

## 3 VARIABILI DI COMODO E CAMBIAMENTI STRUTTURALI

### Indice del capitolo

3.1	Estensioni del modello lineare classico e test di malaspecificazione .....	2
3.2	Le variabili di comodo.....	5
	<i>Variabili di comodo per l'intercetta</i> .....	5
	<i>Variabili di comodo per i coefficienti angolari</i> .....	8
	<i>Tendenze segmentate</i> .....	8
	<i>Le osservazioni anomale</i> .....	9
	<i>Il test di normalità dei residui di Jarque e Bera</i> .....	10
3.3	Variabili di comodo stagionali.....	16
	<i>La depurazione stagionale con il criterio dei minimi quadrati</i> .....	17
	<i>Stagionalità variabile</i> .....	20
	<i>Stagionalità additiva o moltiplicativa</i> .....	22
	<i>La conservazione dei volumi</i> .....	23
	<i>Destagionalizzazione in presenza della tendenza</i> .....	26
	<i>La serie destagionalizzata con la correzione della tendenza</i> .....	27
	<i>La stima del modello con le serie destagionalizzate</i> .....	28
3.4	Un test di cambiamento strutturale per il modello lineare semplice.....	32
	<i>Test di cambiamento di struttura con le variabili di comodo</i> .....	34
3.5	Un test di cambiamento strutturale per il modello lineare multiplo .....	40
	<i>Il test preliminare di uguaglianza delle varianze</i> .....	41
	<i>Test di invarianza strutturale per un sottoinsieme di parametri</i> .....	41
	<i>Test di cambiamento di struttura con le variabili di comodo</i> .....	42
3.6	Bibliografia .....	46

### 3.1 Estensioni del modello lineare classico e test di malaspecificazione

Possiamo riassumere ancora una volta le ipotesi alla base del modello lineare esposte all'inizio del capitolo, presentandole per semplicità in forma scalare, cioè con riferimento al modello di regressione semplice:<sup>1</sup>

- 1) il campione è omogeneo ed i parametri  $\beta_1$  e  $\beta_2$  sono invariabili nel tempo;
- 2) i valori di  $x_t$  sono noti, cioè non aleatori;
- 3)  $m_{xx} - \bar{x}^2 \neq 0$  (3.1.1)
- 4)  $E(\tilde{u}_t) = 0$  ,  $E(\tilde{u}_t \cdot \tilde{u}_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases} \quad \forall t, s$
- 5)  $\tilde{u}_t \sim N(0, \sigma^2)$   $\forall t$

L'ipotesi di non correlazione (4) insieme all'ipotesi di normalità (5) implica che i residui del modello siano stocasticamente indipendenti. Queste due ipotesi possono essere condensate nel modo seguente

$$\tilde{u}_t \sim niid(0, \sigma^2) \quad \forall t \quad (3.1.2)$$

dove l'acronimo *niid* sta per *normally independently and identically distributed*, cioè identicamente e indipendentemente distribuite in modo normale.<sup>2</sup>

Le (3.1.1) sono le cosiddette *ipotesi stocastiche forti* sul modello,<sup>3</sup> sulle quali si basa, come abbiamo visto nel capitolo precedente, l'inferenza statistica nel modello lineare classico. Buona parte dei capitoli successivi, a partire da questo, verrà dedicata ad approfondire il significato di queste ipotesi. In particolare, ci interessa:

<sup>1</sup> La loro espressione in termini matriciali è stata fornita nel capitolo 2.

<sup>2</sup> Talora si impiega l'acronimo più sintetico *nid*, omettendo la *i* corrispondente a *identically* (che è ridondante, soprattutto se si specifica  $\forall t$ ). Dall'ipotesi di identica distribuzione scaturisce quella di costanza della varianza  $\sigma^2$ . Dall'ipotesi di indipendenza segue la non correlazione, e quindi la nullità di tutte le covarianze secondo la (4) delle (3.1.1)

<sup>3</sup> Alcuni autori considerano come "ipotesi stocastiche" solo le (4) e (5), poiché queste si riferiscono direttamente ai residui, che sono l'unica componente esplicitamente stocastica del modello. In effetti, anche la variabile dipendente è aleatoria (si veda la (1.6.2)), quindi l'ipotesi di omogeneità del campione può essere ricompresa fra le ipotesi stocastiche. Inoltre, l'ipotesi (2), che specifica la natura non aleatoria delle variabili esplicative, è essa stessa un'ipotesi stocastica in quanto riferita alla natura del processo stocastico che genera le  $x_t$ .

1. comprendere se esse si prestano a rappresentare adeguatamente la realtà economica
2. verificare con test statistici se esse siano o meno rispettate dai dati
3. investigare le proprietà statistiche del modello nel caso in cui esse non siano verificate
4. studiare delle possibili estensioni o correzioni del modello che tengano conto della violazione o dell'indebolimento delle ipotesi (3.1.1).

Ricordiamo che sia la nostra analisi delle proprietà della stimatore OLS, sia la possibilità di effettuare inferenze statistiche su di esso, si basano sulle ipotesi (3.1.1). Di conseguenza la violazione di una o più di esse ci impedisce di formulare giudizi attendibili circa la significatività dei parametri del modello o di loro combinazioni lineari.

Più precisamente, le ipotesi (3.1.1) dalla (1) alla (4) (dette ipotesi “deboli” o “classiche” o “di Gauss-Markov”) sono alla base della non distorsione dello stimatore OLS (1.6.17) e del teorema di Gauss-Markov (1.8.2), mentre tutte le (3.1.1) prese congiuntamente (cioè le ipotesi forti) sono alla base dei teoremi distribuzionali visti nel capitolo 2, che ci consentono di effettuare test  $t$ ,  $F$  e  $\chi^2$  sui parametri del modello.

I test statistici con i quali si verifica se le ipotesi stocastiche forti sono o meno rispettate dai dati vanno sotto il nome collettivo di *test di malaspecificazione o diagnostici*.<sup>4</sup>

In questo capitolo ci occupiamo della prima fra le ipotesi elencate, quella che il campione sia omogeneo e che quindi i parametri strutturali del modello siano invariabili nel tempo.<sup>5</sup> Ora, viceversa, ipotizziamo che uno o più di questi parametri possano assumere valori diversi nel periodo campionario, in conseguenza di non omogeneità del campione.

La non omogeneità può dipendere dal verificarsi, all'interno del campione, di uno o più eventi eccezionali che determinano scostamenti della variabile dipendente dal proprio valore teorico. Si parlerà in questo caso di osservazioni

---

<sup>4</sup> In lingua inglese: *misspecification tests* o anche *diagnostic tests*, in quanto intendono diagnosticare eventuali “patologie” del modello, ovvero scostamenti dalle ipotesi stocastiche sottostanti ad esso.

<sup>5</sup> Più esattamente, in questa sezione concentriamo la nostra attenzione sui coefficienti di regressione  $b$ . Anche la varianza  $\sigma^2$  è un parametro del modello, ma l'ipotesi della sua costanza nel tempo verrà analizzata separatamente nel successivo capitolo 4.

anomale, la cui presenza determina, in genere, asimmetrie e quindi non normalità nella distribuzione dei residui. Questo argomento è affrontato nel paragrafo 3.2.

Una particolare forma di non costanza dei parametri si presenta nel caso di andamenti stagionali delle variabili. Le oscillazioni determinate dal ciclo stagionale si traducono in slittamenti dell'intercetta del modello, che quindi non è più costante. I problemi legati alla stagionalità vengono introdotti nel paragrafo 3.3.

Una forma particolarmente interessante di non omogeneità è quella del cambiamento di struttura, che si verifica quando il campione considerato abbraccia un periodo storico all'interno del quale si sono manifestati due regimi nell'andamento del fenomeno oggetto di studio, caratterizzati da due distinti insiemi di parametri strutturali.<sup>6</sup> Ciò può accadere, ad esempio, quando in tale periodo sono presi provvedimenti istituzionali che modificano normativamente le funzioni di comportamento: un'equazione che voglia spiegare un tasso di cambio può avere caratteri diversi a seconda che si consideri il regime di cambi fissi oppure quello di cambi flessibili, per cui occorre inserire nell'equazione stessa un meccanismo che le permetta di valutare in maniera differenziata i due regimi. Un altro caso riguarda il deposito previo cui sono costretti gli importatori in periodo di forte disavanzo della bilancia dei pagamenti commerciale, per cui una equazione delle importazioni deve discriminare tra dati rilevati in tempi in cui vige il deposito previo e dati rilevati in tempi senza deposito. Un terzo esempio può concernere il movimento dei capitali, la cui equazione deve poter tener conto di vincoli posti dall'autorità monetaria in certi particolari periodi di tempo.

Nel paragrafo 3.2 vediamo come è possibile rappresentare per mezzo di variabili di comodo la presenza di cambiamenti di struttura, mentre l'analisi effettuata nei paragrafi 3.4 e 3.5 fornisce un primo insieme di strumenti statistici per valutare la loro presenza all'interno del campione.

---

<sup>6</sup> L'analisi può ovviamente estendersi al caso generale di  $r$  cambiamenti di struttura. L'interesse per questo tipo di non omogeneità è stato rafforzato dalla nota critica di Lucas, secondo la quale i parametri strutturali dei modelli economici sono in effetti una mistura di parametri "profondi", rappresentanti le preferenze degli agenti economici, la tecnologia prevalente nel sistema economico e il processo di formazione delle aspettative, per cui essi sono soggetti a mutamenti ogni volta che gli agenti economici sono chiamati ad adattare il proprio comportamento in risposta a mutamenti del quadro macroeconomico indotti, ad esempio, da misure di politica economica.

### 3.2 Le variabili di comodo

I meccanismi che vengono inseriti nelle equazioni affinché queste tengano conto dei cambiamenti di regime fanno uso delle cosiddette *variabili di comodo* e possono essere anche complessi, in funzione dei tipi di influenza che le diverse strutture economiche funzionanti nel periodo campionario producono sulla specificazione dei modelli. In questo paragrafo vediamo come usare le variabili di comodo per rappresentare cambiamenti di struttura nei parametri del modello (distinguendo fra intercetta e coefficienti angolari), segmentazioni nella tendenza delle variabili, e presenza di osservazioni anomale.

I meccanismi presentati in questa sezione possono essere facilmente estesi e combinati per tener conto di cambiamenti di regime più complessi di quelli considerati in queste pagine.

#### *Variabili di comodo per l'intercetta*

Supponiamo in primo luogo che i regimi diversi possibili siano due e che influenzino soltanto l'intercetta: valgono allora i due modelli

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-1} x_{k-1t} + \beta_k + u_t$$

nel regime I (ad esempio nei tempi  $t$  con tasso di cambio fisso), e

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-1} x_{k-1t} + \beta_{k+1} + u_t$$

nel regime II (ad esempio nei tempi  $t$  con tasso di cambio flessibile), che possono essere riassunti nel seguente

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-1} x_{k-1t} + \beta_k d_{1t} + \beta_{k+1} d_{2t} + u_t \quad (3.2.1)$$

dove

$$d_{1t} = \begin{cases} 1 & \text{nel regime I} \\ 0 & \text{nel regime II} \end{cases} \quad d_{2t} = \begin{cases} 0 & \text{nel regime I} \\ 1 & \text{nel regime II} \end{cases}$$

In questo modo le serie storiche  $\{d_{1t}\}$  e  $\{d_{2t}\}$  sono composte da zero e da uno a seconda dei tempi  $t$  e non valgono mai contemporaneamente uno o zero. La matrice  $\mathbf{X}$  è allora formata nel modo seguente

$$\mathbf{X} = [\mathbf{X}_1 \mathbf{d}_1 \mathbf{d}_2]$$

dove  $\mathbf{X}_1$  è la matrice delle  $x_{it}$ ,  $i = 1, 2, \dots, k-1$ ,  $t = 1, 2, \dots, n$ , ed i vettori colonna  $\mathbf{d}_1$  e  $\mathbf{d}_2$  sono costituiti da zeri e da uno. L'equazione (3.2.1) può essere stimata con il criterio dei minimi quadrati ordinari producendo stime per le  $\beta_i$ ,  $i = 1, 2, \dots, k-1$ , uguali nei due regimi, mentre l'intercetta vale  $\hat{\beta}_k$  nel regime I e  $\hat{\beta}_{k+1}$  nell'altro.

Le variabili  $d_{1t}$  e  $d_{2t}$  sono chiamate *di comodo*<sup>7</sup> in quanto non derivano direttamente da speculazioni teoriche ma vengono aggiunte nella specificazione delle equazioni semplicemente per trattare periodi campionari con regimi economici diversi. Nell'esempio precedente se ne usano soltanto due, ma evidentemente il loro numero dipende da quanti sono i regimi differenti da considerare.

Il modello (3.2.1) deve essere stimato *senza* l'intercetta, che è sostituita dai due valori alternativi  $\beta_k$  e  $\beta_{k+1}$ ; se si aggiungesse il termine noto, cui corrisponde nella matrice il vettore di uno  $\mathbf{i}$ , si avrebbe

$$\mathbf{i} = \mathbf{d}_1 + \mathbf{d}_2$$

cioè una colonna di  $\mathbf{X}$  sarebbe combinazione lineare delle altre, la matrice  $\mathbf{X}$  sarebbe singolare per il teorema IV-1.4 e non esisterebbe  $(\mathbf{X}'\mathbf{X})^{-1}$ .

Se si è costretti a stimare l'intercetta (ad esempio a causa di un programma di calcolo automatico che la considera obbligatoriamente), è conveniente sostituire all'equazione (3.2.1) la seguente

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-1} x_{k-1t} + \alpha_1 d_{1t} + \alpha_2 + u_t \quad (3.2.2)$$

che corrisponde alla prima se si pone in essa

$$\alpha_1 + \alpha_2 = \beta_k, \quad \alpha_2 = \beta_{k+1} \quad (3.2.3)$$

ma contiene il termine noto  $\alpha_2$ .

*Osservazione 3.1* – La (3.2.2) con le posizioni (3.2.3) è perfettamente equivalente alla (3.2.1) nonostante contenga un diverso insieme di parametri. Per questo motivo si dice, in termini tecnici, che la (3.2.3) è ottenuta *riparametrizzando* la (3.2.1), ovvero che è una riparametrizzazione (o una diversa parametrizzazione) di quest'ultima. La possibilità di esprimere la medesima equazione con parametrizzazioni differenti torna molto utile nella teoria della verifica delle ipotesi. Accade infatti spesso che un test di esecuzione relativamente complessa diventi molto semplice riparametrizzando il modello.

Sulle stime dei parametri associati con le variabili di comodo è utile effettuare il test della  $t$  di Student per verificare che siano diverse da zero. Se si accetta una sola delle ipotesi nulle

$$H_0: \beta_k = 0, \quad H_0: \beta_{k+1} = 0 \quad (3.2.4)$$

---

<sup>7</sup> In lingua inglese: *dummy*.

si conclude che durante uno dei regimi l'intercetta è nulla e quindi che vi è differenza dall'altro, nel quale l'intercetta è significativamente diversa da zero. D'altro canto, le due intercette possono differire l'una dall'altra essendo entrambe diverse da zero, per cui è utile sottoporre a verifica l'ipotesi lineare

$$H_0: \beta_k - \beta_{k+1} = 0 \quad (3.2.5)$$

che può essere considerata come un caso particolare della (2.4.1). Si noti che nella parametrizzazione (3.2.2) questa ipotesi si riduce alla seguente

$$H_0: \alpha_1 = 0 \quad (3.2.6)$$

Se questa ipotesi è accettata non sussiste differenza tra i due regimi, subordinatamente alla validità della rappresentazione (3.2.1).

*Osservazione 3.2* – Possiamo ora esemplificare concretamente il fatto che parametrizzazioni diverse rendono più o meno agevole l'esecuzione di determinati test. Ad esempio, se si utilizza la parametrizzazione (3.2.1) i due test (3.2.4) sono di immediata esecuzione (le loro statistiche corrispondono ai rapporti  $t$  dei rispettivi coefficienti), mentre con la (3.2.2) solo il secondo dei test (3.2.4) non presenta difficoltà. Il primo, invece, comporta che nella  $t$  di Student si calcoli l'errore standard dato da

$$\begin{aligned} \bar{\sigma}[\text{Var}(\hat{\alpha}_1 + \hat{\alpha}_2)]^{1/2} &= \bar{\sigma}[\text{Var}(\hat{\alpha}_1) + \text{Var}(\hat{\alpha}_2) + 2\text{Cov}(\hat{\alpha}_1, \hat{\alpha}_2)]^{1/2} = \\ &= \bar{\sigma}(a_{kk}^2 + a_{k+1,k+1}^2 + 2a_{kk+1}^2)^{1/2} \end{aligned} \quad (3.2.7)$$

dove le  $a_{ij}^2$  sono gli elementi  $(i,j)$  della matrice  $(\mathbf{X}'\mathbf{X})^{-1}$  e  $\bar{\sigma}$  è la radice quadrata aritmetica della stima (1.7.2). Naturalmente, ove possibile il ricercatore per eseguire la coppia di test (3.2.4) preferirà stimare il modello (3.2.1).

