

2. RICERCA STATISTICA DI MODELLI ESTIMATIVI: UN'APPLICAZIONE DELLE COMPONENTI PRINCIPALI

ERNESTO MILANESE

1. Premessa - 2. Il campione e le variabili considerate - 3. Analisi preliminari dei dati - 4. L'analisi delle componenti principali - 5. L'equazione di regressione - 6. Il metodo seguito - 7. I risultati - 8. Conclusioni.

1. Premessa

Con il presente lavoro si vuole apportare un contributo allo sviluppo di modelli estimativi costruiti con l'ausilio di metodi statistici descrittivi (analisi multivariata), secondo un orientamento che è ormai generalmente condiviso dai cultori di estimo, pur con giustificate cautele¹, e che anzi è progredito sino a farne intravedere la promozione a 'procedimento' di stima (cfr. MISSERI (1985) [15], p. 46).

L'applicazione concerne in maniera particolare il metodo delle *componenti principali*, considerato secondo due diversi orientamenti:

- il primo, più generico, ha più che altro lo scopo di fare apparire i legami spaziali temporali naturali economici ecc., che potrebbero sfuggire a un semplice esame diretto, quando si agisca in settori di studio o ambienti ancora poco conosciuti, ed è quindi rivolto piuttosto alle analisi di natura economica e territoriale;
- il secondo, più specifico, porta ad indagare sui fattori (parametri) caratterizzanti un certo mercato fondiario, al fine d'individuare la relazione funzionale con il prezzo; e ciò, sia in fase di primo approccio, quando appunto si tenta di identificare le variabili che presentino il maggior interesse estimativo², sia successivamente, nella costruzione della 'migliore' equazione di regressione, quale ausilio per una opportuna scelta delle variabili da introdurre o da escludere³.

¹ A titolo esemplificativo, e limitatamente ai più recenti contributi degli economisti agrari, v. GAGGIATI e GALLERANI (1984) [3], GAGGIATI et al. (1982) [2], ANANIA (1981) [1], CHUDLEGH (1979) [4], SIMONOTTI (1979) [18], SMITH (1979) [19], MALACARNE (1977) [12], pp. 153 e 164.

² Che siano cioè *estimativamente valide* (MALACARNE [12], p. 164). Sulla *identificazione dei parametri* v. GRILLENZONI (1968) [7], p. 7.

³ Tale scelta può essere d'aiuto per ridurre la multicollinearità, grazie all'introduzione di nuovi parametri — le componenti principali — che sono linearmente indipendenti per costruzione (ortogonalità); o anche solo per diminuire il numero delle variabili indipendenti, semplificando così il modello senza apprezzabile perdita di informazione.

In questa sede, si è data prevalenza agli aspetti metodologici e si presentano solo i risultati meno legati alla particolare natura dei beni (fondi agricoli), in modo che ne appaia l'interesse applicativo anche per il settore territoriale e forestale; in ciò confortati dall'opinione di MERLO (1985 [13], p. 77) che gli aspetti valutativi assumanò ivi un'importanza anche maggiore rispetto al campo propriamente agricolo.

Si spera anzi che risulti in maniera convincente la capacità rappresentativa di tali analisi numeriche pure per le indagini di natura economico-territoriale, non appena si disponga di campioni di adeguata numerosità⁴, oltre che per quelle propriamente estimative.

2. Il campione e le variabili considerate

Il campione comprende 909 compere-vendite avvenute in comune di Lamezia Terme (CZ) nel triennio 1979-81, identificate a partire dai rispettivi contratti; tale numero è pari al 76% di tutte le note di voltura presentate all'UTE nel medesimo periodo di tempo⁵.

Sono stati considerati solo fondi o appezzamenti comprendenti una unica qualità di coltura; poiché di 562 appezzamenti non è stata possibile la localizzazione, né direttamente né sulle carte dell'IGMI né sulle mappe catastali, l'analisi completa è stata eseguita su 347 casi.

Gli elementi rilevati — le variabili e altre grandezze relative alla qualità dei beni o a specifiche caratteristiche della contrattazione, dette 'fattori' — tutti scelti sulla base della consuetudine, delle analisi conosciute⁶ e della situazione locale, sono elencati, con l'indicazione della fonte, nelle tabb. 1 e 2.

