

5. Alcuni problemi di analisi econometrica su panel data

di Bruno Contini, Silvio Cuneo e Riccardo Revelli

1. Nell'analisi econometrica su panel-data, alla grande ricchezza di dati si contrappone sovente una notevole dose di problematicità nella stima dei modelli considerati appropriati. In questo lavoro verranno posti alcuni interrogativi e presentate esperienze empiriche che li sorreggono e li esemplificano. Per contro – tutto sommato – verranno fornite poche risposte a problemi che, a nostro avviso, è opportuno considerare del tutto aperti sotto il profilo metodologico.

In particolare ci soffermeremo su due tipi di problemi, sovente adombrati in letteratura, ma raramente affrontati con l'enfasi che – a nostro parere – meriterebbero.

2. Il primo ha a che fare con i vincoli posti dalle strutture di generazione dei dati che si osservano: nel caso di dati di impresa si tratta – sovente – di strutture contabili. Consideriamo un modello standard in cui si voglia indagare sui determinanti di un indicatore di performance aziendale (y_{it}); le variabili esplicative (x_{it}) sono indicatori di input, di prezzi relativi dei fattori, di caratteristiche interne di ciascuna impresa:

$$[1] \quad y_{it} = \alpha_i + \beta x_{it} + u_{it} \quad \begin{array}{l} i = 1, 2, \dots I \text{ (imprese)} \\ t = 1, 2, \dots T \text{ (tempo)} \end{array}$$

Nel modello compaiono degli effetti individuali α_i specifici di impresa e invarianti nel tempo, che compendiano tutto quello che è caratteristico dell'impresa ma non è osservabile (ad esempio, l'abilità del singolo imprenditore). È noto che trascurare l'effetto degli α_i introduce distorsioni nelle stime dei coefficienti β ; è ovvio, altresì, che tenerne conto per mezzo di variabili dummy (una per ogni impresa) è del tutto inopportuno quando il numero di imprese osservate è elevato (cosa assai frequente in questo tipo di indagine).

Un metodo per risolvere il problema è quello di stimare operando le differenze prime o una trasformazione «within», in modo che gli effetti individuali α_i vengano eliminati, senza conseguenze sui parametri β :

$$[2] \quad \Delta y_{it} = \beta \Delta x_{it} + (u_{it} - u_{i, t-1})$$

Quando, come nel nostro caso, il numero di osservazioni temporali T è uguale a 2, il metodo delle differenze prime (FDE) e gli stimatori «within» sono del tutto equivalenti.

La natura delle variabili esplicative x_{it} fa sì però, che esse siano fortemente autocorrelate. Ad esempio: il costo del lavoro per l' i -ma impresa al tempo t non può differire molto da quello che la stessa impresa paga al tempo $t + 1$; lo stesso vale per il rapporto capitale-prodotto, per la quota di acquisti esterni su fatturato, etc. Lo schema generativo più semplice potrebbe essere, il seguente

$$[3] \quad x_{it} = \rho x_{i,t-1} + \varepsilon_{it} \quad \text{con } |\rho| < 1$$

Utilizzando [3] nella [1], e passando alle differenze prime per eliminare α_i si ottiene quanto segue:

$$[4] \quad \Delta y_{it} = \beta(\rho - 1) x_{i,t-1} + (u_{it} - u_{i,t-1} + \varepsilon_{it})$$

Secondo questa specificazione Δy_{it} dovrebbe essere regredito sui valori ritardati di x_{it} ; i coefficienti stimati sono $\beta(\rho - 1)$, tanto più vicini a zero quanto maggiore è il coefficiente di correlazione ρ nella [3].

Se, d'altra parte, si stima direttamente sulle differenze prime [2], i regressori Δx_{it} sono tanto più vicini ad essere variabili casuali *pure* (white noise) quanto maggiore è ρ (per $\rho \rightarrow 1$, $\Delta x_{it} \rightarrow \varepsilon_{it}$). Il che significa che utilizzando il modello [2], tutti i coefficienti sono biased verso zero, come in un modello di errori nelle variabili.

Noi faremo qui riferimento ad alcune esperienze realizzate su un panel di imprese provenienti dall'Indagine Annuale sul Valore Aggiunto delle Imprese Manifatturiere, osservate in tre soli periodi di tempo: 1973, 1977 e 1981. Tutte le imprese del campione sono di piccole dimensioni (tra 20 e 100 addetti) ed appartengono a otto settori industriali (class. NACE a 3 cifre):

	(numerosità del campione)
248 - piastrelle in ceramica	(198)
316 - utensileria e oggetti in metallo	(540)
322 - macchine utensili per la trasformazione dei metalli	(249)
431 - tessitura e filatura della lana	(256)
436 - maglieria	(433)
453 - abiti confezionati in serie	(554)
467 - mobilio in legno	(419)
483 - prodotti in plastica	(421)

Il modello oggetto di indagine era volto a misurare i determinanti di lungo periodo della produttività dell'industria italiana, ed in particolare l'impatto di alcuni mutamenti di carattere organizzativo colle-

gabili alla struttura di integrazione verticale delle imprese manifatturiere [Contini 1989].

