

La valutazione dell'(in)efficienza produttiva nel
settore delle case per anziani in Svizzera: stima
econometrica di una frontiera di costo stocastica.

Autore:

DIEGO LUNATI

Candidato al diploma di:

PH.D. IN ECONOMICS

Commissione d'esame:

PROF. MASSIMO FILIPPINI

(Direttore di tesi),

PROF. RICO MAGGI

(Presidente della commissione d'esame),

PROF. ROSELLA LEVAGGI

(Membro esterno della commissione d'esame)

Facoltà di Economia, Università della Svizzera Italiana

Lugano, Svizzera

Febbraio 2009

Abstract

In every European country health care expenditure containment is a big issue. That is true also in Switzerland, where every year health care absorbs steadily more than 10% of GDP. Nursing homes industry, counting for more than 12% of the expenditures, has a considerable weight on the total cost and for this reason it is interesting to address and empirically evaluate the cost efficiency of the sector. Following the stochastic frontier analysis literature, in particular Greene (2005), Schmidt and Sickles (1984) and Pitt and Lee (1981), we estimated a (Translog) total cost function for a hypothetical nursing home production process, designed considering both previous international literature and Swiss peculiarities. The sample used for the estimates is a balanced panel consisting of 354 nursing homes over 5 years (1998-2002) for a total of 1770 observation, scattered among whole Switzerland. From a methodological point of view, the issue of the unobserved heterogeneity is widely discussed and a comparison between the different estimation models is presented. The main empirical results are: (1) the presence of economies of scale at the sample median (around 50 beds), (2) using the more conservative econometric model a non-negligible cost inefficiency of 8% is detected, and (3) as far the inefficiency rankings of the nursing homes are concerned, the need of high-quality data to ease the problem of sensitiveness of the results to the econometric technique.

Riassunto (Summary)

In Svizzera la spesa sanitaria cresce ogni anno in modo marcato ed ha raggiunto nel 2006 un livello pari a 52,7 miliardi di franchi (10,8% del PIL svizzero). Di queste risorse, circa un quinto sono assorbite dal settore sociosanitario, all'interno del quale la maggior parte del peso grava sugli istituti per anziani. L'aumento della spesa sociosanitaria, che cresce a ritmi più sostenuti rispetto alla componente sanitaria, e la complessa organizzazione del settore, accentuata dalla presenza del federalismo, rendono le case per anziani svizzere un campo di studio molto interessante.

Il tema del contenimento dei costi nelle case per anziani riguarda alcuni aspetti economici cruciali come l'uso delle risorse (scarse), i rapporti tra il mercato e l'azione pubblica, gli equilibri tra efficienza ed equità, che coinvolgono direttamente o indirettamente tutti gli attori di questo settore. Dal punto di vista dell'uso delle risorse il discorso è molto lineare e si basa sul principio logico che l'efficienza economica permette di eliminare (o almeno ridurre) gli sprechi. Questa esigenza, che sempre accompagna le attività produttive realizzate attraverso il mercato, è sentita con grande forza sia dagli istituti per anziani privati sia da quelli pubblici. La promozione dell'efficienza produttiva interessa in particolar modo lo Stato, molto presente come erogatore di servizi/regolatore/finanziatore, che deve confrontarsi con forti pressioni nel contenimento della spesa pubblica. Il settore delle case per anziani è caratterizzato da un eccesso di domanda cronico che, osservando l'attuale ritmo di crescita della popolazione anziana rispetto all'incremento di offerta di nuovi istituti, è destinato a peggiorare. In considerazione di questa problematica, è lecito attendersi che l'orientamento della società civile e delle autorità pubbliche cantonali e comunali si indirizzi maggiormente (e in parte lo sta già facendo) verso il potenziamento dell'offerta dei servizi agli anziani.

In un simile contesto, quindi, la misurazione empirica dell'efficienza produttiva si può trasformare in uno strumento per valutare i costi delle case per anziani svizzere, fornendo informazioni utili per chi nel settore deve prendere decisioni importanti. Nell'erogazione dei servizi di assistenza agli anziani, infatti, il problema di produrre al costo minimo non rappresenta l'unica preoccupazione. Vi sono infatti molte dimensioni (i connotati di bene meritorio, i problemi legati alle asimmetrie informative, gli aspetti di efficacia del servizio, la qualità delle cure, l'equità di accesso, etc.) che potrebbero prevalere sulle ragioni dell'efficienza. Una conoscenza empirica dell'importanza di quest'ultimo aspetto può quindi contribuire alla formazione di giudizi maggiormente fondati e consapevoli.

Muovendo da queste premesse, la presente ricerca si prefigge di analizzare l'efficienza produttiva delle case per anziani svizzere, attraverso la stima econometrica di un modello di costo, con un duplice scopo: (1) determinare empiricamente il valore dell'inefficienza di costo e delle economie di scala; (2) indagare i problemi metodologici relativi alla corretta misurazione di queste grandezze. Per formulare giudizi normativi e prendere decisioni basate sull'evidenza empirica, è molto importante che i risultati delle analisi siano affidabili e poco sensibili alla metodologia impiegata. Per questo motivo, oltre a produrre le informazioni sulla struttura dei costi del settore delle case per anziani, discuteremo anche le caratteristiche ed i limiti dei metodi utilizzati per generare i risultati delle stime.

Per analizzare l'efficienza produttiva delle case per anziani svizzere, abbiamo optato per una funzione Translog, nella quale i costi totali sono specificati secondo un approccio neoclassico in senso ampio. Nel settore sanitario, in modo particolare in quello ospedaliero, la letteratura ha dibattuto a lungo ed in profondità il problema connesso alla specificazione del modello di costo, riconoscendo la necessità di affiancare all'output ed ai prezzi dei fattori produttivi anche altre dimensioni come la qualità, il case-mix ed alcune caratteristiche specifiche degli istituti. Nel nostro caso i costi totali di una generica casa per anziani svizzera sono descritti come una funzione dell'output (misurato dalle giornate di residenza), del prezzo del lavoro e del capitale, del fabbisogno medio di cure degli ospiti (*proxy* per il *case-mix*), di due *proxy* per la qualità delle cure (personale medico-infermieristico per residente e numero di servizi attivi presso l'istituto). Oltre ai regressori menzionati, il modello si completa con un trend lineare e con alcune variabili dicotomiche (*dummies*) che considerano altri aspetti legati all'organizzazione del personale, alle modalità dell'offerta e al posizionamento geografico degli istituti.

Dal punto di vista metodologico, la funzione di costo viene stimata impiegando alcuni modelli econometrici, sviluppati nell'ambito della *panel data econometrics*, che mirano alla corretta misurazione dell'inefficienza produttiva. L'analisi delle frontiere stocastiche, traducendo dall'inglese *Stochastic Frontier Analysis* (SFA), è un filone di ricerca fondato sul presupposto che i residui di una regressione celino informazioni di diversa natura, concernenti: in parte l'eterogeneità non osservata nel modello (problema delle variabili esplicative omesse), in parte l'inefficienza delle singole imprese (variabile che interessa misurare) ed in parte un disturbo stocastico. Grazie all'uso di queste tecniche, si riesce a definire una frontiera di costo che funga da termine di paragone per valutare empiricamente l'inefficienza delle imprese, pur tenendo conto della natura stocastica del processo produttivo.

Tra i numerosi modelli elaborati dalla letteratura, in questa tesi impiegheremo cinque diverse specificazioni: (1) il modello “*maximum likelihood POOLed cross-section*” (POOL) nella versione originale proposta da Aigner, Lovell and Schmidt (1977), (2) il *Maximum Likelihood Random Effect Model* (ML-REM), proposto da Pitt and Lee (1981), (3) il *Fixed Effect Model* (FEM) e (4) il *Random Effect Model* (REM), entrambi sviluppati da Schmidt and Sickles (1984), da ultimo (5) il recente modello proposto da Greene (2005), definito dall'autore “*True*” *Random Effect Model* (TREM). I diversi modelli, che si possono raggruppare in due categorie a seconda di come viene considerata l'inefficienza rispetto al tempo (costante o variabile), presentano caratteristiche econometriche differenti ed una diversa capacità di risolvere il problema della separazione tra inefficienza ed eterogeneità non osservata. Questo problema, che risulta intimamente legato alla struttura temporale attribuita alle due componenti, non è ancora completamente risolto in letteratura, dal momento che non esiste un modello in grado di separare in modo perfetto le tre componenti (stocastica, inefficienza ed eterogeneità) del termine di errore. A questo proposito il modello di Greene (2005) è certamente il più interessante, poiché permette di affinare la precisione delle stime individuali di inefficienza (supposta variabile nel tempo), filtrando parzialmente il contributo dell'eterogeneità non osservata (ipotizzata costante nel tempo).

Per effettuare le stime abbiamo utilizzato un campione di dati che proviene da un rilevamento sistematico (annuale) su tutti gli istituti non ospedalieri svizzeri, fornito dall'Ufficio Federale di Statistica, e riguarda gli anni dal 1998 fino al 2002. Utilizzando la definizione di casa per anziani data dall'Ufficio Federale di Statistica, abbiamo isolato le osservazioni pertinenti, eliminando dal campione quelle strutture che presentavano dati mancanti nelle

variabili chiave del modello. Per sfruttare al meglio le informazioni temporali dei dati, abbiamo inoltre bilanciato il panel, che risulta quindi costituito da 5 anni, per i quali abbiamo a disposizione 354 case per anziani. Pertanto, il campione finale utilizzato nelle stime è composto da 1'770 osservazioni che riguardano gli istituti per anziani svizzeri.

I risultati dell'analisi empirica, condotta con la stima di una frontiera di costo stocastica per il settore delle case per anziani, forniscono tre indicazioni principali difficili da ignorare, anche dopo aver considerato tutti i limiti e le attenuanti del caso: (1) il livello medio di inefficienza di costo rilevato nel settore non è trascurabile (2) alla mediana del campione (51 letti) si osserva la presenza di economie di scala non completamente sfruttate e (3) l'eterogeneità non osservata è un elemento molto importante da considerare quando si discutono i risultati empirici, dal momento che potrebbe incidere sulla stabilità dei coefficienti, ma soprattutto dei punteggi di inefficienza.

La prima considerazione richiama l'attenzione sull'importanza di considerare l'inefficienza di costo, nonostante il livello misurato da alcuni modelli (in particolare il modello TREM) non sia eccessivo. La crescente difficoltà nel finanziare i modelli di *welfare* attualmente presenti in Svizzera ha determinato, a molti livelli, un ripensamento di alcune scelte di fondo ed una maggior attenzione all'efficienza nell'uso delle risorse e non vi è ragione di credere che il settore dell'assistenza agli anziani sia impermeabile a questi sforzi tesi a promuovere l'efficienza. L'eccesso di costo, misurato con il modello TREM ritenuto più preciso (Greene, 2005), è risultato in media pari all'8,1%. Viste le proprietà econometriche del modello TREM, questo valore si può considerare una stima conservativa dell'effettivo valore medio di (in)efficienza. L'analisi dell'inefficienza di costo posiziona il settore delle case per anziani ad una distanza tutto sommato accettabile dalla "*best practice*", costituita dalla frontiera di costo, ma questo non significa che l'inefficienza di costo sia un fenomeno trascurabile.

La seconda osservazione empirica riguarda la necessità di attribuire il giusto peso alla dimensione di un istituto che, secondo la stima delle economie di scala, non raggiunge la dimensione ottimale in più della metà delle osservazioni del campione. In Svizzera le economie di scala nel settore dell'assistenza agli anziani sono molto rilevanti per gli istituti di piccole dimensioni e diminuiscono progressivamente fino a svanire completamente una volta raggiunta la dimensione ottimale, individuata nell'intervallo 70-75 posti letto. L'importanza di raggiungere la scala efficiente va certamente considerata con maggiore cautela per quelle zone isolate che presentano un'elevata dispersione abitativa. In questi frangenti, infatti, il conflitto tra efficienza (di scala) ed equità (di accesso) si accentua fino a spostare (soprattutto per ragioni

politiche) l'accento sull'opportunità di avere un beneficio locale (la casa per anziani piccola) in cambio di un costo collettivo (l'eccesso di costo provocato dalla bassa capacità produttiva). Va comunque precisato che la dimensione ottimale stimata (70-75 letti) non dovrebbe comportare disagi ed anonimità diversi, rispetto ad istituti di piccola dimensione, inefficienti dal punto di vista della scala.