D'altro canto, nella parametrizzazione (3.2.1) il test di uguaglianza dei regimi viene espresso come (3.2.5) e quindi richiede l'uso di una statistica  $F$ , mentre nella (3.2.2) esso si riduce alla forma (3.2.6) per cui può essere effettuato semplicemente leggendo la  $t$  di  $\alpha_1$ . In tal caso quindi il ricercatore preferirà stimare il modello (3.2.2).

Il parametro  $\alpha_1$  della (3.2.2) misura lo spostamento (in lingua inglese: *shift*) dell'intercetta nel regime I rispetto al regime II:  $\beta_k - \beta_{k+1} = \alpha_1$ . Per questo motivo la  $d_{1t}$  nel modello (3.2.2) viene detta *shift dummy variable*.

*Variabili di comodo per i coefficienti angolari*

In un secondo meccanismo si suppone che due regimi diversi influenzino sia l'intercetta che il coefficiente angolare  $\beta_{k-1}$  della variabile esplicativa  $x_{k-1,t}$ . In questo caso si ha

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-2} x_{k-2,t} + \beta_{k-1} (d_1 x_{k-1,t}) + \beta_k (d_2 x_{k-1,t}) + \beta_{k+1} d_{1t} + \beta_{k+2} d_{2t} + u_t \quad (3.2.8)$$

con le variabili di comodo che hanno i significati precedenti. Con una procedura analoga a quella sopra esposta, la (3.2.8) può essere sostituita dalla

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-2} x_{k-2,t} + b_1 d_{1t} x_{k-1,t} + b_2 x_{k-1,t} + \alpha_1 d_{1t} + \alpha_2 + u_t \quad (3.2.9)$$

che corrisponde alla (3.2.8) se si pone

$$\alpha_1 = \beta_{k+1} - \alpha_2, \quad \alpha_2 = \beta_{k+2}, \quad b_1 = \beta_{k-1} - b_2, \quad b_2 = \beta_k$$

*Osservazione 3.3* – Se nella (3.2.8) l'ordinamento temporale delle variabili esplicative e delle  $y_t$  non ha significato (ad esempio in dati sezionali), tale equazione sintetizza le due

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-2} x_{k-2,t} + \beta_{k-1} x_{k-1,t} + \beta_{k+1} + u_t$$

$$y_t = \beta_1 x_{1t} + \dots + \beta_{k-2} x_{k-2,t} + \beta_k x_{k-1,t} + \beta_{k+2} + u_t$$

ciascuna delle quali può essere stimata con i dati relativi ad un solo regime economico.

Le intercette nei due regimi sono uguali se è valida l'ipotesi nulla

$$H_0: \alpha_1 = 0$$

mentre sono uguali i coefficienti angolari  $\beta_{k-1}$  e  $\beta_k$  se è valida l'altra

$$H_0: b_1 = 0$$

ipotesi che possono venire verificate nella (3.2.9) con il test della  $t$  di Student.

*Tendenze segmentate*

Talvolta l'ispezione visiva di una serie storica  $\{y_t\}$  mostra chiaramente l'esistenza di una tendenza che varia in due o più periodi di tempo per cui essa non può essere rappresentata mediante una sola curva polinomiale del tipo (I-3.3.3). È necessario, allora, fare uso delle variabili di comodo per distinguere tendenze diverse in periodi diversi. Esemplichiamo questo uso nel caso della tendenza lineare e di una coppia di periodi; la generalizzazione al polinomio di grado  $p$  generico ed a più di due periodi è immediata.

Supponiamo, dunque, che si debba stimare una equazione che nel periodo I valga



$$y_t = \alpha_0 + \alpha_1 t + \mathbf{b}'\mathbf{x}_t + u_t$$

e nel periodo II valga

$$y_t = \gamma_0 + \gamma_1 t + \mathbf{b}'\mathbf{x}_t + u_t$$

Se si opera con le differenze prime l'equazione diventa semplicemente

$$\Delta y_t = \alpha_1 d_{1t} + \gamma_1 d_{2t} + \mathbf{b}'\Delta \mathbf{x}_t + u_t$$

dove

$$d_{1t} = \begin{cases} 1 & \text{nel periodo I} \\ 0 & \text{nel periodo II} \end{cases} \quad d_{2t} = \begin{cases} 0 & \text{nel periodo I} \\ 1 & \text{nel periodo II} \end{cases}$$

mentre se si desidera continuare ad avere l'equazione specificata sui livelli occorre far uso di uno schema del tipo (3.2.6)

$$y_t = \alpha_0 d_{1t} + \gamma_0 d_{2t} + \alpha_1 (d_{1t} t) + \gamma_1 (d_{2t} t) + \mathbf{b}'\mathbf{x}_t + u_t$$

*Esempio 2.2* – Se la tendenza è crescente nel periodo I e decrescente nella II, si ha  $\alpha_1 > 0$  e  $\gamma_1 < 0$ . Se la tendenza è crescente in ambedue i periodi ma con tassi di crescita diversi,  $\alpha_1$  e  $\gamma_1$  sono ambedue positivi ma di valore diverso.

#### *Le osservazioni anomale*

Un uso molto frequente delle variabili di comodo è fatto nel trattamento dei *dati anomali* (o *eccezionali*)<sup>8</sup> che, di nuovo, determinano la non omogeneità del campione. Questi dati sono costituiti dalle osservazioni che nel campione sono localizzate molto lontano dalla loro media, generalmente a causa di disomogeneità repentine nel funzionamento del sistema economico. Dati anomali nelle serie dei prezzi possono, ad esempio, essere prodotti da shock petroliferi; altri nelle serie della produzione industriale possono essere indotti da periodi prolungati di scioperi massicci, e così via.

Questi dati possono provocare notevoli danni nella costruzione della parte sistematica dei modelli econometrici. In particolare, in presenza di osservazioni anomale non sarà generalmente più valida l'ipotesi di normalità dei residui. Le osservazioni anomale comprese nel campione delle variabili endogene o di quelle esogene, infatti, a meno che non compaiano in modo equivalente nelle prime e nelle seconde, comportano elementi di squilibrio anche forte nella componente sistematica delle equazioni e producono disturbi ben netti nella omogeneità dei residui. La loro presenza si traduce quindi all'atto della stima in residui anomali, cioè in episodi nei quali il modello sottostima o sovrastima notevolmente il

---

<sup>8</sup> In lingua inglese: *outliers*.

fenomeno che intende rappresentare. Se i residui anomali hanno tutti lo stesso segno, la distribuzione dei residui sarà asimmetrica (positivamente o negativamente) e quindi si allontanerà dalla normale, che è una distribuzione simmetrica. Se invece i residui anomali si ripartiscono in modo approssimativamente equo fra valori positivi e negativi, allora la distribuzione dei residui sarà platicurtica, avrà cioè le code più “alte” della normale e di conseguenza un coefficiente di curtosi maggiore di tre.<sup>9</sup> Dato che le procedure di inferenza studiate fin qui si basano sull’ipotesi stocastica forte di normalità dei residui, la prima conseguenza della non normalità è appunto che i test di ipotesi sui parametri del modello non daranno più risultati affidabili, almeno in piccoli campioni.<sup>10</sup>

Naturalmente la violazione dell’ipotesi di normalità non influisce sulle proprietà dello stimatore OLS che non dipendono da essa. In particolare, sarà sempre valido il teorema di Gauss-Markov (che richiede solo le ipotesi stocastiche deboli) e quindi lo stimatore OLS sarà sempre un BLUE. Non sarà tuttavia possibile effettuare inferenze affidabili su di esso.

Conviene pertanto eliminare gli effetti di disturbo delle osservazioni anomale facendoli rappresentare da specifiche variabili di comodo che valgono uno nel tempo in cui si manifesta l’osservazione anomala e zero negli altri tempi. Ovviamente ad ogni dato considerato anomalo corrisponde una variabile di comodo, per cui è necessario essere molto parsimoniosi nella loro valutazione per non perdere, nella stima delle equazioni, troppi gradi di libertà.

Non sempre è possibile valutare a priori il carattere anomalo di alcuni dati nel campione delle osservazioni, per cui ci si accorge di questa anomalia soltanto dopo aver stimato le equazioni. Alla luce della discussione precedente gli strumenti diagnostici che possono tornare utili a questo scopo sono due: un test di normalità dei residui e l’analisi del grafico dei residui.

### *Il test di normalità dei residui di Jarque e Bera*

Per illustrare il funzionamento del test di normalità richiamiamo i coefficienti di asimmetria e di curtosi di una variabile aleatoria. Sia

---

<sup>9</sup> Il termine curtosi viene dal greco *κυρως*, che significa “curvo” (lat. *curvus*) e indica appunto quanto è “arcuata” la distribuzione di probabilità. Il termine platicurtica è composto con l’aggettivo *πλατυς* che significa “piatto” e denota una distribuzione più appiattita della normale (cioè con moda più bassa e code più alte). Il coefficiente di curtosi è definito formalmente nel paragrafo XXI-(1.3).

<sup>10</sup> Per i dettagli si veda Spanos [1986, par.21.2].

$$\mu_r = E(x - \mu)^r$$

il momento centrale  $r$ -esimo della variabile considerata, per cui  $\mu_2 = \sigma^2$  è la varianza della sua distribuzione. Il coefficiente di asimmetria è allora dato da

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\mu_3}{\sigma^3} \quad (3.2.10)$$

e quello di curtosi da

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} \quad (3.2.11)$$

I test di normalità si basano sul fatto che in una distribuzione normale è  $\gamma_1 = 0$  (come in ogni distribuzione simmetrica) e inoltre  $\beta_2 = 3$  (peculiarità questa della sola normale). Per questo motivo il coefficiente di curtosi viene ridefinito centrandolo sul valore che assume nella normale

$$\gamma_2 = \beta_2 - 3 \quad (3.2.12)$$

Il coefficiente (3.2.12) viene definito "curtosi eccedente" ed è ovviamente nullo nel caso della normale.

Il test sfrutta le controparti campionarie delle (3.2.10) e (3.2.12), date rispettivamente da

$$\hat{\gamma}_1 = \frac{\frac{1}{n} \sum_{t=1}^n \hat{u}_t^3}{\left( \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 \right)^{\frac{3}{2}}} \quad (3.2.13)$$

e

$$\hat{\gamma}_2 = \frac{\frac{1}{n} \sum_{t=1}^n \hat{u}_t^4}{\left( \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 \right)^2} - 3 \quad (3.2.14)$$

Si dimostra che in campioni sufficientemente grandi i due coefficienti campionari sono stocasticamente indipendenti e si distribuiscono in modo normale

$$\hat{\gamma}_1 \sim N\left(0, \frac{6}{n}\right) \quad (3.2.15)$$

$$\hat{\gamma}_2 \sim N\left(0, \frac{24}{n}\right)$$

Ne consegue che, sempre per campioni sufficientemente grandi, la statistica ottenuta dalla somma delle (3.2.13) e (3.2.14) standardizzate con i rispettivi scarti quadratici medi ed elevate al quadrato si distribuisce come un  $\chi^2$  con due gradi di libertà (indichiamo questa statistica con le lettere  $JB$ , iniziali di Jarque e Bera, i due econometrici che hanno per primi proposto questo test<sup>11</sup>)

$$JB = \frac{n}{6} \hat{\gamma}_1^2 + \frac{n}{24} \hat{\gamma}_2^2 \sim \chi^2_2 \quad (3.2.16)$$

Quando la statistica (3.2.16) eccede il valore soglia della distribuzione del  $\chi^2_2$ , pari a 5.99 al livello di significatività  $\alpha = 0.05$ , concludiamo che i residui sono non normali, il che generalmente sarà dovuto alla presenza di una o più osservazioni anomale.

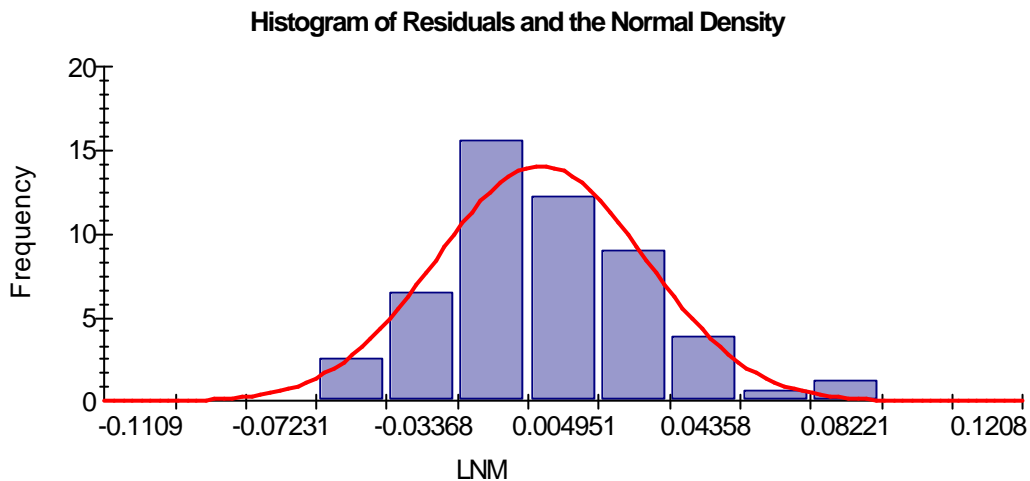


Figura 3.1- L'istogramma dei residui del modello (1.10.2) confrontato con la distribuzione normale. Si noti l'asimmetria positiva dei residui.

Naturalmente la (3.2.16) non ci indica in quale periodo si verificano le anomalie. Un'analisi attenta del grafico dei residui è pertanto necessaria al fine di individuare i tempi delle anomalie e di neutralizzarne gli effetti. Euristicamente, sotto l'ipotesi nulla di normalità dei residui, è possibile ritenere anomali i dati che producono residui al di fuori dell'intervallo  $[-2\sigma, +2\sigma]$ , poiché questi si verificherebbero con probabilità molto piccola, inferiore al 5%. Ovviamente questo intervallo di accettazione di osservazioni non anomale può essere ampliato o ristretto in funzione delle opinioni del ricercatore.

<sup>11</sup> Si veda Jarque e Bera [1980].

Un altro utile strumento diagnostico è dato dall'istogramma dei residui, che approssima la forma della distribuzione di probabilità di questi e quindi, se confrontato con una distribuzione normale di varianza pari a quella dei residui stimati, fornisce una indicazione sintetica delle cause di non normalità (ad esempio, aiuta a stabilire se questa è determinata da asimmetria – poche osservazioni anomale dello stesso segno – o da eccesso di curtosi – molte osservazioni anomale di entrambi i segni).

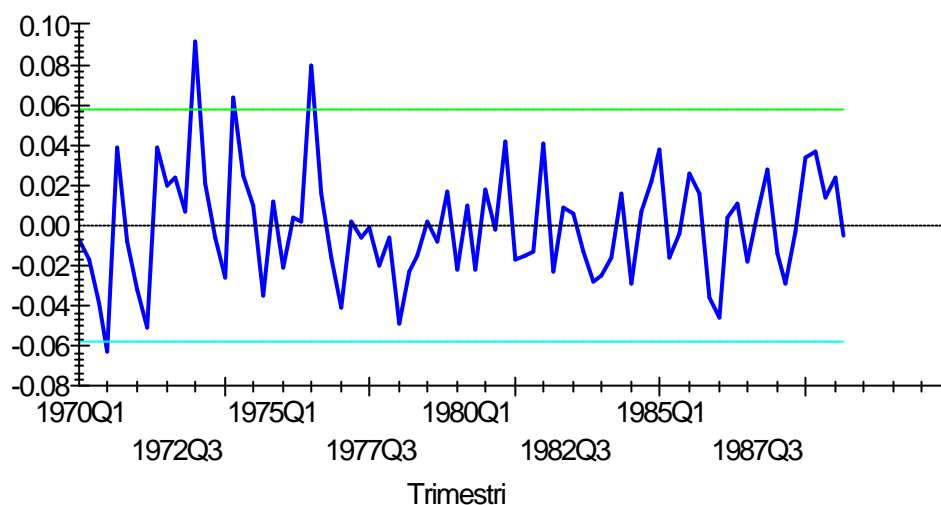


Figura 3.2 – I residui OLS del modello 1.10.2. Il grafico riporta la banda di ampiezza  $4s$  centrata sullo zero. L'asimmetria positiva di essi è rivelata dalla presenza di tre residui anomali positivi, contro uno solo negativo (cfr. la figura 1.5).