Le variabili sono le seguenti: NUME è il *numero d'ordine* del contratto; SUPE è la *superficie totale* (catastale) del fondo; DIST la *distanza* dal centro amministrativo (ex comuni di Sambiase, Sant'Eufemia e Lamezia); ASCI e ORDI sono l'ascissa e l'ordinata *cartografiche* del reticolo chilometrico UTM delle 'tavole' dell'IGMI, da dove è stata rilevata pure l'*altitudine* ALTE; PREZ è il valore dichiarato nell'atto, mentre PREZ1 è il relativo *prezzo*; INFLA è un coefficiente di *adeguamento monetario* (esprime i valori in lire '81); VAM sono i *valori agricoli medi*

⁴ Quello della carenza di dati — sia consentita l'osservazione — è un lamento che da anni si leva nelle ns. riunioni, e che forse dipende meno da noi che dai pubblici uffici dove le informazioni numeriche si accumulano troppo spesso inutilmente. Sull'argomento v. gli atti del XV *Incontro* Ce.S.E.T. (Viterbo, 1985), e in particolare la mozione finale, a p. 189; e inoltre C. VANZETTI, « Il mercato fondiario in Italia », *Giorn. degli Economisti e Annali Econ.*, 1965, nn. 3-4, p. 179; M. GRILLENZONI, « L'utilizzazione dei modelli statistici nella pratica estimativa », *Gen. Rur.*, XXXI (1968), n. 4, p. 290.

⁵ Tutti i dati furono raccolti nel 1983-84 per l'Istituto di Estimo rurale e Contabilità dalla dott.ssa M. T. ARCIERI, che qui ringrazio per la collaborazione.

⁶ Oltre gli studi già citati, GRILLENZONI (1982) [9], e (1969) [8], PANATTONI (1969) [17], MILANO (1968) [14].

Tab. 1
ELEMENTI RILEVATI PER OGNI FONDO

	NUME	Num. d'ordine	Min.	Max.	Media	D.S.	Deviaz.	Curtosi	Fonte
SUPE	Superficie	ha	0,053 0,036	907 909	449,3 455,0	258,2	-0,21	2,107	R
DIST	Distanza	km	0,1 0	7,1 7,0	1,575 1,470	2,40	0,820	2,884	C
ASCI	Ascissa cartograf.	hm	45 5	202 202	119,1 118,6	29,39	-1,90	11,13	C
ORDI	Ordin. cartograf.	hm	26 26	196 196	119,0 118,5	23,01	-2,85	19,58	C
ALTE	Alt. s.l.m.	m	5,1 5	800,1 800	144,6 144,5	141,9	1,90	6,34	C
PREZ	Valore dich.	.000 L	303 303	82.852 136.346	8.483 7.936	12.133	3,52	16,83	R
PREZ 1	Prezzo	.000 L/ha	2.991	20.000	12.215	5.774	-0,437	2,33	
INFLA	Andam. prezzi	L/L _{an}	1,0	1,438	1,177	0,181	-0,024	0,0005	V
VAM	Val. agr. medi	.000 L/ha	1.800 1.800	15.000 15.000	9.460 9.432	4.070	-0,593	2,88	V
VUR	Val. uff. reg.	id.	2.400 0	22.608 22.608	13.185 13.073	5.800	-0,583	2,72	V
RDV	Redd. dom. 39	Lite	150 0	1.600 1.600	1.043 1.026	505,0	-0,982	3,44	V
RAV	Redd. agr. 39	id.	42 0	360 360	208,6 196,7	122,8	-0,366	2,61	V
RDN	Redd. dom. 79	.000 L	50 0	330 330	196,7 188,4	100,0	-0,185	2,37	V
RDA	Redd. agr. 79	id.	16 0	120 120	78,69 76,76	35,6	-1,33	5,34	V

R = rogito; C = cartografia; V = uffici pubblici e atti ufficiali

delle regioni agrarie 13 o 19 (legge 22-X-1971 n. 865) riferiti all'anno della compravendita; similmente, VUR sono i valori di riferimento ufficialmente adottati nei tre anni dall'*Ufficio del Registro* (D.P.R. 26-X-1972 n. 634); RDV, RAV, RDN, RAN sono rispettivamente le tariffe di reddito dominicale e agrario del 1939 e del 1979.

I 'fattori' e i loro livelli di definizione sono riportati nella tab. 2.