La misurazione delle due grandezze empiriche discusse (inefficienza di costo ed economie di scala) ci porta alla terza riflessione, di carattere metodologico, che riguarda l'importanza di verificare la sensibilità dei risultati empirici rispetto alla specificazione econometrica impiegata. A tale proposito, l'analisi dell'inefficienza di costo ha evidenziato in modo molto chiaro come i modelli che considerano con maggior precisione l'eterogeneità non osservata indichino livelli di inefficienza minori. Tuttavia, il problema non riguarda solo la sensibilità della scala di inefficienza, quanto la possibile inconsistenza delle graduatorie dei punteggi dei singoli istituti. A questo riguardo bisogna riconoscere il contributo fondamentale che la qualità dei dati apporta nel processo di stima, risolvendo alla fonte il problema dell'affidabilità dei risultati. In considerazione dei problemi metodologici che affliggono le stime ottenute con i diversi modelli, l'approccio di presentare un ventaglio di risultati, invece che un'indicazione puntuale, appare preferibile come base di ragionamento per formulare raccomandazioni di politica sanitaria.

La presente tesi di dottorato si congeda dal lettore discutendo alcune posizioni critiche espresse dalla letteratura, in particolare Newhouse (1994), Skinner (1994) e Folland et al. (2001), sul valore e l'affidabilità delle analisi empiriche sull'efficienza produttiva nel settore sanitario, con lo scopo di precisare, valorizzandolo, il contributo di questo lavoro e tracciare i possibili sviluppi della ricerca in questo ambito.

KEY WORDS: Efficiency, Economies of Scale, Stochastic Frontier Analysis, Nursing Homes, Translog Cost Function, Panel Data, Unobserved Heterogeneity.

INDICE

Abstract

Riassunto (summary)

Indice delle Figure	iv
Indice delle Tabelle	v
Ringraziamenti	vii
Premessa	xi
Elenco Abbreviazioni	xvii

1 Introduzione	0
1.1 Il problema economico	2
1.2 Gli obiettivi della tesi.....	9
1.3 La metodologia	13
1.4 L'inefficienza produttiva nella letteratura economica	21
1.5 La struttura della tesi.....	26
2 Il settore delle case per anziani in Svizzera	30
2.1 Analisi descrittiva delle case per anziani in Svizzera	31
2.1 Tassonomia: "Altersheime", "Krankenheime" o "Pflegeheime"?.....	32
2.1.1 Forme giuridiche e assetto proprietario	38
2.1.2 Dimensione degli istituti.....	43
2.1.3 Ventaglio di prestazioni offerte	46
2.2 Differenze cantonali nelle caratteristiche delle case per anziani	50
2.2.1 La capacità produttiva cantonale	51
2.2.2 L'organizzazione istituzionale.....	53
2.2.3 Il profilo dei residenti ed il ventaglio dei servizi erogati	55
2.2.4 Il costo medio per giornata di cura	59

3	Elementi di teoria economica della produzione: la funzione di costo e l'efficienza produttiva.	65
3.1	Elementi scelti di microeconomia della produzione	66
3.1.1	Tecnologia, funzione di produzione ed isoquanti.....	67
3.1.2	La funzione di costo (totale) di lungo periodo.....	74
3.1.3	La funzione di costo (variabile) di breve periodo.....	80
3.1.4	Economie di scala.....	84
3.2	L'(in)efficienza produttiva	86
3.2.1	L'(in)efficienza tecnica	88
3.2.2	L'(in)efficienza di costo	89
3.2.3	L'(in)efficienza allocativa	91
3.2.4	L'(in)efficienza di scala.....	93
4	La stima parametrica di una frontiera di costo stocastica: aspetti metodologici	97
4.1	Metodi non parametrici e parametrici a confronto: DEA vs. SFA.....	98
4.1.1	Metodi non parametrici: Data Envelopment Analysis (DEA).....	98
4.1.2	I metodi parametrici e la Stochastic Frontier Analysis (SFA).....	100
4.1.3	DEA vs. SFA: vantaggi e svantaggi nel misurare l'efficienza produttiva	101
4.2	Elementi per la stima di una frontiera di costo stocastica	103
4.2.1	Modello di costo.....	104
4.2.2	Forma funzionale.....	106
	4.2.2.1 Funzione di costo Cobb-Douglas	110
	4.2.2.2 Funzione di costo Translog	113
4.3	Stima econometrica di una frontiera di costo.....	118
4.3.1	L'approccio deterministico all'analisi delle frontiere.....	119
4.3.2	L'approccio stocastico all'analisi delle frontiere (SFA).....	123
4.4	Principali modelli econometrici in uso nella SFA	126
4.4.1	Modelli Cross-Section	126
4.4.2	Modelli Panel-Data.....	131
	4.4.2.1 Modelli Panel con inefficienza costante nel tempo	131
	4.4.2.2 Modelli Panel con inefficienza variabile nel tempo	139
	4.4.2.3 Modelli Panel che considerano l'eterogeneità non osservata	143
4.4.3	Pro e contro nella scelta del modello econometrico	150

5	Definizione del modello di costo per il settore delle case per anziani in Svizzera	158
5.1	Rassegna della letteratura sulla struttura dei costi nel settore delle case per anziani	161
5.2	Il processo produttivo di una generica casa per anziani	183
5.2.1	La misura dell'attività produttiva	185
5.2.2	La misura della qualità e delle caratteristiche ambientali	188
5.2.3	La misura dei prezzi dei fattori produttivi	191
5.2.4	La banca dati disponibile ed il suo contenuto informativo	196
5.3	Specificazione del modello di costo utilizzato per le stime	201
5.3.1	La scelta della specificazione di lungo periodo	202
5.3.2	Il modello di costo utilizzato nelle stime	207
6	Economie di scala ed inefficienza di costo: analisi dei risultati empirici	219
6.1	L'analisi empirica	220
6.1.1	La scelta del campione utilizzato per le stime	220
6.1.2	I risultati delle stime	226
6.1.3	Discussione critica su alcune scelte di specificazione del modello di costo	238
6.2	Misurazione e analisi dell'inefficienza di costo.....	242
6.2.1	Confronto dell'(in)efficienza di costo nei diversi modelli.....	243
6.2.2	(In)efficienza di costo ed eterogeneità.....	250
6.3	Analisi delle economie di scala.....	255
7	Conclusioni	264
7.1	Riepilogo.....	264
7.2	Riflessioni conclusive sui risultati dell'analisi empirica	268
7.2.1	Confronto con la letteratura e aspetti di <i>policy</i>	268
7.2.2	Impiego delle frontiere stocastiche nella regolamentazione	272
7.3	Possibili sviluppi futuri della ricerca	275
	BIBLIOGRAFIA	278

Indice delle Figure

Figura 1.1 - Posizionamento della letteratura economica sul tema dell'inefficienza.	23
Figura 2.1 - Criterio di ripartizione degli stabilimenti non ospedalieri nelle categorie UFS.	36
Figura 2.2 - Classificazione delle case per anziani secondo l'assetto proprietario e lo statuto giuridico.....	40
Figura 2.3 - Distribuzione delle case per anziani in base allo statuto giuridico (anno 2002).	42
Figura 2.4 - Distribuzione degli istituti secondo il numero di posti letto.	44
Figura 2.5 - Box-plot delle case per anziani, suddivise secondo la forma istituzionale.	45
Figura 2.6 - Densità di letti (per 1000 abitanti) nelle case per anziani: differenze dalla media svizzera.	52
Figura 2.7 - Densità di letti (per 100 persone con più di 65 anni): differenze dalla media svizzera.	53
Figura 2.8 - Numero di istituti suddivisi secondo la forma istituzionale (anno 2002).....	54
Figura 2.9 - Livello del fabbisogno medio di cure e/o assistenza per istituto (anno 2002).	57
Figura 2.10 - Costi medi per giornata di residenza: evoluzione e tasso di crescita (2002-1998).....	60
Figura 2.11 - Data di nascita delle leggi cantonali concernenti le case per anziani.....	62
Figura 3.1 - Rappresentazione grafica di un isoquante (2 input ed 1 output).....	71
Figura 3.2 - Forme di isoquanti convessi a confronto (2 input ed 1 output).....	73
Figura 3.3 - Condizioni di involuppo e costi medi di breve e di lungo periodo.	82
Figura 3.4 - Elasticità di scala e di dimensione.	85
Figura 3.5 - Scomposizione dell'efficienza di costo (2 input ed 1 output).....	92
Figura 3.6 - L'inefficienza di scala.....	95
Figura 4.1 - Famiglie di modelli per l'analisi delle Frontiere Stocastiche.....	125
Figura 6.1 - Eterogeneità territoriale dei punteggi di inefficienza.	251
Figura 6.2 - Rappresentazione grafica della funzione di costo medio.	260
Figura 6.3 - Rappresentatività del punto di approssimazione dei dati a livello cantonale.....	262

Indice delle Tabelle

Tabella 2.1 - Suddivisione degli istituti secondo la forma istituzionale (anno 2002).....	43
Tabella 2.2 - Numero medio di servizi erogati, secondo la forma istituzionale	47
Tabella 2.3 - Grado di diffusione di particolari servizi erogati dalle case per anziani svizzere.	48
Tabella 2.4 - Esempio di costruzione dell'indice del fabbisogno medio giornaliero di cure.....	56
Tabella 2.5 - Numero medio di servizi erogati a livello cantonale (anno 2002).....	58
Tabella 4.1 - Confronto tra i principali metodi per la misura delle performance.	102
Tabella 4.2 - Riepilogo dei modelli econometrici da considerare per le stime.....	157
Tabella 5.1 - Riepilogo dei modelli di costo specificati per diversi paesi.	181
Tabella 6.1 - Statistiche descrittive delle principali variabili utilizzate nelle stime.....	225
Tabella 6.2 - Stima dei parametri della funzione di costo con 5 diversi modelli.....	229
Tabella 6.3 - Stabilità delle stime rispetto alla forma funzionale.	239
Tabella 6.4 - Stabilità delle stime rispetto alla specificazione delle variabili ausiliarie.	241

Ringraziamenti

Una tesi di dottorato sul tema dell'efficienza produttiva non può ignorare la relazione esistente tra output e fattori produttivi ed è per questo motivo che, giunti al momento di finalizzare il lavoro (output) dopo anni di fatiche, si devono riconoscere i molti debiti intellettuali ed affettivi (input) accumulati lungo il percorso. Una tesi empirica, tuttavia, non può ignorare l'esistenza della legge (empirica) dei rendimenti decrescenti degli input e, sfortuna vuole, la parte dedicata ai ringraziamenti viene scritta per ultima, quando cioè appare impossibile ripagare tutte le persone che hanno contribuito a rendere possibile questo lavoro, e quando la vena letteraria per farlo si va esaurendo. Se aggiungiamo che si tratta della prima porzione del lavoro con la quale il lettore viene a contatto, emerge una certa difficoltà, mista ad un po' di imbarazzo, nella redazione di queste pagine.

Una tesi di dottorato sul tema dell'efficienza produttiva, per coerenza con il titolo, non dovrebbe neppure soffermarsi troppo su preamboli e ringraziamenti, poiché da un punto di vista produttivo costituiscono uno spreco deliberato di risorse. Tuttavia, l'emozione di sentirsi ormai prossimi alla meta, rende irrefrenabile il desiderio di condividere questo momento con chi ha contribuito a renderlo possibile. Inoltre, completare questa sezione è il segnale perentorio che la ricerca è terminata e che si deve voltare pagina, a prescindere dai risultati ottenuti e dalle questioni rimaste irrisolte.