La specificazione adottata per l'equazione della produttività – ampiamente discussa in Contini e altri [1989] – è la seguente:

$$[5] \quad \left(\frac{V}{L}\right)_{it} = \lambda \left(\frac{M}{L}\right)_{it} + b(1+\lambda) \left(\frac{W}{L}\right)_{it} + \beta \left(\frac{\Delta K}{L}\right)_{it} + \gamma \left(\frac{B}{L}\right)_{it} + \delta \left(\frac{H}{L}\right)_{it} + \eta S_{it} + u_{it}$$

con le variabili che assumono il seguente significato:

- V/L = valore aggiunto per addetto
- M/L = acquisti di materie prime e secondarie per addetto
- W/L = costo del lavoro per addetto (al lordo di contribuzioni)
- $\Delta K/L$ = investimenti per addetto nel quadriennio precedente
- B/L = rapporto tra operai e forza-lavoro totale
- H/L = ore lavorate per addetto
- S = fatturato totale (proxy per la scala di produzione)

tutte quelle in valore sono riportate a valori costanti 1977.

Tale specificazione deriva da un modello di determinazione di prezzo in regime non concorrenziale, basato sul mark-up sui costi variabili. In particolare è possibile identificare esattamente i seguenti parametri:

- λ = mark-up sui costi variabili
- b = quota di costo del lavoro variabile sul costo del lavoro totale.

Gli altri parametri del modello (β , γ , δ , η) non pongono particolari problemi di identificazione, essendo relativi a variabili che – in questo ambito – possono essere considerate esogene, esplicative di variazioni della produttività fisica, e non correlate con variazioni nella struttura del mercato (la quale viene invece colta dal mark-up λ).

Il nostro interesse era mirato a verificare l'esistenza di modifiche strutturali intervenute nei due quadrienni 1973-77 e 1977-81: da un lato il più elevato grado di apertura dei mercati mondiali, a cui dovrebbe corrispondere una riduzione del mark-up λ ; dall'altro la maggiore rigidità del mercato del lavoro (e quindi della struttura dei costi di impresa), i cui effetti dovrebbero manifestarsi attraverso una riduzione della quota b di costo del lavoro variabile.

Per questo motivo tutte le nostre stime sono state costruite tenendo separati i due quadrienni 1973-77 e 1977-81: il numero di osservazioni temporali è quindi uguale a 2, e lo scopo dell'esercizio è quello di confrontare le stime dei parametri ottenute nel primo quadriennio con quelle relative al secondo. Le stime sono disponibili per sei settori su otto. Come commenteremo più avanti, troviamo scarse verifiche per la prima ipotesi, molto robuste per la seconda.

Si consideri la tab. 1 che riporta le stime dei coefficienti nell'equazione [3] per ciascuno dei regressori che compaiono nell'equazione della produttività. In questo caso il ritardo tra x_t e x_{t-1} è di 4 anni – costante per tutte le imprese del panel. Nonostante la lunghezza del ritardo tra osservazioni contigue, tutti i coefficienti stimati sono vicini a 1 (non si riportano, per brevità, gli R², sempre superiori a 0,9), con l'unica eccezione per la variabile $\Delta K/L$, che riflette gli investimenti delle imprese.

TAB. 1. *Stima dei coefficienti di autocorrelazione ρ (OLS su equaz. [3])*

	(M/L)		(W/L)		(H/L)		(B/L)		(S)		($\Delta K/L$)
Industria	1973-1977	1977-1981	1973-1977	1977-1981	1973-1977	1977-1981	1973-1977	1977-1981	1973-1977	1977-1981	1973-1977
316	0,966	0,972	1,076	1,008	0,849	0,956	0,958	0,971	1,191	1,19	0,071
323	0,953	0,881	1,051	1,032	0,988	0,959	0,964	0,972	0,998	1,10	0,807
436	0,895	1,124	1,243	1,122	0,961	0,916	0,975	0,975	0,990	1,23	0,568
453	0,804	0,642	1,132	1,106	0,645	0,928	0,989	0,972	0,908	1,11	0,704
467	0,883	1,057	1,269	0,965	0,351	0,932	0,974	0,970	1,043	1,32	0,844

L'esistenza di effetti individuali è stata oggetto di verifica con il test di specificazione χ^2 proposto da Hausman-Taylor: tale test serve a verificare l'ipotesi che effetti individuali e regressori siano correlati. I valori del test – non riportati in questa sede – conducono ad un rifiuto sistematico dell'ipotesi nulla: si deve quindi concludere che effetti individuali e regressori siano correlati, e conviene pertanto adottare procedure di stima adeguate.

