



Munich Personal RePEc Archive

## **2SLS with principal components: estimation of a nonlinear model of the Italian economy**

Carlo Bianchi and Giorgio Calzolari and Franco Sartori

Centro Scientifico IBM, Pisa, Italy, ISPE, Roma, Italy

1982

Online at <http://mpa.ub.uni-muenchen.de/22665/>

MPRA Paper No. 22665, posted 13. May 2010 20:41 UTC

CARLO BIANCHI - GIORGIO CALZOLARI - FRANCO SARTORI

**Stime 2SLS con componenti principali di un modello  
non lineare dell'economia italiana**

## Stime 2SLS con componenti principali di un modello non lineare dell'economia italiana

### 1. Introduzione

I problemi relativi alla stima di sistemi di equazioni simultanee, in particolare quelli connessi alle stime ottenute col metodo dei minimi quadrati ordinari (Ordinary Least Squares: OLS), sono stati per la prima volta affrontati da Haavelmo nel 1943 [20]. Da allora sono stati proposti altri metodi di stima che partono dalla considerazione [19] che, non essendo in un sistema di equazioni simultanee soddisfatte le condizioni del teorema di Gauss-Markov sui minimi quadrati ordinari, l'applicazione di tale metodo ad ogni singola equazione (stocastica) fornisce stime che possono essere distorte e inconsistenti. Questi metodi di stima alternativi possono essere divisi in due classi [18, pp. 346-347]:

- metodi ad informazione limitata, o stimatori di equazione singola;
- metodi ad informazione completa, o stimatori di sistema.

Le crescenti dimensioni dei modelli macroeconomici e la presenza sempre più frequente di strutture non lineari rendono spesso tecnicamente inutilizzabili i metodi di stima ad informazione completa (si vedano ad esempio [13] e [28, p. 140, nota 2]); d'altra parte la pur limitata esperienza mostra che, anche quando sono applicabili, i metodi ad informazione completa sono molto sensibili agli errori di specificazione, così da sconsigliarne l'uso già per i modelli di modeste dimensioni [15, p. 601]. In pratica, quindi, « si continuano a utilizzare i minimi quadrati ordinari salvo qualche eccezione in cui si usano i minimi quadrati a due stadi e ancor più raramente a tre stadi » [29, p. 83].

Anche l'uso di tali metodi di stima ad informazione limitata, come ad esempio il metodo dei minimi quadrati a due stadi (2 Stage Least Squares: 2SLS), è soggetto ad alcune limitazioni. Tali limitazioni sono legate soprattutto al fatto che questi metodi passano attraverso la stima delle equazioni di forma ridotta; già per modelli di modeste dimensioni, facilmente può accadere che il numero delle osservazioni disponibili sia troppo piccolo rispetto al numero delle variabili predeterminate. O si usano, allora, nuovi metodi che non richiedono la stima della forma ridotta come quelli proposti da Brundy e Jorgenson [7] e da Dhrymes [9] (per una esposizione teorica con applicazioni a un modello lineare dell'economia italiana si rimanda a Visco [33]), oppure si ricorre a stimatori « promiscui » [2] che, pur con formalizzazioni eleganti e persuasive, hanno

destato e destano perplessità anche tra chi scettico non è circa la costruzione e l'uso dei modelli econometrici. Al proposito, infatti, Ball [2, p. 8] dice: « ... I have a suspicion that the application of one technique rather than another (such as the promiscuous use of two stage least squares) has been a genuflection toward the altar of statistical respectability rather than a belief in the increased power of one method over another ... Unless one can clearly demonstrate the immediate practical differences in simulation and forecasting performance of alternative estimation methods applied to models built on comparatively short data samples, one is inclined to be unimpressed by theoretical properties that can be shown to be superior for infinite size samples ... ».

Seguendo questo ultimo suggerimento, scopo del presente lavoro è appunto quello di confrontare, principalmente in simulazione, stime alternative di un modello non lineare dell'economia italiana: il modello ISPE [30]; in particolare saranno messe a confronto stime OLS e stime dove il metodo dei minimi quadrati a due stadi con componenti principali (2 Stage Least Squares with Principal Components: 2SPC) è usato con diverse varianti relative ai criteri di selezione e al numero delle componenti principali stesse.

Va inoltre osservato che la capacità interpolante non è l'unica caratteristica interessante per l'utilizzazione di un modello. Come il costruttore non accetta i risultati della stima di una equazione strutturale se i coefficienti non sono significativamente diversi da zero, così il politico economico non dovrebbe aver fiducia in un determinato strumento se il relativo moltiplicatore non è significativo; o ancora non dovrebbe confidare nel conseguimento di un obiettivo se il corrispondente errore di previsione non ha una distribuzione sufficientemente concentrata. Per procedere alla derivazione di queste informazioni, è necessario disporre di informazioni complete relative alla distribuzione di tutti i coefficienti strutturali del sistema, in particolare è necessario disporre di una stima dell'intera matrice di covarianza. Tale informazione può essere ricavata solo da metodi di stima che tengano conto della simultaneità del modello come, appunto, quello dei minimi quadrati a due stadi. Da questo punto di vista, il presente lavoro può essere riguardato come preliminare per ulteriori studi di convalida (si vedano, al riguardo, [3], [4] e [6]).

### 2. Il metodo dei minimi quadrati a due stadi (2SLS)

Nella stima di un sistema di equazioni simultanee, il metodo dei minimi quadrati ordinari (OLS) fornisce generalmente stime distorte e inconsistenti in quanto viene violata una delle ipotesi fondamentali del modello di regressione lineare, quella, cioè, per cui i regressori sono variabili non stocastiche o comunque indipendenti dal termine di errore [8, p. 146].

La metodologia econometrica fornisce, però, stimatori che, sotto le stesse ipotesi, pur essendo ancora distorti, sono tuttavia almeno consistenti. Dato che la causa dell'inconsistenza delle stime OLS risiede nella correlazione tra le variabili endogene esplicative e i disturbi strutturali, si può pensare di eliminarla depurando preventivamente le endogene esplicative della loro componente casuale.

Il metodo dei minimi quadrati a due stadi (2SLS) [8, pp. 177-180] può essere usato per questo scopo. In un primo stadio si regrediscono le variabili endogene esplicative, correlate con i termini di disturbo, sulle variabili predeterminate, nel secondo stadio si applicano i minimi quadrati ordinari alle equazioni strutturali dove, però, al posto del valore osservato di ogni endogena, come regressore, viene sostituito il valore calcolato nel primo stadio. Questo metodo di stima è ampiamente trattato nella letteratura econometrica, relativamente ai modelli lineari. Si tralasciano, pertanto, considerazioni sulle ipotesi e sul problema della identificazione delle equazioni stocastiche e ci si limita ad introdurre brevemente quei simboli che ricorreranno nella parte di questo lavoro dedicata ai modelli non lineari.

Dato un modello lineare di  $q$  equazioni simultanee con  $q$  variabili endogene e  $p$  variabili predeterminate, una tipica equazione strutturale può essere scritta come

$$(1) \quad y = Y_1 a + X_1 b + u$$

dove:  $y$  è il vettore  $T \times 1$  di osservazioni sulla variabile endogena spiegata dall'equazione;

$Y_1$  è la matrice  $T \times m$  di osservazioni sulle  $m$  variabili endogene incluse nella equazione;

$X_1$  è la matrice  $T \times n$  di osservazioni sulle  $n$  variabili predeterminate (esogene ritardate, endogene ritardate) incluse nella equazione;  $a$  e  $b$  sono i vettori  $m \times 1$  e  $n \times 1$  dei parametri da stimare;

$u$  è il vettore  $T \times 1$  dei disturbi casuali.

Si assume che  $p > n$  e si rappresenta la matrice  $X$ , di dimensioni  $T \times p$ , contenente tutte le variabili predeterminate del modello, come  $X = [X_1 \ X_2]$ , dove  $X_2$  è la matrice  $T \times (p-n)$  delle variabili predeterminate non incluse nella equazione da stimare. Supponendo che  $(X'X)$  sia non singolare, nella stima 2SLS della (1) si procederà come segue.

Primo stadio — Ogni variabile endogena della matrice  $Y_1$  è regredita su tutte le variabili predeterminate  $X$  del sistema, si stimano cioè col metodo OLS le equazioni di forma ridotta; in

notazione matriciale il valore calcolato ( $\hat{Y}_1$ ) sarà:

$$(2) \quad \hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

Secondo stadio — I valori  $\hat{Y}_1$ , sono usati come regressori in (1) al posto di  $Y_1$  e, applicando nuovamente il metodo OLS, si ottengono le stime 2SLS dei vettori dei parametri  $a$  e  $b$ ; sarà quindi:

$$(3) \quad \begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} \hat{Y}_1' \hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1' \hat{Y}_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1' \\ X_1' \end{bmatrix} y$$

Allo stesso modo si procederà per le rimanenti equazioni stocastiche del modello. Il metodo, così come è stato esposto, risulta applicabile solo

quando la matrice  $X$  ha rango  $p$ , il che implica necessariamente che il numero di osservazioni disponibili ( $T$ ) non può essere inferiore al numero di variabili predeterminate ( $p$ ). Sfortunatamente, già per modelli di modeste dimensioni, questa condizione può non essere rispettata, nel qual caso  $X'X$  è una matrice singolare e l'equazione (3) non può essere risolta.