Forse, un modo indolore per farlo è proprio quello di ripercorrere all'indietro, con la mente e con l'anima, il tempo e le energie dedicate a questo progetto, alla ricerca di tutti coloro che mi hanno aiutato, in modo consapevole o meno, nei più svariati modi. Scrivere i ringraziamenti assume quindi un connotato nuovo, che non conoscevo, ossia provare soddisfazione rispetto al proprio lavoro, attraverso il riconoscimento dei meriti altrui.

Iniziando con l'ambito accademico e professionale, è mio desiderio ringraziare i colleghi dell'istituto Mecop (Università della Svizzera Italiana – USI) e del Dipartimento di Scienze Aziendali e Sociali (Scuola Universitaria Professionale della Svizzera Italiana – SUPSI) con i quali ho avuto il privilegio e la fortuna di confrontarmi. Tra questi, mi accingo a menzionare, senza un particolare ordine, solo pochi nomi.

Un ringraziamento speciale va al Professor Massimo Filippini, per i suoi suggerimenti, la sua pazienza e capacità di guida, per avermi insegnato l'importanza della fiducia e della chiarezza nelle relazioni professionali (praticandole in prima persona nei miei confronti). Lo ringrazio, inoltre, per avermi dato l'opportunità di svolgere il dottorato in economia sotto la sua direzione, contribuendo in modo decisivo alla mia crescita personale. Da ultimo, ma non meno importante, desidero esprimere una sincera gratitudine per la libertà di cui ho goduto nello svolgimento di questa tesi e per il sostegno ricevuto nei confronti delle scelte professionali compiute nel recente passato.

Al Professor Amedeo Fossati sono riconoscente per le sue lezioni di “Scienza delle Finanze” e per aver scritto del “pesce sciamano”, cose che, a sua insaputa, mi hanno instillato la curiosità e la passione nei confronti dell'economia pubblica.

Ringrazio poi Loa Buchli, per l'aiuto che mi ha dato all'inizio della mia avventura svizzera e per la sincera amicizia. Una menzione di gratitudine va a tutti i miei “compagni di ufficio”, presenti e passati, per avermi sopportato (ma anche supportato). I loro nomi sono: Rubens, Fabio, Barbara, Ilaria, Diego, Laura, Luca, Fabiano, Laura, Chiara, Delia e Marta)

Ringrazio il Professor Luca Crivelli che, con il suo contagioso entusiasmo, ha saputo trasmettermi una passione sincera nei confronti della didattica e della ricerca e

la Professoressa Wilma Minoggio per l'incoraggiamento e per la comprensione dimostratami durante le fasi finali della tesi.

Desidero esprimere la mia gratitudine nei confronti del Professor Rico Maggi, per alcuni preziosi consigli e per la sua amicizia, così come sono riconoscente nei confronti di Marco Alderighi, Oscar Gonzalez e Jurg Mägerle per le piacevoli (e irripetibili) discussioni "serali".

Un ringraziamento va ai Proff. Harold Fried, William Greene e Subal Kumbhakar per aver contribuito, durante i corsi dottorali da loro impartiti, a migliorare in modo sostanziale le mie conoscenze specifiche sul tema dell'analisi dell'efficienza produttiva.

Operando un'autocensura ai ringraziamenti accademico-professionali, desidero riconoscere alcuni meriti *sui generis* a determinati elementi che hanno contribuito al completamento di questa tesi. Mi riferisco ad esempio ai dati che, col passare del tempo, da cross-section sono diventati panel, permettendomi di applicare altre metodologie e fornendomi (come la carota per l'asinello) un costante incentivo a condurre nuove analisi. Un riconoscimento va anche all'Ufficio Federale di Statistica che, insieme alla carota, ha fornito pure il bastone, nella forma di una banca dati spinosa da trattare e che mi ha riservato più di un dispiacere, rafforzando in me la tempra del ricercatore.

Questo intermezzo ironico (nei limiti delle doti umoristiche di un economista) funge da linea di demarcazione tra i debiti di gratitudine nei confronti del mondo professionale, e la riconoscenza verso la sfera degli affetti privati. Di questa fanno parte tutti i parenti e gli amici che in molti modi diversi hanno saputo sostenermi o consigliarmi, meritandosi un esplicito ringraziamento.

Agli amici, nessuno escluso, va il merito di aver alimentato costantemente le energie emotive, che sono una risorsa preziosa per riuscire a resistere all'onda d'urto provocata dallo svolgimento di un dottorato, in parallelo all'attività professionale. Alla cerchia dei parenti va il mio sincero ringraziamento per avermi sempre circondato di affetto ed attenzioni.

In modo particolare desidero ringraziare i miei genitori, Doriana e Mauro, per l'amore incondizionato, che da loro ho sempre ricevuto, e per avermi trasmesso quella fiducia e serenità che mi hanno consentito di arrivare fino alla fine di questo (estenuante) dottorato.

Un ringraziamento speciale va ovviamente alla mia famiglia, ed in particolare ai miei bimbi Sofia e Matteo che, dal giorno della loro nascita, mi hanno dato un motivo in più ed uno stimolo decisivo per terminare la tesi. A mia moglie Susy sarò eternamente grato per la pazienza, l'amore ed il sostegno che quotidianamente mi dimostra, ma soprattutto per i sacrifici che ha fatto in questi anni per costruire la nostra famiglia, alleggerendomi di buona parte degli oneri e delle fatiche.

Da ultimo, il mio ringraziamento va a Barbara Antonioli-Mantegazzini per la preziosa ed attenta rilettura, che ha permesso di migliorare sensibilmente la qualità del testo. Resta inteso che tutti gli errori o le imprecisioni ancora presenti nella versione finale della tesi sono unicamente responsabilità dell'autore.

Lugano, febbraio 2009

Diego Lunati

*Aequam memento rebus in
arduis servare mentem, non
secus in bonis.*

(Quinto Orazio Flacco, Poeta)

Premessa

Il termine efficienza discende dal verbo latino *eficere* che letteralmente significa portare a compimento. L'efficienza presuppone quindi l'esistenza di un obiettivo da raggiungere e fornisce una valutazione del modo con il quale l'obiettivo viene raggiunto, dove il modo è un criterio ben preciso scelto a priori. In altre parole, l'efficienza esprime intrinsecamente un giudizio valutativo su una determinata azione che, con successo, raggiunge uno scopo desiderato. Anche se la terminologia con la quale si esprime questo concetto è molto varia nelle diverse discipline, il concetto di efficienza è sempre il medesimo ed indica un rapporto tra risultati e risorse a disposizione.

In economia esistono diversi concetti di efficienza ma, in ambito produttivo, questo termine si riferisce a quelle imprese che riescono ad ottenere la massima quantità possibile di prodotto (output) con una data quantità di fattori produttivi ed una certa tecnologia a disposizione. Questa relazione tra input e output è definita come efficienza tecnica e richiama il concetto ingegneristico di rendimento. Peraltro, in economia non ci si confronta solamente con grandezze fisiche, poiché quasi tutte le risorse sono associate ad un prezzo, che gli agenti economici devono pagare per assicurarsene. Ecco che il concetto di efficienza si può esprimere in termini monetari, valutando il rapporto tra l'output prodotto ed il costo sostenuto.

Dopo aver chiarito che l'efficienza economica non significa nulla se non si specifica esattamente **cosa** si desidera misurare e **come** si intende farlo, appare necessario porre una domanda più intrigante, ossia:

*perché la scienza economica si dovrebbe preoccupare
della definizione e misurazione dell'inefficienza?*

La risposta al quesito non può essere unanime e soprattutto è soggetta alla naturale evoluzione che caratterizza tutte le scienze, in particolar modo quelle sociali.

Tutto l'impianto teorico dell'economia politica si appoggia su ipotesi ragionevoli e condivisibili sotto il profilo scientifico (individualismo metodologico), che si devono confrontare, e spesso scontrare, con la realtà dei fatti. Potremmo allora affermare che l'efficienza, sta alla teoria economica, come l'inefficienza sta al mondo reale. Parlare di inefficienza e cercare di misurarla significa gettare un solido ponte tra l'impalpabile perfezione astratta della teoria e le sue "imperfette" manifestazioni che concretamente osserviamo attraverso i dati economici.

Nel testo impiegheremo spesso l'espressione "(in)efficienza", ad indicare come l'idea stessa di misurare l'efficienza sia un'attività inutile se non concediamo anche dal punto di vista teorico la sua possibile esistenza. Assumendo che nel mondo reale si possano verificare degli sprechi, dotarsi di uno strumento concettuale di misura (l'efficienza) permette di rilevare (ove presente) l'inefficienza. Ecco che il tema dell'(in)efficienza diventa interessante non soltanto come argomento dell'economia applicata, quale misurazione delle performance economiche in un determinato settore, ma anche come verifica empirica della solidità delle teorie economiche tradizionali e stimolo alla formulazione di nuovi modelli teorici.

La prospettiva con la quale si dovrebbe guardare l'inefficienza non è quella di rifiutare ciò che si discosta dalla perfezione (e quindi dall'efficienza). Piuttosto è quella di riconoscere che una misurazione dello spreco di risorse non è utile intrinsecamente, ma lo diventa pensando al costo opportunità ad esso associato. Ad esempio, l'aspetto interessante nel rilevare un eccesso di costo in una casa per anziani non è la quantificazione del denaro sprecato, quanto piuttosto del numero di giornate di assistenza agli anziani che si potevano erogare utilizzando le risorse in modo efficiente.

Oltre a misurare il costo opportunità delle risorse sprecate, la valutazione dell'(in)efficienza di un'impresa permette di giudicare le prestazioni produttive conseguite sia in termini assoluti che in termini relativi.

Un parallelo tra le prestazioni economiche di un'impresa e quelle sportive di un'atleta può essere molto istruttivo per inquadrare il significato che un'analisi dell'efficienza può assumere per gli attori economici coinvolti in un certo mercato. La "legge dello sport" stabilisce in modo imperativo le gerarchie dei partecipanti ad una competizione, così che alla fine si possa individuare: il migliore, il secondo classificato e via dicendo. La filosofia che anima lo sport è proprio il confronto delle prestazioni con lo scopo di valutare in termini assoluti e relativi il risultato ottenuto da ciascun atleta. In pratica nello sport si redigono classifiche, nelle quali è possibile leggere le prestazioni assolute, misurate con un certo criterio cardinale (punti, secondi, metri, numero di vittorie,...). Ancora più importante è però stabilire in senso ordinale le prestazioni per capire in modo comparativo la posizione di ciascun atleta nella graduatoria (primo, secondo, etc.). Da sempre, chi si occupa di sport sposa la filosofia della competizione (che si spera leale) e della misurazione delle prestazioni (che si spera precisa). Per concludere il parallelo potremmo descrivere l'attività economica delle imprese come una competizione sportiva (competizione di mercato) nella quale ogni atleta (impresa) fa del proprio meglio per primeggiare, ma dove alla fine il terreno di gioco (mercato) decide l'ordine di classifica.

Secondo questa visione, ma utilizzando una terminologia più consona, la "legge del mercato" garantisce che solo le imprese efficienti riusciranno a sopravvivere alla concorrenza esercitata dalla competizione reciproca. Questa interpretazione economica del principio darwiniano della selezione naturale è distillata nel paradigma neoclassico, che sembra sminuire l'importanza della misurazione delle singole prestazioni delle imprese, assumendo implicitamente la scomparsa nel tempo di eventuali comportamenti o situazioni inefficienti.

Nella realtà ciò che possiamo osservare è semplicemente una serie successiva di periodi, all'interno dei quali coesistono e si sovrappongono imprese che vivono situazioni aziendali profondamente diverse. In altre parole, pur accettando il principio neoclassico dell'efficienza dinamica, è possibile che le decisioni delle imprese

osservate in un dato momento si concretizzano in prestazioni più o meno efficienti all'interno dello stesso mercato.

Un'analisi statica delle prestazioni delle imprese, che in pratica corrisponde ad una fotografia "istantanea", fornisce indicazioni molto interessanti per capire e rilevare quali siano le imprese che ottengono i migliori risultati. Inoltre, la misurazione delle prestazioni delle imprese (analisi dell'efficienza produttiva), non solo costituisce il punto di partenza per valutare l'evoluzione di un settore economico attraverso il confronto dei valori di (in)efficienza registrati in periodi di tempo successivi, ma fornisce dei segnali e delle indicazioni potenzialmente utili alle imprese stesse.