TAB. 2
DEFINIZIONE DEI 'FATTORI' E LORO CODIFICAZIONE

-
- 1) ANNO [R] - Anno del contratto:
1979 = 79; 1980 = 80; 1981 = 81
 - 2) QUAL [R] - Qualità catastale:
Vigneto = 1; Seminativo = 2; Sem. arb. = 3; Sem. irr. = 4; Sem. irr. arb. = 5;
Uliveto = 6
 - 3) CLAS [R] - Classe catastale:
I = 1; II = 2; III = 3
 - 4) ATTI [I] - Attitudine:
Agricola = 1; Urbana = 2; Mista = 3
 - 5) APPE [I] - Appetibilità:
Alta = 1; Media = 2; Bassa = 3
 - 6) COMU [R] - Zona amministrativa (ex comuni):
Nicastro = 1; Sambiasi = 2; Sant'Eufemia = 3
 - * 7) ALTI [I] - Zona altimetrico-economica:
Montagna = 1; Collina = 2; Pianura = 3
 - ** 8) FABB [R] - Presenza fabbricati:
Si = 1; No = 2
 - ** 9) VEND [R] - Categoria venditore:
Agrario = 1; Contad. propr. = 2; Non agrario = 3
 - ** 10) ACQU [R] - Categoria acquirente:
Agrario = 1; Contad. propr. = 2; Non agrario = 3; Emigrante = 4; Contad. non propr. = 5
 - ** 11) PAGA [R] - Forma di pagamento:
Contanti = 1; Differito = 2

* Significativo con P = .05

** Significativo con P = .01

Le lettere tra [] richiamano la fonte (v. tab. 1)

3. Analisi preliminare

Nella tab. 1 figurano il *minimo*, il *massimo*, la *media*, la *deviazione standard*, l'*asimmetria* e la *curtosi* di tutte le variabili (i valori della prima riga si riferiscono ai 347 casi completi, quelli della seconda a tutto il campione).

Si può notare facilmente come molte di esse siano asimmetriche (generalmente a sinistra), anche fortemente, e quasi tutte platicurtiche o leptocurtiche. Il *test* con il 'chi quadrato' porta a riconoscere che solamente una (il VAM) può essere considerata di distribuzione normale ($P=0.05$).

Circa i 'fattori', sempre nella tab. 2 e con l'usuale terminologia, è stato indicato se l'analisi della varianza risulti o meno significativa.

Tralasciando, come detto nella premessa, ogni altra considerazione circa le distribuzioni della superficie, dei prezzi, dei valori predeterminati, ecc., che appaiono conformi a quelle attese, e pure le molte suggerite dalla tabulazione delle variabili per livello dei 'fattori', o dall'esame della matrice di correlazione⁷, si osserva però che:

1. le *distanze* (come pure le *quote*) devono essere ridotte tutte di 100 m, essendo state alterate per necessità di calcolo; la distribuzione è tale da confermare la prevalenza di destinazione urbana, giacché circa la metà dei fondi sta in un raggio di meno di 300 m dal centro degli abitati principali;
2. le distribuzioni dell'*ascissa*, dell'*ordinata*, e dell'*altitudine* concordano con il maggior numero di contrattazioni avvenute in prossimità di Nicastro (56% del totale);
3. molte *correlazioni parziali* risultano elevate tra alcuni gruppi di variabili, come quelle catastali, i VAM, i valori del Registro, o quelle legate a evidenti caratteri geografici, ad es. ordinata e altitudine;
4. la più alta correlazione parziale del *valore* si ha con la superficie, mentre il *prezzo* unitario appare più sensibile all'evoluzione temporale e, in subordine, alla superficie;
5. sempre alte sono le correlazioni tra *prezzo* e valori fiscali, in particolare con i VUR.

⁷ L'analisi dei coefficienti di correlazione parziale, sempre utile come preliminare, è stata proposta da CHUDLEIGH [4] come tecnica professionale di *routine* anche da sola.

4. L'analisi delle componenti principali

Si è già detto (v. nota 3) che l'analisi delle componenti principali è stata proposta anche come tecnica ausiliaria per la riduzione della multicollinearità⁸; ma forse essa, come l'analisi fattoriale, riveste piuttosto una funzione eminentemente classificatoria (cfr. CONTINI (1973) [5], pp. 279-80), e consente di rilevare l'esistenza di associazione tra variabili di cui si ignorano le interconnessioni, come pure di farsi un'idea di quante fonti indipendenti di variazione siano 'realmente' in gioco nelle variabili analizzate (MADDALA (1981) [11], pp. 193-94).

È stato anche osservato che può talora apparire alquanto incerto il significato da attribuirsi in concreto alle componenti stesse (per tutti, v. MADDALA, *loc. cit.*), o che spesso risulta difficile caratterizzare estimativamente la corrispondenza tra le componenti e le variabili originarie, almeno nel senso indicato alla nota 2; ma è vero che alle tecniche di questo tipo viene riconosciuta una particolare utilità proprio nelle applicazioni microeconomiche e nelle indagini di mercato (cfr. CONTINI [5], p. 280).