La tab. 2 pone a confronto i risultati ottenuti per due sotto-periodi 1973-77 e 1977-81, con tre procedure di stima diverse:

a) OLS (minimi quadrati ordinari) sulle due cross-sections 1973-77 e 1977-81;

b) OLS sulle due cross-sections di cui sopra, introducendo tra i regressori anche il valore ritardato (4 anni) della variabile dipendente (V/L);

c) FDE (stimatori «within», differenze prime 1977-73 e 1977-81).

Porre a confronto gli estimatori OLS (in cui si trascurano gli effetti individuali) e quelli FDE (teoricamente adeguati per eliminare l'effetto distorsivo degli effetti individuali) è del tutto naturale¹. Restano i

¹ Occorre ricordare, tuttavia, che – indipendentemente da qualsiasi altra considerazione – il metodo FDE comporta un peggioramento dell'efficienza delle stime. Infatti

dubbi derivanti dalla forte autocorrelazione riscontrata in tutti i regressori, e degli effetti distorsivi (verso zero) che possono inficiare le stime FDE. Questi effetti sono, invero, molto frequenti e vistosi nei parametri di (M/L) , (H/L) , $(\Delta K/L)$, meno frequenti nei parametri di (W/L) ; sembrano invece assenti per quanto riguarda i parametri di (B/L) e (S) .

Nella tab. 2 sono riportate anche stime OLS con $(V/L)_{-1}$ tra i regressori: l'ipotesi sottostante questa prova è che $(V/L)_{-1}$ possa costituire una buona «proxy» degli effetti individuali. A priori l'ipotesi sembra verosimile, nel senso che – come vari studi hanno rilevato² – si manifestano forti persistenze nei comportamenti delle imprese, in particolare tra quelle di piccola dimensione (come sono quelle oggetto della nostra indagine). Se il motivo della persistenza è riconducibile alla presenza di un imprenditore dotato di certe caratteristiche personali (positive o negative), qualche indicatore di performance di impresa potrebbe fare all'uopo.

I risultati mostrano che le stime che si ottengono presentano differenze abbastanza marginali rispetto a quelle ricavate con gli OLS nelle cross-sections 1973-77 e 1977-81. Le differenze sono invece notevoli con le stime FDE. Il fatto che l'introduzione di $(V/L)_{-1}$ non sposti i risultati verso quelli delle stime «within» potrebbe essere interpretato, a prima vista, come un'indicazione che $(V/L)_{-1}$ non sia una buona «proxy» per gli effetti individuali. Se, tuttavia, si tiene in considerazione quanto osservato in precedenza sul probabile bias verso zero delle stime FDE (per effetto della struttura di autocorrelazione di molti regressori) la conclusione diventa quanto meno dubbia. A nostro parere il problema resta del tutto aperto.

Prima di passare al secondo punto metodologico, conviene riassumere brevemente le indicazioni di carattere economico che emergono da queste stime:

1) non sembra esservi una riduzione del mark-up λ tra il 1977 e il 1981 così come avevamo ipotizzato. Utilizzando le stime OLS e OLS_L, in tre casi su sei il mark-up aumenta nel 1981. Considerando invece le stime FDE (con tutte le cautele del caso), non sembra esservi alcun mutamento sostanziale nella struttura del mercato;

2) sembra invece sistematicamente confermata (tab. 3) l'ipotesi

dalla [2], si ottiene

$$E(u_{it} - u_{it-1})^2 = 2 \sigma_u^2$$

nell'ipotesi che gli u_{it} non siano autocorrelati. In effetti la tab. 2 conferma che – in 5 casi su 6 – s_u^2 (OLS) è circa la metà di s_u^2 (FDE), e di conseguenza tutti i singoli coefficienti FDE hanno varianza maggiore di quelli OLS.

² In particolare, D. C. Mueller [1986], e – per l'Italia – B. Contini [1989].