Chiarita la motivazione che spinge a misurare empiricamente le prestazioni produttive delle imprese di un determinato settore economico, sorge spontanea una seconda importante domanda:

*Quali sono gli strumenti che permettono in termini
concreti di misurare l'efficienza produttiva?*

La valutazione delle prestazioni di un'impresa, effettuata attraverso un'analisi dell'efficienza, si configura come una graduatoria o una classifica che fornisce indicazioni relative alla posizione di ciascun'impresa nei confronti di altre imprese ed anche una misura del livello medio delle prestazioni di tutte le aziende. Per riuscire a redigere una graduatoria occorre misurare (in modo preciso) tutte le prestazioni delle imprese coinvolte e stabilire un termine di paragone.

Nel caso del settore dell'assistenza agli anziani, la prestazione delle imprese si può ragionevolmente misurare con il costo di produzione, riferendolo poi alla *best practice* del settore, definita empiricamente dai dati.

L'econometria e la ricerca operativa forniscono diverse strategie per individuare una frontiera di costo da utilizzare come pietra miliare nei confronti. Questo, tuttavia, non risolve definitivamente il problema, poiché l'uso di tecniche diverse sposta la discussione sulla validità dei risultati ottenuti con metodi di misura differenti. Le indicazioni di efficienza che scaturiscono da un'analisi empirica devono, quindi, essere valutate con la consapevolezza che alcune dimensioni del problema potrebbero

mancare, o che la misura effettuata può avere dei limiti, obbligando ad interpretare i risultati con molta cautela.

Se si trova un terreno metodologico comune per definire gli indicatori sui quali poggia il confronto, i risultati delle analisi di efficienza possono servire in modo particolare alle imprese, per migliorarsi attraverso la comparazione con altri istituti, ed allo Stato nel processo di regolamentazione. Ci riferiamo alla possibilità di utilizzare i risultati di un'analisi dell'efficienza nell'ambito di una regolamentazione incentivante (*yardstick regulation*) di questo settore che potrebbe ridurre le asimmetrie di informazione tra regolatore e regolato.

In armonia con lo stile interrogativo fin qui seguito, desideriamo porre ancora una domanda:

Qual è il senso di questa premessa?

I concetti presentati non rappresentano una novità nell'ambito della scienza economica, così come i problemi legati alla metodologia sono ben noti a chi si occupa di ricerca applicata. La motivazione per introdurre il tema dell'efficienza partendo da così lontano risponde quindi ad altre necessità.

In primo luogo, si desidera manifestare in modo esplicito l'interesse dell'autore nell'effettuare un'analisi empirica legata ai temi dell'efficienza produttiva. Analizzare la complessità che si cela dietro alle cose apparentemente semplici (come il concetto di efficienza) è ciò che ripaga un ricercatore delle fatiche sostenute per soddisfare la propria curiosità intellettuale, anche quando questa dovesse rivolgersi a temi molto specifici.

Un secondo motivo per presentare in termini generali il tema dell'inefficienza è quello di cogliere l'intreccio che nella ricerca applicata inevitabilmente si forma tra: teoria, verifica empirica e metodologia. In altre parole, desideriamo sottolineare come, pur restando nell'ambito dell'analisi dell'efficienza produttiva, si potrebbero condurre diverse ricerche tutte interessanti, ma tutte diverse a dipendenza di quale aspetto del problema si decide di approfondire. Di conseguenza il problema dell'efficienza, indotto dalla scarsità relativa delle risorse, si pone ancora prima di iniziare la ricerca in ragione del *trade-off* che si crea tra ampiezza e profondità nella scelta dei contenuti da

trattare. Per questo motivo, prima di incanalare il discorso verso gli argomenti che sostanzieranno la presente ricerca, sembrava pertinente una riflessione preliminare da una prospettiva molto ampia.

Un ultimo ragionamento che desideriamo proporre riguarda lo spirito critico e la determinazione ai quali si è ispirato l'intero percorso di ricerca. Il punto di partenza, scelto per questa tesi, è solo un piccolo ramo dell'albero della ricerca economica, che si innesta comunque su altri rami, via via più grossi, fino ad arrivare al tronco. Partire dal generale per scendere poi nel dettaglio, è un approccio che evidenzia maggiormente le alternative disponibili, favorendo il senso critico nei confronti del proprio lavoro. Di conseguenza, la consapevolezza di aver valutato compiutamente diverse opzioni, aumenta anche la determinazione nel difendere le proprie scelte. Discutere, in questa sede ed in questo modo, il tema della definizione dell'efficienza produttiva e della sua misurazione, può assumere anche una valenza evocativa di quello che è stato il *modus operandi* di questa tesi. Per necessità editoriali, infatti, esporremo solo il percorso principale seguito durante la ricerca, senza dare spazio alle numerose analisi e riflessioni svolte che hanno comunque contribuito a produrre il documento finale.

In quel che segue, ci accingiamo a presentare i risultati dell'analisi empirica sull'(in)efficienza produttiva nel settore delle case per anziani in Svizzera, con la serenità di aver lavorato scrupolosamente e la consapevolezza che in ogni ricerca vi sono dei limiti. Saremmo quindi sorpresi e delusi di non suscitare nel lettore alcun rilievo critico. Sorpresi, per aver raggiunto un livello di piena efficienza che, come discusso all'inizio della premessa, non appartiene al mondo reale. Delusi, perché l'assenza di critiche, escludendo a priori la possibilità che non vi siano osservazioni da muovere, significherebbe aver scritto una tesi dagli argomenti incomprensibili.

La sentenza al lettore!

Elenco Abbreviazioni (ordine di comparsa)

POOL	maximum likelihood POOL ed corss-section
ML-REM	Maximum Likelihood Random Effect Model
FEM	Fixed Effect Model
REM	Random Effect Model
TREM	True Random Effect Model
TFEM	True Fixed Effect Model
USI	Università della Svizzera Italiana
SUPSI	Scuola Universitaria Professionale della Svizzera Italiana
PIL	Prodotto Interno Lordo
SFA	Stochastic Frontier Analysis
GLS	Generalized Least Square
MLE	Maximum Likelihood Estimation
LSDV	Least Square Dummy Variables
LAMal	Legge sull'Assicurazione Malattia
UFS	Ufficio Federale di Statistica
NOGA	Nomenclatura Generale delle Attività economiche
FdP	Funzione di Produzione
CTM	Costi Totali Medi
EC	Efficienza di Costo
EA	Efficienza Allocativa
RSND	Rendimenti di Scala Non Decrescenti
RSNC	Rendimenti di Scala Non Crescenti
RSC	Rendimenti di Scala Costanti
DEA	Data Envelopment Analysis
TFP	Total Factor Productivity (produttività totale dei fattori)
RSV	Rendimenti di Scala Variabili
CES	Constant Elasticity of Substitution
OLS	Ordinary Least Squares (minimi quadrati ordinari)
COLS	Corrected Ordinary Least Squares
MOLS	Modified Ordinary Least Squares
GLS	Generalized Least Squares
FE	Fixed Effect
RE	Random Effect
ML	Maximum Likelihood
SUR	Seemingly Unrelated Regression
CON	Certificate Of Need
SNF	Skilled Nursing Facility
INF	Intermediate Nursing Facility
RUG	Resources Utilization Groups
SC	Skilled Care
IC	Intermediate Care
TQM	Total Quality Management
FTE	Full Time Equivalent
AI	Assicurazione Invalidità
AVS	Assicurazione Vecchiaia Superstiti
DSS	Dipartimento di Sanità e Socialità

“It's a job that's never started that takes the longest to finish.”

(John Ronald Reuel Tolkien, Scrittore)

1 Introduzione

Il termine efficienza è senza dubbio di facile comprensione visto l'impiego che siamo soliti farne nei più svariati ambiti quotidiani. Il concetto di efficienza, in termini molto generali, esprime la capacità di fare una certa cosa nel miglior modo possibile. Appare chiaro come l'idea stessa di efficienza presupponga un obiettivo da raggiungere ed un mezzo per farlo. Il tema di questa tesi è l'analisi dell'efficienza produttiva di un campione di case per anziani svizzere. In particolare, siamo interessati alla stima con metodi econometrici dell'inefficienza di costo e di scala delle case per anziani considerate nell'analisi. L'attività svolta dagli istituti residenziali per anziani appare, agli occhi di un osservatore esterno, molto semplice nei suoi tratti essenziali: ospitare e prendersi cura di persone che, per scelta o per necessità, non vivono in abitazioni private. La definizione dell'efficienza produttiva e la sua misurazione riguardano, in modo molto generale, i costi dell'attività produttiva e il volume di beni e servizi offerti. A ben vedere, alcune cose che in apparenza sembrano semplici ed inequivocabili, si rivelano complesse e affascinanti se analizzate con un grado di dettaglio che caratterizza la ricerca universitaria. Il presente lavoro prende spunto da questi due elementi (l'attività di una casa per anziani e l'efficienza produttiva), cercando di intrecciarli e riferirli al contesto svizzero.

Nel settore dell'assistenza agli anziani, la dimensione economica legata all'erogazione dei servizi ha assunto una crescente importanza in tutte le società industrializzate, compresa la Svizzera. La rapida ascesa della spesa sanitaria, e con essa la parte generata dall'erogazione dei servizi di assistenza agli anziani, ha richiamato grande interesse sul tema sia in ambito accademico sia da parte degli attori che animano il settore. In modo particolare si è posto il problema del possibile spreco di risorse causato da una dimensione ridotta degli istituti (economie di scala non sfruttate) e da una gestione non improntata alla minimizzazione dei costi (inefficienza di costo).

L'attuale contesto, caratterizzato da una crescente domanda di servizi residenziali per anziani e da un'insufficiente offerta di posti letto, rende molto importante un'indagine empirica sulla misurazione dell'effettivo grado di efficienza raggiunto in Svizzera. In particolare, la corretta definizione della capacità produttiva di una casa per anziani è un aspetto cruciale, in considerazione del lungo periodo di ammortamento degli investimenti effettuati e gli alti costi associati ad un'eventuale modifica della dimensione di un istituto. Pertanto, misurare il valore delle economie di scala presenti nel settore significa fornire un'informazione utile soprattutto ai decisori coinvolti nei futuri progetti di realizzazione delle nuove strutture. Indipendentemente dalla scala adottata, sprechi di risorse sono possibili sotto forma di un eccessivo costo di produzione.

In questo settore si registra un forte intervento del settore pubblico nell'erogazione di servizi, nel finanziamento degli stessi e nella regolamentazione del mercato. La triplice veste assunta dal settore pubblico sottolinea l'importanza di un'analisi dell'inefficienza di costo, poiché le crescenti difficoltà finanziarie in cui versano molti comuni, responsabili di erogare servizi agli anziani, enfatizzano l'importanza di ridurre al minimo l'inefficienza di costo. Il ragionamento si applica anche nei confronti delle molte strutture private, rispetto alle quali lo Stato, in qualità di fornitore dei mezzi finanziari e di regolatore, mira a promuovere una gestione improntata a criteri di efficienza. Una misurazione dell'inefficienza di costo può quindi contribuire all'attuale dibattito sul tema delle case per anziani con una valutazione empirica delle prestazioni produttive degli istituti.

A seguire inquadriamo meglio il problema economico che fa da sfondo a questa analisi, contestualizzando la ricerca rispetto alle caratteristiche del settore delle case per

anziani svizzere. La parte centrale del capitolo sarà dedicata alla precisa definizione degli obiettivi della tesi seguita da una breve riflessione sui problemi metodologici da risolvere nella sezione dedicata alle verifiche empiriche.

A chiusura di questa introduzione presenteremo brevemente l'evoluzione del concetto di efficienza nella letteratura e illustreremo la struttura del lavoro, descrivendo sinteticamente gli argomenti contenuti nei capitoli che compongono la presente ricerca.