## 2.1 Il metodo dei minimi quadrati a due stadi nei modelli non lineari

Nel caso di modelli non lineari nelle variabili endogene, una tipica equazione strutturale può essere in generale scritta come segue [22, p. 373]:

$$(4) \quad y = Fa + X_1 b + u$$

dove la  $F$  è una matrice  $T \times m$  di osservazioni su  $m$  funzioni incluse nella equazione considerata. La matrice  $F = (f_1, f_2, \dots, f_m)$  potrebbe anche essere espressa come  $F = [Y_{m-k} \ F_k]$ , dove  $Y_{m-k}$  sono le  $m-k$  variabili endogene  $y_j$  ( $j=1, \dots, m-k$ ) presenti senza trasformazioni, mentre le  $F_k$  sono  $k$  funzioni  $f_h = f_h(Y, X)$  ( $h=1, \dots, k$ ) di almeno una variabile endogena. Dal punto di vista dell'intero modello, ogni trasformazione (anche lineare) di variabili endogene definisce una nuova variabile endogena e quindi comporta l'aggiunta [12, p. 28] di una nuova equazione definizionale (« latente » [19]) all'intero modello; funzioni di sole variabili predeterminate debbono, invece, essere considerate come nuove variabili predeterminate [12, p. 27].

Le difficoltà nell'applicazione del metodo 2SLS a modelli non lineari sono essenzialmente collegate alla stima della forma ridotta (primo stadio). Infatti la forma ridotta di modelli non lineari è di difficile derivazione analitica ed anche quando questa derivazione è possibile essa è, in generale non lineare nei parametri [19]. Per evitare questa difficoltà [19, p. 116], una approssimazione alla parte non stocastica della forma ridotta può essere ottenuta con un polinomio ( $P$ ) di grado  $r$  nelle variabili predeterminate, rappresentante i primi termini di una espansione in serie di Taylor della forma ridotta stessa; tale approssimazione, in generale, migliorerà al crescere di  $r$ . Il modello di regressione sottostante a questa ipotesi sarà, allora:

$$(5) \quad f_j = P_j^r(X) + e_j$$

dove:  $f_j$  è la  $j$ -esima variabile endogena esplicativa della equazione da stimare;

$P_j$  è un polinomio di grado  $r$  nelle variabili predeterminate ( $X$ ) del modello;

$e_j$  è il disturbo casuale della (5), che in questo caso è chiamata « pseudo-forma ridotta » [19, p. 115].

Come è stato notato [21, pp. 14-15], [22, p. 374], essendo ancora verificate le ipotesi circa la identificazione, le stime OLS

$$(6) \quad \hat{f}_j = \hat{P}_j^r(X)$$

saranno asintoticamente incorrelate col termine di disturbo  $u_j$ ; le  $\hat{f}_j$  saranno, quindi, utilizzabili nel secondo stadio della procedura 2SLS.

## 2.2 Il metodo dei minimi quadrati a due stadi con componenti principali (2SPC).

La procedura esposta alla fine del paragrafo 2.1, pur garantendo la consistenza del metodo 2SLS [22] (per i modelli dinamici, però, si vedano le considerazioni in Gallant [16, pp. 73-74]), raramente può essere utilizzata in pratica. Infatti, molto spesso, il numero delle variabili predeterminate ( $p$ ) è di gran lunga superiore alle dimensioni del campione ( $T$ ) e, quindi, già nel caso  $r=1$ , la matrice  $(X'X)$  sarà singolare; d'altra parte anche quando  $T > p$  (sempre nell'ipotesi  $r=1$ ), la multicollinearità, che è molto frequente quando si usano serie di dati economici, comporta che la matrice  $(X'X)$  sia prossima alla singolarità, così da ridurre l'affidabilità delle stime di forma ridotta del primo stadio.

Questa difficoltà, unita alla considerazione che la variabilità delle variabili predeterminate può essere approssimata da un ridotto insieme delle prime loro componenti principali, ha suggerito (si vedano [1] e [25]) l'introduzione di un nuovo metodo di stima a due stadi dove alcune componenti principali delle variabili predeterminate vengono impiegate nella stima della forma ridotta (primo stadio del metodo 2SLS). Questo procedimento permette di applicare il 2SLS anche quando il numero di osservazioni ( $T$ ) è piccolo rispetto al numero di variabili predeterminate ( $p$ ) (si può applicare anche con  $T < p$ ) ed evita il problema della multicollinearità nella stima di forma ridotta, essendo le componenti principali tra di loro ortogonali.

Le componenti principali sono tante quante sono le variabili linearmente indipendenti nell'insieme delle variabili originarie e sono generalmente ottenute [10] in un ordine tale che la prima componente spiega la maggior quota di variabilità (e di varianza generalizzata delle variabili originarie, la seconda componente (che è ortogonale alla prima) spiega la maggior quota di variabilità residua e così di seguito.

La matrice di tutte le componenti principali ( $Q$ ) è, una trasformazione ortogonale della matrice delle variabili ( $X$ ) e la varianza generalizzata di  $Q$  è uguale a quella di  $X$ .

Come si vedrà nel paragrafo 3, nel caso del modello ISPE le prime 21 componenti principali spiegano il 100% della varianza generalizzata delle 62 variabili predeterminate, ma già le prime 4 spiegano il 91% e la prima da sola il 71% della variabilità totale.

Dal punto di vista computazionale, la stima 2SPC di un sistema di equazioni simultanee procede nel modo seguente:

- i — Si definiscono le  $p$  variabili predeterminate del modello e, dopo averle standardizzate, si organizzano in una matrice  $X$  di dimensioni  $T \times p$ .
- ii — Si derivano le componenti principali delle variabili organizzate nella matrice  $X$ , ordinandole secondo la quota, decrescente, di varianza generalizzata spiegata (questo ordinamento corrisponde all'ordinamento secondo gli autovalori di  $X'X$ ).
- iii — Per ogni equazione si scelgono le componenti più « adatte » alla stima della forma ridotta (primo stadio) delle variabili endogene

esplicative (matrice  $F$  della equazione 4) presenti in quella equazione, ottenendosi per mezzo di stime OLS quella matrice  $\hat{F}$  che, sostituita nella (4), permette di ottenere (secondo stadio) le stime dei vettori  $a$  e  $b$ .

Il punto (i) non è di ovvia applicazione pratica nei modelli non lineari. Infatti, se la sopra accennata espansione in serie di Taylor risolve il problema della forma funzionale con cui le variabili predeterminate entrano nella forma ridotta, per essere sicuri di ottenere una buona approssimazione il polinomio approssimante  $P$  dovrebbe avere un grado abbastanza elevato; ciò comporta la definizione di nuove variabili con conseguenti problemi di precisione numerica nella derivazione delle componenti principali.

D'altra parte, anche se questa procedura permette talvolta di trascurare alcune delle specificazioni funzionali di variabili esogene, rimane il problema di come definire e contare tutte le variabili predeterminate del modello. In particolare, come considerare le variabili endogene ritardate? Devono essere certamente incluse con il ritardo con cui compaiono nella forma strutturale, ma non appena si consideri che il modello viene spesso (certamente in previsione) simulato dinamicamente, perchè non introdurre (come suggerito in [15]) le stesse variabili con ritardi più lunghi? A questo proposito va subito rilevato che ciò comporta una riduzione delle dimensioni del campione e quindi questo approccio è sconsigliabile, almeno per modelli a base annua.

Ancora, di alcune variabili che compaiono nel modello solo e sempre come tassi di variazione, in simulazione viene spesso richiesto il valore in livelli; deve, in questo caso, la variabile ritardata essere inclusa tra le predeterminate del modello?

La soluzione a questi problemi è stata data in modo non univoco in varie applicazioni empiriche ([11], [21] e [23]): generalmente, oltre ad assumere uguale a uno il grado del polinomio, si riduce il numero di variabili predeterminate per non compromettere l'efficienza computazionale delle procedure di calcolo usate nella derivazione delle componenti principali.

Onde ridurre la collinearità tra componenti principali e variabili predeterminate usate nel primo stadio di stima, è stato suggerito di derivare le componenti principali soltanto per le variabili predeterminate non incluse nella equazione (metodo I di Kloek e Mennes, [25, p. 50]). Questa proposta è teoricamente collegata al rispetto di condizioni sufficienti per la consistenza. Infatti (si vedano [1, pp. 286-287] e [7, pp. 216-217]), la consistenza di tutti i metodi 2SLS troncati (come il 2SPC) prevede come condizione sufficiente, anche se non necessaria [27], che la forma ridotta includa tra i regressori anche le variabili predeterminate della equazione; nel primo stadio del 2SPC vi debbono essere, quindi, come regressori alcune componenti principali e le variabili predeterminate della equazione.

L'approccio ora descritto potrebbe essere rilevante per modelli lineari di modeste dimensioni, ma non lo è certamente quando il numero delle predeterminate è elevato e, a maggior ragione, quando il modello è non lineare; in questo caso, infatti, il numero di variabili esogene che si introducono nella pseudo-forma ridotta è, per le considerazioni fatte sopra, abbastanza elevato. D'altra parte, le ridotte dimensioni campionarie

e il problema dei gradi di libertà comportano che la specificazione delle equazioni stocastiche coinvolga un numero ridotto di variabili e, quindi, già per modelli di modeste dimensioni, il numero di predeterminate di ogni equazione è molto basso rispetto al numero di tutte le predeterminate del sistema. Nel modello ISPE ad esempio, per ogni equazione stocastica vi sono generalmente 2 o 3 predeterminate, termine costante incluso, mentre le predeterminate conteggiate ed incluse nella derivazione delle componenti principali sono 62.