Tab. 2. *Equazione della produttività [S]: diversi metodi di stima*

Metodo di stima	S _a	M/L	W/L	H/L	B/L	ΔK/L	S	(V/L) ₋₁
Industria 316								
1977								
— OLS	3,78	0,129	1,065	1,245*	-0,701 (no)	0,0014 (no)	0,0004*	
— OLS _L	3,71	0,120	0,963	1,075*	-1,066 (no)	0,0003 (no)	0,0003 (no)	0,135
— FDE	5,49	0,047*	0,915	-0,034*	1,215 (no)	-0,0049 (no)	0,0018	
1981								
— OLS	4,72	0,098	0,627	1,731*	3,483	0,1261*	0,0008	
— OLS _L	4,56	0,096	0,521	1,350*	1,550 (no)	0,0730*	0,0007	0,255
— FDE	5,09	0,091	0,597	-0,086 (no)	-2,399 (no)	-0,0097 (no)	0,0016	
Industria 322								
1977								
— OLS	3,53	0,066*	1,148	0,480 (no)	-0,0672 (no)	0,2530	0,0006*	
— OLS _L	3,51	0,070*	1,113	0,324 (no)	-0,2591 (no)	0,2428	0,0005**	0,061*
— FDE	6,93	0,000 (no)	0,680*	1,153 (no)	-4,789 (no)	-0,0668 (no)	0,0036	
1981								
— OLS	4,58	0,070*	0,465	3,968	1,192 (no)	0,1274*	0,0015	
— OLS _L	4,43	0,066*	0,3672*	3,171*	-0,206 (no)	0,0855*	0,0012	0,264
— FDE	4,94	-0,049 (no)	0,607	-0,0297 (no)	-9,976*	-0,0675**	0,0027	
Industria 436								
1977								
— OLS	2,61	0,008	1,002	0,223 (no)	0,434 (no)	0,1003*	0,0009	
— OLS _L	2,61	0,087	0,986	0,227 (no)	0,363 (no)	0,0988*	0,0009	0,027 (no)
— FDE	4,43	0,044*	0,881	-0,136 (no)	-5,681*	0,109*	0,0016	
1981								
— OLS	3,82	0,101	0,521	2,534	0,201 (no)	0,1903	0,0001	
— OLS _L	3,75	0,097	0,473	2,323	-7,81 (no)	0,1722	0,0010	0,177
— FDE	4,29	0,098	0,489	0,284 (no)	-4,253**	0,1870	0,0013	

Industria 453										
1977										
— OLS	3,05	0,164	1,007	1,027*	-2,287*	0,1071 (no)	0,0006			
— OLS _L	2,97	0,153	0,909	0,991*	-2,278*	0,0577 (no)	0,0006			0,126
— FDE	4,117	0,105	1,068	-0,247 (no)	-12,625	-0,0229 (no)	0,0006			
1981										
— OLS	5,28	0,221	0,376	4,564	-2,811*	0,2333**	0,0006*			
— OLS _L	5,14	0,216	0,332	4,293	-4,070*	0,1642 (no)	0,0005*			0,234
— FDE	5,49	0,143	0,452	1,943*	-15,449	0,360*	0,0009*			
Industria 467										
1977										
— OLS	3,28	0,044*	1,053	0,989*	-0,934 (no)	0,1169*	0,0009			
— OLS _L	3,28	0,044*	1,044	0,980*	-0,976 (no)	0,1158*	0,0009			0,014 (no)
— FDE	4,61	0,046*	1,135	-0,236**	-2,776 (no)	0,0789**	0,0012			
1981										
— OLS	5,05	0,140	0,492	3,654	-1,414 (no)	0,0025 (no)	0,0009			
— OLS _L	5,02	0,138	0,430	3,292	-1,878 (no)	0,0021 (no)	0,0008			0,1334*
— FDE	6,12	0,047*	0,309*	1,234**	-6,065**	0,0015 (no)	0,0013			
Industria 483										
1977										
— OLS	10,24	0,184	1,087	4,780*	-4,928 (no)	0,1530*	0,0004 (no)			
— OLS _L	10,09	0,177	1,047	4,612*	-5,899**	0,1193*	0,0003 (no)			0,168
— FDE	10,58	0,105	1,289	0,911 (no)	-6,743 (no)	0,2466*	0,0024			
1981										
— OLS	8,07	0,045*	0,639	5,360	-3,930*	0,2109*	0,0012			
— OLS _L	7,79	0,029**	0,532	4,384	-6,357*	0,1605*	0,0013			0,155
— FDE	11,63	0,097*	1,358	3,343*	2,709 (no)	-0,1910*	0,0008*			

Test «t» di Student:

— nessun segnalatore: $|t| > 4$ — *: $2 < |t| \leq 4$

— **: $1,65 < |t| \leq 2$ — no: $|t| < 1,6$

TAB. 3. *Stime di λ e b dalle equazioni di produttività (cfr. tab. 2)*

Industria	$\hat{\lambda}$				$\hat{b}$			
	1977		1981		1977		1981	
	OLS	FDE	OLS	FDE	OLS	FDE	OLS	FDE
316	0,129	0,047*	0,098	0,091	0,94	0,87	0,57	0,54
322	0,066*	0,0 (no)	0,070*	-0,049 (no)	1,07	0,68	0,43	0,60
436	0,088	0,044*	0,101	0,098	0,92	0,84	0,47	0,45
453	0,164	0,105	0,221	0,143	0,86	0,97	0,31	0,40
467	0,044*	0,046*	0,140	0,047*	1,00	1,08	0,43	0,30
483	0,184	0,105	0,045*	0,097*	0,92	1,17	0,61	1,24

che la quota di lavoro variabile b si sia sensibilmente ridotta tra il quadriennio 1973-77 e quello successivo 1977-81, a testimonianza di un irrigidimento della struttura dei costi di impresa. Le indicazioni al riguardo sono identiche e molto robuste, sia che si considerino le stime OLS che quelle FDE.