1.1 Il problema economico

In Svizzera la spesa sanitaria ogni anno cresce in modo marcato ed ha raggiunto nel 2006 un livello pari a 52,7 miliardi di franchi¹. Questa cifra assume particolare rilievo se confrontata con un indicatore della produzione del paese (PIL), che nel medesimo anno ammontava a (circa) 487 miliardi di franchi. Il costo complessivo del sistema sanitario equivale quindi al 10,8% del PIL svizzero e, nel confronto internazionale, questo indicatore posiziona la Svizzera al secondo posto (a pari merito con la Francia), dietro agli Stati Uniti (15,0%) e davanti alla Germania (10,2%).

I costi generati dall'intero sistema sanitario svizzero si possono ripartire in tre grandi categorie secondo il criterio dei fornitori di prestazione: (1) ospedali, (2) servizi ambulatoriali e (3) istituti non ospedalieri. Nell'ultima categoria confluiscono istituti ed associazioni che, pur occupandosi di attività molto diverse, erogano prestazioni sanitarie (ad esempio i centri di fisioterapia, le imprese che offrono cure a domicilio), sociali (ad esempio consultori, centri per minorenni) e sociosanitarie (case per anziani, istituti per il recupero delle tossicodipendenze, istituti per invalidi e portatori di handicap).

In Svizzera, l'importanza del settore sociosanitario in termini di spesa è tutt'altro che trascurabile e nel 2006 rappresentava circa un quinto della spesa sanitaria complessiva. All'interno del settore sociosanitario la maggior parte delle spese è dovuta alle case per anziani che nel 2006 generavano circa il 12% del totale della spesa sanitaria. In Svizzera il settore dell'assistenza agli anziani è caratterizzato dalla presenza di tre tipologie di strutture diverse sotto il profilo della forma istituzionale: case per

¹ Tutti i dati sulla sanità svizzera presentati in questo capitolo provengono dall'Ufficio Federale di Statistica e si riferiscono all'anno più recente a disposizione.

anziani pubbliche, case per anziani private non profit e case per anziani for profit. La presenza dello Stato non si limita, tuttavia, alla sola erogazione diretta del servizio (case per anziani pubbliche), ma comprende anche l'attività di sussidio alle istituzioni private (soprattutto quelle non profit) e di regolamentazione.

Il ruolo dello Stato assume grande centralità in questo settore e, dal punto di vista della regolamentazione, le case per anziani vengono associate per analogia al settore ospedaliero, condividendone lo stesso regime di attenzione ai costi. La spinta politica ad operare con una maggiore efficienza produttiva, ossia a ridurre o eliminare le inefficienze di costo e di scala, è ancor più marcata a causa del progressivo invecchiamento della popolazione e della coscienza che i costi per la presa a carico sanitaria e sociale aumentano considerevolmente con l'aumentare dell'età.² Nei prossimi decenni saremo verosimilmente confrontati con un aumento notevole della domanda di prestazioni sanitarie e assistenziali per persone anziane. In considerazione di questo fatto, si può affermare che l'efficienza delle case per anziani è un tema capace di risvegliare, anche in Svizzera, un notevole interesse sul piano politico, pari a quello dell'efficienza nel settore ospedaliero.

Un aspetto critico riguarda la disponibilità di posti letto nelle case per anziani, che risulta insufficiente, creando difficoltà (liste d'attesa) all'utenza nella ricerca di un posto in una casa per anziani. Viste le attuali tendenze demografiche e l'aumento della speranza di vita, il problema è destinato ad acuirsi se non verrà potenziata l'offerta di servizi. Le pressioni dal lato della domanda stanno stimolando la risposta da parte della società, sotto forma di iniziative private o pubbliche.

La tendenza in atto si può facilmente osservare attraverso i dati statistici del recente passato, che evidenziano una crescita del numero totale di posti letto per persone anziane offerti a livello svizzero.³ Le dichiarazioni programmatiche dei responsabili della politica sanitaria a livello cantonale, ed i numerosi progetti comunali

² Secondo Bjørnerud et al. (2006), le proiezioni per il 2050 relative al settore *Long Term Care*, indicano un impatto sui costi causato dall'invecchiamento della popolazione compreso tra 0,7 e 1,4 punti percentuali del PIL.

³ Si veda la pubblicazione dell'Ufficio Federale di Statistica Stat-Santé (anni 2000-2008) sul rilevamento degli istituti non ospedalieri, che può essere scaricata gratuitamente dal seguente indirizzo: <http://www.bfs.admin.ch/bfs/portal/fr/index/themen/14/03/02/data/01.html>.

per la costruzione di nuovi istituti per anziani, sono un ulteriore segnale della rilevanza dell'argomento.

L'interesse ad effettuare una valutazione empirica dell'efficienza produttiva è quindi molto elevato e risponde all'esigenza concreta di avere un supporto per poter prendere le decisioni in modo informato. Per questa ragione molti studi empirici del settore si sono concentrati sulla misurazione delle prestazioni produttive, con obiettivi marcatamente di policy.⁴ Inoltre, la letteratura empirica, maggiormente concentrata sulla misurazione ed analisi dell'efficienza nel settore sanitario e dei servizi pubblici, ha messo in evidenza molte analogie sul fronte metodologico, ma anche alcune differenze sul piano dei risultati.⁵ La discussione è dunque ancora aperta, valorizzando il significato di un nuovo studio empirico sul tema.

A questo proposito la misurazione delle economie di scala appare (*ex ante*) di grande rilevanza, visto che in Svizzera la maggioranza delle case per anziani presenta una dimensione ridotta (inferiore a 60 posti letto). Inoltre, gli istituti per anziani, pur svolgendo un'attività molto simile (concettualmente identica), mostrano un livello dei costi unitari per giornata di residenza piuttosto eterogeneo. Con queste premesse, in chiave dinamica, conoscere il valore delle economie di scala e della dimensione ottima di un istituto (in termini di posti letto) fornisce un contributo importante, soprattutto in sede di progettazione di una nuova struttura, dove un corretto dimensionamento della casa per anziani permette di evitare sprechi per molti anni a venire.

Se il problema della scala produttiva è più rilevante per gli istituti di nuova costruzione, l'efficienza di costo è un tema che interessa tutte le imprese del settore, poiché si riferisce allo spreco di risorse connesso con l'attività produttiva corrente. A prescindere da considerazioni legate all'età di un istituto (vecchio vs. nuovo) o alla forma istituzionale (pubblica vs. privata), le imprese che erogano servizi agli anziani sono confrontate con un vincolo di bilancio più o meno stringente che dovrebbe portarle ad operare secondo criteri di efficienza. Determinare empiricamente la *best practice* del settore permette di confrontare le singole prestazioni delle case per anziani, in modo

⁴ Si vedano ad esempio Nyman (1983), Gertler and Waldman (1992), Gulley and Santerre (2003) e Farsi and Filippini (2008).

⁵ Si vedano ad esempio Jacobs et al. (2006), Fox (2002). Per una recente rassegna sui temi dell'efficienza nei servizi pubblici in Italia si vedano gli studi della Banca d'Italia (2003, 2008).

relativo (tra loro) ed assoluto (rispetto alla frontiera di costo minimo), fornendo indicazioni utili tanto ai dirigenti degli istituti, quanto al regolatore.

Con riferimento all'attività dello Stato svolta nel ruolo di produttore di servizi, più che in quello di arbitro del mercato, emerge un potenziale conflitto di obiettivi tra equità ed efficienza che rende interessante una valutazione empirica delle prestazioni produttive delle case per anziani. Nell'ottica della società, non è possibile stabilire una scala di priorità rispetto ai problemi di efficienza e di equità menzionati senza fare ricorso a giudizi di valore. Le scelte normative, che la società è chiamata a fare in questo settore, potrebbero quindi essere agevolate dalla conoscenza dei risultati di un'analisi positiva sull'efficienza.⁶

In questo settore, ad esempio, una delle argomentazioni principali per legittimare l'intervento pubblico risiede proprio nella particolare natura del servizio di assistenza agli anziani, definito bene meritorio.⁷ Dietro al concetto di bene meritorio ritroviamo, quindi, una serie di valori quali: l'universalità nell'accesso al servizio, l'equità territoriale, l'equità (spesso intesa come progressività) di finanziamento, l'equità di trattamento in termini di cure ricevute, che lo Stato cerca di realizzare con il proprio intervento.

Nel settore dell'assistenza agli anziani l'equità di finanziamento esaspera il problema dell'eccesso di domanda ponendo ancora maggiori pressioni sulle finanze pubbliche (e quindi sull'importanza dell'efficienza di costo), mentre l'equità di accesso solleva alcuni problemi economici legati alla dimensione degli istituti.

In merito a quest'ultimo punto rileviamo come la realizzazione dell'equità di accesso implichi una distribuzione dei servizi sociosanitari relativamente capillare sul territorio, in modo che ogni potenziale utente, la cui mobilità è tipicamente molto bassa,

⁶ Per un approfondimento su questi temi si veda Borzaga et al. (1998)

⁷ Il termine nasce ed è utilizzato nell'ambito dell'economia pubblica (Musgrave, 1959), per indicare quella particolare categoria di beni e servizi il cui consumo è ritenuto dalla collettività particolarmente desiderabile e meritevole, ma che il consumatore individuale non riconosce. La promozione di questi beni necessita quindi un approccio paternalistico da parte dello Stato.

possa raggiungere il servizio con facilità.⁸ In presenza di economie di scala l'obiettivo dell'efficienza produttiva si scontra con l'equità territoriale. Il principio dell'efficienza di scala non tiene conto dell'eventuale bassa densità della domanda, caratteristica delle zone periferiche e delle valli, prescrivendo l'organizzazione dell'offerta di servizi per anziani in pochi istituti di grandi dimensioni. Per contro, in simili situazioni, l'equità territoriale si realizza quando la distanza tra gli ospiti e le strutture presenti sul territorio è contenuta, prediligendo un'offerta frammentata in molte case per anziani di piccole dimensioni. Per fare un esempio estremo, la chiusura in regioni periferiche di piccole case per anziani, se da un lato riduce l'inefficienza di scala, dall'altro acuisce il problema dell'equità, in quanto la prospettata chiusura potrebbe rendere meno agevole l'accesso ai servizi residenziali alla popolazione anziana di quelle regioni.

La stima dell'efficienza produttiva può quindi fornire elementi di riflessione importanti (misurazione delle economie di scala) per contribuire alla presa di decisioni consapevoli. La riflessione appena terminata sui trade-off che esistono tra equità ed efficienza di scala in questo settore, non deve tuttavia trarre in inganno. In questa tesi di dottorato ci occuperemo quasi esclusivamente dell'efficienza produttiva delle case per anziani, limitando al minimo indispensabile le considerazioni legate all'equità o ad altri aspetti. La nostra è quindi un'analisi parziale, le cui raccomandazioni, prima di poter essere tradotte in una strategia di politica sanitaria, dovranno essere valutate in considerazione del loro probabile impatto su altre dimensioni rilevanti (equità, qualità, regolamentazione, consenso politico, etc.) del problema dell'erogazione di servizi per anziani. Fatta questa precisazione, desideriamo evidenziare gli aspetti legati alla misurazione dell'efficienza produttiva, problema che assume un ruolo cruciale in una ricerca empirica.

Da un punto di vista metodologico, un'analisi empirica dell'efficienza produttiva di un campione di case per anziani presenta alcuni problemi, che devono essere affrontati. In particolare, un problema molto importante è quello dell'eterogeneità dei processi produttivi e dell'organizzazione e regolamentazione delle case per anziani. La

⁸ Il problema della mobilità non si può analizzare compiutamente guardando solo il lato della domanda. Intuitivamente le ragioni che frenano gli anziani ad allontanarsi dal domicilio sono riconducibili al desiderio di vicinanza con il territorio e con l'ambiente familiare di provenienza. Occorre però considerare che nel settore sono presenti lunghe liste di attesa (eccesso di domanda), la cui gestione decentralizzata scoraggia ulteriormente la mobilità.

stima di una funzione di costo si basa, infatti, sul confronto delle informazioni quantitative rilevate dalla banca dati, in merito ai singoli istituti presenti nel campione. Le singole osservazioni vengono confrontate solamente rispetto alle variabili incluse in un modello di costo con la conseguenza che, da un punto di vista strettamente empirico, possiamo distinguere l'eterogeneità osservata da quella non osservata. La prima rende conto delle differenze tra gli istituti rispetto alle dimensioni incluse nell'analisi e misurate dal ricercatore attraverso le variabili scelte. La seconda forma di eterogeneità, che in misura più o meno grande è sempre presente in una banca dati, riguarda tutti quegli aspetti che, pur caratterizzando il processo produttivo di una casa per anziani, non sono stati rilevati durante la fase di raccolta dati o, peggio, non si riescono a misurare.