Applichiamo ora questi concetti all'equazione delle importazioni (1.10.1), le cui stime sono riportate nella (1.10.2). Vale la pena di sottolineare che utilizziamo il modello non vincolato (1.10.1) perché i test per le ipotesi di uguaglianza delle elasticità e di omogeneità di grado zero nei prezzi, condotti nel paragrafo 2.4, respingono le rispettive ipotesi nulle, suggerendo quindi che le stime vincolate presentate nel paragrafo 1.11 sono inappropriate. Naturalmente se i residui risultassero non normali il risultato di questi test andrebbe revocato in dubbio. Una corretta prassi vorrebbe quindi che essi venissero reiterati dopo aver individuato e corretto la causa della non normalità.

Il test di normalità (3.2.16) sui residui della (1.10.2) dà una statistica pari a 6.92, superiore al valore soglia. Concludiamo che i residui sono non normali. Il loro istogramma è riportato nella figura 3.1 che lo confronta con una distribuzione normale di pari varianza. La distribuzione dei residui è positivamente asimmetrica. L'asimmetria è determinata in particolare da alcune osservazioni

anomale attorno a 0.08. Per individuare la data nella quale si realizzano consultiamo il grafico dei residui della (1.10.2), già visto nella figura 1.4, che riportiamo nella figura 3.2 insieme alle relative bande di confidenza date da più e meno due scarti quadratici medi.

Il grafico conferma l'asimmetria positiva dei residui (ci sono tre residui maggiori di  $2\sigma$  contro uno solo minore di  $-2\sigma$ ) e indica che l'osservazione anomala più marcata si verifica nel primo trimestre del 1973. Stimiamo quindi nuovamente il modello (1.10.1) aggiungendo la variabile  $d_{73,t}$  che vale 1 nel primo trimestre del 1973 e zero altrove. Formalmente, questa variabile può essere definita utilizzando la funzione indicatore di un evento, cioè la funzione  $I(.)$  che vale 1 se l'evento specificato fra parentesi è vero e zero in caso contrario. Nel nostro caso quindi

$$d_{73,t} = I(t=1973:1) \quad (3.2.17)$$

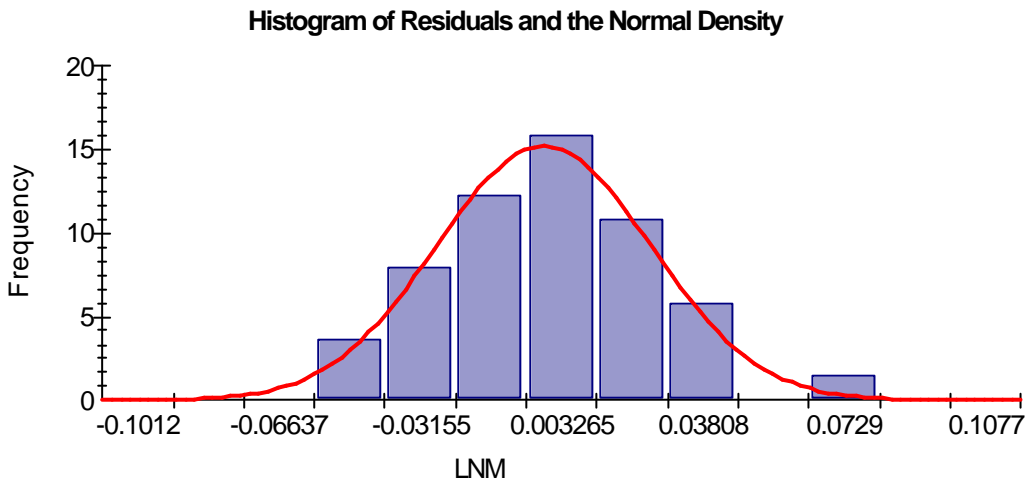


Figura 3.3 – L'istogramma dei residui della (3.2.19).

Aumentando la (1.10.1) con la variabile definita dalla (3.2.17) otteniamo il modello:

$$\ln y_t = \beta_1 + \beta_2 \ln x_{1t} + \beta_3 \ln x_{2t} + \beta_4 \ln x_{3t} + \beta_5 \ln x_{4t} + \delta d_{73,t} + u_t \quad (3.2.18)$$

dove  $\delta$  è il coefficiente della variabile di comodo. La stima della (3.2.18) fornisce i seguenti risultati

$$\hat{\ln y}_t = -6.08 \quad + 0.570 \ln x_{1t} \quad + 0.908 \ln x_{2t} \quad - 0.156 \ln x_{3t} \quad + 0.097 \ln x_{4t} \quad (3.2.19)$$

(-6.6)                      (7.1)                      (23.9)                      (-8.6)                      (3.3)

$$+ 0.096 d_{73,t}$$

(3.4)

$$n = 80, R^2 = 0.986, R^2_c = 0.985, RSS = 0.054, SEE = 0.027, JB = 2.07$$

Il coefficiente della variabile di comodo è significativo, con una  $t$  pari a 3.4, e la statistica del test di normalità scende ora a 2.07, all'interno della regione di accettazione. La correzione sembra quindi aver avuto successo. Si noti che il coefficiente di determinazione corretto è aumentato rispetto alla (1.10.2).

L'istogramma dei residui della (3.2.19) è riportato nella figura 3.3 e appare ora più simmetrico. Il grafico dei residui (figura 3.3) mostra il permanere di alcune osservazioni al di fuori della banda di confidenza. Tuttavia esse non sono tali da indurre una significativa asimmetria dei residui (il test  $JB$  non rifiuta) e quindi si può presumere che non influenzino le proprietà del modello.

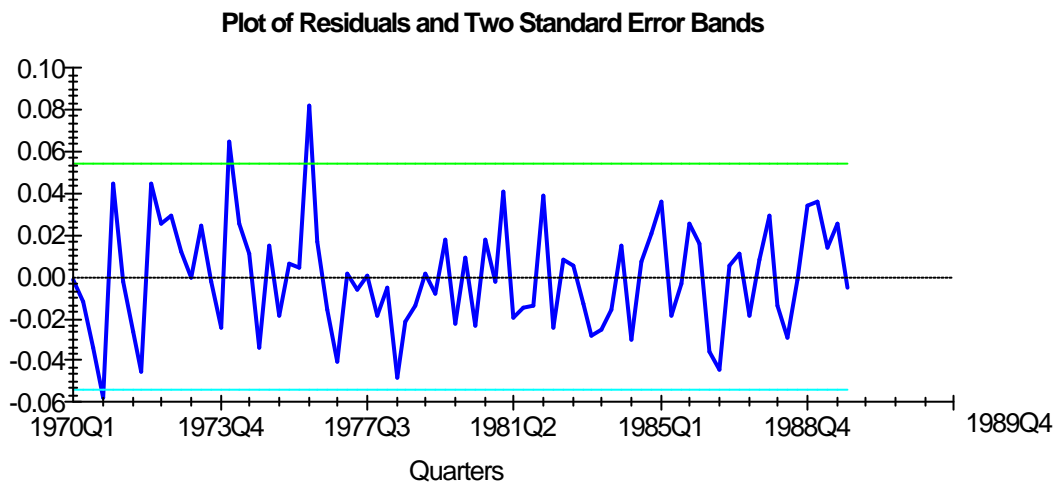


Figura 3.4 - Il grafico dei residui della (3.2.19).

Possiamo ora ripetere i test per le ipotesi di uguaglianza delle elasticità e di omogeneità di grado zero nei prezzi. Ad esempio, la devianza dei residui nel modello con il vincolo  $H_0: \beta_2 = \beta_3$  e la dummy puntuale è pari a 0.063, per cui applicando la prima delle (2.5.4) otteniamo

$$\frac{0.063 - 0.054}{0.054} \frac{74}{1} = 12.3$$

da confrontare con il valore soglia di una  $F_{1,74}$  che al livello 0.05 è pari a 3.8. Di conseguenza l'ipotesi di uguaglianza delle elasticità viene respinta anche dopo aver corretto le osservazioni anomale e ripristinato la normalità dei residui.

### 3.3 Variabili di comodo stagionali

Le variabili di comodo possono essere utilmente adoperate nella cosiddetta *depurazione stagionale*. Per motivare ed illustrare questa procedura osserviamo che gran parte dei fenomeni economici posseggono un andamento *stagionale*, cioè una conformazione di diminuzioni ed aumenti di valore che si ripetono in misura simile ogni anno per cause diverse, ad esempio la preferenza o l'abitudine dei lavoratori a concentrare le ferie in uno stesso periodo, il *window dressing* degli istituti di credito che produce un incremento dei depositi nel mese di dicembre, o anche la convenienza che alcuni operatori della distribuzione sembrano avere nell'aumentare i prezzi dei prodotti distribuiti in particolari periodi dell'anno. Nel primo esempio si ha una configurazione stagionale nelle serie della produzione, nel secondo una nelle serie dei depositi bancari, nel terzo l'andamento stagionale nei prezzi.

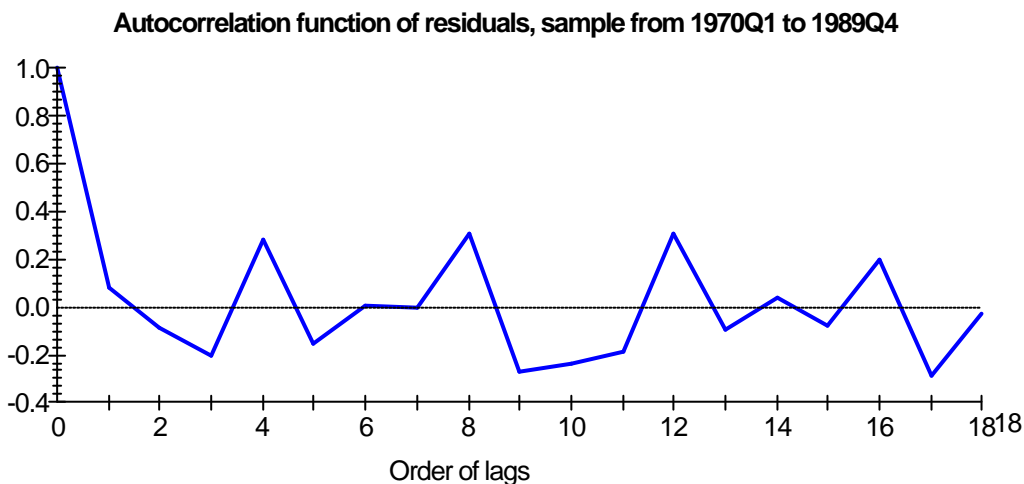


Figura 3.5 – Il correlogramma dei residui del modello (3.2.19).

La presenza delle stagionalità nelle osservazioni campionarie può distorcere notevolmente la significatività delle equazioni che vengono stimate. Se è stagionale la variabile endogena ma questo carattere non è rappresentato dall'insieme delle esogene, accade che la configurazione stagionale venga trasferita sui residui, che quindi presentano una correlazione nonnulla *al ritardo stagionale*: quattro se le serie storiche adoperate sono trimestrali, dodici se le serie sono mensili. In altre parole la ripetizione annuale delle diminuzioni e degli aumenti di valore, che equivale ad una periodicità di quattro o di dodici unità temporali a seconda della cadenza delle serie storiche campionarie, si traduce in una covarianza nonnulla dei



residui intervallati di quattro o di dodici tempi. La terza delle ipotesi deboli (1.6.10) sui residui viene quindi a cadere.

A titolo di esempio si osservi nella figura 3.5 il correlogramma dei residui dell'equazione (3.2.19), rappresentati a loro volta nella figura 3.4. Il correlogramma 3.5 presenta un picco a tutti i ritardi stagionali (cioè ai ritardi multipli interi della cadenza stagionale, che per dati trimestrali è uguale a quattro) e rende quindi evidente un fenomeno che già l'osservazione diretta del grafico 3.4 lascia intuire, ovvero l'esistenza di cicli stagionali nei residui. Questo fenomeno è prevedibile. La figura 1.3 infatti mostra che le variabili coinvolte nella stima hanno tutte una marcata stagionalità, ma il profilo di questa stagionalità differisce dall'una all'altra: ad esempio, la serie dei consumi ha un andamento più livellato di quella delle importazioni e degli investimenti. Queste differenze si scaricano sul residuo dell'equazione, determinandone la stagionalità.

Può, d'altro canto, succedere che la conformazione stagionale dell'endogena sia ben rappresentata da un carattere analogo presentato dalla parte sistematica dell'equazione. In questo caso i residui posseggono covarianze nulle ai ritardi stagionali ma può sussistere il problema che la significatività dell'equazione sia in gran parte (o totalmente) dovuta proprio alle stagionalità simili presenti nella variabile endogena e nell'insieme delle esogene. La bontà della relazione stimata viene quindi, in questa situazione, a dipendere dalle caratteristiche stagionali e non da una effettiva associazione economica.

#### *La depurazione stagionale con il criterio dei minimi quadrati*

Per ovviare a questi problemi è necessario *depurare* le serie campionarie delle stagionalità prima di utilizzarle nella costruzione di equazioni econometriche. Per eliminare le stagionalità nella serie storica  $\{y_t\}$  si può aggiungere ai suoi elementi una costante  $b$  che varia da trimestre a trimestre o da mese a mese a seconda della cadenza: a titolo esemplificativo, nel primo caso si ha

$$\begin{aligned} y_1 &= b_1 + y_1^d \\ y_2 &= b_2 + y_2^d \\ y_3 &= b_3 + y_3^d \\ y_4 &= b_4 + y_4^d \\ y_5 &= b_1 + y_5^d \\ &\dots \end{aligned} \tag{3.3.1}$$

con le quattro costanti che si ripetono ogni anno. La variabile con l'indice  $d$  non presenta più stagionalità e forma la serie *destagionalizzata*  $\{y_t^d\}$  se vale l'ipotesi di configurazione stagionale costante nel tempo; se la stagionalità è variabile la  $\{y_t^d\}$  è destagionalizzata soltanto approssimativamente.

Utilizzando quattro variabili di comodo (*stagionali*)  $d_{1t}$ ,  $d_{2t}$ ,  $d_{3t}$ ,  $d_{4t}$ , le (3.3.1) possono essere definite tramite l'equazione

$$y_t = b_1 d_{1t} + b_2 d_{2t} + b_3 d_{3t} + b_4 d_{4t} + y_t^d \quad (3.3.2)$$

con

$$d_{it} = \begin{cases} 1 & t = i, i+4, i+8, \dots \\ 0 & t \neq i, i+4, i+8, \dots \end{cases} \quad i = 1, 2, 3, 4 \quad (3.3.3)$$

La (3.3.2) può essere scritta nella forma compatta

$$\mathbf{y} = \mathbf{S}\mathbf{b} + \mathbf{y}^d \quad (3.3.4)$$

dove  $\mathbf{y}^d$  è il vettore delle  $y_t^d$  destagionalizzate,  $\mathbf{b} = [b_1 \ b_2 \ b_3 \ b_4]'$  è il vettore dei *fattori di destagionalizzazione* e la matrice  $\mathbf{S}$  di ordine  $n \times 4$  è formata dalle quattro variabili di comodo

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (3.3.5)$$

*Osservazione 3.4* - A causa della presenza delle variabili di comodo, nella (3.3.2) manca l'intercetta. Se la si volesse inserire, le variabili di comodo sarebbero soltanto tre.

Ovviamente, se la cadenza dei dati è mensile, le variabili di comodo nella (3.3.2) sono dodici e l'ordine della matrice  $\mathbf{S}$  è  $n \times 12$ .

La stima di  $\mathbf{b}$  è facilmente ottenuta utilizzando il criterio dei minimi quadrati ordinari sull'equazione (3.3.2) dove la serie dei residui  $\{y_t^d\}$  costituisce la variabile destagionalizzata e dove l'intercetta è data dall'insieme dei coefficienti delle variabili di comodo; considerando la (3.3.4) si ha

$$\hat{\mathbf{b}} = (\mathbf{S}'\mathbf{S})^{-1} \mathbf{S}'\mathbf{y}$$

per cui

$$\hat{\mathbf{y}}^d = \mathbf{y} - \mathbf{S}\hat{\mathbf{b}} = \mathbf{y} - \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1} \mathbf{S}'\mathbf{y} = \mathbf{M}\mathbf{y} \quad (3.3.6)$$

dove

$$\mathbf{M} = \mathbf{I} - \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' \quad (3.3.7)$$

Queste ultime due relazioni indicano che la serie destagionalizzata è ottenuta come combinazione lineare, mediante la matrice  $\mathbf{M}$ , delle serie originale  $\{y_t\}$ .