Per completezza – ed a conferma delle indicazioni suggerite dalla specificazione [4] – si riportano in tab. 4 gli stimatori di $\beta(\varrho-1)$ ottenuti regredendo le differenze prime di (V/L) sui valori ritardati di tutti i regressori. In quasi tutti i casi i coefficienti stimati non sono significativamente diversi da zero. Da tali coefficienti è comunque possibile ricavare stimatori di β dividendo ciascuno per $(\varrho-1)$, dove ϱ è quello stimato dalla [3] e riportato in tab. 1. La tab. 5 riporta gli stimatori così ottenuti dei coefficienti β : si tratta di numeri palesemente insensati da un punto di vista economico. Ciò fornisce conferma, ancorché solo intuitiva, della scarsissima attendibilità di stimatori ottenuti da una specificazione del tipo [4], quando ϱ è prossimo a 1.

3. Il secondo problema cui vorremmo accennare trova origine in alcuni recenti studi empirici su panel-data di impresa: un approccio che si rileva utile come strumento di analisi esplorativa, preliminare all'applicazione di modelli specificati in modo rigoroso, è quello di suddividere le osservazioni campionarie in gruppi, omogenei al loro interno rispetto ad alcune caratteristiche prefissate o latenti, molto differenziate tra i gruppi.

Le tecniche che danno luogo a questi raggruppamenti sono numerose: l'analisi discriminatoria, la cluster analysis, o anche – più semplicemente – la suddivisione del campione in quartili rispetto alla funzione di ripartizione di qualche indicatore campionario di performance.

Un esempio è quello del raggruppamento delle imprese a seconda della persistenza di utili lordi: all'interno di ogni settore produttivo le imprese sono state suddivise a seconda della loro posizione nella distribuzione di frequenza congiunta del rapporto tra utili lordi e valore aggiunto negli anni 1977 e 1981. Sia $u(t)$ tale rapporto. Si sono

Tab. 4. Stime OLS nel modello generalizzato [4]: $\Delta y_{it} = \gamma X_{i,t-1} + (u_{it} - u_{i,t-1} + \varepsilon_t)$

Industria	H/L		B/L		M/L		W/L		S		$\Delta K/L$		S_u	
	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81
316	0,283..	-0,567..	7,31*	7,31*	0,036..	-0,06*	-0,644	-0,756	-0,005..	0,094*	-0,001*	0,0005**	6,27	
322	-4,51*	2,75*	0,927..	13,67*	-0,026..	0,04..	-0,308..	-0,573*	0,105..	-0,044..	-0,001*	0,0002..	7,52	
436	-0,751..	-0,329..	3,61..	6,91**	0,013..	0,028..	-0,277..	-0,436*	0,157*	0,201	-0,001*	-0,00001..	5,16	
453	-0,133..	-2,82*	15,35*	-3,70..	-0,033..	-0,099*	-0,954	-0,461*	-0,103..	0,443*	0,001*	-0,0002..	5,50	
467	-0,145..	-0,75..	-0,36..	8,32**	-0,041..	0,085*	-0,405*	-0,725	0,091..	0,0002..	-0,001**	-0,0002..	5,42	

Tab. 5. $\hat{\beta} = \hat{\gamma} / (\hat{\rho} - 1)$: stimatori ricavati da [4] e da [3]

Industria	H/L		B/L		M/L		W/L		S		$\Delta K/L$	
	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81	1973-77	1977-81
316	-1,88	12,96	-176,7	-253,1	-1,085	2,365	-8,41	-93,46	-0,004	0,003	0,006	
322	374,4	-68,05	-25,9	-497,4	0,568	-0,332	-6,05	-18,05	1,66	0,0024	-0,544	
436	19,55	3,93	-146,6	-276,8	-0,124	0,228	-1,13	-3,59	0,073	-0,00005	-0,364	
453	0,375	39,24	-1475,6	133,5	0,169	0,278	-7,23	-4,34	-0,014	-0,004	0,347	
467	0,223	11,16	14,3	-277,4	0,352	1,48	-1,93	20,93	-0,014	-0,0008	-0,584	

riclassificate le imprese secondo il seguente schema a doppia entrata: $Q1$, $Q2$, $Q3$ e $Q4$ sono i quartili delle distribuzioni di u (1977) e u (1981). Le imprese nella cella ($Q1$, $Q2$) – ad esempio – sono quelle che cadono nel primo quartile di $f[u(1977)]$ e nel secondo quartile di $f[u(1981)]$.