La qualità, ad esempio, è un elemento difficile da misurare, ma che potrebbe incidere sul livello dei costi di una casa per anziani. L'impatto sulle prestazioni produttive di un istituto, provocato dal contesto sociale, geografico o regolatorio, è un altro fattore che poco si presta ad essere quantificato. Nelle ricerche empiriche basate su dati secondari, rilevati quindi con finalità non sempre coincidenti, può risultare impossibile reperire queste informazioni ed anche surrogarle con variabili *proxy*. Nello specifico di questa ricerca, esistono delle chiare difficoltà nel quantificare in modo preciso la qualità delle prestazioni erogate agli anziani e gli aspetti legati alla regolamentazione (problema dell'osservabilità di un'informazione), oltre all'impossibilità di reperire quelle informazioni che non sono state rilevate nell'inchiesta originale svolta dall'UFS.

In ogni caso, anche ammesso di avere a disposizione dati molto dettagliati, si pone un problema di specificazione del modello a causa dal trade-off che esiste tra precisione nel modello e comparabilità delle osservazioni. Esasperando il discorso, potremmo arrivare ad una specificazione empirica così particolareggiata che ogni singola osservazione nel campione sarebbe unica e non comparabile con tutte le altre! In Svizzera, per quanto attiene alla regolamentazione del settore delle case per anziani, il problema emerge in modo evidente. L'intervento dello Stato si configura, infatti, come un insieme di 26 sistemi di regolamentazioni cantonali, con basi legali e filosofie

sensibilmente diverse fra loro.⁹ Queste differenze, frutto della sovranità e dell'autonomia dei cantoni contemplate dalla Costituzione federale, determinano un panorama molto eterogeneo delle condizioni quadro nelle quali operano i diversi istituti.

Appare dunque ragionevole, nonostante la marcata caratterizzazione cantonale di questo settore, descrivere a livello nazionale il processo produttivo di una generica casa per anziani svizzera. L'ipotesi sottostante è che le differenze presenti nei dati, ma non specificate dal modello di costo, non siano tali e tante da impedire un confronto. Nella pratica, quindi, già in sede di specificazione del modello si accetta la presenza di un certo grado di eterogeneità non osservata, che impone di considerare gli aspetti metodologici legati a questo problema, poiché la mancata considerazione delle differenze presenti, ma non osservabili, potrebbe alterare i risultati delle stime. Il problema dell'eterogeneità non osservata è ancora più rilevante per la stima di una frontiera (e non semplicemente una funzione) di costo, strumento utilizzato per confrontare, dopo aver tenuto conto di tutte le differenze considerate dal modello, i livelli di efficienza dei diversi istituti. La stima dei punteggi di efficienza può risultare quindi molto sensibile al modo nel quale la presenza di eterogeneità viene considerata dal modello econometrico.

Nel settore sociosanitario, una discussione sui problemi econometrici riscontrati nella stima dell'efficienza di costo in presenza di eterogeneità non osservata si può trovare in due recenti studi di Farsi et al. (2005) e Farsi e Filippini (2008). Gli autori, considerata un'identica funzione di costo, mettono a confronto i risultati delle stime ottenute con diversi modelli econometrici, mostrando l'importanza di utilizzare una specificazione econometrica in grado di rilevare e (almeno parzialmente) di filtrare l'eterogeneità non osservata. Una seconda indicazione da considerare è che l'eterogeneità non osservata manifesta i suoi effetti soprattutto sulle stime dell'efficienza di costo che dipendono in modo cruciale dalla struttura assegnata ai residui della regressione dai diversi modelli di frontiera di costo.

Vista la natura e le caratteristiche del settore dell'assistenza agli anziani in Svizzera e considerati i limiti della banca dati a nostra disposizione, ci troveremo

⁹ Per una recente panoramica sulle caratteristiche e sulle differenze cantonali del sistema sanitario svizzero, si veda il rapporto OECD (2006) "Reviews of Health System – Switzerland".

confrontati certamente con un certo grado di eterogeneità non osservata. Rispetto a tale problema, la discussione sulla sensibilità dei risultati al tipo di metodo utilizzato costituisce un aspetto centrale in un'analisi empirica dell'efficienza produttiva ed è per questo motivo che discuteremo a fondo, sia da un punto di vista metodologico, sia da un punto di vista empirico, il problema dell'eterogeneità non osservata. L'affidabilità e la precisione di uno strumento di misura sono requisiti fondamentali per condurre analisi empiriche di tipo positivo che, in modo più o meno accentuato, potrebbero tradursi in raccomandazioni di policy o precetti normativi. A questo riguardo, il dibattito sui vantaggi e gli svantaggi connessi all'utilizzo dei diversi approcci metodologici per misurare l'efficienza produttiva è da sempre molto sentito ed è tutt'ora in corso. Mentre in passato, però, la distinzione tra i diversi filoni di ricerca era piuttosto marcata, oggi assistiamo invece ad uno sforzo verso la convergenza ¹⁰

Alla luce delle considerazioni presentate, emergono in modo chiaro gli aspetti sui quali si focalizzerà maggiormente il lavoro e le motivazioni che lo animano, mentre restano da specificare in modo più concreto gli obiettivi puntuali di questa ricerca.

1.2 Gli obiettivi della tesi

La presente tesi di dottorato rientra nell'ambito della ricerca economica applicata, affiancando gli obiettivi a carattere prevalentemente scientifico-metodologico, con altri maggiormente indirizzati alla verifica empirica di determinate caratteristiche economiche del settore delle case per anziani svizzere.

L'analisi empirica sull'efficienza produttiva delle case per anziani svizzere sarà impostata secondo un approccio di tipo positivo, cercando quindi di presentare ed interpretare i risultati ottenuti senza scendere sul terreno dei giudizi di valore (analisi normativa). Il peso che lo Stato ricopre in questo mercato rende difficile la separazione tra l'analisi economica del settore delle case per anziani e le considerazioni di politica sanitaria che ne scaturiscono. La presenza del federalismo, poi, rende quasi inevitabile fare riferimento alle diverse situazioni cantonali ed alle scelte pubbliche che le hanno determinate.

¹⁰ Per una rassegna recente di entrambi gli approcci si veda ad esempio Fried et al. (2008).

Il presente lavoro, tuttavia, né si orienta verso un'analisi minuziosa della politica sanitaria in materia di case per anziani, né ha intenti di suggerire questa o quella ricetta per migliorare il sistema. Detto con una terminologia inglese, questa tesi di dottorato non è *policy oriented*, anche se l'analisi empirica sull'efficienza delle case per anziani produce dei risultati che, inevitabilmente, contengono molte implicazioni rilevanti per la politica economica di questo settore.

L'oggetto di studio principale di questa tesi si concentra sull'analisi dell'efficienza produttiva nel settore delle case per anziani svizzere. Per raggiungere questo obiettivo empirico abbiamo stimato un modello di costo impiegando alcune recenti tecniche econometriche ad una banca dati (panel) messa a disposizione dall'Ufficio Federale di Statistica. In pratica si cerca di costruire una rappresentazione generale del settore delle case per anziani (modello di costo) che sia in grado di descrivere la tecnologia produttiva tipica del settore, fornendo risposte, empiricamente fondate, ad una serie di domande di carattere economico.

La prima informazione che desideriamo “estrarre” dai dati riguarda la determinazione dei parametri che contraddistinguono una funzione di costo per una generica casa per anziani. Si tratta quindi di risolvere i problemi legati all'econometria applicata (specificazione del modello teorico, scelta delle variabili da utilizzare, forma funzionale, etc.), per riuscire a spiegare in modo quantitativo i fattori che determinano i costi delle case per anziani. La stima empirica della funzione di costo permette, infatti, di verificare l'eventuale presenza di economie di scala in questo settore.¹¹

La stima di una funzione (frontiera) di costo consente, inoltre, di determinare empiricamente l'efficienza di costo delle singole case per anziani e questa valutazione costituisce un secondo importante obiettivo della ricerca. L'efficienza di costo, come avremo modo di chiarire in seguito, esprime la capacità di utilizzare al meglio le risorse per produrre al costo minore possibile una determinata quantità di prodotto. Nel settore delle case per anziani significa riuscire a garantire un certo numero di giornate di

¹¹ Nella letteratura economica della produzione si distingue tra economie di scala ed economie di dimensione. Questa differenza concettuale e terminologica, tratta da Chambers (1988), sarà presentata nel capitolo 3.

residenza ad un prefissato standard qualitativo dei servizi erogati, generando il più basso costo totale possibile.

Per misurare il livello di efficienza raggiunto dai singoli istituti in accordo con la definizione appena enunciata, occorre conoscere, oltre al costo totale fornito dai dati, il termine di paragone, espresso come il livello di costo minimo teorico che si potrebbe raggiungere con l'attuale livello produttivo. Il valore del limite inferiore dei costi, che si definisce frontiera di costo e che si può ricavare dai risultati delle stime econometriche, consente di misurare quante risorse vengano "sprecate" nel processo produttivo di ogni casa per anziani. Riuscire a determinare il livello di inefficienza di costo, sia per il settore sia per i singoli istituti, rappresenta un altro tassello per la valutazione dell'efficienza produttiva delle case per anziani svizzere.

La valutazione empirica dell'efficienza produttiva non può prescindere dall'impiego di uno strumento di misura. Da un punto di vista metodologico, questa tesi si prefigge di esplorare i vantaggi nella misurazione dell'efficienza produttiva dovuti all'utilizzo di alcuni modelli econometrici proposti recentemente da Greene (2005). Si tratta di alcuni modelli (*panel data*) di frontiere stocastiche che permettono di considerare maggiormente il problema dell'eterogeneità non osservata, problema molto presente nel settore delle case per anziani.¹²

La stabilità e la precisione delle stime sono, infatti, un punto cruciale per ogni valutazione empirica dal momento che i risultati ottenuti con un modello econometrico potrebbero cambiare utilizzando un modello diverso. Il problema risulta ancora più accentuato in presenza di forte eterogeneità non osservata, aspetto che alcuni modelli econometrici non sono in grado di considerare. Per questa ragione è interessante operare un confronto tra i risultati ottenuti con diversi modelli panel per verificarne l'affidabilità in un contesto empirico caratterizzato dal problema dell'eterogeneità non osservata.

L'immediata conseguenza che discende da una misurazione del livello di inefficienza di costo degli istituti per anziani è la possibilità di formulare e verificare empiricamente alcune spiegazioni circa i fattori che determinano e influenzano i valori

¹² Per un'applicazione riferita al contesto sanitario e sociosanitario svizzero si veda Farsi and Filippini (2004; 2008).

di stimati. Alcuni confronti tra inefficienza e determinate caratteristiche degli istituti risultano molto intuitivi come, ad esempio, il livello di inefficienza medio registrato nei diversi cantoni o l'impatto della forma istituzionale o della dimensione (misurata dal numero di posti letto) sull'inefficienza.