Queste ultime considerazioni e quella già sottolineata circa la naturale correlazione esistente tra serie di dati economici, comportano che le componenti principali delle variabili predeterminate non incluse nella equazione abbiano, da un certo punto di vista, (principalmente in termini di varianza), le stesse caratteristiche di quelle calcolate sulle predeterminate dell'intero modello (inclusendo cioè anche le predeterminate dell'equazione), così da non giustificare una derivazione separata.

Seguendo questa linea, nella scelta delle variabili predeterminate da cui derivare le componenti principali si potrebbe estendere quanto suggerito da Fisher [15, pp. 624-625], che propone di considerare le catene causali esistenti all'interno del modello. Tenendo conto della struttura e delle causalità del modello, ogni endogena porta alla identificazione di un insieme di predeterminate da cui dipende in modo prevalente; su queste si potrebbero calcolare le componenti principali, evitando così di dare eccessivo rilievo a variabili predeterminate poco importanti. Questa procedura, a parte le difficoltà [17] legate all'individuazione della catena causale per ogni endogena (basti pensare ad una endogena funzione non lineare di altre endogene), non garantisce la consistenza delle stime, in quanto variabili endogene appartenenti alla stessa equazione, avendo catene causali diverse, saranno regredite, nel primo stadio, su un diverso insieme di componenti principali [27].

Se nel primo stadio si introducono anche le predeterminate della equazione (cioè le variabili  $X_i$  della (1) o della (4)), un metodo interessante, utile per evitare la eventuale collinearità esistente tra componenti principali e dette variabili  $X_i$ , è stato suggerito da Kloek e Mennes [25, p. 51, metodo 2]. Seguendo questo metodo, per ogni equazione le componenti principali sono derivate dall'insieme dei residui ( $W$ ) delle regressioni delle variabili predeterminate escluse dalla equazione (cioè le variabili  $X_j$ ) sulle variabili predeterminate incluse nella equazione stessa (e cioè le  $X_i$ ). E' evidente che in questo caso, essendo i residui  $W$  ortogonali ai regressori  $X_i$ , altrettanto lo saranno le componenti principali di  $W$ , che altro non sono se non combinazioni lineari di detti residui.

Questa procedura, che non è di nessuna utilità quando non si introducono le variabili  $X_i$  nel primo stadio della stima, comporta, in generale, tante derivazioni di componenti principali quante sono le equazioni da stimare (per una applicazione con simulazione del modello si veda [21]).

Identificato l'insieme (o gli insiemi) di variabili su cui derivare le componenti principali (i) e calcolate le componenti principali stesse (ii: si veda, ad esempio, Dixon [10] per il metodo di calcolo), rimane il

problema della scelta di quante e quali (iii) componenti introdurre nella stima della forma ridotta. Qui di seguito saranno discussi alcuni criteri generali, mentre nel paragrafo 3 saranno presentati i criteri seguiti nei diversi esperimenti sul modello ISPE.

Va innanzitutto rilevato che definizione delle variabili predeterminate e scelta delle componenti principali sono problemi direttamente collegati nel senso che già dall'insieme delle variabili su cui le componenti sono derivate si hanno suggerimenti su come procedere alla scelta delle componenti stesse, e l'introduzione o meno delle predeterminate della equazione nella stima della forma ridotta influenza i criteri di scelta.

Come è stato notato in [21, p. 23] e [25, p. 50], non ci sono precise regole a priori o criteri univoci per decidere quante componenti ( $k$ ) debbano essere usate nelle stime 2SPC. Esiste tuttavia un limite inferiore richiesto per il rispetto delle condizioni per la identificabilità (si vedano [1] e [25]). In particolare se nella forma ridotta si introducono anche le predeterminate presenti nella equazione, dovrà essere  $k \geq m$ , dove  $m$  è il numero delle variabili endogene esplicative ( $F_i$ ) presenti nella equazione; non introducendo invece le predeterminate  $X_i$  nella stima della forma ridotta, dovrà essere  $k \geq m+n-1$ , dove  $n$  è il numero delle predeterminate (termine costante compreso) presenti nella equazione. Come è stato rilevato in [27], questo secondo metodo non è consistente se il numero di componenti principali usate come regressori nel primo stadio è fisso all'aumentare delle dimensioni campionarie.

Un limite superiore a  $k$  è dato dalla attendibilità della stima, per cui sono le dimensioni campionarie  $T$  che diventano vincolanti. Quindi  $(k+n)/T$  non dovrà essere troppo grande, ad esempio non dovrà superare  $1/3$ , come suggerito in [25].

Nel rispetto di questi vincoli, la scelta del numero  $k$  è un fatto prevalentemente sperimentale; in pratica, partendo dal numero minimo si potrebbero aggiungere componenti e fermare la sperimentazione quando non vi fossero modifiche significative negli errori standard dei coefficienti stimati, oppure, restando nell'ipotesi di introdurre le componenti secondo la quota crescente di varianza spiegata, si potrebbe fermare la sperimentazione quando ulteriori componenti non contribuissero significativamente alla spiegazione della variabilità totale delle variabili sulle quali sono calcolate.

Quando si introducono le variabili predeterminate  $X_i$  nella forma ridotta (primo stadio) delle stime 2SPC, a meno di non usare il metodo 2 proposto da Kloek e Mennes, appare di nuovo il problema della collinearità tra i regressori. Gli stessi Kloek e Mennes hanno proposto un altro metodo (metodo 4, [25, p. 53]), computazionalmente più semplice e che, in pratica, dovrebbe avere gli stessi vantaggi del metodo 2. In sintesi il metodo proposto consiste nel derivare le componenti principali su tutte le predeterminate  $X$  e nell'ordinare, per ogni equazione, le componenti non secondo gli autovalori, ma secondo la relazione

$$(7) \quad \theta_i = v_i (1 - R_i)$$

dove  $v_i$  è l'autovalore associato alla componente  $i$ -esima e  $R_i$  è il coefficiente di correlazione multipla della regressione della componente  $i$ -esima

sulle variabili  $X_i$  della equazione che si sta stimando. Saranno scelte quindi le  $k$  componenti con i valori di  $\theta_i$  più grandi.

### 5. Applicazione al modello ISPE

La versione del modello ISPE utilizzato per la stima simultanea è sostanzialmente quella presentata in [30]. E' stato, tuttavia, opportuno apportare alcune modifiche che, pur non alterando le proprietà e le caratteristiche teoriche del modello, hanno, da un lato, comportato un leggero miglioramento della capacità previsiva del modello stimato col metodo OLS e, dall'altro, reso più semplice la procedura di stima simultanea. Tali modifiche riguardano:

- l'estensione del periodo di stima dell'equazione delle esportazioni di prodotti manufatti, di quella dei prezzi impliciti nelle esportazioni di prodotti manufatti, e di quella degli investimenti in abitazioni;
- l'eliminazione di alcune funzioni dei prezzi dal lato dell'offerta;
- una diversa specificazione dei ritardi nella funzione degli investimenti.

L'allungamento del periodo di stima di alcune equazioni si è reso necessario per ragioni di omogeneità; per tutte le equazioni, le stime qui presentate si riferiscono al periodo 1955-1976. Le nuove stime OLS hanno conservato all'incirca lo stesso grado di affidabilità di quelle relative al periodo meno esteso presentate in [30]: ciò ha consentito di non apportare modifiche alla specificazione delle equazioni di cui al punto (a).

La determinazione dei prezzi dal lato dell'offerta era stata originariamente introdotta in attesa che altri settori del modello, in particolare quello della domanda e il settore pubblico, ricevessero una trattazione più dettagliata attraverso una analisi esplicita della distribuzione del reddito. L'eliminazione delle funzioni dei prezzi impliciti nel valore aggiunto del settore terziario privato ed una spiegazione dei prezzi dal lato della domanda direttamente con i costi del lavoro per unità di prodotto hanno permesso una semplificazione formale del modello senza modificare le ipotesi teoriche sottostanti.

L'ultima modifica è stata apportata alla funzione degli investimenti fissi nel settore industriale ed in quello terziario privato. La non-convergenza, in soluzione, del modello negli anni 1955-1957 è sembrata dipendere da una instabilità dovuta alla forte interrelazione tra investimenti fissi e capacità produttiva nel settore industriale: se, ad esempio, in una prima iterazione del processo di soluzione gli investimenti fossero risultati superiori al loro valore di soluzione, anche la capacità utilizzata sarebbe stata più alta del valore di soluzione; ciò era sufficiente per allontanare, nell'iterazione successiva, ancor più gli investimenti dal valore di soluzione e così via. Si è trattato, quindi, di modificare questa interrelazione, rendendola meno esplosiva nel periodo corrente. Ciò è stato fatto specificando diversamente la funzione degli investimenti fissi: in particolare, si è cambiata la struttura dei ritardi con cui le diverse variabili esplicative determinano gli investimenti.

Il modello, nel periodo campionario 1955-1976, è stato stimato con il metodo OLS e con diverse varianti del metodo 2SPC. Nelle stime le componenti principali sono state derivate da tutte le variabili predeterminate del modello, come suggerito da vari autori (si vedano, ad esempio, [11], [14], [17], [21], [23], [25], [28]); ciò ovviamente comporta una sola derivazione delle componenti principali stesse. Seguendo quanto esposto nel paragrafo 2.2, le variabili predeterminate del modello ISPE risultano essere 62, in queste includendo esogene correnti, esogene ritardate, endogene ritardate e loro trasformazioni non lineari.