		$f(u_{1981})$			
		$Q1$	$Q2$	$Q3$	$Q4$
$f(u_{1977})$	$Q1$	III			
	$Q2$		II		
	$Q3$				
	$Q4$				I

Poiché nello studio eravamo particolarmente interessati a questioni riguardanti la persistenza dei profitti delle imprese, abbiamo limitato la ricognizione a tre gruppi di imprese:

gruppo I · imprese appartenenti alla cella ($Q4$, $Q4$), denominate imprese «brave»;

gruppo II · imprese appartenenti alle celle intermedie ($Q2$, $Q2$), ($Q2$, $Q3$), ($Q3$, $Q2$) e ($Q3$, $Q3$), denominate «medie»;

gruppo III · imprese appartenenti alla cella ($Q1$, $Q1$), denominate imprese «somare».

Sono quindi escluse dallo studio le imprese appartenenti ai dieci blocchi fuori dalla diagonale (che ne include altri sei). Un'indagine preliminare aveva appurato che oltre il 70% delle imprese restavano così sotto osservazione contro un 37,5% teorico nell'ipotesi di indipendenza nelle graduatorie di u (1977) e u (1981). In sé questo risultato costituisce evidenza di un notevole grado di persistenza nei livelli di performance delle imprese.

La tab. 6 riassume, per il solo 1981, le medie campionarie osservate in ciascuno dei tre gruppi per una serie di indicatori economici di impresa. Le differenze più vistose si trovano nelle colonne dei «ricavi per addetto» e degli «acquisti per addetto»: l'appartenenza ai tre gruppi discrimina in modo eclatante i valori medi. Le imprese «brave» hanno ricavi per addetto e acquisti per addetto enormemente più elevati delle imprese «medie», e queste, a loro volta, mostrano valori assai maggiori di quelle «somare». Vi sono notevoli differenze anche in altre colonne: la prima (rapporto tra utili lordi e valore aggiunto) riflette semplicemente il criterio di costruzione dei gruppi, e quindi non può destare sorpresa. La seconda (prodotto per addetto) ne è quasi una conseguenza. Sono invece notevoli e riflettono la gerarchia

TAB. 6. Medie campionarie di alcuni indicatori micro-economici nei gruppi I, II, III

Industria	Gruppo	Frequenze	Utili lordi	Prodotto per addetto	Ricavi per addetto	Acquisti per addetto	Rapporto tra operai e addetti (%)	Investi- menti per addetto nel qua- driennio 1977-81
			Valore aggiunto					
248	I	12	0,56	34,1	82,3	39,9	0,82	16,3
	II	74	0,34	21,7	52,3	23,6	0,88	7,1
	III	41	0,04	12,3	23,9	9,0	0,90	3,1
316	I	74	0,54	29,8	85,7	42,8	0,81	10,3
	II	151	0,31	19,3	48,9	20,9	0,83	4,3
	III	52	0,03	15,4	37,7	15,8	0,85	2,9
322	I	24	0,54	34,1	70,9	28,4	0,79	13,1
	II	66	0,32	22,4	45,6	15,8	0,79	5,7
	III	25	0,04	18,0	34,5	11,7	0,81	3,1
431	I	36	0,55	31,4	125,8	56,6	0,81	10,6
	II	50	0,31	19,3	43,9	14,0	0,87	6,6
	III	9	0,11	12,8	21,3	5,4	0,85	2,8
436	I	46	0,54	23,5	92,5	43,7	0,82	5,6
	II	153	0,27	14,7	40,8	16,9	0,88	2,7
	III	60	-0,12	10,7	22,1	7,6	0,92	1,4
453	I	42	0,49	22,9	71,6	29,0	0,82	2,1
	II	124	0,20	12,1	29,8	12,9	0,91	1,3
	III	46	-0,18	9,8	18,2	5,7	0,93	0,8
467	I	41	0,51	24,9	74,6	35,6	0,82	6,4
	II	181	0,30	17,6	44,7	20,8	0,86	9,6
	III	50	-0,35	12,3	29,5	13,1	0,86	2,2
483	I	54	0,58	36,9	125,0	71,8	0,81	11,0
	II	125	0,32	21,4	59,7	29,7	0,82	6,5
	III	39	-0,10	14,8	44,2	21,1	0,98	3,9

dei gruppi le differenze nell'ultima colonna (investimenti) ed, entro certi limiti, anche nella penultima (rapporto tra operai e addetti).

Non si notano invece differenze di rilievo tra le medie di gruppo del costo del lavoro, come ci si attenderebbe. Differenze del tutto analoghe si notano tra gli stessi gruppi nel 1977 (non riportato in tabella per brevità).

In Contini [1989] questi dati sono stati oggetto di analisi dettagliata. In sintesi, le conclusioni che se ne possono trarre sono le seguenti:

i) le imprese «brave» operano su segmenti di mercato più qualificato, e spuntano prezzi medi assai superiori a quelli delle imprese «medie»; queste, a loro volta, fanno meglio di quelle «somare»;

ii) le imprese «brave» sono notevolmente diverse anche sotto il profilo organizzativo: in particolare sono maggiormente disintegrate verticalmente (acquistano dall'esterno molto di più di quanto non

facciano le altre), ed il livello di disintegrazione verticale sembra un indicatore inequivocabile di buona performance aziendale.