D'altronde, l'uso delle componenti principali nella stima dei fattori da introdurre nelle regressioni è ormai tecnica corrente (v. ANANIA [1]), proprio allo scopo di sostituire le variabili originarie con un minor numero di nuove variabili che riuniscano in sé una quota 'sufficiente' della informazione complessivamente contenuta nei dati, pur sussistendo disparità di opinioni sui criteri con cui operare la selezione: generalmente si fa riferimento alla varianza residua globale (cfr. GRECO (1983) [6], p. 49).

Secondo GRECO (*loc. cit.*), i principi raccomandabili e che occorrerebbe tenere presenti durante il procedimento sono i seguenti, dai più generali a quelli specificamente riferiti al metodo delle componenti principali:

1. La descrizione della caratteristica complessa oggetto dello studio sarà tanto più accurata quanto più numerose sono le rilevazioni;
2. La scelta delle variabili, quando si badi alla loro capacità di essere rappresentative, può essere fatta solo in grazia delle conoscenze specifiche della disciplina che indaga il fenomeno⁹;
3. Le componenti (o nuovi fattori) possono essere interpretate come espressione di caratteristiche generali degli individui analizzati che non si manifestano come tali, ma che più o meno direttamente influenzano i valori delle variabili originarie;

⁸ Per una specifica applicazione estimativa v. MORTON (1977) [16].

⁹ Risulta un'altra volta confermata l'importanza degli apporti personali del ricercatore, che deve scegliere con accorto giudizio le variabili da porre per prime nel modello, e controllare passo a passo se l'equazione di regressione rispecchi i corrispondenti presupposti economici (TRIPPI (1974) [20], p. 566).

4. È lecito sintetizzare le informazioni in un minor numero di nuove variabili solo se queste riescano da sole a 'spiegare' con sufficiente approssimazione *ciascuna* delle variabili (originarie) che sia ritenuta basilare per la descrizione del fenomeno studiato.

Volendo poi tentare un'interpretazione economica ed estimativa di questi 'mostri' matematicamente evocati (rammento che ogni componente — per mantenere il linguaggio figurato — è 'montata' prelevando parti delle variabili originarie), mi pare che si possa fare riferimento a precedenti assai diffusi nella pratica: si pensi a concetti quali la 'panoramicità' di un fondo, o l' 'eleganza' di un quartiere cittadino, che appaiono come un insieme, un incastro, di altre singolari caratteristiche in vario modo 'dosate' dal mercato (o dallo stimatore), e con quella certa etichetta contrassegnate.

5. *L'equazione di regressione*

La regressione multipla è considerata da SMITH ([19], p. 248), che riporta l'esperienza dell'*U.S. Fish and Wildlife Service*, uno strumento « in grado di saggiare obiettivamente un mercato e di determinare quali variabili (fattori) agiscono sui prezzi di scambio, e quale peso debba essere assegnato a ognuna ».

Il procedimento, idealmente, consisterebbe nello scegliere le variabili che meglio spiegano il fenomeno, con alta significatività e con il minimo errore (statistico) possibile. Il perito dovrebbe però sempre potere spiegare *logicamente* la scelta delle variabili da lui impiegate nella regressione: il valore stimato (di *predizione*) così ottenuto potrà allora essere definito come il più probabile, nelle circostanze date.

Molte però sono le difficoltà che si incontrano nelle applicazioni, sia per le incertezze nella definizione e nella misura delle variabili, sia per la presenza della multicollinearità¹⁰. È inoltre piuttosto raro il caso che risulti soddisfatta la condizione di multinormalità, cosicché vengono a mancare i presupposti per le normali procedure di inferenza statistica. Ciononostante, si è generalmente disposti ad accettare qualche rischio in tal senso, in cambio dei vantaggi che un'equazione 'standard' può offrire in alcuni casi di stima, tanto più che non risulta difficile ai periti 'aggiustare' i risultati secondo opportunità, ossia rendere aderente al singolo caso il valore 'normale', 'ordinario' così ottenuto, come si fa abitualmente con le aggiunte e detrazioni, i comodi e gli scomodi.

¹⁰ Secondo MADDALA ([11], p. 183), la questione da porsi circa la multicollinearità, in questo tipo di problemi, non è tanto se essa sia o meno presente, ma 'quanto' lo sia e disturbi la risoluzione del nostro problema.