La proporzione cumulativa della varianza totale delle variabili predeterminate spiegata dalle prime dieci componenti principali è la seguente:

.710 .845 .884 .912 .935 .957 .972 .980 .986 .990

Come è facile rilevare, della varianza totale le prime 4 componenti spiegano più del 90%, le prime 6 più del 95% e le prime 10 il 99%. Le prime 21, ovviamente, spiegano il 100% (delle 62 predeterminate, al massimo 21 possono considerarsi linearmente indipendenti). Le varianti del 2SPC sperimentate possono essere divise in due categorie:

Metodo 2SPC1: nel primo stadio della stima di ogni equazione sono state introdotte solo le componenti principali.

Metodo 2SPC2: nel primo stadio sono state incluse, oltre alle componenti principali, anche le variabili predeterminate dell'equazione stessa.

Del metodo 2SPC1 sono state sperimentate le seguenti varianti:

2SPC1 a: per ogni equazione viene introdotto, oltre alla costante, il numero minimo di componenti principali ( $k$ , variabile da equazione a equazione) richiesto per il rispetto delle condizioni per la identificabilità (si vedano [1, p. 285] e [25, p. 50]); in questo caso, per ogni equazione, sarà  $k=m+n-1$ , dove  $m$  e  $n$  sono, rispettivamente, il numero di endogene esplicative e di predeterminate (costante inclusa) presenti nella equazione.

2SPC1 b: per ogni equazione vengono introdotte tante componenti principali quante necessarie per spiegare almeno il 90% della varianza generalizzata: quindi  $k=4$ , ma allo scopo di non violare le condizioni per la identificabilità, nella equazione 14 ne sono state impiegate cinque.

2SPC1 c: come nel caso (b), ma spiegando almeno il 95% della varianza generalizzata:  $k=6$ .

Per spiegare il 99% della varianza sarebbero necessarie 10 componenti, ma con ciò si supererebbe il limite posto nel paragrafo 2.2; è stata pertanto eseguita un'altra prova limitando a 7 il numero delle componenti principali:

2SPC1 d: come nei casi (b) e (c), ma con  $k=7$ .

Il metodo 2SPC2 è stato sperimentato nelle seguenti varianti:

- 2SPC2 a: oltre alle variabili predeterminate, vengono introdotte nel primo stadio della stima di ogni equazione tante componenti principali quante necessarie, al minimo, per il rispetto delle condizioni per la identificabilità; cioè  $k=m$ , dove  $m$  è il numero delle variabili endogene esplicative nell'equazione stessa.
- 2SPC2 b: come al punto (a), ma selezionando le componenti principali secondo la tecnica proposta nel paragrafo 2.2, cioè ordinandole secondo  $\theta_i = v_i (1-R_i)$ .
- 2SPC2 c: oltre alle variabili predeterminate, vengono introdotte 4 componenti principali ordinate secondo l'autovalore.
- 2SPC2 d: come al punto (c), ma con 5 componenti principali.
- 2SPC2 e: oltre alle variabili predeterminate, vengono introdotte 4 componenti principali ordinate secondo  $\theta_i = v_i (1-R_i)$  (par. 2.2).
- 2SPC2 f: come al punto (e), ma con 5 componenti principali.

I due metodi, con le loro varianti, sono stati sperimentati sulle equazioni strutturali del modello ISPE. Due equazioni strutturali (8 e 18) coinvolgono solo variabili predeterminate e sono state, quindi, stimate soltanto col metodo OLS.

Nella tabella 1 sono riportati, nell'ordine, i risultati relativi ai metodi OLS e 2SPC2 f; quest'ultimo come sarà detto in seguito, è stato scelto perchè fornisce, tra tutte le varianti del 2SPC, i migliori risultati di simulazione sul periodo campionario. Sotto ogni coefficiente, in parentesi, è riportata la stima dell'errore standard (asintotico, nelle stime a due stadi); in parentesi quadra è riportata la stima dell'errore standard dell'equazione. Nel caso 2SPC2 f le stime degli errori standard non sono state corrette per i gradi di libertà, seguendo il suggerimento di Theil [32, p. 451].

Per il significato delle variabili e delle equazioni definizionali riportate alla fine di tabella 1, si rimanda a tabella 2.

Pur raccomandando la massima cautela nel confrontare risultati ottenuti con stimatori diversi quando, come nella maggior parte dei casi, manchino conoscenze a priori sul valore dei parametri da stimare, va tuttavia osservata, in tabella 1, una sostanziale concordanza in valore, segno e relativo errore standard dei coefficienti stimati; d'altra parte, a conforto di quanto rilevato, nei casi in cui, come per le equazioni 13, 14 e 19, le stime di alcuni coefficienti non sono confrontabili, spesso si tratta di stime statisticamente non significative per tutti i metodi applicati.

#### 4. Analisi dei risultati di simulazione

Il modello, stimato con i vari metodi alternativi, è stato simulato sul periodo campionario di stima 1955-1976 sia staticamente, sia dinamicamente.

Non tutti i risultati di simulazione vengono qui presentati, sia perchè non sempre si è raggiunta la convergenza in soluzione (come per le

varianti 2SPC1 a, 2SPC1 b, 2SPC2 a e 2SPC2 b), sia perchè alcuni metodi forniscono risultati molto simili ad altri riportati (2SPC2 e è molto simile a 2SPC2 f).

Sul periodo di simulazione, onde facilitare la valutazione del comportamento del modello (si vedano, ad esempio, [11], [17], [21], [23], [33]), sono state calcolate, per ogni variabile endogena, le seguenti misure sintetiche dell'accuratezza delle previsioni:

- Radice dell'Errore Quadratico Medio (Root Mean Square Error: RMSE [24, p. 242])
- Errore Assoluto Medio Percentuale (Mean Absolute Percentage Error: MAPE [24, p. 242])
- Coefficienti di Disuguaglianza di Theil (U [31, p. 28])

Si noti che RMSE e MAPE sono indicatori della bontà di interpolazione « in livelli », mentre il coefficiente di disuguaglianza di Theil è un indicatore della capacità del modello di prevedere le variazioni di livello e quindi, in particolare, i punti di svolta. Per ogni variabile e per ogni misura, ogni stimatore è stato successivamente ordinato (come in [11] e [21]) secondo il valore della misura in esame; la somma dei valori dello stesso indice per variabili diverse fornisce un indicatore sintetico del comportamento dell'intero modello in relazione al metodo di stima.

Le simulazioni statiche hanno mostrato risultati sostanzialmente simili per tutti i metodi di stima; i risultati non vengono riportati. Nel caso di simulazioni statiche si potrebbe, quindi, concludere che i vantaggi derivanti dall'uso di un metodo di stima che tenga conto della simultaneità del sistema non sembrano tali da consigliare l'uso del metodo stesso, tanto più se si pensa alla semplicità di ottenimento di stime OLS rispetto a stime 2SPC. Tenendo tuttavia presenti la struttura dinamica e le modalità di utilizzo del modello, da un punto di vista empirico sono certamente più interessanti e ricchi di contenuto informativo confronti su simulazioni dinamiche (come, ad esempio, in [11], [17], [21] e [23]).

Nella tabella 3, relativamente a MAPE, RMSE e U di Theil, sono riportati i risultati di simulazione dinamica su tutto il periodo campionario per le principali variabili endogene. Dato il limitato numero di cifre utilizzato nel riportare il valore degli indicatori, risulta talvolta impossibile ordinare i vari stimatori; è perciò il caso di ricordare che, di fatto, l'ordinamento è stato ottenuto utilizzando un maggior numero di cifre decimali, tali da distinguere tra loro, in ogni caso, i valori degli indicatori.

Da un primo confronto tra stime OLS e stime 2SPC è immediato rilevare una netta superiorità del 2SPC; per il modello ISPE, all'interno del metodo 2SPC è possibile, cioè, trovare alcune varianti che, quanto a « bontà della previsione » su tutto il periodo campionario (o capacità interpolante del modello), forniscono risultati decisamente più accurati di quelli ottenibili col metodo OLS. Questo è confermato sistematicamente da tutti gli indicatori presentati. Infatti, ad eccezione della U di Theil per la variabile PING, nei casi 2SPC1 d, 2SPC2 d e 2SPC2 f tutti gli indicatori sono, per tutte le variabili, sistematicamente migliori di quelli relativi alle stime OLS.



