_{ij}}{(TC)_{ij}} = S_{ij} \varepsilon_{ij} \qquad (2.3)$$

Dato il suo significato matematico, il rapporto espresso dalla (2.3) consente l'eliminazione dell'influenza esercitata dalla componente trend-ciclo sul valore osservato per ciascun dato e, quindi, il confronto del fenomeno in periodi diversi al netto delle fluttuazioni dovute alla stagionalità.

L'analisi, anno per anno, dei rapporti lordi di stagionalità consente di rilevare il tipo di stagionalità presente nei dati.

Per testare l'**ipotesi nulla di assenza di stagionalità** nei dati si utilizza quella particolare tecnica statistica nota come **analisi della varianza**.

Al fine della verifica dell'ipotesi, si distribuiscono i rapporti lordi di stagionalità in dodici gruppi corrispondenti ai dodici mesi e se ne considerano i valori medi, per cui l'ipotesi nulla è del tipo:

$$H_0 : \mu_G = \mu_F = \ldots = \mu_D = \mu$$

In assenza di stagionalità nella serie, la differenza tra i valori medi dei rapporti lordi non è significativa.

Assumendo che i valori della componente accidentale si distribuiscano normalmente con media nulla e varianza costante, le medie aritmetiche dei differenti rapporti lordi di stagionalità relativi a uno stesso mese (o in generale periodo) di diversi anni rappresentano i **rapporti netti di stagionalità** e sono depurati della componente casuale.

I dati destagionalizzati della serie si ottengono dividendo per S_{ij}:

$$x^*_{ij} = \frac{(TC)_{ij}\, S_{ij} \varepsilon_{ij}}{S_{ij}} = (TC)_{ij}\, \varepsilon_{ij}$$

2. Stima del trend-ciclo in assenza di stagionalità

Supponendo che le osservazioni nella serie storica non siano offerte da oscillazioni stagionali, è possibile procedere ad una stima del trend-ciclo, utilizzando due metodi alternativi:

— metodo analitico;
— metodo delle medie mobili ponderate.

Con il **metodo analitico** si individua una funzione matematica del tempo valida per l'**orizzonte temporale** cui la serie si riferisce (**trend globale**). Le fasi per la stima del trend-ciclo sono le seguenti:

— scelta della forma funzionale;
— stima dei parametri della funzione;
— verifica della validità del modello scelto.

Differenze successive

Per individuare la specifica funzione matematica del tempo si ricorre al concetto di *differenza prima* e *seconda*. Data una serie storica x_t si definisce:

— **differenza prima**, e si indica con $\Delta(x_t)$, la serie delle variazioni intervenute tra x_t e x_{t-1}, al variare di $t = 2, 3, \ldots$. In pratica:

$$\Delta(x_t) = x_t - x_{t-1}$$

— **differenza seconda**, e si indica con $\Delta_2(x_1)$, la differenza prima applicata due volte ad una serie storica, ed è pari a:

$$\Delta_2(x_t) = \Delta(x_t) - \Delta(x_t - 1) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

In generale, si parla di **differenza successiva** quando si applica più volte la differenza prima ad una serie storica.

La **funzione** prescelta è:

— una retta di equazione $f(t) = \beta_0 + \beta_1 t$ se $\Delta(x_t) \approx$ costante;

— una parabola di equazione $f(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ se $\Delta_2(x_t) \approx$ costante;

— una curva esponenziale semplice di equazione $f(t) = \beta_0 e^{\beta_1 t}$ se $\dfrac{x_t}{x_{t-1}} \approx$ costante;

— una curva esponenziale modificata di equazione $f(t) = k + \beta_0 e^{\beta_1 t}$ se $\dfrac{\Delta(x_t)}{\Delta(x_{t-1})} \approx$ costante;

— una curva di Gompertz $f(t) = \dfrac{1}{k + \beta_0 \beta_1 t}$ se $\Delta(\log x_t) \approx$ costante.

La fase successiva alla scelta della forma funzionale è quella relativa alla **stima dei parametri**; essa è effettuata, per le prime due tipologie di funzioni, passando ai logaritmi per la terza tipologia di funzioni, attraverso il **metodo dei minimi quadrati**. Per le altre due funzioni si ricorre a **tecniche di stima non lineari**.

Con il **metodo delle medie mobili ponderate**, invece, si individua un polinomio interpolante valido per un gruppo di osservazioni successive (**trend locale**). Le fasi per la stima del trend - ciclo sono le seguenti:

— si suddivide la serie storica in gruppi di r osservazioni successive:

$$x_{-m}, x_{-m+1}, \ldots, x_0, \ldots, x_{m-1}, x_m$$

dove si è scelto $r = 2m + 1$, ossia r dispari, per un motivo che vedremo nel punto successivo;

— si individua un polinomio di grado p non superiore a $r - 1 = 2m$, tale da stimare il valore del trend-ciclo del termine centrale del gruppo.

Pertanto, il polinomio interpolante sia:

$$\beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_p t^p$$

Se il metodo di stima dei parametri applicato è quello dei minimi quadrati, deve essere:

$$\sum_{t=-m}^{m} \left(x_t - \beta_0 - \beta_1 t - \beta_2 t^2 - \ldots - \beta_p t^p\right)^2 = \min \qquad (2.4)$$

Si noti che se si pone nella (2.4) $t = 0$, il valore del polinomio interpolante è pari a β_0, per cui $(TC)_0 = \beta_0$, ed è pari ad una media ponderata con pesi che dipendono da r e da p.