6. Il metodo seguito

Come già detto, la tecnica di analisi multivariata qui seguita è quella delle *componenti principali*, come comunemente adottata (cfr. ANANIA [1]), e orientata alla ricerca del modello estimativo; ossia, più concretamente, alla costruzione dell'equazione di regressione (MORTON [16]).

Essa può essere così schematizzata:

1. i fattori iniziali vengono individuati attraverso l'analisi delle componenti principali¹¹;
2. una parte di essi, a partire dal primo e fino a un certo livello scelto a seconda delle circostanze, viene poi ruotata con il metodo « *varimax* », in modo da facilitare la successiva scelta delle variabili;
3. ognuna delle variabili originarie viene assegnata a uno solo dei fattori ruotati, e precisamente a quello con il quale è massima la correlazione;
4. viene infine costruita l'equazione di regressione, con l'avvertenza che una sola variabile per gruppo (cioè per fattore) viene utilizzata: quando ne è entrata una, tutte le altre del gruppo vengono escluse¹²;
5. si prosegue (procedura *stepwise*) sino a quando vi siano variabili utilizzabili, e la regressione su di esse sia significativamente migliore¹³ rispetto a quella determinata nel passo precedente del calcolo (generalmente al livello del 5%).

Circa la rotazione, il numero dei fattori da ruotare è stabilito o sulla base degli autovalori, ad es. fino a quando sono superiori all'unità (cfr. ANANIA [1]), oppure controllando la connessione con le variabili originarie, in modo da non abbandonarne alcuna considerata valida, ed eliminando quelli con dei valori praticamente irrilevanti (KENDALL (1980) [10], p. 101). Si rammenta che i fattori ruotati sono a loro volta combinazioni lineari dei fattori non ruotati, ottenuti in modo da rendere massima la somma delle varianze associate alle variabili per *ogni colonna* della matrice dei fattori; ciò ha l'effetto di generare *più alte* e *più basse* correlazioni nelle colonne della matrice dei dati, e quindi rende più facile la separazione delle variabili.

¹¹ La prima componente è quella combinazione lineare delle variabili originarie per la quale risulta massima la quota assorbita di variabilità della matrice dei dati; il secondo fattore è determinato similmente, a partire dalla variabilità residua; e così via per le successive. Tutte le componenti sono tra loro ortogonali per costruzione (cfr. nota 3).

¹² Questa procedura riduce sistematicamente la probabilità che variabili altamente correlate entrino insieme nell'equazione di regressione.

¹³ Questa condizione viene comunemente espressa richiedendo che le variabili presentino tutte un F significativo (secondo FISCHER, 1924), con $P=0.05$.

Quando l'indagine si trova ancora in una fase iniziale, di 'esplorazione', magari non interessa la ricerca del legame funzionale (equazione di regressione) e ci si limita allora a identificare le componenti principali e a constatare le correlazioni (parziali) esistenti tra le variabili; l'analisi può allora estendersi a *tutte* le grandezze in gioco, comprendendo la variabile *dipendente*.

7. I risultati

Tralasciando, per brevità, l'esame delle 15 componenti principali e le loro correlazioni con le variabili, nella tab. 3 si mostra come le 13 variabili indipendenti ritenute vengano a ripartirsi — secondo le regole sopra indicate (ossia della massima correlazione assoluta, il cui valore è riportato accanto) — tra le prime tre componenti ruotate, le quali da sole assorbono il 72,6% della variabilità totale¹⁴.

Tab. 3
VARIABILI ASSEGNATE A OGNUNO DEI FATTORI RUOTATI

<i>I fattore</i>			
SUPE -232	VAM -872	RDV -976	RAV -.968
RDN -988	RAV -.1	VUR -.825	
<i>II fattore</i>			
INFLA -.944	NUME .943	DIST -.271	
<i>III fattore</i>			
ALTE -.920	ORDI -.907	ASCI -.433	

Il primo fattore (la prima componente ruotata) appare legato al gruppo dei valori 'predeterminati', e ciò non dovrebbe stupire quando si consideri da un lato l'ampio uso che di tali valori si fa nella pratica, dall'altro il significato di parametro tecnico che essi vengono ad assumere nella fattispecie¹⁵; anche la superficie è qui compresa, pur con una correlazione piuttosto bassa, ma che è del medesimo ordine di grandezza di

¹⁴ Le variabilità associate a ognuna delle prime tre componenti sono infatti 41,9%, 15,9% e 14,8%.