Tuttavia, cogliere il legame tra l'eccesso di costo ed altre caratteristiche del processo produttivo può rivelarsi difficile e, metodologicamente parlando, ricco di insidie.¹³ Senza sottrarci di principio alla responsabilità di effettuare questo delicato esercizio di caratterizzazione dell'inefficienza stimata, cercheremo di inquadrarlo in una cornice adeguatamente circoscritta, per evitare di conferirgli un peso superiore a quello che la precisione delle stime, ed il buon senso, autorizzerebbero a darle. Alla luce di quanto appena esposto sembra opportuno riassumere, in forma schematica ed interrogativa, i 6 quesiti principali ai quali si cercherà di dare risposta:

- D1.** *Gli istituti per anziani attivi in Svizzera sono contraddistinti mediamente da un elevato grado di efficienza di costo o sono per lo più inefficienti?*
- D2.** *Esistono economie di scala in questo settore? (Se sì) Qual è la dimensione ottimale, in termini di posti letto, di una casa per anziani?*
- D3.** *I risultati delle stime dei parametri della funzione di costo sono molto sensibili alla scelta del modello econometrico oppure forniscono indicazioni molto simili sulla tecnologia produttiva del settore?*
- D4.** *L'impiego delle tecniche econometriche proposte da Greene (2005), considerando maggiormente l'eterogeneità non osservata, permette di misurare l'inefficienza di costo in modo più preciso?*
- D5.** *Dai risultati dell'analisi empirica emergono relazioni statisticamente significative tra alcune caratteristiche delle case per anziani ed il rispettivo livello di inefficienza di costo?*
- D6.** *Quali sono le principali implicazioni di politica sanitaria connesse con i risultati dell'analisi dell'efficienza produttiva svolta per il settore delle case per anziani svizzere?*

¹³ Per una discussione sul problema di determinare i fattori che spiegano l'inefficienza, seguendo l'approccio a "due stadi" (*two-stage regression*), si veda ad esempio Deprins and Simar (1989).

Scorrendo gli obiettivi fissati nel presente lavoro, emerge come il contenuto informativo di un'analisi sull'inefficienza produttiva coinvolga molti aspetti, afferenti ad aree diverse. Troviamo, infatti, aspetti metodologici (econometria dei panel data e frontiere stocastiche di costo), misurativi (economie di scala e inefficienze di costo), valutativi (*benchmarking* degli istituti e livello medio di efficienza del settore), di politica sanitaria (raccomandazioni basate sull'analisi positiva) e formativi (divulgazione della cultura del confronto e del miglioramento). La ricerca si profila, tuttavia, come un'analisi parziale del settore delle case per anziani, con un marcato interesse agli aspetti empirico-metodologici legati all'efficienza produttiva. Per questa ragione molti temi meritevoli di approfondimento come l'equità, la qualità delle cure, l'efficacia dell'intervento pubblico, non troveranno spazio in questa tesi, nonostante la loro importanza. Come più volte ribadito, questa tesi non mira ad indagare il settore delle case per anziani a 360°, bensì cerca di legare, in una presentazione organica, gli obiettivi di ricerca menzionati sopra, presentandoli con chiarezza e lucidità, ma con la consapevolezza che si tratta in ogni caso di una porzione di un puzzle molto più grande.

Nel prossimo paragrafo, visto il carattere prevalentemente applicato del lavoro, desideriamo anticipare alcuni aspetti legati alla metodologia utilizzata nella parte empirica della tesi.

1.3 La metodologia

Le tecniche a disposizione per quantificare l'inefficienza produttiva di un settore economico (e delle imprese che lo costituiscono) sono numerose, anche se gli approcci maggiormente seguiti si possono raggruppare in due famiglie: (1) i metodi non parametrici e (2) i metodi parametrici. La caratteristica che accomuna i due approcci è quella di specificare un modello di produzione o di costo ed applicare ai dati disponibili uno strumento quantitativo (econometria per i metodi parametrici e programmazione lineare per quelli non parametrici), per valutare le prestazioni economiche e/o le caratteristiche della tecnologia produttiva.¹⁴

¹⁴ La discussione sulla scelta del metodo è presentata nel capitolo 3. Per una rassegna dettagliata dei metodi per la misurazione empirica delle prestazioni produttive si veda Fried et al. (2008).

In questa tesi, utilizzando esclusivamente l'approccio parametrico, definiremo un modello di costo econometrico per una generica casa per anziani svizzera, con l'obiettivo di stimare i parametri associati alle variabili che lo costituiscono e che ne descrivono il processo produttivo. Il fine ultimo dei metodi parametrici è stabilire e quantificare l'effetto sui costi, generato da un eventuale cambiamento delle variabili esplicative del modello.

Da un punto di vista empirico, parlare di inefficienza produttiva significa determinare i parametri di una funzione di costo che esprima la minor spesa possibile, associata ad ogni livello di output ed alle caratteristiche osservate nel processo produttivo. La stima di questa funzione di costo, denominata frontiera di costo, permette quindi di confrontare il valore dei costi effettivamente sostenuti da ogni singola impresa con quelli previsti teoricamente (sulla base del modello e dei parametri stimati) per un'omologa impresa efficiente. Inoltre, i parametri della funzione stimata vengono utilizzati per calcolare le economie di scala nel settore delle case per anziani.

In altre parole, dalla stima econometrica dei parametri della funzione di costo e dai residui della regressione è possibile ricavare una misurazione dell'inefficienza di costo delle singole imprese. Nascosti dietro a questa semplice visione del processo empirico di stima di una frontiera si trovano tre elementi di grande importanza: (1) la specificazione del modello, (2) la scelta della forma funzionale e (3) la scelta del modello econometrico da utilizzare per le stime. Rimandando ai capitoli 4 e 5 la discussione approfondita di questi aspetti, desideriamo introdurre alcune considerazioni generali di carattere econometrico, per concludere il paragrafo discutendo il problema della specificazione del modello di costo.

In letteratura gli approcci sviluppati per la stima parametrica delle frontiere di costo si possono suddividere sostanzialmente in due grandi categorie: metodi stocastici e metodi deterministici.¹⁵ Questi ultimi considerano la differenza tra il costo empiricamente osservato (dati) e quello teoricamente stimato (previsione del modello) come inefficienza di costo.

¹⁵ Per una presentazione delle metodologie e delle tecniche di stima si vedano ad esempio: Jacobs et al. (2006), Daraio et al. (2008) e Fried et al. (2008).

In altre parole, i costi che il modello non è in grado di giustificare, catturati quindi dai residui della regressione, vengono considerati come uno spreco di risorse. I metodi deterministici non ammettono, quindi, che i residui della regressione econometrica possano comprendere anche una componente stocastica non attribuibile alla mancata minimizzazione dei costi.

Come discusso in precedenza, sembra improbabile che un modello di costo riesca a descrivere in modo esatto le caratteristiche del processo produttivo di tutte le imprese osservate. In aggiunta a questo, occorre sottolineare come le prestazioni potrebbero essere influenzate da fattori casuali che sfuggono al controllo delle imprese. Infine, le osservazioni empiriche, attraverso le quali si determina la frontiera di costo, potrebbero contenere alcune imprecisioni dovute al processo di raccolta e rielaborazione dei dati.

Per superare questo limite alla fine degli anni '70 si sviluppa, partendo dai lavori di Aigner et al. (1977) e Meeusen and Van der Broeck (1977), un filone di ricerca interessato agli aspetti econometrici specificamente legati alla corretta misurazione dell'inefficienza, noto con il termine di analisi delle frontiere stocastiche (SFA), dall'inglese *Stochastic Frontier Analysis*.¹⁶

Queste tecniche, che si inseriscono nell'ambito dell'econometria dei *Panel Data*, impiegano dati riferiti ad unità statistiche distinte (dati *cross-section*), ma osservati per una serie successiva di periodi temporali (dati *time-series*).¹⁷ L'incrocio di queste due dimensioni permette di estrarre dalla funzione di costo, anche informazioni sull'inefficienza produttiva.

Tutti i modelli sviluppati nell'area della SFA partono dal presupposto che i residui di una regressione celino informazioni di diversa natura, concernenti in parte l'eterogeneità non osservata nel modello (problema delle variabili esplicative omesse), in parte l'inefficienza delle singole imprese ed in parte un disturbo stocastico. Grazie all'uso di queste tecniche, si riesce a definire una frontiera di costo che funga da termine di paragone per valutare empiricamente l'inefficienza delle imprese, pur tenendo conto della natura stocastica del processo produttivo.

¹⁶ Per una presentazione sul tema delle frontiere stocastiche (SFA) si veda ad esempio Kumbhakar and Lovell (2000).

¹⁷ Per una analisi dettagliata delle tecniche econometriche panel si veda ad esempio Baltagi (2008).

Una rassegna dei principali modelli di frontiera stocastica, sviluppati ed applicati nelle verifiche empiriche dell'efficienza produttiva, si trova nel capitolo 4, mentre qui di seguito desideriamo discutere le linee di sviluppo della recente letteratura sulla stima di frontiere di costo basate su dati panel, che toccano più da vicino gli aspetti metodologici trattati nella tesi. Ci riferiamo sostanzialmente a due problemi: (1) la specificazione del comportamento dell'inefficienza di costo rispetto al tempo e (2) il ruolo dell'eterogeneità non osservata.

In merito al primo aspetto, viene fatta una distinzione tra i modelli che considerano costante nel tempo l'inefficienza di ogni singola impresa e modelli che non pongono questa condizione.

Tra i modelli ad inefficienza costante nel tempo, in letteratura sono molto impiegati il *Maximum Likelihood Random Effect Model* (ML-REM), proposto da Pitt and Lee (1981), il *Fixed Effect Model* (FEM) ed il *Random Effect Model* (REM), entrambi sviluppati da Schmidt and Sickles (1984).¹⁸ I due modelli ad effetti casuali REM e ML-REM si possono distinguere rispetto alla tecnica di stima, rispettivamente *Generalized Least Square* (GLS) e *Maximum Likelihood Estimation* (MLE), e rispetto alle ipotesi sulla distribuzione del termine di errore. Dal punto di vista concettuale sono invece assimilabili, poiché considerano le caratteristiche individuali delle singole imprese come la manifestazione di una variabile casuale (da qui il termine "effetti casuali").

Discorso diverso va fatto per il modello ad effetti fissi FEM, noto anche come *Least Square Dummy Variables* (LSDV). In questo modello si ipotizza una specificità delle caratteristiche individuali delle imprese, catturata sostituendo all'intercetta comune un vettore di *dummies*. Il termine "effetti fissi" è quindi riferito all'unicità del valore associato ad ogni singola impresa e determinato dalla stima dei parametri delle intercette individuali.

La famiglia dei modelli *Panel* che considera l'inefficienza delle singole imprese come una grandezza che varia nel tempo è molto articolata. In questa sede ci preme solamente citare alcuni lavori di riferimento quali Cornwell et al. (1990), Kumbhakar (1990) e Battese and Coelli (1992).

¹⁸ Per la presentazione e discussione di questi modelli si rimanda al capitolo 4.

Il modello di Cornwell et al. (1990) si può considerare un'estensione dei modelli FEM e REM, poiché ne condivide la strategia di stima GLS e l'assenza di ipotesi restrittive sulla distribuzione del termine di errore. Dal punto di vista della procedura di stima, i modelli di Kumbhakar (1990) e Battese and Coelli (1992) sono invece assimilabili al ML-REM, poiché si basano sulla tecnica di massima verosimiglianza. I tre modelli descritti, pur caratterizzando in modo diverso la struttura temporale dell'inefficienza, riescono a determinare empiricamente il valore dell'inefficienza (variabile nel tempo) delle singole imprese.

Con riferimento alla famiglia di modelli ad inefficienza variabile nel tempo, Greene (2005) ha recentemente proposto alcune specificazioni, definite dall'autore "*True*" *Random Effect Model* (TREM) e "*True*" *Fixed Effect Model* (TFEM). Questi modelli permettono di affinare la precisione delle stime individuali di inefficienza (variabile nel tempo), filtrando parzialmente il contributo dell'eterogeneità non osservata, ma supposta costante nel tempo.

La struttura di questi modelli è simile a quella del ML-REM, ma in questo caso il termine di errore contempla sia una parte simmetrica (*white noise*), sia la componente di inefficienza (variabile nel tempo e tra gli individui). Inoltre, grazie alle tecniche di simulazione, è possibile scomporre i residui della regressione nelle tre componenti desiderate: inefficienza (variabile nel tempo), errore casuale (variabile nel tempo) ed eterogeneità non osservata (costante nel tempo). Questi modelli risolvono in parte il problema della separazione tra inefficienza ed eterogeneità, anche se tendono a sottostimare la prima, dal momento che ogni forma di inefficienza costante nel tempo viene classificata come eterogeneità.