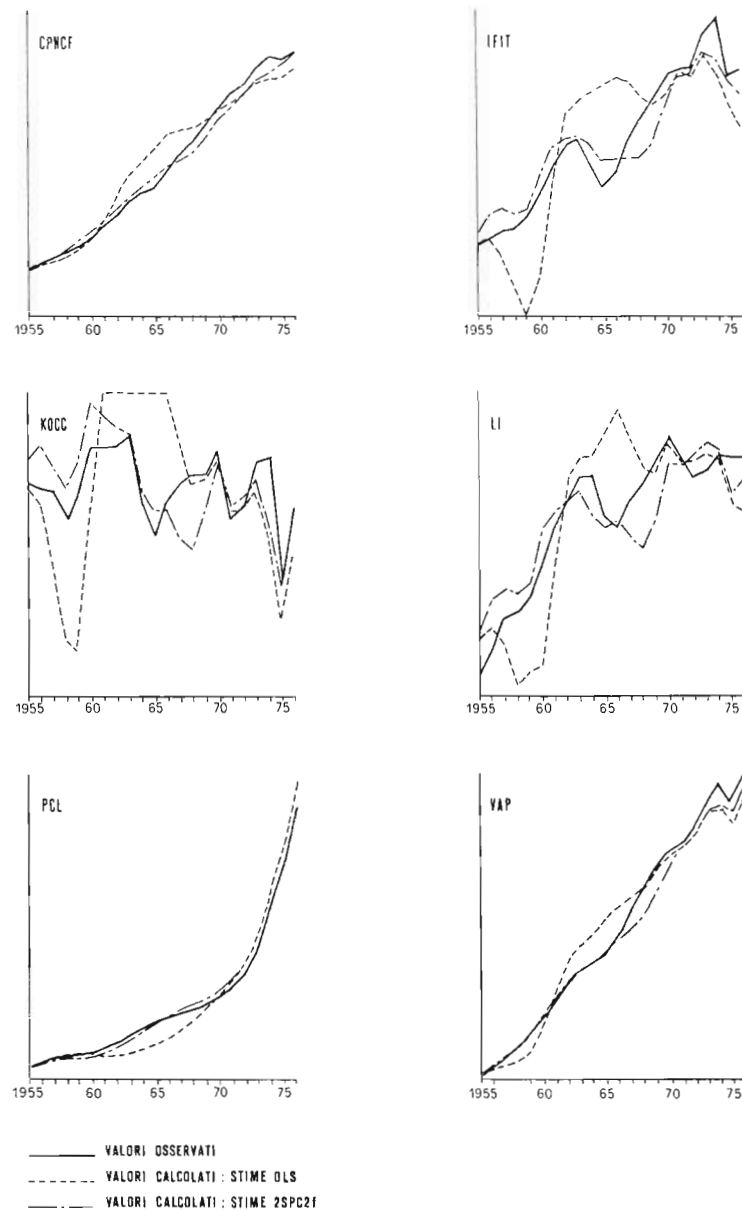


Fig. 1

La capacità interpolante dei metodi OLS e 2SPC nella variante 2SPC2 f (gli stessi, cioè, di cui alla tabella 1) è visualizzata nella figura 1 dove, per le variabili CPNCF, IFIT, KOCC, LI, PCL e VAP, sono confrontati graficamente i valori storici e i valori (relativi ai due stimatori) calcolati in simulazione dinamica su tutto il periodo campionario. Oltre al maggior accostamento ai valori storici nei casi di stime 2SPC (da cui più bassi valori di RMSE e MAPE), è interessante notare come queste stesse stime colgano meglio, rispetto alla stima OLS, i punti di svolta (e quindi più basso valore della U di Theil).

Limitando ora l'analisi ai due metodi 2SPC1 e 2SPC2 e alle loro varianti, la prima considerazione che emerge riguarda la sensibilità dei metodi stessi relativamente al numero delle componenti principali usate nel primo stadio della stima. Infatti, se con il numero minimo di componenti principali necessarie per il rispetto delle condizioni per la identificabilità non convergono in simulazione (o quanto meno non convergono a valori ragionevoli) né 2SPC1 né 2SPC2, aumentando progressivamente il numero di componenti principali si registra un deciso miglioramento delle prestazioni dei due metodi. Anche se per brevità non sono stati riportati i risultati di tutti gli esperimenti effettuati, particolarmente significativi ed evidenti a questo riguardo sono i risultati del 2SPC1; infatti, come mostrano gli indicatori riportati nella tabella 3, si hanno decisi miglioramenti passando da 4 a 5 e a 7 componenti principali.

Per le varianti 2SPC1 b e 2SPC2 c si ha un comportamento molto erratico ed instabile in simulazione (confermato dal fatto che riducendo il numero di componenti principali non si raggiunge la convergenza), per cui, almeno nel caso del modello ISPE, si può dire che è consigliabile superare i limiti suggeriti dal rispetto delle condizioni necessarie per soddisfare le condizioni per la identificabilità.

I risultati delle varianti 2SPC1 d, 2SPC2 d e 2SPC2 f non permettono in generale un netto ordinamento tra le varianti stesse; una analisi accurata dei valori degli indicatori mostra, infatti, che essi sono più vicini di quanto non appaia dall'analisi delle posizioni relative. D'altra parte, anche considerando il numero totale di variabili usate nel primo stadio della stima, queste varianti sono abbastanza simili; infatti, nel caso 2SPC1 d sono state usate 7 componenti principali, mentre nel 2SPC2 d e 2SPC2 f sono state usate 5 componenti più le predeterminate (mediamente due) presenti nella equazione da stimare.

Particolarmente interessanti sembrano i risultati relativi alle varianti 2SPC2 d e 2SPC2 f che, come rilevato nel paragrafo 3, differiscono soltanto per il criterio con cui le componenti principali sono selezionate: in particolare nella variante 2SPC2 f si è seguito l'ordinamento suggerito da Kloeck e Mennes [25, p. 53, metodo 4]. Dato il numero di componenti coinvolte (5 in questo caso), soltanto per 2 equazioni le 5 componenti introdotte non risultano essere le stesse 5 che tradizionalmente si sarebbero introdotte secondo l'ordine relativo agli autovalori (variante 2SPC2 d); non ci si poteva, quindi, attendere modifiche sostanziali nei risultati, e questo è confermato dai risultati empirici. Tuttavia, rispetto agli indicatori calcolati, si ha quasi sempre un miglioramento sistematico, anche se di lieve entità,

della capacità interpolante del modello nel caso di stime 2SPC2 f rispetto a stime 2SPC2 d.

L'esame degli indici di capacità previsiva permette anche di ricavare alcune considerazioni di interesse economico, con particolare riferimento alla interdipendenza tra settori o tra variabili del modello. Infatti, se in simulazione dinamica per alcune varianti del 2SPC si ha sempre e per tutte le variabili un miglioramento rispetto alla capacità previsiva del modello stimato col metodo OLS, non sempre si può osservare una netta superiorità nella capacità previsiva tra differenti varianti: singole variabili o singoli settori del modello sono previsti meglio con una certa variante del metodo di stima simultanea, altre variabili con un'altra variante. Tali differenze possono essere dovute a problemi di specificazione di qualche equazione del modello.

Per semplificare l'analisi ci limiteremo a confrontare gli indicatori di accuratezza delle previsioni delle varianti 2SPC1 d e 2SPC2 f nella tabella 3. Si può notare che la variante 2SPC1 d prevede meglio della variante 2SPC2 f il valore degli investimenti fissi nell'industria e nel terziario, il livello delle importazioni di merci, il grado di capacità utilizzata nell'industria e l'occupazione industriale. La variante 2SPC2 f, rispetto alla 2SPC1 d, prevede invece meglio le variabili relative alle variazioni delle remunerazioni da lavoro dipendente, le variabili relative alle variazioni dei deflatori di prezzo e il livello dei consumi privati nazionali.

Tenendo presenti le relazioni causali del modello, è possibile interpretare i risultati precedentemente illustrati. Nella variante 2SPC1 d la principale fonte di errori sembra essere la funzione delle remunerazioni da lavoro dipendente nel settore industriale. Infatti, gli errori provocati da questa funzione si diffondono, attraverso il settore prezzi, alle previsioni dei consumi privati: nella variante 2SPC2 f, nella quale i salari sono previsti meglio a causa di errori compensativi risultanti nella equazione dell'occupazione, si hanno risultati migliori in tutto il settore prezzi e nei consumi privati.

Gli errori di previsione degli investimenti derivano, in massima parte, da quelli della capacità utilizzata e quindi da quelli del valore aggiunto industriale. Su quest'ultimo rifluiscono tutti gli errori delle funzioni di domanda e di offerta: l'errore di gran lunga più elevato è quello della funzione delle importazioni, com'è facile notare considerando anche i risultati della variante 2SPC2 c.

Tali considerazioni possono essere utili per avere indicazioni sulle equazioni da specificare meglio per apportare al modello ulteriori miglioramenti.

##### 5. Conclusioni

Si è cercato, in questo lavoro, di fornire una valutazione sperimentale, cioè su un modello operativo, del metodo di stima 2SPC, confrontandolo col metodo OLS. I risultati sperimentali sul modello ISPE suggeriscono le seguenti conclusioni.

— La scelta e il numero delle componenti principali rimane un fatto eminentemente sperimentale; si è notata una elevata sensibilità dei risultati al numero di componenti usate.

— L'analisi comparata di alcune varianti del metodo di stima 2SPC potrebbe fornire indicazioni per una più corretta specificazione delle equazioni.

— Il confronto puro e semplice delle diverse stime dei coefficienti strutturali e le simulazioni statiche non forniscono sufficienti indicazioni circa la superiorità di un metodo di stima sull'altro.

— I risultati di simulazione dinamica permettono, invece, una netta discriminazione tra gli stimatori considerati. In particolare, analogamente alle conclusioni in [23], vi sono varianti del 2SPC che forniscono risultati decisamente superiori a quelli del metodo OLS.