¹⁵ Si rammenti inoltre che i valori sono quelli *dichiarati*.

quella che appare nella matrice di correlazione tra di essa e i 'valori predeterminati'¹⁶.

Il secondo fattore è temporale, e riflette appunto l'incremento dei valori di vendita dal 1979 al 1981; risulta a prima vista strano il legame con la distanza, che però già si manifestava nella matrice di correlazione dei dati, e che potrebbe essere spiegato con l'ipotesi che le compravendite siano iniziate dai terreni più prossimi ai centri.

Il terzo fattore, infine, è nettamente legato a caratteristiche geografiche, e si adegua all'aumento di quota a mano a mano che si procede verso Nord.

Sulla base dei fattori così ottenuti, e seguendo la procedura sopra riportata (§ 6), si è poi costruita l'equazione di regressione, come riportata nella tab. 4 (i valori sono espressi in migliaia di lire).

TAB. 4
REGRESSIONE STEP-WISE 'FATTORIALIZZATA'

Y (variabile dipendente): PREZ (prezzo)			
*** Coefficienti della regressione ***			
	Stima	S.E.	t
Costante	10801	2540	4.25
SUPE (superficie)	5341	213	25.03
INFLA (inflazione)	-6046	2140	-2.83

*** Analisi della varianza ***			
	GL	SQ	MQ
REGRESS.	2	3.260E 10	1.630E 10
RESIDUI	344	1.789E 10	5.201E 7
TOTALE	346	5.049E 10	1.459E 8

Erre quadro (varianza spiegata) 64,4%

S.E.: errore standard; GL: gradi di libertà; SQ: somma dei quadrati; MQ: media dei quadrati.

Per confronto, si riportano nelle tabb. 5 e 6 altri due modelli di regressione, e precisamente:

- a) quella ottenuta attraverso una *stepwise* usuale, ossia senza scegliere le variabili ognuna a partire da un solo fattore;
- b) quella che conserva tutte e 13 le variabili.

¹⁶ La correlazione della superficie è invece di 0,798 nei confronti del *valore*.

TAB. 5
STEP-WISE USUALE

Y (variabile dipendente): PREZ
*** Coefficienti della regressione ***

	Stima	S.E.	t
Costante	-4051	930	-4.35
SUPE (superficie)	5688	197	28.82
VUR (Uff. Registro)	0.5647	0.0613	9.21

*** Analisi della varianza ***

	GL	SQ	MQ	F = 420
REGRESS.	2	3.581E 10	1.791E 10	
RESIDUI	344	1.468E 10	4.268E 7	
TOTALE	346	5.049E 10	1.459E 8	

Erre quadro (varianza spiegata) 70.8%

Per i simboli v. tabella precedente.

TAB. 6
REGRESSIONE CON TUTTE LE VARIABILI

Y (variabile dipendente): PREZ
*** Coefficienti della regressione ***

	Valore	S.E.	t
Costante	17214	8477	2.03
NUME	-9.19	3.65	-2.52
SUPE	5663	199	28.45
DIST	-61	153	-0.40
ASCI	3.2	12.4	0.26
ORDI	-25.5	27.5	-0.93
ALTE	-0.72	4.52	-0.16
VAM	0.149	0.259	0.58
VUR	0.348	0.220	1.58
RDV	4.58	3.53	1.30
RAV	7.1	18.2	0.39
RDN	-28.5	25.3	-1.13
RAN	11.3	60.2	0.19
INFLA	-12164	5417	-2.25

*** Analisi della varianza ***

	GL	SQ	MQ	F = 66
REGRESS.	13	3.642E 10	2.802E 9	
RESIDUI	333	1.407E 10	4.226E 7	
TOTALE	346	5.049E 10	1.459E 8	

Erre quadro (varianza spiegata) 71.0%

Per i simboli, v. tabella 4.

8. Conclusioni

Nei limiti di impostazione della presente memoria, e con le ulteriori avvertenze che

- a) sulla capacità rappresentativa del modello certamente influiscono gli aspetti di eterogeneità del campione (i 'fattori' della tab. 2) che sono stati qui ignorati per semplicità, e i cui effetti dovrebbero apparire anche dall'analisi dei residui che si spera di poter effettuare quanto prima;
- b) migliori risultati quindi si dovrebbero ottenere con la costruzione di più equazioni o con l'introduzione di variabili *dummy* che tengano conto della presenza di *fabbricati*, della forma di *pagamento*, della categoria di *acquirenti e venditori*, che sul mercato di Lamezia Terme e nel periodo 1979-81 hanno evidenziato un effetto statisticamente rilevante;

si possono tuttavia trarre le seguenti conclusioni:

1. Viene confermata la capacità della tecnica delle componenti principali di mostrare le associazioni tra le variabili e il loro legame con 'fattori' dai quali in qualche modo dipendono; fattori che spesso non si manifestano come tali e in modo esplicito, ma egualmente potrebbero avere un significato economico, perché appunto stanno dietro alle variabili studiate, e che nel caso in esame sono i 'valori pre-determinati', il 'tempo' e lo 'spazio'.
2. Ne segue che può essere facilitata e abbreviata la ricerca del 'migliore' modello di regressione, giacché, conoscendo dall'inizio quanti siano i 'fattori' veramente indipendenti che spiegano il fenomeno (quelli almeno evidenziabili dal campione studiato), appare inutile aggiungere variabili in numero superiore, che renderebbero anzi meno affidabile il modello stesso — quando se ne voglia estendere l'utilizzazione a casi analoghi¹⁷ — per l'aumentare della multicollinearità, e pertanto con gli usuali effetti che questa comporta: inversione dei segni di alcune variabili (rispetto alle attese del perito); forti oscillazioni dei coefficienti dell'equazione all'entrata e all'uscita di ognuna delle variabili nella *stepwise*.
3. Anche se nel caso particolare, la 'fattorializzazione' delle variabili non esplica forse tutta la sua potenzialità, a motivo del già piccolo numero di termini dell'equazione ottenuta con la *stepwise* usuale, si può notare tuttavia:

¹⁷ Naturalmente questo aspetto non avrebbe rilevanza se si cercasse solamente di ottenere la massima capacità 'previsiva' dell'equazione per la situazione in esame; in tale ipotesi diverrebbe anzi preferibile l'equazione con il massimo *erre quadro*, senza preoccupazioni per il livello di significatività (F o t).

— che il segno della costante diviene positivo, quindi più conforme alla realtà del mercato;

— che viene forzato nel modello l'andamento temporale dei prezzi (inflazione), al posto di un parametro (VUR) che si lo rappresentava, ma non esplicitamente e duplicando parzialmente la variabile SUPE.

Riprendendo infine quanto accennato nella premessa intorno alla capacità rappresentativa dei modelli ottenuti mediante procedimenti statistici e all'interesse che i medesimi possono assumere come ausilio per il perito, almeno in alcune delle situazioni operative più complesse e ripetibili, si aggiungono le seguenti considerazioni.

- a) Una delle critiche più frequenti e giustificate all'impiego della regressione in campo estimativo è stata quella che difficilmente e raramente le equazioni così calcolate sono utilizzabili al di fuori dello specifico ambito (beni e mercato) originario, risultando in conseguenza di assai limitato interesse applicativo.

La or ora vista costruibilità di modelli semplificati, e perciò più affidabili, e soprattutto il fatto che alle originarie variabili indipendenti, nettamente individuate e specifiche per ogni caso in esame, vengano a sostituirsi i 'fattori' derivati dalle componenti principali, porta a credere che il segnalato inconveniente non abbia più a manifestarsi nella maggioranza dei casi, dato che tali 'fattori' alle variabili sono legati, ma risultano al contempo di contenuto più generico e pertanto meno dipendenti dalla particolare e contingente situazione locale.

È infatti ben diverso, ad es., applicare altrove un modello avente come parametri obbligati certi valori agricoli medi (VAM) e un coefficiente di adeguamento monetario prefissato nel suo valore, o invece sapere che in una situazione simile dovremmo riuscire a spiegare una buona parte, o una gran parte (70-80%), del prezzo di una data superficie di terreno mediante due o tre più generici parametri, il primo appartenente al gruppo dei 'valori predeterminati' (ma che potremo scegliere in questo gruppo secondo opportunità), il secondo che rappresenti l'andamento dei prezzi nel tempo, l'ultimo di natura geografica o spaziale.

- b) Un'altra questione che è stata sollevata è quella della effettiva applicabilità di queste tecniche nella pratica professionale.

In effetti, al momento, sarebbe irrealistico pensare che le procedure di calcolo occorrenti per l'analisi multivariata siano alla portata di un perito estimatore o di un normale studio professionale, giacché sono coinvolte risorse relativamente potenti e sofisticate: così nel caso in esame, sono stati utilizzati un linguaggio statistico specializzato, il GENSTAT, e un elaboratore *Digital VAX*.