La matrice  $\mathbf{M}$  è del tipo (1.7.4), quindi simmetrica ed idempotente; si vede, inoltre, immediatamente che

$$\mathbf{MS} = 0 \quad (3.3.3)$$

L'idempotenza di  $\mathbf{M}$  è una caratteristica interessante della depurazione stagionale operata con variabili di comodo, poiché permette di stabilire che tale procedura è influente se eseguita su di una serie già destagionalizzata; infatti, se questa è  $\mathbf{My}$ , si ha

$$\hat{y}^d = \mathbf{M}(\mathbf{My}) = \mathbf{My} \quad (3.3.9)$$

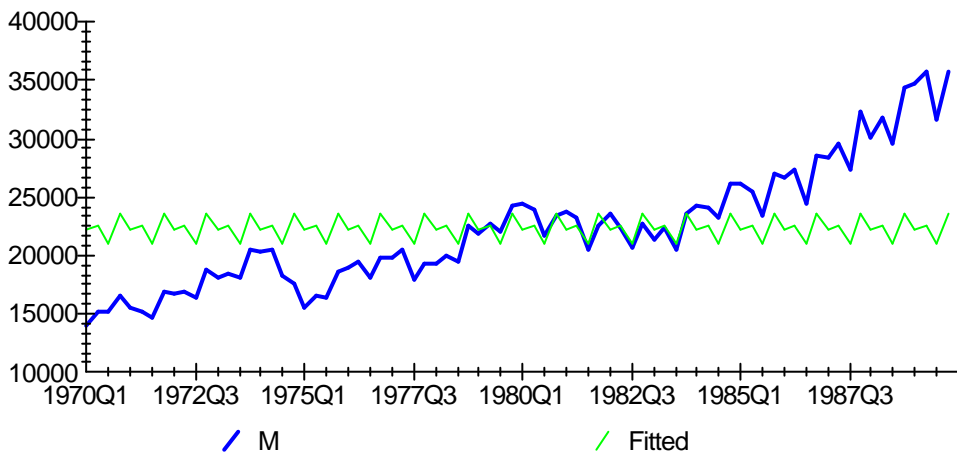


Figura 3.6 – La serie storica non destagionalizzata delle importazioni di beni e servizi a prezzi 1980 e la stima dei fattori di destagionalizzazione effettuata con la (3.3.2).

A titolo di esempio, possiamo destagionalizzare con la (3.3.4) la serie storica delle importazioni, che costituisce la variabile dipendente della (3.2.18) ed è già stata rappresentata nella figura 1.2. Stimando la (3.3.2) otteniamo

$$\hat{y}_t = 22189.7 d_{1t} + 22568.7 d_{2t} + 20966.5 d_{3t} + 23567.8 d_{4t}$$

(18.8)                      (19.1)                      (17.7)                      (19.9)

$$n = 80, R^2 = 0.031, R^2_c = -0.006, RSS = 2.12E+09, SEE = 5275.8, JB = 6.79$$

Si noti che tutti i fattori di destagionalizzazione sono fortemente significativi, anche se la bontà complessiva del modello è scarsa, con un coefficiente di

determinazione corretto addirittura negativo. Questo è un risultato ovvio e non preoccupante, essendo determinato dal fatto che la (3.3.4) si propone di catturare solo un aspetto della serie, cioè la sua stagionalità. Di conseguenza, se il contributo di questa alla variabilità complessiva della serie è relativamente piccolo, come accade per serie dotate di forte tendenza, la devianza non spiegata dalla (3.3.4) sarà relativamente grande e il coefficiente di determinazione piccolo.

La figura 3.6 propone il grafico dei valori storici e di quelli stimati, che costituiscono una stima delle stagionalità della serie. Il residuo stimato, che corrisponde alla serie destagionalizzata, è rappresentato nella figura 3.7. Il confronto con la figura 3.6 rende evidente come l'applicazione della (3.3.4) abbia in effetti contribuito a livellare alquanto il profilo stagionale della serie.

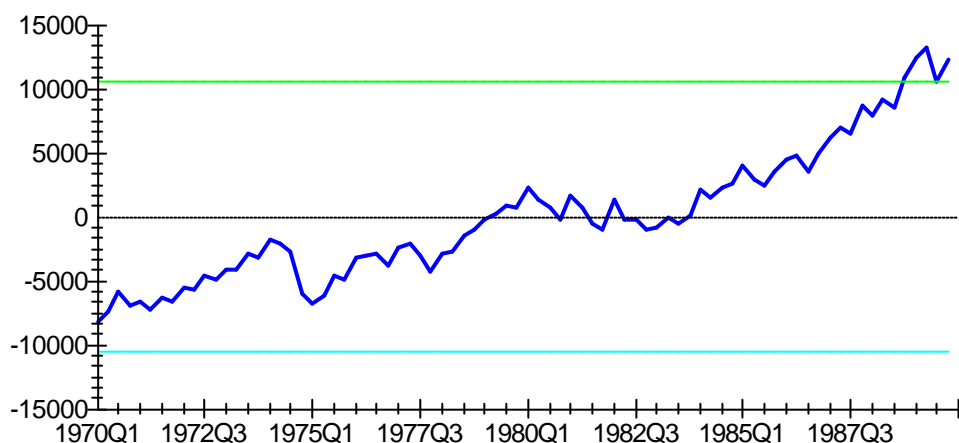


Figura 3.7 – La serie storica delle importazioni di beni e servizi a prezzi 1980 destagionalizzata con i fattori stimati mediante la (3.3.2) mostrati nella precedente figura 3.6.

### Stagionalità variabile

In molti casi gli effetti della stagionalità, sia pure periodici, non sono costanti nel tempo ma variano, ad esempio aumentando o diminuendo il modo lineare. Questo fenomeno si presenta con grande frequenza nel caso di serie storiche espresse in termini nominali, la cui variabilità, anche stagionale, aumenta nel tempo per effetto dell'inflazione. La figura 3.8 esemplifica questo andamento con riferimento alla serie dell'indice dei salari nominali nel settore manifatturiero in Italia (dati trimestrali grezzi dal 1979:1 al 1988:4).

In questo caso la (3.3.2) deve essere sostituita con l'equazione seguente

$$y_t = b_1(\alpha_0 + \alpha_1 t)d_{1t} + b_2(\alpha_0 + \alpha_1 t)d_{2t} + b_3(\alpha_0 + \alpha_1 t)d_{3t} + b_4(\alpha_0 + \alpha_1 t)d_{4t} + y_t^d \quad (3.3.10)$$

nella quale si vede chiaramente come i fattori di destagionalizzazione varino linearmente nel tempo con lo stesso andamento.

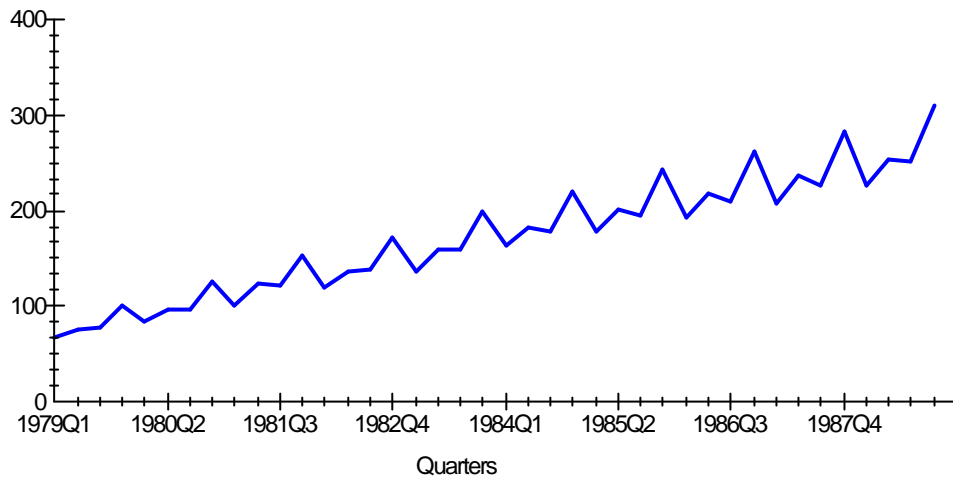


Figura 3.8 - L'indice dei salari nominali nel settore manifatturiero in Italia (dati trimestrali grezzi dal 1979:1 al 1988:4). Si tratta di un tipico caso di serie con ciclo stagionale di ampiezza crescente.

La (3.3.10) può essere scritta nel modo

$$y_t = \sum_{i=1}^4 \beta_i d_{it} + \sum_{i=1}^4 \gamma_i q_{it} + y_t^d \quad (3.3.11)$$

dove le variabili di comodo  $d_{it}$  sono ancora definite dalla (3.3.3) e le altre  $q_{it}$  dalla

$$q_{it} = t \cdot d_{it} \quad i=1, 2, 3, 4 \quad (3.3.12)$$

Per stimare i fattori di destagionalizzazione nella (3.3.11) si può utilizzare ancora il criterio dei minimi quadrati applicati alla forma (3.3.4) nella quale ora  $\mathbf{b} = [\beta_1 \beta_2 \beta_3 \beta_4 \gamma_1 \gamma_2 \gamma_3 \gamma_4]'$  e la matrice  $\mathbf{S}$  di ordine  $n \times 8$  è data da

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 4 \\ 1 & 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 6 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \quad (3.3.13)$$

Nel caso della serie dei salari rappresentata nella figura 3.8 la procedura di destagionalizzazione con la (3.3.11) fornisce la serie rappresentata nella figura 3.9.

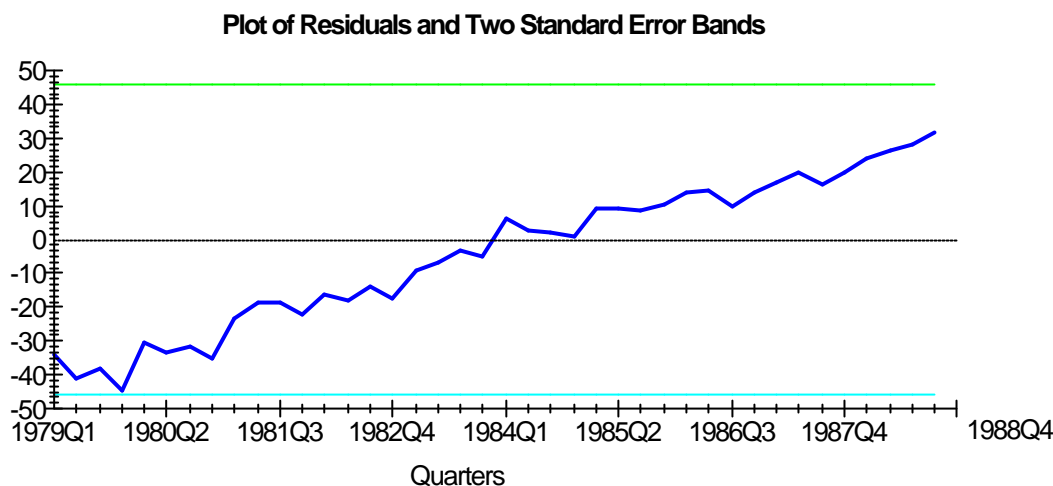


Figura 3.9 – L'indice dei salari nominali della figura 3.8 destagionalizzato con la (3.3.11).

#### Stagionalità additiva o moltiplicativa

Nella (3.3.2) i fattori di destagionalizzazione sono *additivi* nel senso che aggiungono o sottraggono ad  $y_t$  una quantità variabile a seconda della sua conformazione stagionale. In alcune circostanze, tuttavia, è più conveniente rappresentarli nella forma *moltiplicativa*, che ritorna ad essere del tipo (3.3.2) se si prendono i logaritmi dei due membri

$$y_t = b_1^{d_{1t}} b_2^{d_{2t}} b_3^{d_{3t}} b_4^{d_{4t}} y_t^d \quad (3.3.14)$$

La generalizzazione della stagionalità moltiplicativa al caso variabile del punto precedente è lasciata al lettore.

È interessante osservare che il modello moltiplicativo fornisce una rappresentazione più parsimoniosa, ma spesso ugualmente valida, delle stagionalità variabili. Ciò deriva dal fatto che la trasformazione logaritmica, “schiacciando” in modo più che proporzionale i valori più elevati, contribuisce a stabilizzare la varianza della serie, e quindi anche quella del suo ciclo stagionale. A titolo di esempio si osservi, nella figura 3.10, il grafico del logaritmo dell'indice dei salari nominali grezzi rappresentato nella figura 3.8. La serie logaritmizzata ha stagionalità di ampiezza stabile, le quali quindi si prestano ad essere rappresentate mediante il modello (3.3.2), che ora diventa

$$\ln y_t = b_1 d_{1t} + b_2 d_{2t} + b_3 d_{3t} + b_4 d_{4t} + y_t^d$$

anziché con il più complicato (3.3.11). Ma prendendo gli antilogaritmi dell'ultima formula otteniamo appunto il modello moltiplicativo (3.3.14).

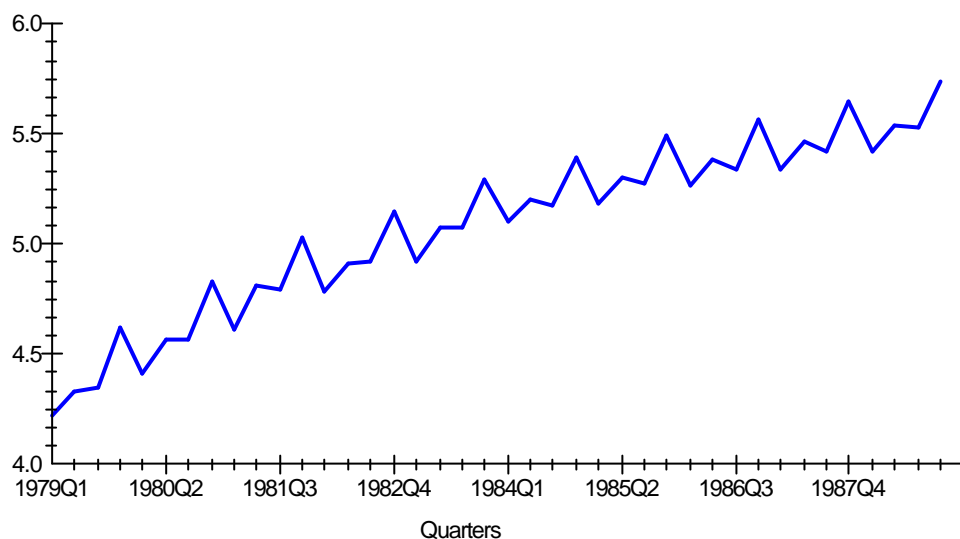


Figura 3.10 – Il logaritmo dell'indice dei salari nominali grezzi rappresentato nella figura 3.8. È evidente come la trasformata logaritmica stabilizzi la varianza del ciclo stagionale. Di converso, ciò significa che le stagionalità della variabile in unità naturali possono essere rappresentate adeguatamente dal modello moltiplicativo (3.3.14).

Stimando il modello (3.3.2) con i logaritmi dei salari nominali grezzi otteniamo

$$\ln \hat{y}_t = 4.92 d_{1t} + 5.05 d_{2t} + 5.04 d_{3t} + 5.27 d_{4t}$$

(39.9)      (41.0)      (40.9)      (42.8)

$$n = 80, R^2 = 0.105, R^2_c = 0.030, RSS = 5.46, SEE = 5.07, JB = 3.13$$

e i relativi valori destagionalizzati sono rappresentati nella figura 3.11, dalla quale si vede che il modello moltiplicativo contribuisce a livellare, ma non elimina del tutto, le ciclicità stagionali.

#### La conservazione dei volumi

In effetti, non è conveniente utilizzare  $\{\hat{y}_t^d\}$  come serie destagionalizzata se la sua media è diversa da zero oppure se essa possiede una tendenza. Nel primo caso, infatti, pur se  $\sum_{t=1}^n y_t / n = \bar{y} \neq 0$ , si ha che  $\sum_{t=1}^n \hat{y}_t^d = 0$  in virtù della (1.4.18) e considerando che i coefficienti delle variabili di comodo formano l'intercetta. Questo azzeramento della somma dei dati destagionalizzati è un difetto della procedura

poiché per ragioni di contabilità è generalmente utile che la somma dei dati destagionalizzati coincida con la somma dei dati originali (*conservazione dei volumi*); si usa porre, allora,

$$\hat{y}_t^c = \hat{y}_t^d + \bar{y} \quad t = 1, 2, \dots, n \quad (3.3.15)$$

ottenendosi una serie che al tempo stesso è destagionalizzata e conserva i volumi.