CARLO BIANCHI \* - GIORGIO CALZOLARI \* - FRANCO SARTORI \*\*

\* *Centro Scientifico IBM, Pisa*

\*\* *ISPE, Roma*

Tabella 1

Stime OLS - 2SPC2f

## 1-Consumi Privati Nazionali

CPNL/RDP = $\frac{\Delta(RDP/PCL)}{(RDP/PCL)}$	T	1.	
-.3229 (.1009)	-.00299 (.00029)	.8016 (.0079)	[.0079]
-.2435 (.1393)	-.00290 (.00030)	.7965 (.0099)	[.0075]

## 2-Investimenti Privati Fissi Non-Residenziali nell'Industria e nel Terziario

IFIT = $.4(1-KOCC) + \frac{(.6FMI+.4PMI_{-1})}{PIFIT} (VAI+VAT)_{-1}$	1.		
-13290. (2166.)	.4069 (.1272)	.1132 (.0131)	2151. (239.1)
-12440. (2196.)	.2501 (.1647)	.1274 (.0162)	1986. (247.4)

## 3-Investimenti in Abitazioni

IAB = $PCL/PIAB$	INIZ <sub>-1</sub>	V <sub>-1</sub>	IAB <sub>-1</sub>	1.
2065. (398.6)	.0056 (.0010)	-747.7 (124.0)	.3411 (.0917)	683.5 (495.2)
1914. (380.5)	.0057 (.0009)	-715.0 (113.8)	.3436 (.0810)	774.0 (445.6)

## 4-Esportazioni di Prodotti Manifatturati

LnXM\$ = $\ln(.5DXML/ERL$ + .5(DXML/ERL$)_{-1})$	LnXMM\$	Ln(XM\$ / XM\$)	Ln(.5DXMM\$ + .5DXMM\$ <sub>-1</sub> )	1.
-1.953 (.4458)	1.468 (.0853)	-1.009 (.2908)	1.380 (.5051)	-10.10 (2.874)
-1.928 (.5274)	1.472 (.0966)	-1.020 (.3073)	1.353 (.5943)	-9.940 (3.400)

## 5-Importazioni di Merci

LnM = $\ln(.5(CPNC+INT+IF+X-IAB) + .5(CPNC+INT+IF+X-IAB)_{-1})$	KOCC	Ln(.5PM/PING + .5(PM/PING) <sub>-1</sub> )	DMEC	1.
1.627 (.0349)	.9803 (.2289)	-.2145 (.1384)	.1417 (.0369)	-9.389 (.6815)
1.639 (.0331)	1.103 (.2542)	-.1536 (.1325)	.1429 (.0342)	-9.222 (.6380)

## 6-Prodotto Lordo del Settore Terziario Privato (Escluse le Abitazioni)

VAT/(VAP+MT) = $1/(VAP+MT)$	T	1.	
3262. (330.4)	.00517 (.00050)	.1595 (.0161)	[.0035]
3191. (319.9)	.00506 (.00048)	.1629 (.0156)	[.0033]

## 7-Occupazione Industriale (Equazione Tecnica)

Ln(LI/VAI) = $\ln(VAI/VAI_{-1})$	LnKOCC	LnVAIP	1.
-.4471 (.1728)	-.2495 (.1949)	-.8473 (.0135)	7.492 (.1404)
-.3800 (.1824)	-.2289 (.1957)	-.8438 (.0129)	7.456 (.1358)

segue Tabella 1

## 8-Occupati Dipendenti Nell'Industria (Equazione Tecnica. Solo Stima OLS)

LDI/LI = $KOCC_{-1}$	T	1.	
.1184 (.0439)	.0030 (.0002)	.6932 (.0405)	[.0067]

## 9-Occupati nel Settore Terziario Privato (Equazione Tecnica)

LTAT = $VAT$	LTAT <sub>-1</sub>	1.	
.0397 (.0143)	.6478 (.1373)	1607. (611.4)	[80.28]
.0351 (.0147)	.6905 (.1408)	1421. (624.9)	[75.49]

## 10-Occupati Dipendenti nel Settore Terziario Privato (Equazione Tecnica)

LDT = $LTAT$	1.	
.6928 (.0189)	-732.2 (113.8)	[48.43]
.6953 (.0181)	-747.1 (109.5)	[46.20]

## 11-Salari e Stipendi per Dipendente nell'Industria al Netto dei Contributi Sociali

ΔSUINE = $\Delta PCL$	ZDCO	DIS <sub>-1</sub>	1.
.7699 (.0858)	7.587 (1.202)	-.0046 (.0018)	7.984 (1.357)
.8069 (.0803)	7.576 (1.093)	-.0042 (.0017)	7.547 (1.254)

## 12-Salari e Stipendi per Dipendente nel Settore Terziario Privato

ΔSUDT = $\Delta SUINE$	ΔSUINE <sub>-1</sub>	1.	
.4956 (.0778)	.3501 (.0818)	1.999 (.8362)	[1.752]
.4738 (.0886)	.3651 (.0839)	2.088 (.8062)	[1.632]

## 13-Indice dei Prezzi all'Ingrosso

ΔPING = $\Delta KOCC$	$\Delta \left( \frac{SUINE(1+ACSIM)}{VAI/LI} \right)$	Δ(PM\$.ERL\$)	ΔPVAA	1.
.1786 (.1339)	.1878 (.0909)	.5640 (.0440)	.0924 (.1015)	.4653 (.6329)
.4881 (.2104)	.3899 (.1358)	.5242 (.0494)	-.0322 (.1207)	-.0635 (.7001)

## 14-Prezzi Impliciti nella Variabile CIT (Consumi)

LnPCIT = $\ln CLUPI$	LnCLUPT	Ln(CIT/CIT <sub>-1</sub> )	LnPM\$ + LnERL\$	LnPCIT <sub>-1</sub>	1.
.1378 (.0749)	.5191 (.0961)	.2066 (.1107)	.0709 (.0211)	.2842 (.0785)	.1584 (.0157)
.0779 (.1130)	.6938 (.1659)	.0731 (.1813)	.0631 (.0245)	.1487 (.1240)	.1758 (.0197)

## 15-Prezzi Impliciti negli Invest. Privati Fissi Non-Residenziali Industria e Terziario

LnPIFIT = $\ln(.8CLUPI + .2CLUPI_{-1})$	LnKOCC	Ln(DXMM\$.ERL\$)	1.
.6102 (.0286)	.2328 (.1038)	.4115 (.0377)	-2.422 (.2562)
.6116 (.0267)	.2067 (.1145)	.4087 (.0355)	-2.406 (.2400)

16-Prezzi Impliciti negli Investimenti in Abitazioni				
PIAB =	.8CLUPI+.2CLUPI <sub>-1</sub>	IAB	PM\$.ERL\$	1.
	1.062 (.0934)	.000058 (.000016)	.2636 (.0546)	-.3060 (.0476)
	1.032 (.0969)	.000069 (.000016)	.2786 (.0560)	-.3340 (.0487)

17-Prezzi Impliciti nelle Esportazioni di Prodotti Manufatti				
LnDXML =	Ln(VAI/LI)	Ln(SUINE. (1+ACSIM))	Ln(DXMW\$.ERL\$)	1.
	-.7307 (.1254)	.3001 (.0801)	.7566 (.0778)	-4.359 (.5595)
	-.9507 (.1797)	.8402 (.1143)	.6337 (.1052)	-3.447 (.7688)

18-Imposte Dirette (Solo Stima OLS)				
LnTD =	LnRNLD <sub>-1</sub>	DUMMY76. LnRNLD <sub>-1</sub>	DUMMY70	1.
	1.199 (.0150)	.0173 (.0038)	-.1040 (.0406)	-4.776 (.1561)

19-Contributi per la Sicurezza Sociale				
LnCS =	Ln(SUINE. (1+ACSIM).LDI+ SUDT.LDT)	Ln(LI+LTAT)	LnACSIM	1.
	.9134 (.0402)	1.127 (.5231)	.3692 (.2374)	-10.64 (4.744)
	.8443 (.0521)	2.090 (.6985)	.0800 (.2739)	-19.35 (6.329)

$$CLUPI = SUINE.(1 + ACSIM) / (VAI / LI)$$

$$CLUPT = SUDT / (VAT / LTAT)$$

$$CIT = CPNCF - VAA - VAF$$

$$PCIT = (PCPNCF.CPNCF - VAA.PVAA - VAF.PVAF) / (CPNCF - VAA - VAF)$$

$$CPNCF = CPNL / PCL$$

$$IF = IFIT + IAB + IES$$

$$XT = 627.16 XM\$ + XAML + XS$$

$$MT = M + MS$$

$$VAP = CPNCF + CNT + IF + DSC + XT - MT + APA$$

$$VAI = VAP - VAT - VAA - VAF + SBI$$

$$VAI = VAIP \text{ se } VAI > VAIP$$

$$KOCC = VAI / VAIP$$

$$DIS = FORLA - LTAT - LI - LA - LPA$$

$$PVAP = (CPNCF.PCPNCF + CNT.PCNT + APA.PAPA + IFIT.PIFIT + IAB.PIAB + IES.PIES + DSC.PDSC + XS.PXS + 627.16 XM$.DXML + XAML.DXAML - M.PM - MS.PMS) / VAP$$