Il modello appropriato per «spiegare» differenze tra gruppi come quelli I, II e III (ma anche tra altri gruppi formati con criteri diversi) potrebbe scriversi nella seguente forma:

$$[6] \quad y_{igt} = \alpha'_g + \alpha''_i + \beta_g x_{igt} + \varepsilon_{igt}$$

$$\begin{cases} i = \text{imprese} \\ g = \text{gruppo} \\ t = \text{tempo} \end{cases}$$

i gruppi (g) sono a volte latenti, a volte frutto di una scelta a priori del ricercatore. Gli effetti α'_g sono specifici di gruppo; quelli α''_i sono effetti individuali. Ambedue sono invarianti nel tempo.

Non vi è motivo di ritenere che i coefficienti β_g siano gli stessi tra un gruppo e l'altro (ad esempio, è presumibile che i fattori determinanti la produttività tra le imprese «brave» siano diversi da quelli caratteristici delle imprese «somare»). Se li costringiamo ad essere uguali ci precludiamo qualsiasi possibilità di *capire* perché certe imprese riescono bene ed altre male. Allo stesso modo i comportamenti delle imprese razionate dal lato della domanda sono diversi da quelli delle imprese razionate dal lato dell'offerta.

In cosa consiste il problema? Molto semplicemente, nel fatto che – qualsiasi sia il criterio di raggruppamento – esso farà – in qualche modo – riferimento alle variabili y_{igt} e x_{igt} che compaiono nel modello.

Se il raggruppamento è «robusto», la variabilità di alcune x_{igt} (o combinazioni lineari di esse) *all'interno* dei gruppi sarà molto bassa; sarà invece elevata *tra* i gruppi.

La tab. 6 dimostra chiaramente la «robustezza» dei raggruppamenti. Un'analisi della varianza ne rafforza semplicemente le conclusioni. Una stima di una versione generalizzata della [5] *dentro* ai gruppi rischia quindi di dare luogo a risultati molto instabili per problemi di multicollinearità; una stima *tra* i gruppi potrebbe invece essere statisticamente significativa, ma inutile perché media i coefficienti di diversi gruppi, esattamente quello che volevamo evitare.

La tab. 7 riporta i risultati delle stime OLS relative alla solita equazione della produttività, effettuate «dentro» ai gruppi I, II e III, nelle cross-sections 1977 e 1981.

La significatività delle stime relative a tutti i regressori che maggiormente «discriminano» tra i gruppi – (M/L), (K/L), (S) – è quasi sempre nulla. Le cose vanno un po' meglio per il coefficiente di (W/L) che – come si è detto prima – non mostra particolare differen-

Tab. 7. Equazioni della produttività [5]: stima OLS all'interno di tre gruppi di imprese, riclassificate secondo il criterio della persistenza dei profitti nel 1977-81