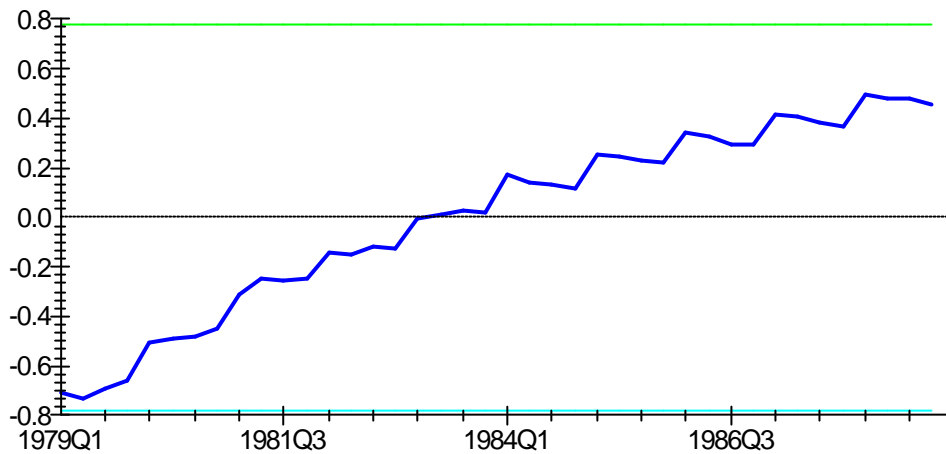


Figura 3.11 - La serie storica dei salari nominali (in logaritmi) destagionalizzata col modello moltiplicativo.

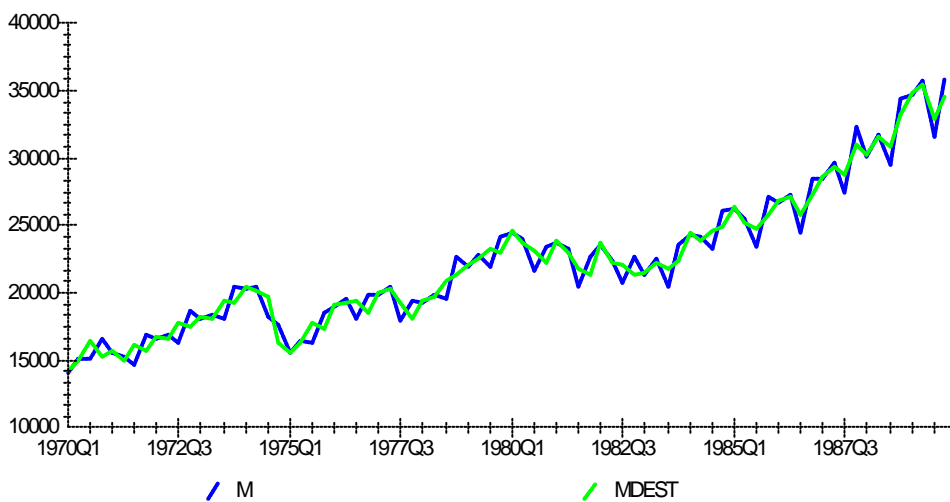


Figura 3.12 - La serie delle importazioni destagionalizzata con la conservazione dei volumi.



La figura 3.12 rappresenta la serie storica delle importazioni destagionalizzata e trasformata con la (3.3.15).

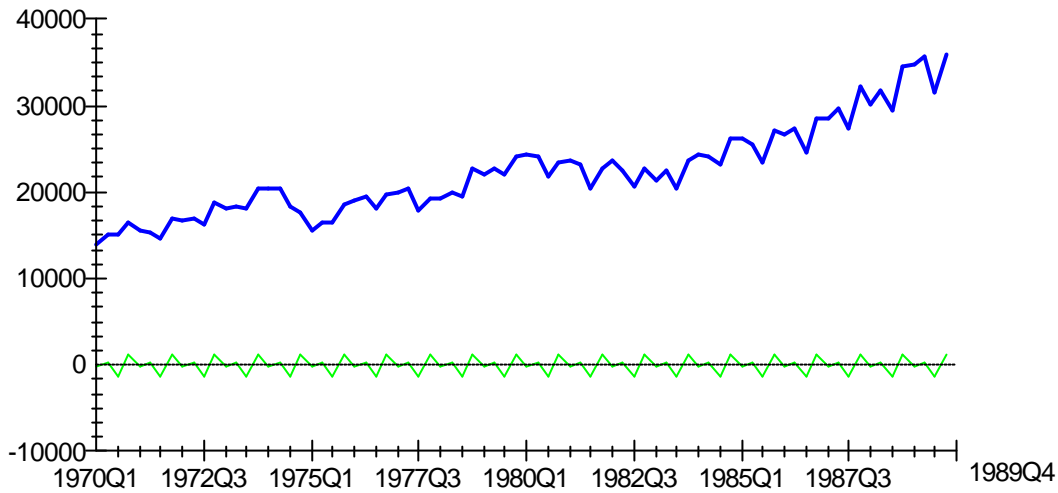


Figura 3.13 – Il grafico dei valori storici e stimati delle importazioni ottenuti con il modello (3.3.17). I valori stimati coincidono in questo caso con il ciclo stagionale centrato sullo zero, in virtù delle proprietà delle variabili di comodo centrate (3.3.16). Sottraendo ai valori storici questa stima centrata del ciclo stagionale si ottiene una serie destagionalizzata che conserva i volumi (si veda la figura 3.14).

Lo stesso risultato di conservazione dei volumi si ottiene utilizzando al posto delle variabili di comodo definite dalla (3.3.3) quelle centrate, definite come

$$\bar{d}_{it} = \begin{cases} 0.75 & t = i, i + 4, i + 8, \dots \\ -0.25 & t \neq i, i + 4, i + 8, \dots \end{cases} \quad i = 1, 2, 3, 4 \quad (3.3.16)$$

Le  $\bar{d}_{it}$  si ottengono sottraendo alle  $d_{it}$  la loro media, che nel caso il campione comprenda un numero intero di anni è pari a 0.25. Si noti che ognuna delle  $\bar{d}_{it}$  è perfettamente collineare alle altre tre. Si ha, ad esempio,  $\bar{d}_{1t} = -(\bar{d}_{2t} + \bar{d}_{3t} + \bar{d}_{4t})$ . Di conseguenza solo tre delle  $\bar{d}_{it}$  possono essere inserite contemporaneamente in una regressione e la (3.3.2) diventa

$$y_t = b_1 \bar{d}_{1t} + b_2 \bar{d}_{2t} + b_3 \bar{d}_{3t} + \hat{y}_t^c \quad (3.3.17)$$

dove ora il residuo conserva i volumi, dato che le variabili di comodo, essendo centrate, non sottraggono alla  $y_t$  la propria media campionaria.

Nel caso delle importazioni la stima del modello con variabili di comodo (3.3.17) centrate fornisce i seguenti risultati

$$\hat{y}_t = -1378.1 \bar{d}_{1t} - 999.1 \bar{d}_{2t} - 2601.3 \bar{d}_{3t}$$

(-0.1)
(-0.1)
(-0.3)

$$n = 80, R^2 = -18.2, R^2_c = -18.1, RSS = 4.20E+10, SEE = 23349.8, JB = 55.74$$

Il grafico dei valori storici e stimati è riportato nella figura 3.13 e da esso risulta evidente che la (3.3.17) fornisce una stima del ciclo stagionale centrata sullo zero. Sottraendo questa stima alla serie originale otteniamo il residuo della (3.3.17), cioè la serie destagionalizzata che conserva i volumi (la stessa rappresentata nella figura 3.12).

#### *Destagionalizzazione in presenza della tendenza*

Nel secondo caso citato all'inizio del punto precedente, ovvero quando la variabile possiede una tendenza, sussiste il difetto di una distorsione delle stime dei fattori di destagionalizzazione indotta dalla presenza della tendenza nella serie originale. Infatti, sempre nel caso della cadenza trimestrale dei dati, facilmente estendibile a quello della cadenza mensile, se  $n$  è multiplo di 4, come accade generalmente, si ha

$$\mathbf{S} = [\mathbf{I}_4 \mathbf{I}_4 \dots \mathbf{I}_4]'$$

$$\mathbf{S}'\mathbf{S} = [\mathbf{I}_4\mathbf{I}_4 + \mathbf{I}_4\mathbf{I}_4 + \dots + \mathbf{I}_4\mathbf{I}_4] = \langle n/4, n/4, n/4, n/4 \rangle = (n/4) \cdot \mathbf{I}_4$$

per cui

$$\hat{\mathbf{b}} = \frac{4}{n} \cdot [\mathbf{I}_4 \mathbf{I}_4 \dots \mathbf{I}_4] \mathbf{y} = [\bar{y}_1 \bar{y}_2 \bar{y}_3 \bar{y}_4]'$$

dove  $\bar{y}_i$ ,  $i = 1, 2, 3, 4$ , indica la media aritmetica delle  $y_t$  relative al solo trimestre  $i$ -esimo. In tale maniera, la presenza della tendenza nella serie originale aumenta di per sé i valori dei fattori stagionali man mano che si passa dal primo al quarto, se è crescente; li diminuisce se decrescente.

Per depurare correttamente una serie delle stagionalità, allora, è necessario eliminare anche la tendenza; ad esempio, mediante un polinomio del tipo (I-2.8.3). Combinando questo e la (3.3.2) si ottiene

$$y_t = a_1 t + a_2 t^2 + \dots + a_p t^p + b_1 d_{1t} + b_2 d_{2t} + b_3 d_{3t} + b_4 d_{4t} + y_t^* \quad (3.3.18)$$

dove  $\{ \hat{y}_t^* \}$  è la serie depurata sia della tendenza che delle stagionalità. Si osservi che nella (3.3.18) è stato tolto il termine di grado zero in  $t$  a causa della presenza delle variabili di comodo.

Scrivendo la (3.3.18) per tutte le  $t$  si ottiene, in forma sintetica

$$\mathbf{y} = \mathbf{Pa} + \mathbf{Sb} + \mathbf{y}^*$$

dove  $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_p]'$  e

$$\mathbf{P} = \begin{bmatrix} 1 & 1^2 & \dots & 1^p \\ 2 & 2^2 & \dots & 2^p \\ 3 & 3^2 & \dots & 3^p \\ \dots & \dots & \dots & \dots \\ n & n^2 & \dots & n^p \end{bmatrix} \quad (3.3.19)$$

ed ancora

$$\mathbf{y} = [\mathbf{PS}] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} + \mathbf{y}^*$$

dalla quale si traggono le stime dei minimi quadrati ordinari sia per i coefficienti  $\mathbf{a}$  che per gli altri  $\mathbf{b}$

$$\begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \end{bmatrix} = [(\mathbf{PS})'(\mathbf{PS})]^{-1}(\mathbf{PS})'\mathbf{y} = \begin{bmatrix} (\mathbf{P}') \\ (\mathbf{S}') \end{bmatrix} (\mathbf{PS}) \begin{bmatrix} (\mathbf{P}') \\ (\mathbf{S}') \end{bmatrix}^{-1} \begin{bmatrix} (\mathbf{P}') \\ (\mathbf{S}') \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{P}'\mathbf{P} & \mathbf{P}'\mathbf{S} \\ \mathbf{S}'\mathbf{P} & \mathbf{S}'\mathbf{S} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}'\mathbf{y} \\ \mathbf{S}'\mathbf{y} \end{bmatrix} \quad (3.3.20)$$

*La serie destagionalizzata con la correzione della tendenza*

Poiché si può dimostrare in algebra matriciale che data la matrice partizionata nella seguente maniera

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad (3.3.21)$$

dove  $\mathbf{A}_{11}$  ed  $\mathbf{A}_{12}$  sono matrici quadrate non singolari, l'inversa della  $\mathbf{A}$  può essere scritta nella forma

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ -\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}_{22} \end{bmatrix} \quad (3.3.22)$$

con  $\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$ , il vettore  $\hat{\mathbf{b}}$  nella (3.3.22) vale

$$\begin{aligned} \hat{\mathbf{b}} &= -\mathbf{B}_{22}(\mathbf{S}'\mathbf{P})(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{y} + \mathbf{B}_{22}\mathbf{S}'\mathbf{y} = \mathbf{B}_{22}\mathbf{S}'[\mathbf{I} - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}']\mathbf{y} = \mathbf{B}_{22}\mathbf{S}'\mathbf{Q}\mathbf{y} = \\ &= [\mathbf{S}'\mathbf{S} - \mathbf{S}'\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{S}]^{-1}\mathbf{S}'\mathbf{Q}\mathbf{y} = [\mathbf{S}'\mathbf{Q}\mathbf{S}]^{-1}\mathbf{S}'\mathbf{Q}\mathbf{y} \end{aligned} \quad (3.3.23)$$

avendo posto

$$\mathbf{Q} = \mathbf{I} - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'$$

Allora la serie  $\{y_t\}$  soltanto destagionalizzata vale

$$\hat{\mathbf{y}}^d = \mathbf{y} - \mathbf{S}\hat{\mathbf{b}} = [\mathbf{I} - \mathbf{S}[\mathbf{S}'\mathbf{Q}\mathbf{S}]^{-1}\mathbf{S}'\mathbf{Q}]\mathbf{y}$$

ancora pari ad una trasformazione lineare di  $\mathbf{y}$ .

Nel caso della serie delle importazioni, che effettivamente presenta una tendenza crescente (come risulta dalla figure 3.6 o 3.13) possiamo stimare il modello (3.3.18) nel quale poniamo inizialmente  $p = 2$  (tendenza parabolica) per tener conto del fatto che la tendenza appare non perfettamente rettilinea. La stima fornisce i seguenti risultati

$$\hat{\ln y}_t = 16582.1 d_{1t} + 16758.0 d_{2t} + 14948.3 d_{3t} + 17337.5 d_{4t} + 25.36 t + 2.24 t^2$$

(26.4)                      (26.5)                      (23.5)                      (27.1)                      (0.8)                      (5.9)

$$n = 80, R^2 = 0.913, R_c^2 = 0.907, RSS = 1.89E+08, SEE = 1600.1, JB = 2.71$$

dai quali si vede come il termine di secondo grado in  $t$  sia significativo, confermando così l'ipotesi che la tendenza sia parabolica. Il grafico dei valori storici e di quelli stimati è riportato nella figura 3.14, mentre la 3.15 riporta i residui stimati, cioè la serie delle importazioni contemporaneamente detrendizzata e destagionalizzata.

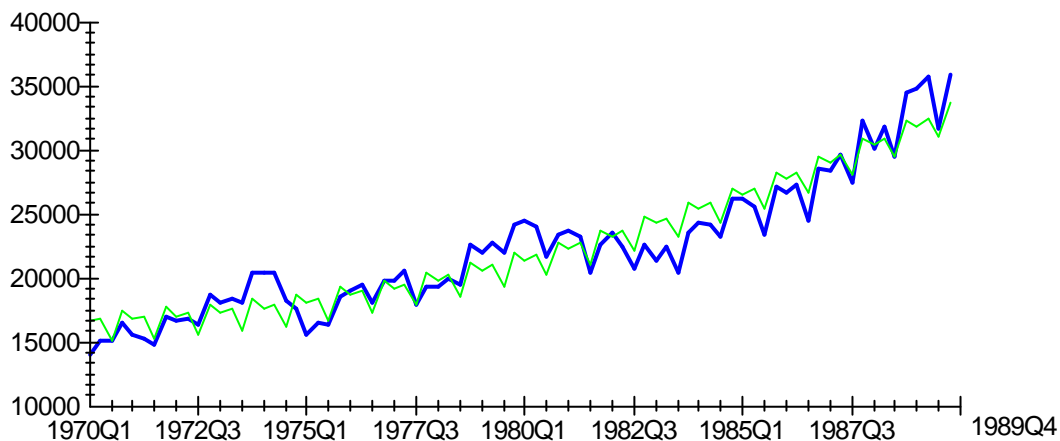


Figura 3.14 – Valori storici e stimati del modello (3.3.18) stimato con la serie delle importazioni ponendo  $p = 2$  (tendenza parabolica).