$$PCL = PCPNCF + (TIL - TRI) / CPNCF$$

$$PILL = VAP.PVAP + SUPA.LPA + AMMPA + TIL - TRI$$

$$RNLD = PILL + RETNE + TINE$$

$$TIL = ATI.CPNL$$

$$RDP = RNLD - (TNPA + TD + CS + TIL + TINE - TRI)$$

$$SPA = RNLD - RDP - (SUPA.LPA + AMMPA + APA.PAPA)$$

## VARIABILI ENDOGENE

CIT	- Consumi Privati Nazionali al Costo dei Fattori (al Netto della Spesa per Consumi Agricoli e della Spesa per Abitazioni)
CLUPI	- Costo del Lavoro per Unita' di Prodotto nell'Industria
CLUPT	- Costo del Lavoro per Unita' di Prodotto nel Settore Terziario Privato
CPNCF	- Consumi Privati al Netto delle Imposte Indirette
CPNL	- Consumi Privati al Lordo delle Imposte Indirette
CS	- Contributi Sociali
DIS	- Disoccupazione
DXML	- Prezzi Impliciti nelle Esportazioni di Prodotti Manufatti (1970=1)
IAB,IF	- Investimenti in Abitazioni e Investimenti Fissi Totali
IFIT	- Investimenti Fissi Privati Non-Residenziali nell'Industria e nel Terziario Privato
KOCC	- Indico di Utilizzazione della Capacita' Produttiva nell'Industria
LDI,LDT	- Occupati Dipendenti nell'Industria e nel Terziario Privato (Migliaia)
LI	- Occupati nell'Industria (Migliaia)
LTAT	- Occupati nel Settore Terziario Privato (Migliaia)
M,MT	- Importazioni, rispettivamente, di Merci e di Mercati e Servizi
PCIT	- Prezzi Impliciti nella Variabile CIT
PCL	- Prezzi Impliciti nei Consumi Privati al Lordo delle Imposte Indirette (1970=1)
PCPNCF	- Prezzi Impliciti nei Consumi Privati al Netto delle Imposte Indirette (1970=1)
PIAB	- Prezzi Impliciti negli Investimenti in Abitazioni (1970=1)
PIFIT	- Prezzi Implic. negli Investim. Privati nell'Industria e Terziario Privato (1970=1)
PILL	- Prodotto Interno Lordo
PING	- Indico dei Prezzi all'Ingresso (1970=1)
PVAP	- Prezzi Impliciti nel Prodotto Lordo del Settore Privato (1970=1)
RDP	- Reddito Lordo Disponibile del Settore Privato
RNLD	- Reddito Nazionale Lordo Disponibile
SPA	- Risparmio Lordo del Settore Pubblico
SUDT	- Salari e Stipendi per Dipendente, Settore Privato dei Servizi (Milioni Lire Corr.)
SUINE	- Salari e Stipendi per Dipendente nell'Industria al Netto dei Contributi Sociali
TD,TIL	- Imposte Dirette e Indirette
VAI,VAIP	- Prodotto Lordo nell'Industria e nel Settore Privato
VAT	- Prodotto Lordo nel Settore Privato dei Servizi (Esclusi Fabbricati)
XM\$,XT	- Esportaz. di Prodotti Manufatti (Miliardi Dollari USA 1970) e di Mercati e Servizi

## VARIABILI ESOGENE

ACSIM	- Aliquota Contributi Sociali degli Operai nell'Industria Manifatturiera
AMMPA	- Ammortamenti della Pubblica Amministrazione
APA	- Consumi Intermedi della Pubblica Amministrazione
ATI	- Aliquota delle Imposte Dirette
CNT	- Consumi Netti del Non-Residenti
DMEC	- Variabile Dummy per il Mercato Comune (=1 dal 1960)
DSC	- Investimenti in Scorte
DUMMY70	- Variabile Dummy per l'Equazione delle Imposte Dirette (1970=1)
DUMMY76	- Variabile Dummy per l'Equazione delle Imposte Dirette (1976=1)
DXAML	- Prezzi Impliciti nelle Esportazioni di Prodotti Non-Manufatti (1970=1)
DXMWS	- Valori Medi Unitari Esportaz. Manufatti Paesi Industriali in Dollari USA (1970=1)
ERL\$	- Saggio di Cambio Lira / Dollaro USA
FMI	- Finanziamenti a Medio e Lungo Termine
FORLA	- Forza Lavoro
IES	- Investimenti Esogeni (=IA+IPA, [30])
INIZ	- Costruzioni Residenziali Iniziate (Migliaia di Metri Cubi)
LA,LPA	- Occupazione nel Settore Agricolo e nel Settore Pubblico (Migliaia)
MS	- Importazioni di Servizi
PAPA	- Prezzi Impliciti nei Consumi Intermedi della Pubblica Amministrazione (1970=1)
PCNT	- Prezzi Impliciti nei Consumi Netti dei Non-Residenti (1970=1)
PDSC	- Prezzi Impliciti negli Investimenti in Scorte (1970=1)
PIES	- Prezzi Impliciti negli Investimenti Esogeni (=IA.PIA+IPA.PIPA)/(IA+IPA), [30]
PM\$	- Prezzi Impliciti nelle Importazioni di Mercati in Dollari USA (=PM/ERL\$, [30])
PMS\$	- Prezzi Impliciti nelle Importazioni di Servizi in Dollari USA (=PMS/ERL\$, [30])
PVAF	- Prezzi Impliciti nel Prodotto Lordo dei Fabbricati
PVAA	- Prezzi Impliciti nel Prodotto Lordo del Settore Agricolo (1970=1)
PXS	- Prezzi Impliciti nelle Esportazioni di Servizi (1970=1)
RETNE	- (=VAI+RNE+TNE, [30])
SBI	- Servizi Bancari Imputati
SUPA	- Costo del Lavoro per Occupato nel Settore Pubblico (Milioni di Lire Correnti)
T	- Variabile di Trend (1951=1, 1952=2,...)
TINE	- Imposte Indirette Netto alla Comunita' Economica Europea
TRI	- Contributi alla Produzione
V	- Velocita' di Circolazione della Moneta
VAA,VAIP	- Prodotto Lordo nel Settore Agricolo e Prodotto Lordo dei Fabbricati
VAF	- Prodotto Lordo Potenziale nell'Industria
XAML	- Esportazioni di Prodotti Non-Manufatti
XS	- Esportazioni di Servizi
XMW\$	- Indico di Quantita' Esportaz. Prodotti Manufatti dei Paesi Industriali (1970=1)
XMWTS	- Valore di Trend di XMW\$
ZDCO	- Variabile Dummy (=1 nel 1963, 1970 e 1973)