Il settore dei 'piccoli' elaboratori è però in una tale evoluzione, e i programmi statistici esistenti sono già così ampi e di facile im-

piego, che non è illusorio ritenere che entro due-tre anni le procedure di analisi multivariata saranno disponibili sui nuovi *Personal Computers*, e accessibili così a ogni professionista.

Spero che questo mio convincimento trovi il consenso di tutti i cultori della disciplina, e che ad Essi riesca gradito quanto sopra esposto.

RIFERIMENTI BIBLIOGRAFICI

- [1] ANANIA G., « Differenziazioni aziendali e modelli di classificazione: i risultati di un'applicazione di analisi fattoriale », Riv. Econ. Agr., XXXVI (1981), No. 3, pp. 455-479 (25).
- [2] GAGGIATI P. et al., « Un modello di inferenza statistica nella stima dei valori fondiari », Gen. Rur., No. 9 (1982), pp. 13-18 (6).
- [3] GAGGIATI P. e GALLERANI V., « L'analisi dei prezzi di mercato dei beni fondiari in Emilia-Romagna », Ce.S.E.T. Notizie-Aestimum, Nn. 11-12 (1984), pp. 22-35 (14).
- [4] CHUDLEIGH W. H., « The Application of Correlation Matrix Analysis to Real Estate Appraisal », The Appraisal J., (1979), pp. 523-30 (8).
- [5] CONTINI B., *Introduzione alla econometria*, Bologna, Il Mulino, 1973.
- [6] GRECO L., « Sulle proprietà delle componenti principali come metodo per la sintesi delle informazioni ». Statistica, XLIII (1983), No. 1, pp. 49-66 (18).
- [7] GRILLENZONI M., « Introduzione all'analisi dei prezzi di mercato dei beni fondiari », Estr. da: Riv. del Catasto e dei SS.TT.EE. n.s., XXIII (1968), Nn. 5-6, pp. 3-8 (6).
- [8] — « Analisi dei valori fondiari in periodo breve », Estr. da: Riv. del Catasto e dei SS.TT.EE., n.s., XXIV (1969), Nn. 1-2, pp. 3-17 (15).
- [9] — *Fattori che influenzano la proprietà, ecc.* Commissione della Comunità Europea, Bruxelles, 1982, « Inform. Agricola, No. 81 ».
- [10] KENDALL M., *Multivariate Analysis*, 2ª ed. C. Griffin & C., London e High Wycombe, 1980.
- [11] MADDALA G. S., *Econometrics*, Internat. Student Ed. (McGraw-Hill), 1981.
- [12] MALACARNE F., *Lineamenti di teoria del giudizio di stima*, Bologna, Edagricole, 1977.
- [13] MERLO M., « Alcune osservazioni sui contenuti e i metodi dell'estimo ecc. », in: *I contenuti dell'estimo rurale*, a cura di M. Grillenzoni e S. C. Misseri, Ce.S.E.T. (Firenze), 1985, « Quaderni, n. 2 ».
- [14] MILANO G., « Analisi della regressione nella valutazione dei fondi rustici », in: *Annali della Facoltà di Agraria di Bari* (1968), Bari, 1969, pp. 405-423 (19).

Le procedure di calcolo e i programmi (su VAX sotto VMS) sono stati sviluppati dal geom. P. Peruzzi, tecnico coadiutore del Dipartimento Economico-estimativo Agrario e Forestale dell'Università degli Studi di Firenze.

- [15] MISSEI S. C., « I contenuti dell'estimo rurale etc. », in: *I contenuti dell'estimo rurale*, a cura di M. Grillenzoni e S. C. Misseri, Ce.S.E.T. (Firenze), 1985, « Quaderni, n. 2 ».
- [16] MORTON T. G., « Factor Analysis, Multicollinearity, and Regression Appraisal Models », *The Appraisal J.*, (1977), pp. 578-88 (11).
- [17] PANATTONI A., « Una indagine quantitativa sul mercato fondiario nelle provincie di Pisa e Livorno », *Riv. Econ. Agr.*, XXIII (1968), No. 6, pp. 7-34 (28).
- [18] SIMONOTTI M., « Teoria estimativa e statistica », *Piano*, IV (1979), No. 9, pp. 71-88 (18).
- [19] SMITH D. W., « An Appraiser looks at Multiple Regression », *The Appraiser J.*, (1979), pp. 248-52 (5).
- [20] TRIPPI R. R., « A Comparaison of Linear and Nonlinear Models of Residential Real Property Value », *The Appraisal J.*, (1974), pp. 565-71 (7).