#### La stima del modello con le serie destagionalizzate

Per stimare il modello contenente serie storiche tutte destagionalizzate

$$\mathbf{y}^* = [\mathbf{x}_1^* \mathbf{x}_2^* \dots \mathbf{x}_k^*] \mathbf{b} + \mathbf{u} \quad (3.3.24)$$

dove  $\mathbf{x}_i^*$  è il vettore colonna relativo alla  $i$ -esima variabile, non è necessario eliminare preliminarmente la stagionalità in ciascuna serie e poi effettuare la stima. È più conveniente seguire un'altra strada, più immediata: se infatti le serie destagionalizzate sono

$$\begin{aligned} \mathbf{y}^* &= \mathbf{y} - \mathbf{Pa}_0 - \mathbf{Sc}_0 \\ \mathbf{x}_i^* &= \mathbf{x}_i - \mathbf{Pa}_i - \mathbf{Sc}_i \quad i = 1, 2, \dots, k \end{aligned}$$

dove la matrice  $\mathbf{P}$  data dalla (3.3.19) e la  $\mathbf{S}$  dalla (3.3.5), il modello lineare generale destagionalizzato (3.3.24) è equivalente all'altro

$$\mathbf{y} = \mathbf{Xb} - [\mathbf{Pa}_0 \ \mathbf{Pa}_1 \ \dots \ \mathbf{Pa}_k] \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} - [\mathbf{Sc}_0 \ \mathbf{Sc}_1 \ \dots \ \mathbf{Sc}_k] \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} + \mathbf{u}$$

cioè

$$\mathbf{y} = \mathbf{Xb} + (\mathbf{Pa}_0 - \mathbf{Pa}_1^+ - \mathbf{Pa}_2^+ + \dots - \mathbf{Pa}_k^+) + (\mathbf{Sc}_0 - \mathbf{Sc}_1^+ - \mathbf{Sc}_2^+ + \dots - \mathbf{Sc}_k^+) + \mathbf{u} \quad (3.3.25)$$

dove si è posto  $\mathbf{a}_i^+ = \mathbf{a}_i b_i$ ,  $\mathbf{c}_i^+ = \mathbf{c}_i b_i$ , per  $i = 1, 2, \dots, k$ . Ma la (3.3.25) può essere scritta nella forma

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Pa} + \mathbf{Sc} + \mathbf{u} \quad (3.3.26)$$

avendo posto  $\mathbf{a}_0 - \sum_{i=1}^k \mathbf{a}_i^+ = \mathbf{a}$ ,  $\mathbf{c}_0 - \sum_{i=1}^k \mathbf{c}_i^+ = \mathbf{c}$ , ed è quindi conveniente stimare contemporaneamente i parametri  $\mathbf{b}$  del modello lineare, i fattori di destagionalizzazione  $\mathbf{c}$  ed i parametri  $\mathbf{a}$  della tendenza all'interno del modello (3.3.26).

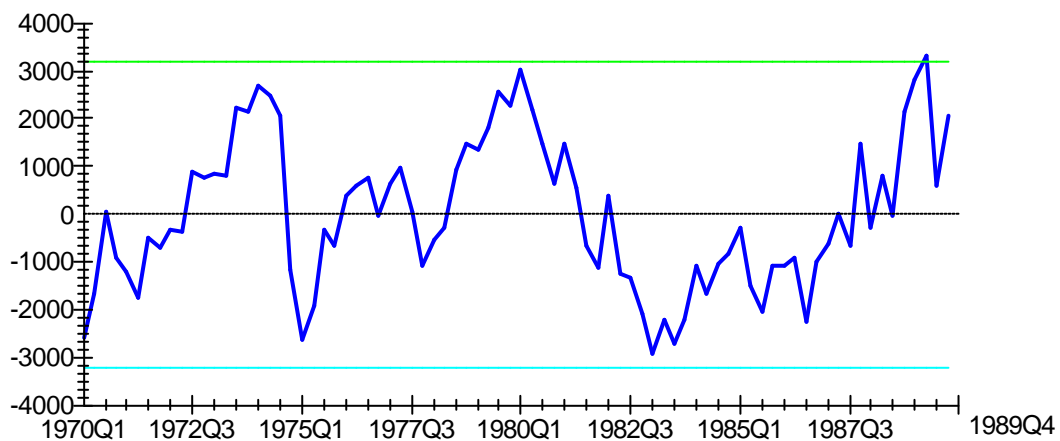


Figura 3.15 - La serie delle importazioni detrendizzata e destagionalizzata con la (3.3.18).

Possiamo ora riprendere la funzione delle importazioni (1.10.1), che nella (3.2.18) è stata aumentata con una *dummy* puntuale per tener conto di un'osservazione anomala nel primo trimestre del 1973. Come abbiamo visto all'inizio di questo paragrafo i residui della (3.2.19) mostrano stagionalità (si veda

la figura 3.5), perché il ciclo stagionale delle importazioni, esaminato in dettaglio in questo paragrafo, non coincide con quello delle variabili esplicative. Aumentiamo quindi il modello (3.2.18) con tre variabili di comodo (si ricordi che il modello contiene l'intercetta, per cui vale l'osservazione 3.4) e con una tendenza parabolica

$$\ln y_t = \beta_1 + \beta_2 \ln x_{1t} + \beta_3 \ln x_{2t} + \beta_4 \ln x_{3t} + \beta_5 \ln x_{4t} + \delta d_{73,t} + a_1 t + a_2 t^2 + b_1 d_{1t} + b_2 d_{2t} + b_3 d_{3t} + u_t \quad (3.3.27)$$

La stima della (3.3.27) fornisce i seguenti risultati

$$\begin{aligned} \ln \hat{y}_t = & -7.31 & + 0.783 \ln x_{1t} & + 0.785 \ln x_{2t} & - 0.081 \ln x_{3t} & - 0.128 \ln x_{4t} + \\ & (-4.1) & (4.0) & (13.4) & (-2.5) & (-2.2) \\ & + 0.057 d_{73,t} & + 0.003 t & + 0.00001 t^2 & + 0.033 d_{1t} & + 0.0002 d_{2t} \\ & (2.4) & (1.22) & (1.13) & (3.21) & (0.02) \end{aligned} \quad (3.3.28)$$

$$- 0.020 d_{3t} \quad (-1.7)$$

$$n = 80, R^2 = 0.992, R^2_c = 0.990, RSS = 0.032, SEE = 0.021, JB = 1.95$$

Nella (3.3.28) i coefficienti della tendenza non sono significativi. Questo risultato è plausibile e dipende dal fatto che la tendenza della variabile dipendente è spiegata da quella delle esplicative. Si noti anche che la variabile  $x_{4t}$ , che misura i prezzi interni, ha ora segno negativo in disaccordo con la teoria (secondo la quale un aumento dei prezzi interni determina un aumento della domanda di beni importati). Questo risultato può essere dovuto al fatto che nel modello sono presenti due variabili irrilevanti (la tendenza lineare e quella quadratica) entrambe fortemente correlate con la serie dei prezzi. Siamo quindi in presenza di una distorsione determinata da multicollinearità.

Per ovviare a questi problemi ristimiamo la (3.3.27) eliminando la tendenza. Otteniamo le stime

$$\begin{aligned} \ln \hat{y}_t = & -9.81 & + 1.007 \ln x_{1t} & + 0.798 \ln x_{2t} & - 0.155 \ln x_{3t} & + 0.013 \ln x_{4t} + \\ & (-8.3) & (7.1) & (13.3) & (-10.2) & (0.4) \\ & + 0.054 d_{73,t} & + 0.047 d_{1t} & + 0.012 d_{2t} & - 0.016 d_{3t} \end{aligned} \quad (3.3.29)$$

$$(2.2) \quad (5.2) \quad (1.3) \quad (-1.4)$$

$$n = 80, R^2 = 0.991, R^2_c = 0.990, RSS = 0.036107, SEE = 0.022, JB = 2.46$$

nelle quali ora i prezzi interni entrano con il segno atteso, anche se hanno un coefficiente non significativo. La correzione della stagionalità migliora la bontà

della rappresentazione fornita dal modello, che passa da un  $R^2_c = 0.985$  nella (3.2.19) a un  $R^2_c = 0.990$ . I valori storici e quelli stimati sono riprodotti nella figura 3.16, che testimonia il buon adattamento ai dati del modello prescelto.

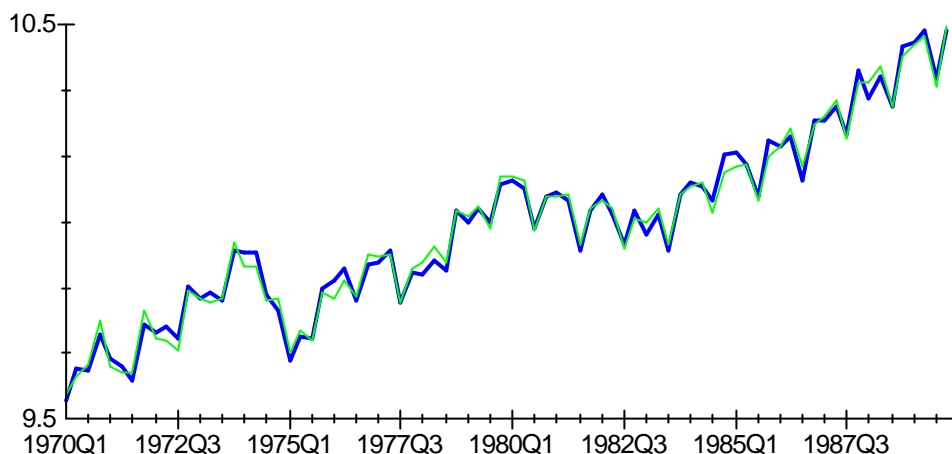


Figura 3.16 - I valori storici delle importazioni e quelli stimati col modello (3.3.29).

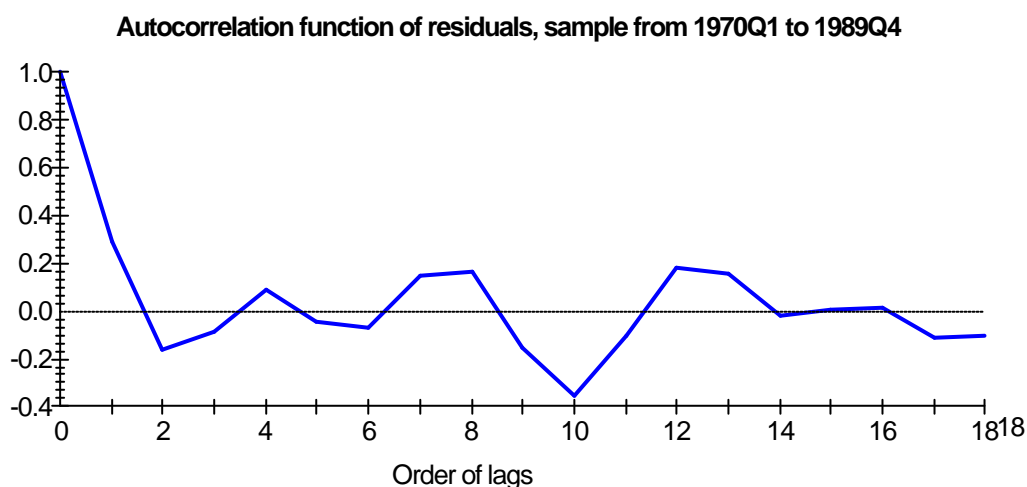


Figura 3.17 - Il correlogramma dei residui della (3.3.29). Dal confronto con la figura 3.5, che riporta il correlogramma dei residui della (3.2.19), si nota come la correzione della stagionalità mediante le variabili di comodo abbia eliminato i picchi stagionali del correlogramma.

Un'altra indicazione del miglioramento apportato al modello si ricava da correlogramma dei residui, raffigurato nella 3.17, dalla quale risulta chiaramente come la correzione con le variabili di comodo abbia eliminato le autocorrelazioni stagionali dei residui.

### 3.4 Un test di cambiamento strutturale per il modello lineare semplice

In questo paragrafo e nel successivo affrontiamo l'ipotesi di invarianza strutturale, considerandola innanzitutto con riferimento alla semplice equazione del consumo (1.3.1) e verifichiamo che i suoi due parametri  $\beta_1$  e  $\beta_2$  rimangano invariati nell'intero campione, contro l'alternativa che essi cambino passando da una prima parte del campione, di ampiezza  $n_1$ , ad una seconda, di ampiezza  $n_2$ . Dal punto di vista economico questo cambiamento ha un significato rilevante, soprattutto per quanto riguarda la propensione marginale al consumo  $\beta_2$ , che raramente rimane costante nel medio-lungo periodo.

In presenza di cambiamento strutturale si ha

$$y_t = \beta_{11} + \beta_{12}x_t + u_t \quad \text{nel primo sottoperiodo} \quad t = 1, 2, \dots, n_1 \quad (3.4.1)$$

$$y_t = \beta_{21} + \beta_{22}x_t + u_t \quad \text{nel secondo sottoperiodo} \quad t = n_1 + 1, \dots, n_1 + n_2 \quad (3.4.2)$$

dove il primo indice di ogni coefficiente indica il regime, e il secondo la variabile cui il coefficiente è riferito.

La forma matriciale (1.4.4) diventa

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{n_1} \\ y_{n_1+1} \\ y_{n_1+2} \\ \dots \\ y_{n_1+n_2} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & x_{n_1} & 0 & 0 \\ 0 & 0 & 1 & x_{n_1+1} \\ 0 & 0 & 1 & x_{n_1+2} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & x_{n_1+n_2} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_{n_1} \\ u_{n_1+1} \\ u_{n_1+2} \\ \dots \\ u_{n_1+n_2} \end{bmatrix}$$

cioè, in termini compatti

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \quad (3.4.3)$$

con ovvie definizioni dei vettori e delle matrici indicati.



La stima dei minimi quadrati ordinari è facilmente calcolata

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{bmatrix} &= \left( \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}' \right)^{-1} \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{X}_1' \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' \mathbf{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1' \mathbf{y}_1 \\ \mathbf{X}_2' \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y}_1 \\ (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{y}_2 \end{bmatrix} \end{aligned} \quad (3.4.4)$$

ed è uguale a quella che si otterrebbe stimando separatamente le due equazioni (3.4.1) e (3.4.2). Le ipotesi sotto le quali è valida la stima OLS sono le (1.4.9) adattate al caso presente

$$n_1 > 2, \quad r(\mathbf{X}_1) = 2; \quad n_2 > 2, \quad r(\mathbf{X}_2) = 2$$

Se invece il cambiamento strutturale non vale, si verifica l'ipotesi nulla

$$H_0: \beta_{11} = \beta_{21} = \beta_{01}, \quad \beta_{12} = \beta_{22} = \beta_{02} \quad (3.4.5)$$

composta da due relazioni lineari che possono essere scritte nella forma (2.4.1)

$$H_0: \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

essendo in questo caso  $\mathbf{b} = [\mathbf{b}_1' \mathbf{b}_2']'$ . Supponendo valide le ipotesi forti sui residui in ambedue i sottoperiodi si può utilizzare il test della  $F$  di Fisher con 2 e  $n-4$  gradi di libertà dato dalla (2.4.12), dove  $\hat{\mathbf{b}}$  ed  $\hat{\mathbf{u}}$  vengono calcolate ipotizzando valido il cambiamento di struttura, per cui

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\mathbf{b}}_1' \hat{\mathbf{b}}_2' \end{bmatrix}', \quad \hat{\mathbf{u}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \hat{\mathbf{b}} \quad (3.4.6)$$

Se, invece, si preferisce utilizzare il test della  $F$  di Fisher dato in una delle due forme (2.5.4), occorre considerare che se è valida  $H_0$  si ha che  $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_0$ , la (3.4.3) diventa

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \mathbf{u}_0 \quad (3.4.7)$$

ed i vettori  $\mathbf{b}_0 = [\alpha_0 \beta_0]'$  ed  $\mathbf{u}_0$  vincolati dalle restrizioni lineari (3.4.5) sono

$$\mathbf{b}_0 = \left( \begin{bmatrix} \mathbf{X}_1' & \mathbf{X}_2' \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{X}_1' & \mathbf{X}_2' \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = [\mathbf{X}_1' \mathbf{X}_1 + \mathbf{X}_2' \mathbf{X}_2]^{-1} [\mathbf{X}_1' \mathbf{y}_1 + \mathbf{X}_2' \mathbf{y}_2] \quad (3.4.8)$$

$$\mathbf{u}_0 = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \mathbf{b}_0 \quad (3.4.9)$$

per cui l'applicazione di ciascuna delle (2.5.4) è immediata. In questo caso il numero dei vincoli è ancora  $q=k=2$  e il numero dei gradi di libertà della devianza  $\hat{\mathbf{u}}'\hat{\mathbf{u}}$  è ancora  $n-2k = n-4$  poiché questa è calcolata ipotizzando valido il cambiamento di struttura.