Tabella 3

	Mean Absolute Percentage Errors (MAPE)							Root Mean Square Errors (RMSE)							Coefficienti di Disuguaglianza di Theil (U)						
	OLS	2SPC 1b	2SPC 1c	2SPC 1d	2SPC 2c	2SPC 2d	2SPC 2f	OLS	2SPC 1b	2SPC 1c	2SPC 1d	2SPC 2c	2SPC 2d	2SPC 2f	OLS	2SPC 1b	2SPC 1c	2SPC 1d	2SPC 2c	2SPC 2d	2SPC 2f
CPNCF	5.7	10.	3.1	2.9	5.8	2.8	2.7	1782	3369	1038	986.	1766	932.	886.	.59	1.1	.40	.39	.79	.38	.37
DXML	4.3	8.8	3.2	3.1	7.7	2.8	2.7	0.07	0.14	0.06	0.06	0.10	0.05	0.05	.46	.48	.37	.36	.52	.36	.36
IAB	6.3	7.9	5.5	5.0	6.6	5.2	5.2	219.	316.	202.	191.	270.	201.	201.	.88	.82	.76	.72	.74	.72	.72
IF	11.	36.	4.7	4.4	22.	5.0	5.1	1218	4521	553.	532.	2508	569.	583.	1.4	3.6	.66	.64	2.2	.62	.62
IFIT	21.	73.	9.1	8.3	43.	9.4	9.6	1197	4590	539.	516.	2405	518.	523.	2.9	42.	.82	.78	17.	.73	.73
KOCC	6.0	12.	2.9	2.6	9.2	2.9	3.0	0.07	0.14	0.03	0.03	0.10	0.03	0.03	1.2	1.5	.63	.61	1.5	.60	.59
LDI	4.2	7.7	2.1	2.0	8.8	2.8	2.9	334.	640.	153.	145.	622.	187.	201.	1.6	1.9	.95	.89	2.1	.89	.90
LDT	3.1	2.9	2.5	2.4	1.4	2.2	2.2	119.	120.	94.6	91.3	55.7	85.0	83.8	.72	.67	.65	.64	.64	.63	.63
LI	3.9	6.3	2.1	1.9	7.5	2.4	2.5	361.	635.	176.	166.	633.	198.	210.	1.7	1.9	1.1	1.0	2.2	1.0	1.0
LTAT	1.8	2.2	1.3	1.2	1.3	1.1	1.1	126.	165.	93.5	89.8	96.2	83.2	82.3	.68	.69	.64	.63	.71	.63	.63
MT	15.	35.	6.4	6.3	25.	7.5	7.6	1274	3103	701.	665.	1863	657.	667.	.75	1.4	.37	.36	1.2	.35	.34
PCL	7.3	10.	4.3	4.1	7.5	3.5	3.2	0.11	0.19	0.08	0.08	0.12	0.06	0.05	.38	.67	.23	.21	.44	.18	.17
PIFIT	5.9	9.3	3.7	3.5	6.9	2.9	2.7	0.08	0.16	0.07	0.06	0.09	0.04	0.04	.34	.52	.26	.26	.39	.25	.24
PILL	3.9	5.4	2.1	1.9	4.4	1.6	1.4	2375	4837	2501	2259	3631	1513	1278	.15	.27	.11	.10	.15	.09	.09
PING	3.1	3.9	2.2	2.2	3.2	1.8	1.7	4.15	5.30	3.19	3.29	3.26	2.64	2.56	.19	.23	.20	.19	.18	.20	.20
PVAP	7.2	9.6	4.2	4.0	8.0	3.4	3.1	0.09	0.15	0.07	0.06	0.11	0.05	0.05	.40	.65	.23	.22	.46	.20	.20
SUINE	9.1	14.	5.3	4.9	12.	4.5	4.3	0.16	0.36	0.15	0.13	0.21	0.09	0.08	.34	.50	.18	.18	.10	.17	.18
VAP	4.6	11.	2.1	1.9	7.0	2.1	2.1	2047	5164	1077	1007	3113	1026	1036	.60	1.1	.34	.32	.20	.32	.32
	Stimatori ordinati secondo MAPE							Stimatori ordinati secondo RMSE							Stimatori ordinati secondo U di Theil						
CPNCF	5	7	4	3	6	2	1	6	7	4	3	5	2	1	5	7	4	3	6	2	1
DXML	5	7	4	3	6	2	1	5	7	4	3	6	2	1	5	6	4	2	7	1	3
IAB	5	7	4	1	6	2	3	5	7	4	1	6	2	3	7	6	5	3	4	2	1
IF	5	7	2	1	6	3	4	5	7	2	1	6	3	4	5	7	4	3	6	2	1
IFIT	5	7	2	1	6	3	4	5	7	4	1	6	2	3	5	7	4	3	6	2	1
KOCC	5	7	2	1	6	3	4	5	7	2	1	6	3	4	5	7	4	3	6	2	1
LDI	5	6	2	1	7	3	4	5	7	2	1	6	3	4	5	6	4	1	7	2	3
LDT	7	6	5	4	1	3	2	6	7	5	4	1	3	2	7	6	5	4	3	2	1
LI	5	6	2	1	7	3	4	5	7	2	1	6	3	4	5	6	4	3	7	1	2
LTAT	6	7	4	3	5	2	1	6	7	4	3	5	2	1	5	6	4	3	7	1	2
MT	5	7	2	1	6	3	4	5	7	4	2	6	1	3	5	7	4	3	6	2	1
PCL	5	7	4	3	6	2	1	5	7	4	3	6	2	1	5	7	4	3	6	2	1
PIFIT	5	7	4	3	6	2	1	5	7	4	3	6	2	1	5	7	4	3	6	2	1
PILL	5	7	4	3	6	2	1	4	7	5	3	6	2	1	6	7	4	3	5	2	1
PING	5	7	4	3	6	2	1	6	7	3	5	4	2	1	3	7	4	2	1	5	6
PVAP	5	7	4	3	6	2	1	5	7	4	3	6	2	1	5	7	4	3	6	2	1
SUINE	5	7	4	3	6	2	1	5	7	4	3	6	2	1	5	7	4	3	6	1	2
VAP	5	7	4	1	6	3	2	5	7	4	1	6	2	3	5	7	4	2	6	3	1
Totale	93	123	61	39	104	44	40	93	126	65	42	99	40	39	93	120	74	50	101	36	30

## BIBLIOGRAFIA

- [1] T. AMEMIYA, « On the Use of Principal Components of Independent Variables in Two-Stage Least-Squares Estimation », in *International Economic Review*, 7, 1966, pp. 283-303.
- [2] R.J. BALL, « Some Problems of Empirical Macro-Economic Models », paper presented at the IFAC/IFORS International Conference on Dynamic Modelling and Control of National Economies, Warwick, 1973.
- [3] C. BIANCHI and G. CALZOLARI, « Simulation of a Nonlinear Econometric Model », in *Simulation of System '79*, ed. by L. DEKKER, G. SAVASTANO and G.C. VANSTEENKISTE, Amsterdam, North Holland, 1980, pp. 105-113.
- [4] C. BIANCHI and G. CALZOLARI, « The One-Period Forecast Errors in Nonlinear Econometric Models », in *International Economic Review*, 21, 1980, pp. 201-208.
- [5] C. BIANCHI, G. CALZOLARI and P. CORSI, « A Program for Stochastic Simulation of Econometric Models », in *Econometrica*, 46, 1978, pp. 235-236.
- [6] C. BIANCHI, G. CALZOLARI and P. CORSI, « Estimating Asymptotic Standard Errors and Inconsistencies of Impact Multipliers in Nonlinear Econometric Models », in *Journal of Econometrics*, 16, 1981, pp. 277-294.
- [7] J.M. BRUNDY and D.W. JORGENSON, « Efficient Estimation of Simultaneous Equations by Instrumental Variables », in *The Review of Economics and Statistics*, 53, 1971, pp. 207-224.
- [8] P.J. DHRYMES, *Econometrics: Statistical Foundation and Applications*, New York, Harper & Row, 1970.
- [9] P.J. DHRYMES, « A Simplified Structural Estimator for Large-Scale Econometrics Models », in *The Australian Journal of Statistics*, 13, 1971, pp. 168-175.
- [10] W.J. DIXON, *BMD Biomedical Computer Programs*, Los Angeles, University of California Press, 1971.
- [11] M. DUTTA, and P.L. SHARMA, « Alternative Estimators and Predictive Power of Alternative Estimators: An Econometric Model for Puerto Rico », in *The Review of Economics and Statistics*, 55, 1973, pp. 381-385.
- [12] D.L. EDGERTON, « Some Properties of Two Stage Least Squares as Applied to Nonlinear Models », in *International Economic Review*, 13, 1972, pp. 26-32.
- [13] H. EISENPRESS, « Note on the Computation of Full-Information Maximum Likelihood Estimates of Coefficients of a Simultaneous System », in *Econometrica*, 30, 1962, pp. 343-348.
- [14] M.K. EVANS and L.R. KLEIN, *The Wharton Econometric Forecasting Model*. Economics Research Unit, University of Pennsylvania, 1968.
- [15] F.M. FISHER, « Dynamic Structure and Estimation in Economic-Wide Econometric System », in *The Brookings Quarterly Econometric Model of the United States*, ed. by J.S. DUESENBERY, G. FROMM, L.R. KLEIN

- and E. KUH, Chicago, Rand McNally, 1965, pp. 198-635.
- [16] A.R. GALLANT, « Three-Stage Least-Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations », in *Journal of Econometrics*, 5, 1977, pp. 71-88.
- [17] G. GAMBETTA, « Stima e Simulazione di un Modello Macroeconomico: Un Esempio di Metodologia Econometrica », in *Un Modello per l'Economia Italiana*, Bologna, Il Mulino, 1976, pp. 223-252.
- [18] A.S. GOLDBERGER, *Econometric Theory*, New York John Wiley, 1964.
- [19] S.M. GOLDFELD and R.E. QUANDT, « Nonlinear Simultaneous Equations: Estimation and Prediction », in *International Economic Review*, 9, 1968, pp. 113-136.
- [20] T. HAAVELMO, « The Statistical Implication of a System of Simultaneous Equations », in *Econometrica*, 11, 1943, pp. 1-12.
- [21] J. HIRVONEN, *On the Use of Two Stage Least Squares with Principal Components. An Experiment with a Quarterly Model*, Helsinki, Bank of Finland, 1975.
- [22] H.H. KELEJIAN, « Two Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous variables », in *Journal of the American Statistical Association*, 66, 1971, pp. 373-374.
- [23] L.R. KLEIN, « Estimation of Interdependent Systems in Macroeconomics », in *Econometrica*, 37, 1969, pp. 171-192.
- [24] L.R. KLEIN, *A Textbook of Econometrics*, Englewood Cliffs, Prentice-Hall, 1974.
- [25] T. KLOEK and L.B.M. MENNES, « Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables », in *Econometrica*, 28, 1960, pp. 45-61.
- [26] J. KMENTA, *Elements of Econometrics*, New York, Macmillan, 1971.
- [27] M. MCCARTHY, « Notes on the Selection of Instruments for Two-Stage Least Squares and k-Class Type Estimators of Large Models », in *The Southern Economic Journal*, 37, 1971, pp. 251-259.
- [28] B.M. MITCHELL, « Estimation of Large Econometric Models by Principal Component and Instrumental Variable Methods », in *The Review of Economics and Statistics*, 53, 1971, pp. 140-146.
- [29] G.M. REY, « L'Utilizzo dei Modelli Econometrici per la Politica Economica », in *Soluzione e Impiego di Modelli Econometrici*, a cura di G. PARENTI, Bologna, Il Mulino, 1974, pp. 71-92.
- [30] F. SARTORI, « Caratteristiche e Struttura del Modello », in *Un Modello Econometrico dell'Economia Italiana; Caratteristiche e Impiego*, Roma, *Ispequaderni*, Serie Studi e Ricerche, 1, 1978, pp. 9-36.
- [31] H. THEIL, *Applied Economic Forecasting*, Amsterdam, North Holland, 1966.
- [32] H. THEIL, *Principles of Econometrics*, New York, John Wiley, 1971.
- [33] I. VISCO, « Metodi Alternativi di Stima di un Modello Econometrico: Teoria e Applicazione a un Modello dell'Economia Italiana », in *Giornale degli Economisti e Annali di Economia*, 35, 1976, pp. 67-98.