Industria	H/L			B/L			M/L			W/L			$\Delta K/L$			S			S_w		
	1977	1981	1988	1977	1981	1988	1977	1981	1988	1977	1981	1988	1977	1981	1988	1977	1981	1988	1977	1981	1988
I	1,55 (no)	-0,283 (no)	-14,87*	-20,81*	0,051 (no)	0,070*	1,52	1,698	0,097**	-0,052	0,001*	0,0006 (no)	4,36	4,86							
II	0,45 (no)	0,247 (no)	17,80	19,9	0,016 (no)	0,024*	1,56	1,603	0,906*	0,142*	0,0 (no)	0,0001 (no)	1,20	1,40							
III	-0,739 (no)	0,162 (no)	-19,40	-11,67*	0,033 (no)	0,018 (no)	1,39	0,893	0,0007 (no)	-0,036 (no)	0,001*	0,0004 (no)	1,57	2,22							
I	-3,56 (no)	-0,032	-40,73	-35,01 (no)	-0,029 (no)	-0,055 (no)	1,11*	2,20*	0,410*	-0,102 (no)	0,0009 (no)	0,0001 (no)	2,38	8,81							
II	-0,410 (no)	1,483*	-19,75	-14,35	0,0086 (no)	-0,006 (no)	1,58	1,143	0,080 (no)	0,004 (no)	0,0003 (no)	0,018	1,23	1,45							
III	-1,831*	4,39*	-22,93	-11,99*	0,048 (no)	-0,006 (no)	1,45	0,584*	0,037 (no)	0,163 (no)	0,0024 (no)	0,004 (no)	1,15	2,17							
I	-0,877 (no)	-0,941 (no)	-14,76	-16,11**	0,112	0,096*	1,48	2,11	0,055 (no)	0,275*	0,0006**	-0,0002 (no)	2,39	5,1							
II	0,078 (no)	0,577 (no)	-15,42	-18,35	0,031*	0,037*	1,54	1,403	-0,037**	0,013 (no)	0,0001 (no)	0,0004*	0,98	1,4							
III	-0,147 (no)	2,45	-7,07**	-1,99 (no)	-0,091*	-0,011 (no)	1,079	0,134*	-0,053 (no)	-0,301*	0,0010*	0,002	1,34	1,4							
I	-1,99 (no)	-7,82**	-24,57*	-48,76	0,288	0,098 (no)	1,58*	1,648*	-0,314 (no)	0,021 (no)	-0,0002 (no)	0,0004 (no)	4,75	5,7							
II	-0,097	0,198 (no)	-11,125	15,71	0,025*	0,032*	1,237	1,273	-0,026 (no)	-0,023 (no)	0,0016 (no)	-0,0001 (no)	0,66	0,9							
III	1,106 (no)	3,52	-8,15*	-4,05 (no)	-0,121**	-0,176*	0,846	0,231*	0,349*	-0,486 (no)	0,0014*	0,0021	1,21	1,5							
I	-0,200 (no)	2,94 (no)	-18,13*	-17,11 (no)	0,0004 (no)	0,055 (no)	2,21	2,04*	0,012 (no)	0,263 (no)	-0,0001 (no)	-0,0001 (no)	2,37	5,7							
II	0,028 (no)	0,039 (no)	-17,59	-20,49	0,0377*	0,019**	1,48	1,49	0,007 (no)	0,0004 (no)	0,0001 (no)	0,0001 (no)	1,08	1,3							
III	-0,192 (no)	0,999 (no)	-16,74*	-1,81 (no)	-0,010 (no)	-0,007 (no)	1,38	0,413*	-0,004 (no)	0,061 (no)	0,00 (no)	0,001	1,61	2,0							
I	-2,55 (no)	1,97 (no)	-29,05*	-37,73**	0,023 (no)	-0,035 (no)	2,099	2,049*	0,034 (no)	0,417*	0,00002 (no)	0,0018	4,94	11,							
II	-0,177 (no)	1,33**	-22,28	-26,87	0,011 (no)	0,021**	1,63	1,55	-0,019 (no)	0,006 (no)	0,0001 (no)	0,0001 (no)	1,75	2,6							
III	-1,17 (no)	-1,16 (no)	-18,86*	-16,03*	-0,103*	0,172	1,22	0,525*	0,019 (no)	-0,011 (no)	-0,0001 (no)	0,001*	2,51	2,0							

ze tra i gruppi. Per questi motivi non sono identificabili né il mark-up specifico di gruppo, né il coefficiente h che riflette la flessibilità/rigidità della struttura dei costi di impresa.

Tutto ciò – a ben vedere – non può sorprendere: in fondo il raggruppamento può essere una tecnica rozza ma efficace per evidenziare differenze di comportamento. È pur sempre una tecnica di analisi multivariata. Stimare un modello come [6] *dentro* ai gruppi equivale a spremere i dati ulteriormente con strumenti non molto dissimili da quelli utilizzati in fase esplorativa. È strano che i dati non si lascino «spremere» oltre certi limiti? Noi non lo crediamo. Né crediamo che si possano trarre regole «ottimali» di comportamento per un ricercatore in questo tipo di situazioni (che sono, peraltro, molto più frequenti di quanto si pensi). L'unica raccomandazione generale è quella di sottoporre i dati ad analisi esplorative (di tipo essenzialmente descrittivo) per scoprire se vi siano, esplicite o latenti, relazioni *del tipo* di quelle evidenziate, *prima* di lasciarli alla mercé di sofisticati packages econometrici che – sicuramente – tendono a oscurarle. Meglio ancora se questa esplorazione è guidata da ipotesi sufficientemente robuste, ricavabili dalla teoria economica. Un aggancio più esplicito tra questo tipo di analisi esplorativa e l'analisi econometrica vera e propria sarebbe molto auspicabile. Ma, temiamo, la strada da percorrere è ancora lunga.

Riferimenti bibliografici

- Bouissou, M. B., Laffont, J. J., Vuong, Q. J. (1986), *Disequilibrium Econometrics on Micro Data*, in «Review of Economic Studies», pp. 113-124.
- Contini, B., Cuneo, S., Revelli, R. (1991), *Productivity and Imperfect Competition: Econometric Estimation from Panel-Data*, in «Journal of Economic Behavior and Organizations», di prossima pubblicazione.
- Contini, B. (1989), *Organizations, Markets and Persistence of Profits in Italian Industry*, in «Journal of Economic Behavior and Organization», 12.
- Mueller, D. C. (1986), *Profits in the Long Run*, Cambridge, Cambridge University Press.
- Griliches, Z., Hausman, J. A. (1986), *Error in Variables in Panel Data*, in «Journal of Econometrics», 31, pp. 93-118.