*Osservazione 3.5* - È conveniente, dal punto di vista didattico, rimarcare che l'ipotesi nulla da verificare è relativa all'omogeneità del campione e non al cambiamento strutturale, nel quale consiste, invece, l'ipotesi alternativa. Inoltre, il vettore  $\mathbf{b}_0$  delle stime dei minimi quadrati vincolati, che è necessario per determinare  $\mathbf{u}_0$  ed utilizzare il test della  $F$  di Fisher in una delle due forme (2.5.4), può naturalmente essere calcolato con la formula (1.11.18) che sfrutta la matrice  $\mathbf{R}$  ed il vettore  $\mathbf{r}$ , ma più semplicemente è determinabile tramite la (3.4.8) dato che nel caso di validità del vincolo (ipotesi nulla) il campione  $(\mathbf{y}, \mathbf{X})$  è omogeneo e l'equazione (3.4.3) assume la forma semplice (3.4.7).

#### *Test di cambiamento di struttura con le variabili di comodo*

Nel paragrafo 3.1 abbiamo visto che le variabili di comodo possono essere utilizzate per rappresentare spostamenti (*shift*) nei parametri. Non sorprende quindi che esse giochino un ruolo essenziale anche nella verifica dell'ipotesi di cambiamento di struttura. In sintesi, il test per la (3.4.5) può essere sottoposto a verifica accertando che siano nulle le variabili di comodo che misurano lo spostamento dei parametri fra il primo e il secondo sottoperiodo.

Formalmente, i due modelli (3.4.1) e (3.4.2) vengono fusi nell'unico modello

$$y_t = \beta_{11} + \beta_{12}x_t + \gamma_1 d_t + \gamma_2 (dx_t) + u_t \quad (3.4.10)$$

dove la  $d_t$  è una *shift dummy* definita come  $d_t = I(t > n_1)$  e i parametri  $\gamma_j$  misurano lo spostamento dell'intercetta (per  $j = 1$ ) e della pendenza (per  $j = 2$ ). La (3.4.10) in forma matriciale è

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{n_1} \\ y_{n_1+1} \\ y_{n_1+2} \\ \dots \\ y_{n_1+n_2} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & x_{n_1} & 0 & 0 \\ 1 & x_{n_1+1} & 1 & x_{n_1+1} \\ 1 & x_{n_1+2} & 1 & x_{n_1+2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n_1+n_2} & 1 & x_{n_1+n_2} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_{n_1} \\ u_{n_1+1} \\ u_{n_1+2} \\ \dots \\ u_{n_1+n_2} \end{bmatrix}$$

Il test di cambiamento strutturale viene quindi condotto verificando l'ipotesi nulla che i due parametri  $\gamma_j$  siano contemporaneamente nulli nella (3.4.10). La statistica  $F$  con 2 e  $n-4$  gradi di libertà che si ottiene è identica a quella ottenuta applicando le (2.5.4) alle grandezze (3.4.8)-(3.4.9).<sup>12</sup>

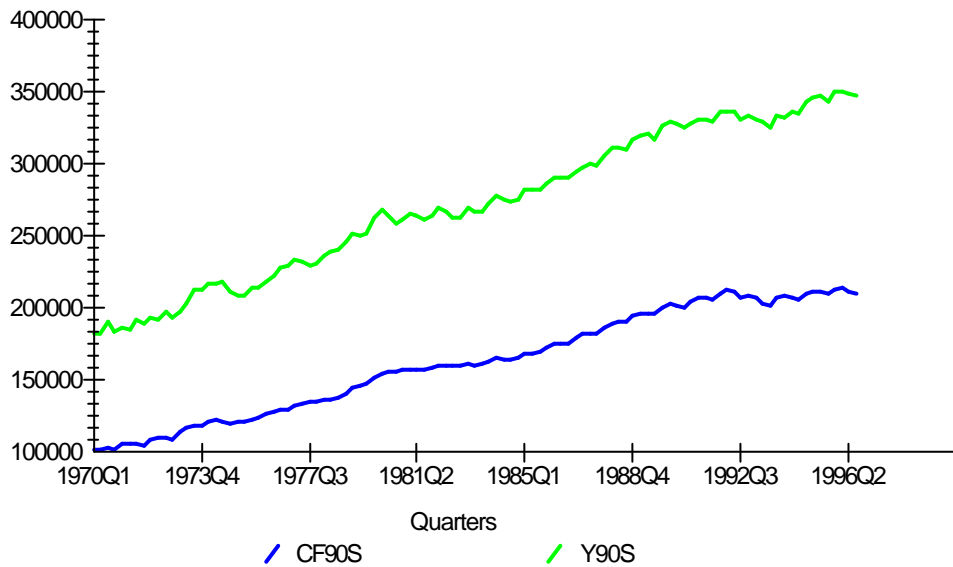


Figura 3.18 – Le serie storiche dei consumi e del PIL destagionalizzate con le variabili di comodo. Il grafico delle serie grezze è riportato nella I-2.2.

<sup>12</sup> L'equivalenza è dimostrata da Gujarati [1970] al quale rinviamo il lettore desideroso di approfondimenti. Questa procedura è utile nel caso in cui non si disponga di un programma di calcolo che effettui automaticamente il test di cambiamento di struttura. La procedura standard infatti richiederebbe la stima di tre regressioni (quella su tutto il campione – la (1.3.1) e le due sui sottocampioni – le (3.4.1)-(3.4.2)), mentre la procedura di Gujarati richiede che se ne stimino due, cioè la (1.3.1) e la (3.4.10).

Per esemplificare questi risultati riprendiamo le serie dei consumi e del reddito rappresentate nella figura I-2.2 e utilizziamole per verificare l'ipotesi di cambiamento di struttura fra la (3.4.1) e la (3.4.2). Dato che queste serie presentano stagionalità, la cui presenza, come sappiamo, può comportare una violazione delle ipotesi stocastiche deboli, prima di stimare il modello (1.2.1) rimuoviamo queste stagionalità con la regressione (3.3.2) reintegrando le rispettive medie campionarie con la (3.3.15) per conservare i volumi.<sup>13</sup> Le serie destagionalizzate sono rappresentate nella figura 3.18, che può essere confrontata con la I-2.2 per apprezzare l'efficacia della procedura.

L'esame della figura 3.18 mostra che soprattutto nel caso del PIL la procedura lascia qualche residuo di ciclicità stagionale. Questo risultato potrebbe dipendere da cambiamenti di struttura nel ciclo stagionale. In effetti la (3.3.2) presuppone che questo ciclo abbia struttura costante su tutto il campione, ma se così non è ed esistono, poniamo, due distinti regimi di stagionalità nel campione (il primo valido da 1 a  $n_1$ , e il secondo da  $n_1+1$  a  $n$ ), i fattori di destagionalizzazione stimati mediante la (3.3.2) vengono ad essere una mistura di quelli relativi ai due regimi e quindi sottraendoli alla serie non si ottiene mai una serie perfettamente destagionalizzata. Ricordiamo dal capitolo I-2 che la teoria economica prevede che il consumo assorba in qualche misura gli shock reali e abbia quindi un profilo più livellato nel tempo. Ne consegue che anche le procedure di destagionalizzazione applicate ad esso avranno maggiore efficacia, come sembra indicare la figura 3.18. Fatte queste riserve, prendiamo comunque per buone le serie destagionalizzate della figura 3.18 e le impieghiamo per applicare il test  $F$  di cambiamento di struttura.

Per verificare l'ipotesi di cambiamento di struttura occorre in primo luogo determinare la data  $n_1$  nella quale ha termine il primo sottoperiodo. In mancanza di informazioni *a priori*, indicazioni in tal senso possono essere ricavate dal grafico dei residui sotto la nulla (cioè in assenza di cambiamento di struttura), cioè dai residui dell'equazione stimata sull'intero campione. Nel nostro caso questa stima produce i seguenti risultati:

---

<sup>13</sup> Come sappiamo, in generale è indifferente se la stagionalità viene rimossa prima della stima del modello (come nella (3.3.24)) o contestualmente ad essa inserendo le *dummy* nella regressione (come nella (3.3.26)). In questo caso però rimuovendo la stagionalità prima di effettuare il test di cambiamento di struttura stiamo implicitamente ipotizzando che i fattori di destagionalizzazione non cambino da un sottoperiodo all'altro. Questa ipotesi viene fatta solo per comodità didattica (cioè per concentrare l'attenzione sui soli parametri  $\alpha$  e  $\beta$ ), ma nei casi concreti potrebbe rivelarsi inappropriata, per cui una corretta prassi prevede che il test di cambiamento di struttura coinvolga anche i fattori di destagionalizzazione e detrendizzazione e venga quindi eseguito sulla regressione completa (3.3.26).

$$\hat{c}_t = -27026.0 + 0.697 y_t$$

(-21.0)      (150.2)

(3.4.11)

$$n = 107, R^2 = 0.995, R^2_c = 0.995, RSS = 640,024,057.05, SEE = 2468.9, JB = 3.18$$

e il grafico dei suoi residui è rappresentato nella figura 3.19, dalla quale risultano distintamente almeno due episodi di cambiamento di struttura, situati approssimativamente all'inizio degli anni '80 (dove i residui, dopo aver toccato un minimo con un'osservazione anomala negativa, cominciano a manifestare oscillazione sensibilmente meno pronunciate di quelle verificatesi negli anni '70), e all'inizio degli anni '90 (dove i residui, dopo aver toccato un massimo, si pongono su un evidente trend decrescente).

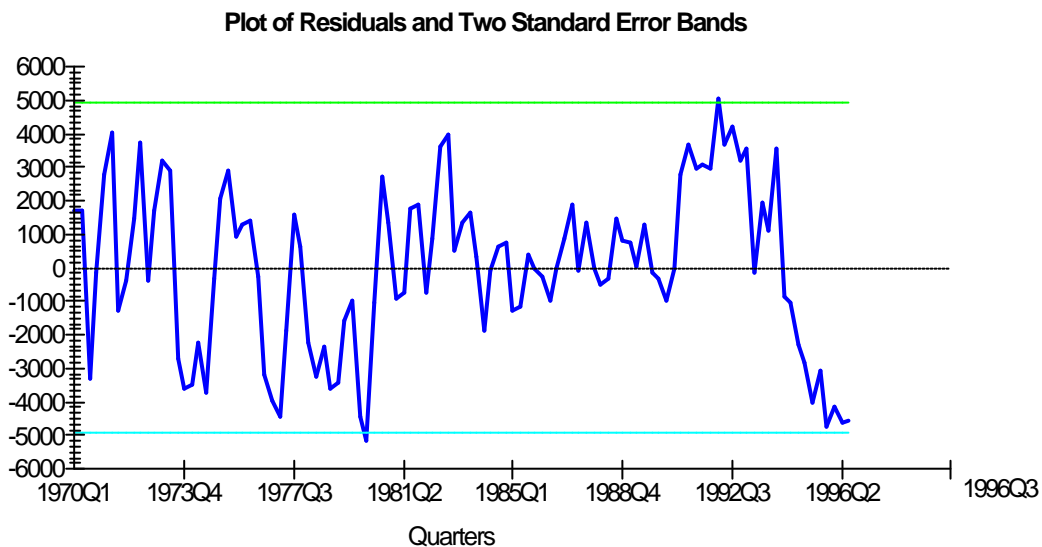


Figura 3.19 - Il grafico dei residui dell'equazione del consumo (3.4.11) stimata sull'intero campione 1970:1-1996:3.

Soffermandoci su questo secondo episodio, nel quale il modello tende vistosamente a sovrastimare la variabile dipendente (generando così residui negativi sempre più grandi), prendiamo come  $n_1$  la data immediatamente precedente al punto di massimo dei residui, ovvero 1991:4 (supponiamo cioè che il secondo regime abbia inizio nel 1992). Riportiamo di seguito la stima delle (3.4.1)-(3.4.2), cioè delle equazioni relative ai due sottocampioni

$$\hat{c}_t = -28274.6 + 0.703 y_t$$

(-20.8)      (134.9)

(3.4.12)

$$n = 88, R^2 = 0.995, R^2_c = 0.995, RSS = 415,181,154.24, SEE = 2197.2, JB = 1.65$$

$$\hat{c}_t = 98259.7 + 0.326 y_t \quad (3.4.13)$$

(5.3)      (5.9)

$$n = 19, R^2 = 0.678, R_c^2 = 0.659, RSS = 58,383,954.08, SEE = 1853.2, JB = 0.90$$

Le stime nei due sottocampioni differiscono in modo così marcato da far pensare che queste differenze non siano effetto della normale variabilità campionaria. Nel secondo sottoperiodo, di 19 osservazioni, la propensione marginale al consumo si dimezza, passando da 0.70 a 0.32, e la significatività complessiva del modello diminuisce sensibilmente, con un  $R_c^2$  che scende da 0.99 a 0.65.

Per effettuare il test  $F$  di cambiamento di struttura la forma più pratica è in questo caso senz'altro quella basata sulle devianze, cioè la prima delle (2.5.4). Nel nostro caso la devianza vincolata (sotto  $H_0$ ) è quella della (3.4.11), pari a 640,024,057.05, mentre la devianza "libera" è data dalla somma delle devianze delle (3.4.12) e (3.4.13), pari a 473,565,108.32. La statistica del test, con 2 e  $n-4 = 103$  gradi di libertà, è quindi:

$$\frac{640,024,057.05 - 473,565,108.32}{473,565,108.32} \frac{103}{2} = 18.10 \quad (3.4.14)$$

e va confrontata con il valore soglia della  $F_{2,103}$  che al livello  $\alpha = 0.05$  è pari a circa 3.09. L'ipotesi di assenza di cambiamento di struttura è quindi rifiutata.

Per concludere, mostriamo che con l'approccio delle variabili di comodo espresso dalla (3.4.10) è possibile effettuare il medesimo test stimando due sole regressioni. Infatti, oltre alla regressione (3.4.11) sotto la nulla, che va comunque stimata, per ottenere la devianza "libera" basterà stimare la (3.4.10), che nel nostro caso fornisce i seguenti risultati

$$\hat{c}_t = -28274.6 + 0.703 y_t + 126534.3 d_t - 0.376 d_t y_t \quad (3.4.15)$$

(-21.3)      (138.2)      (5.9)      (-5.9)

$$n = 107, R^2 = 0.996, R_c^2 = 0.996, RSS = 473,565,108.32, SEE = 2144.2, JB = 1.07$$

dove  $d_t$  è la *shift dummy* costruita come  $d_t = I(t > 1991:4)$  (cioè la variabile che vale zero nel primo sottoperiodo e uno nel secondo).

Si noti come nella (3.4.15) l'intercetta e la propensione marginale al consumo siano quelle della (3.4.12) mentre i parametri delle variabili  $d_t$  e  $d_t y_t$  misurano lo spostamento rispettivamente dell'intercetta e della propensione marginale fra il primo e il secondo sottoperiodo (cioè fra la (3.4.12) e la (3.4.13)). Il test  $F$  sulla significatività congiunta di questi due coefficienti coincide per costruzione con la (3.4.14) e ha il medesimo significato. Si noti anche che questo approccio fornisce più informazioni sulla natura del fenomeno. Ad esempio, se nella (3.4.15) fosse significativa la  $t$  di Student di uno solo dei due spostamenti, ne concluderemmo che

solo la relativa variabile è responsabile del cambiamento di struttura. Queste indicazioni diventano molto utili quando si passa al modello lineare multiplo, nel quale, in presenza di più esplicative, può essere interessante individuare quella che determina i problemi di cambiamento strutturale.

### 3.5 Un test di cambiamento strutturale per il modello lineare multiplo

Il test del paragrafo precedente può essere facilmente esteso al caso di  $k$  variabili esplicative. Se è valido il cambiamento strutturale vale ancora il modello (3.4.3), questa volta con  $\mathbf{b}_1 = [\beta_{11} \ \beta_{12} \dots \beta_{1k}]'$  e  $\mathbf{b}_2 = [\beta_{21} \ \beta_{22} \dots \beta_{2k}]'$  vettori di  $k$  elementi ciascuno e con  $\mathbf{X}_1$  ed  $\mathbf{X}_2$  matrici di ordine appropriato, mentre se il cambiamento non sussiste si verifica l'ipotesi nulla

$$H_0: \beta_{11} = \beta_{21} = \beta_{01}, \beta_{12} = \beta_{22} = \beta_{02}, \dots, \beta_{1k} = \beta_{2k} = \beta_{0k} \quad (3.5.1)$$

composta da  $k$  relazioni lineari descritte nella forma (2.4.1) diventano

$$H_0: \begin{bmatrix} 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (3.5.2)$$

e si può utilizzare il test della  $F$  di Fisher con  $k$  ed  $n-2k$  gradi di libertà dato dalla (2.4.12), dove  $\hat{\mathbf{b}} = [\hat{\mathbf{b}}_1' \hat{\mathbf{b}}_2']$  è la stima dei minimi quadrati data dalla (3.4.4) e  $\hat{\mathbf{u}}$  è la stima dei residui definita dalla seconda della (3.4.6). Ancora, questo test può essere effettuato se valgono le ipotesi (1.4.8), che nei due sottocampioni diventano  $r(\mathbf{X}_1)=k$  e  $r(\mathbf{X}_2)=k$ , e

$$n_1 > k, \quad n_2 > k \quad (3.5.3)$$

Se si preferisce utilizzare il test della  $F$  di Fisher dato in una delle due forme (2.5.4), occorre considerare che se è valida  $H_0$  la (3.4.3) diventa

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \mathbf{b} + \mathbf{u} \quad (3.5.4)$$

ed i vettori  $\mathbf{b}_0 = [\beta_{01}, \beta_{02}, \dots, \beta_{0k}]'$  ed  $\mathbf{u}_0$  vincolati dalle restrizioni lineari (3.5.1) sono dati ancora dalle (3.4.8) e (3.4.9) per cui l'uso di ciascuna delle (2.5.4) è immediato. In particolare, la prima della (2.5.4) diventa

$$F = \frac{\mathbf{u}'_0 \mathbf{u}_0 - (\hat{\mathbf{u}}'_1 \hat{\mathbf{u}}_1 + \hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2)}{\hat{\mathbf{u}}'_1 \hat{\mathbf{u}}_1 + \hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2} \frac{n-2k}{k} \quad (3.5.5)$$

dove  $\mathbf{u}'_0 \mathbf{u}_0$  è la devianza dei residui del modello vincolato ad avere parametri uguali (3.5.1), mentre le altre due devianze  $\hat{\mathbf{u}}'_1 \hat{\mathbf{u}}_1$  e  $\hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2$  sono relative alle equazioni stimate separatamente nel primo e nel secondo sottoperiodo.



*Il test preliminare di uguaglianza delle varianze*

L'ipotesi (3.5.1) che si sottopone verifica riguarda i parametri del modello di regressione e presuppone che nei due sottoperiodi le varianze  $\sigma_1^2$  e  $\sigma_2^2$  dei residui siano uguali. Se, tuttavia, non si è certi di questa uguaglianza è necessario verificare anche l'ipotesi ulteriore

$$H_0^{(1)} : \sigma_1^2 = \sigma_2^2 = \sigma^2 \quad (3.5.6)$$

tramite la statistica

$$F^{(1)} = \frac{\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 / (n_2 - k)}{\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 / (n_1 - k)} \quad (3.5.7)$$

che si distribuisce come una  $F$  di Fisher con  $n_2 - k$  ed  $n_1 - k$  gradi di libertà, essendo pari al rapporto di due variabili aleatorie indipendenti del tipo  $\chi^2$  divise per i rispettivi gradi di libertà. Tale test, classico nell'inferenza statistica, può essere effettuato preliminarmente alla verifica delle ipotesi (3.5.1) in quanto Phillips e McCabe (1983) ne hanno mostrato l'indipendenza.

L'ipotesi alternativa alla  $H_0^{(1)}$  è

$$H_0^{(1)} : \sigma_1^2 < \sigma_2^2 \quad \text{oppure} \quad \sigma_1^2 > \sigma_2^2 \quad (3.5.8)$$

per cui il test di ipotesi viene ad essere bilaterale, con i due estremi della regione di accettazione di  $H_0^{(1)}$  dati dai quantili  $F'$  ed  $F''$ , di probabilità  $\alpha/2$  ed  $1-\alpha/2$  rispettivamente, con  $\alpha$  livello di significatività del test.

Ovviamente la verifica dell'ipotesi (3.5.6) è da svolgersi preliminarmente anche al test del paragrafo precedente, qualora si sia incerti sull'uguaglianza delle varianze nei due sottoperiodi.

*Test di invarianza strutturale per un sottoinsieme di parametri*

Una variante del test precedente si ha quando l'ipotesi nulla di uguaglianza dei coefficienti non riguarda tutti questi ma soltanto un loro sottoinsieme. In altre parole, si vuole verificare che il cambiamento strutturale sia avvenuto interessando soltanto  $q < k$  parametri che, senza perdere in generalità, possiamo supporre essere i primi. Dunque, i primi  $q$  parametri sono diversi nella prima e nella seconda parte del campione: siano  $\mathbf{b}_1$  e  $\mathbf{b}_2$ ; gli altri  $k - q$  parametri siano uguali e formino il vettore  $\mathbf{b}_3$ . Allora il modello (3.4.3), nel quale sussiste il cambiamento strutturale, diventa

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{X}_{13} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{X}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \quad (3.5.9)$$

dove  $\mathbf{X}_1$  è la matrice delle variabili esplicative relative ai parametri variati nella prima parte del campione,  $\mathbf{X}_2$  è la matrice delle esplicative relative parametri variati nella seconda parte, e  $\mathbf{X}_{13}$  e  $\mathbf{X}_{23}$  sono le matrici delle esplicative relative ai parametri invariati nelle due parti del campione, rispettivamente.

La stima dei minimi quadrati del modello (3.5.9) è facilmente ottenuta

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{0} & \mathbf{X}'_1\mathbf{X}_{13} \\ \mathbf{0} & \mathbf{X}'_2\mathbf{X}_2 & \mathbf{X}'_2\mathbf{X}_{23} \\ \mathbf{X}'_{13}\mathbf{X}_1 & \mathbf{X}'_{23}\mathbf{X}_2 & \mathbf{X}'_{13}\mathbf{X}_{13} + \mathbf{X}'_{23}\mathbf{X}_{23} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'_1\mathbf{y}_1 \\ \mathbf{X}'_2\mathbf{y}_2 \\ \mathbf{X}'_{13}\mathbf{y}_1 + \mathbf{X}'_{23}\mathbf{y}_2 \end{bmatrix}$$

e si può ancora utilizzare il test della  $F$  di Fisher con  $q$  ed  $n-(k+q)$  gradi di libertà dato dalla (2.4.12), considerando che l'ipotesi nulla è ancora la (3.5.2) formata tuttavia da  $q$  equazioni lineari, e che  $\hat{b}$  ed  $\hat{u}$  sono calcolate supponendo valido il cambiamento di struttura, per cui

$$\hat{b} = [\hat{b}'_1 \hat{b}'_2 \hat{b}'_3]' \quad , \quad \hat{u} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{X}_{13} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{X}_{23} \end{bmatrix} \hat{b}.$$

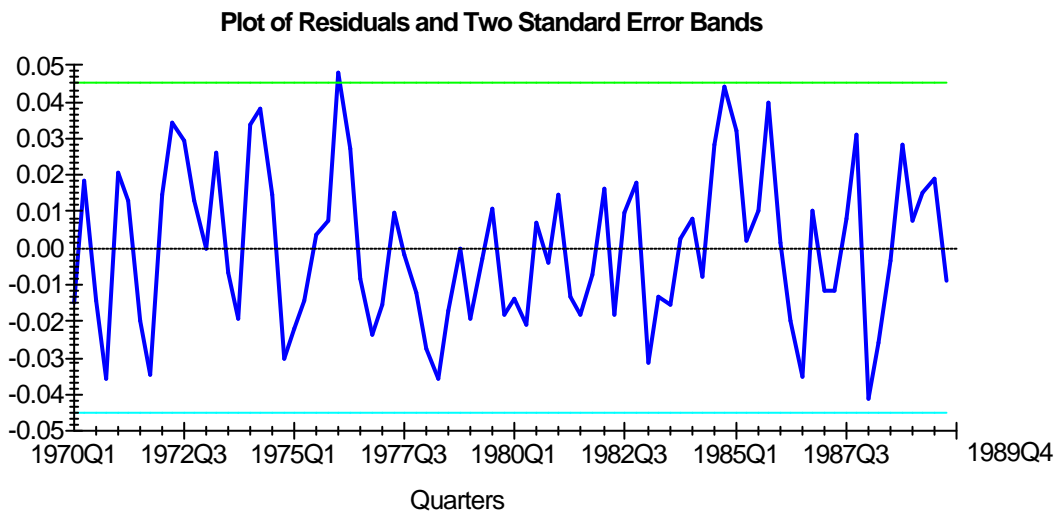


Figura 3.20 – Il grafico dei residui dell'equazione delle importazioni (3.3.29).

#### Test di cambiamento di struttura con le variabili di comodo

Come abbiamo accennato nel paragrafo precedente, l'approccio delle variabili è particolarmente utile nel contesto del modello lineare multiplo perché le  $t$  di Student delle singole *shift dummy* forniscono informazioni utili circa l'intensità con la quale il cambiamento di struttura coinvolge i diversi parametri.

Per esemplificare questo punto applichiamo il test alla (3.3.29), che rappresenta la nostra “migliore” specificazione della funzione delle importazioni. Come nel paragrafo precedente, partiamo da un’analisi dei residui dell’equazione per individuare la data di un possibile cambiamento di struttura. Il grafico è riportato nella figura 3.20 e mostra anche in questo caso tre episodi distinti, il primo e l’ultimo dei quali sono caratterizzati da residui molto grandi, mentre quello centrale (approssimativamente dal 1977 al 1984) mostra residui relativamente contenuti. In particolare, i residui toccano un massimo, corrispondente a un’osservazione anomala, nel primo trimestre del 1976, dopo di che scendono rapidamente. Prendiamo quindi 1976:1 come data terminale del primo sottoperiodo e definiamo la variabile di comodo  $d_t = I(t > 1976:1)$  per mezzo della quale impostiamo la regressione

$$\begin{aligned} \ln y_t = & \beta_1 + \beta_2 \ln x_{1t} + \beta_3 \ln x_{2t} + \beta_4 \ln x_{3t} + \beta_5 \ln x_{4t} + \delta d_{73,t} + b_1 d_{1t} + b_2 d_{2t} + b_3 d_{3t} + \\ & + \gamma_1 d_t + \gamma_2 (d_t \ln x_{1t}) + \gamma_3 (d_t \ln x_{2t}) + \gamma_4 (d_t \ln x_{3t}) + \gamma_5 (d_t \ln x_{4t}) + \\ & + \gamma_6 d_t d_{1t} + \gamma_7 d_t d_{2t} + \gamma_8 d_t d_{3t} + u_t \end{aligned} \quad (3.5.10)$$

corrispondente alla (3.3.27) senza la tendenza parabolica (risultata non significativa nelle analisi precedenti) ma aumentata con le variabili di spostamento (necessarie per effettuare il test). Si noti che non abbiamo spezzato il coefficiente della variabile di comodo puntuale  $d_{73,t}$ , visto che questa assume valori significativi solo nel primo sottoperiodo.<sup>14</sup> Il test di cambiamento di struttura può quindi essere condotto come test di significatività congiunto sugli otto coefficienti  $\gamma_j$  di spostamento dei parametri. La stima della (3.5.10) fornisce i seguenti risultati

$$\begin{aligned} \ln \hat{y}_t = & -10.94 & + 0.925 \ln x_{1t} & + 0.994 \ln x_{2t} & - 0.286 \ln x_{3t} & + 0.229 \ln x_{4t} + \\ & (-3.8) & (2.6) & (7.6) & (-3.7) & (1.2) \\ & + 1.477 d_{73,t} & + 0.104 d_{1t} & + 0.063 d_{2t} & + 0.039 d_{3t} & + 2.151 d_t + \\ & (1.4) & (7.0) & (4.3) & (1.2) & (0.6) \\ & - 0.161 d_t \ln x_{1t} & - 0.035 d_t \ln x_{2t} & + 0.18 d_t \ln x_{3t} & - 0.219 d_t \ln x_{4t} & - 0.082 d_t d_{1t} + \\ & (-0.4) & (-0.2) & (2.2) & (-1.1) & (-4.5) \\ & - 0.078 d_t d_{2t} & - 0.040 d_t d_{3t} & & & \\ & (-4.4) & (-1.2) & & & \end{aligned} \quad (3.5.11)$$

---

<sup>14</sup> Dato che nel secondo sottoperiodo la  $d_{73,t}$  vale costantemente zero, essa è ivi perfettamente collineare all’intercetta. Per questo motivo alcuni programmi di calcolo non riescono ad effettuare il test di cambiamento di struttura in presenza di variabili di comodo puntuali. La regressione (3.5.10) permette invece di superare questa difficoltà pratica.

$$n = 80, R^2 = 0.994, R_c^2 = 0.993, RSS = 0.022193, SEE = 0.018, JB = 0.70$$

Il test di cambiamento di struttura può effettuarsi utilizzando la prima delle (2.5.4), tenendo presente che in questo caso la devianza libera è quella della (3.5.11), pari a 0.022193 con  $n - 2k + 1 = 63$  gradi di libertà,<sup>15</sup> mentre quella vincolata (sotto  $H_0$ ) è quella della (3.3.29), pari a 0.036107, con  $k = 8$  gradi di libertà. Otteniamo così la statistica

$$\frac{0.036107 - 0.022193}{0.022193} \frac{63}{8} = 4.93 \quad (3.5.12)$$

che va confrontata con il quantile di una  $F_{8,63}$  che al livello  $\alpha = 0.05$  è pari a circa 2.09. L'ipotesi nulla di invarianza strutturale viene quindi respinta.

Un'analisi delle  $t$  di Student dei coefficienti di spostamento  $\gamma_j$  mostra che sono significativi solo quelli relativi alle prime due variabili di comodo stagionali e quello della variabile  $\ln x_{3t}$ , che, lo ricordiamo, è il logaritmo del deflatore delle importazioni. Fra il primo e il secondo sottoperiodo quindi abbiamo modifiche significative nel ciclo stagionale e nell'elasticità al prezzo dei beni importati, che diminuisce da circa il 28% a circa il 10%.

Prima di trarre conclusioni da questo test bisogna però notare che esso viene condotto subordinatamente all'ipotesi di costanza della varianza  $\sigma^2$ . Ora, da un'analisi del grafico 3.20 appare evidente che i residui passano da un primo periodo di varianza elevata a un secondo di varianza contenuta, per poi tornare verso la fine del campione a una situazione di elevata variabilità. È quindi opportuno effettuare il test (3.5.7), che in effetti, secondo una corretta prassi operativa, andrebbe eseguito prima del (3.5.5). A questo scopo è necessario effettuare la stima di due regressioni separate per il primo e per il secondo sottoperiodo e confrontare le rispettive devianze dei residui. Ristimando la (3.3.29) sul campione 1970:1-1976:1 otteniamo una devianza dei residui  $\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 = 0.0064562$ , mentre sul campione 1976:2-1989:4 la devianza è di  $\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 = 0.015737$ .<sup>16</sup>

La statistica del test di uguaglianza delle varianze è quindi

<sup>15</sup> Dobbiamo aggiungere una unità ai gradi di libertà del numeratore perché il modello nel primo sottocampione comprende un parametro in più (la dummy puntuale) che non figura nel secondo.

<sup>16</sup> Per la stima nel secondo sottoperiodo dobbiamo togliere la variabile di comodo puntuale  $d_{73,t}$ . Naturalmente la somma di queste due devianze coincide con la devianza dei residui della (3.5.11). Non riportiamo i risultati completi delle regressioni, che il lettore interessato potrà ricavare da sé utilizzando i dati disponibili su <http://econometria.net>.

$$F^{(1)} = \frac{\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 / (n_2 - k)}{\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 / (n_1 - k)} = \frac{0.015737 \cdot 16}{0.0064562 \cdot 47} = 0.82 \quad (3.5.13)$$

e va confrontata con il quantile di una  $F_{47,16}$  che per  $\alpha = 0.05$  è pari a 2.13. L'ipotesi nulla di uguaglianza delle varianze non viene quindi respinta.

Tuttavia occorre tener presente che il grafico 3.20 mette in luce l'esistenza di tre episodi distinti nella variabilità dei residui della (3.3.29). Dato che il test (3.5.13) confronta le varianze di due soli sottoperiodi, può darsi che esso non riesca a captare situazioni nelle quali i cambiamenti di regime della varianza sono più di due. Nel prossimo paragrafo vedremo dei test più generali per la verifica dell'ipotesi di costanza di  $\sigma^2$ .

### **3.6 Bibliografia**

Gujarati, D. [1970], "Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Two Linear Regressions: A Note", *The American Statistician*, **24**, pp. 50-52.

Jarque, C.M., Bera, A.K. [1980], "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals", *Economics Letters*, **6**, pp. 255-9.

Johnston, J. [1984], *Econometric Methods*, III edizione, Singapore: McGraw-Hill.