Il problema di separare l'eterogeneità non osservata dalla stima della componente individuale di inefficienza non è di facile soluzione poiché risulta intimamente legato alla struttura temporale che viene specificata per le due componenti. A ben vedere, sia l'eterogeneità sia l'inefficienza potrebbero essere determinate in parte da fattori strutturali (costanti nel tempo) e in parte da fattori contingenti (variabili nel tempo). Il risultato è la potenziale presenza di 4 componenti distinte che si confondono tra loro e si mescolano con l'elemento stocastico dell'errore, formando il residuo della regressione (unico dato che il ricercatore osserva). I modelli finora proposti dalla letteratura si basano su ipotesi differenti in merito alla struttura dell'inefficienza e presentano

caratteristiche diverse e vantaggi distintivi, ma nessuno è in grado di separare in modo perfetto e stimare le tre componenti (stocastica, inefficienza ed eterogeneità) del termine di errore.¹⁹

Di conseguenza il problema che si pone è quello della stabilità dei risultati, che si ottengono stimando il medesimo modello con diversi modelli econometrici. L'affidabilità e la correttezza delle misure effettuate costituiscono un punto centrale di una ricerca empirica, poiché i valori delle stime rappresentano concetti economici da riferire ad un settore specifico.²⁰ Per poter utilizzare in termini concreti (ad esempio nell'ambito della regolamentazione) i risultati di un'analisi empirica sull'inefficienza produttiva, è necessario che i valori stimati garantiscano un adeguato livello di precisione per evitare di prendere decisioni errate, le cui implicazioni potrebbero avere grande impatto nel mondo reale.

Una possibile soluzione pratica al problema può essere quella di ottenere la "verità" attraverso il confronto tra due misure imprecise, ammesso di poter stabilire la direzione dell'errore. Ripetere l'esercizio di stima della stessa funzione di costo, con gli stessi dati, ma con modelli econometrici diversi, contribuisce a fornire una valutazione dell'inefficienza più affidabile. Alla luce di queste motivazioni, la scelta metodologica si è indirizzata verso un portafoglio di modelli econometrici, simili nell'approccio ma diversi nelle caratteristiche, che presenteremo in dettaglio nel capitolo 4.

La nostra attenzione si rivolge ora a menzionare alcuni aspetti problematici da affrontare nella scelta del modello di costo. Il primo punto riguarda la specificazione dell'orizzonte temporale considerato nel modello di costo. Le funzioni di costo di breve e di lungo periodo descrivono un problema produttivo leggermente diverso. Nel breve periodo, infatti, l'impresa fronteggia dei vincoli (ad esempio la capacità produttiva) e quindi in grado di influenzare solamente i costi variabili rispetto all'output, mentre nel lungo periodo l'impresa può scegliere la combinazione ottimale dei fattori che porta alla minimizzazione dei costi totali.

¹⁹ Una discussione sui vantaggi e svantaggi connessi con ciascuna delle tecniche menzionate si trova nel capitolo 4.

²⁰ Per una discussione sul problema delle condizioni di stabilità delle stime di una frontiera da impiegare nell'ambito della regolamentazione si veda Bauer et al. (1998).

La scelta di stimare una funzione di costo totale o variabile, oltre a descrivere due aspetti diversi del problema produttivo, ha anche delle implicazioni econometriche importanti. Nella misurazione delle economie di scala, che dipende anche dalla specificazione dell'orizzonte temporale scelta, i risultati ottenuti dalle stime di una funzione di costo variabile potrebbero allontanarsi da quelli prodotti specificando una funzione di costo totale.²¹

Un altro aspetto da considerare nella specificazione del modello di costo riguarda la scelta della forma funzionale.²² Le informazioni sulla tecnologia che si possono estrarre dai dati potrebbero essere influenzate dalle caratteristiche della forma funzionale scelta. Ad esempio, l'uso della forma funzionale Cobb-Douglas non permette al valore delle economie di scala di variare rispetto all'output, caratteristica che invece potrebbe essere presente nel settore analizzato. Una forma funzionale flessibile, ad esempio la Translog, supera questo limite, ponendo però un possibile problema di multicollinearità dei regressori, dovuto all'elevato numero di parametri (in modo particolare i prodotti incrociati) da stimare.

L'ultima considerazione metodologica che presenteremo in questa sezione riguarda l'approccio teorico seguito nella specificazione del modello di costo, sottinteso dal tipo di variabili utilizzate nelle stime. Nell'ambito dell'economia applicata si possono distinguere due grandi famiglie di modelli: le specificazioni ad hoc e quelle neoclassiche.

Il termine ad hoc si riferisce al fatto che la relazione tra i regressori e la variabile dipendente è specificata sulla base delle esigenze specifiche della ricerca. Ad esempio, preso un campione di imprese, è possibile spiegare la differenza nei costi unitari, descrivendo questi ultimi come una variabile che dipende da una serie di fattori esplicativi scelti ad hoc dal ricercatore. In pratica si tratta di verificare empiricamente,

²¹ La distorsione introdotta nel valore delle economie di scala, specificando in modo errato l'orizzonte temporale è stata analizzata empiricamente da Aletras (1999), con riferimento al settore ospedaliero.

²² Per una rassegna critica delle forme funzionali maggiormente impiegate nell'analisi empirica delle frontiere stocastiche si veda Kuenzle (2005).

con una regressione multipla, i legami tra le variabili considerate, senza ricondurre la specificazione nell'alveo della teoria microeconomica.²³

Per contro, la funzione di costo neoclassica parte dalla teoria della minimizzazione dei costi dell'impresa, descrivendo empiricamente la struttura dei costi di produzione dell'impresa che, secondo i principi della teoria del duale, devono dipendere dalla quantità degli output e dal prezzo pagato per i fattori produttivi. Seguendo in modo rigoroso la teoria microeconomica, assumendo cioè la minimizzazione dei costi come obiettivo (produttivo) dell'impresa e l'esogenità dell'output e dei prezzi degli input, otteniamo la funzione di costo neoclassica.²⁴

Molti lavori empirici in ambito sanitario affiancano alle variabili del modello di costo (output e prezzi dei fattori produttivi) una serie di variabili "ambientali", che permettono di considerare (almeno parzialmente) l'eterogeneità nelle caratteristiche delle imprese. Il tentativo è quello di conciliare il maggior realismo descrittivo dei modelli ad hoc con la solidità teorica di una funzione basata sulla teoria microeconomica della produzione. In letteratura, questa specificazione "mista" della funzione di costo ha conosciuto diverse denominazioni, ma nel presente lavoro indicheremo questo approccio con il termine generale di "modello microeconomico in senso ampio".

A prescindere dalla specificazione in senso stretto o ampio, la sostanziale differenza delle funzioni neoclassiche rispetto a quelle ad hoc consiste nell'ipotesi, fatta dalle prime, di attribuire alle imprese analizzate un comportamento volto alla minimizzazione dei costi. Questa ipotesi può rivelarsi molto restrittiva o poco realistica, soprattutto quando in un determinato settore esistono vincoli regolatori, imperfezioni di mercato ed una certa componente di inefficienza. Per lavorare con la teoria microeconomica di produzione neoclassica, senza richiedere tutte le condizioni di regolarità della funzione di costo, è stato introdotto il concetto di "funzione di costo *behavioural*".

²³ Per un esempio di questo approccio in ambito ospedaliero si vedano Lave et al. (1972) e Feldstein and Shuttinga (1977), mentre una rassegna sul settore delle case per anziani si trova in Bishop (1980).

²⁴ Per alcune applicazioni in ambito sanitario si vedano ad esempio Cowing and Holtmann (1983), Conrad and Strauss (1983) e McKay (1988).

Nel settore sanitario, in modo particolare in quello ospedaliero, la letteratura ha dibattuto a lungo ed in profondità il problema connesso alla specificazione del modello di costo, riconoscendo il fatto che ogni debolezza nella scelta del modello viene indirettamente catturata al momento di effettuare le stime. La letteratura empirico-econometrica, nonostante gli incessanti sforzi, non ha ancora trovato (e non può trovare) una risposta univoca al problema.²⁵

Per analizzare l'efficienza produttiva delle case per anziani svizzere, abbiamo optato per una funzione Translog, nella quale i costi totali sono specificati secondo un approccio microeconomico in senso ampio. I motivi che ci hanno portato a compiere questa scelta saranno discussi nei capitoli 4 e 5, insieme alla descrizione delle variabili esplicative utilizzate per rappresentare l'attività produttiva del settore analizzato.

Dopo aver introdotto il contesto dal quale nasce la ricerca, i suoi obiettivi e la metodologia, desideriamo presentare una breve digressione sul concetto di inefficienza (soprattutto quella produttiva), declinato secondo alcune correnti di pensiero sviluppatesi nella letteratura.

1.4 L'inefficienza produttiva nella letteratura economica

Il desiderio di misurare empiricamente l'inefficienza produttiva di un settore economico presuppone un aprioristico convincimento del ricercatore che l'inefficienza costituisca un problema reale. Sul piano teorico, tuttavia, il concetto di inefficienza non ha riscosso molto successo nella letteratura neoclassica, che fa comunque da sfondo a questa tesi. Nel paradigma "classico", anche dopo la rivoluzione neoclassica-marginalista, l'inefficienza non viene presa in grande considerazione e la prima vera concessione (sul piano teorico) all'esistenza di comportamenti inefficienti è opera di Pareto che sposta l'accento sul termine "homo", anziché sulla sua supposta virtù di essere "oeconomicus".²⁶ In parole povere gli individui sono razionali nelle intenzioni e

²⁵ Per una discussione sui problemi che si incontrano nella specificazione dell'attività produttiva e nella misurazione corretta dell'efficienza nel settore sanitario, si veda Jacobs et al. (2006).

²⁶ Pareto individua la fonte di irrazionalità dell'agire umano in quell'insieme di sentimenti e pulsioni umane che definisce "residui". Questa parola scelta da Pareto è straordinariamente in linea con la terminologia econometrica sulle frontiere stocastiche, che costruiscono i propri modelli sulla specificazione del termine di errore, empiricamente chiamato residuo della regressione.

irrazionali nella pratica e questo stimola Pareto nella verifica “empirica” delle sue teorie economiche: *“Spinto da desiderio di apportare un complemento indispensabile agli studi di economia politica e soprattutto ispirandomi all’esempio delle scienze naturali, io sono stato indotto a comporre il mio Trattato di sociologia il cui unico scopo – dico unico e insisto su questo punto – è di ricercare la realtà sperimentale per mezzo dell’applicazione alle scienze sociali dei metodi che hanno fatto le loro prove in fisica, chimica, astronomia, in biologia e in altre scienze simili.”*²⁷

Altri autori, pur condividendo il corpus delle teorie microeconomiche, ammettono che nel mondo reale si possono trovare giustificazioni razionali a comportamenti in apparenza poco efficienti. È il caso di Hicks (1936) che, analizzando il mercato di monopolio, sottolinea come con ogni probabilità il comportamento del monopolista tragga *“...maggiori benefici non preoccupandosi di raggiungere una posizione molto vicina al massimo profitto di quanti ne otterrebbe adoperando ogni sforzo possibile per raggiungerla. Il miglior profitto di un monopolista è una vita tranquilla.”*²⁸ L’ipotesi di Hicks è semplice e chiara: l’assenza di competizione oltre ad allontanare i prezzi di vendita dai costi marginali provoca una deviazione sistematica dalla minimizzazione dei costi (inefficienza produttiva e allocativa) causata (e giustificata) dal minor impegno posto nel raggiungimento dell’efficienza.

A partire dagli anni ’50, tra gli economisti che hanno coltivato la teoria microeconomica neoclassica moderna, un numero via via crescente ha continuato ad occuparsi di quella zona d’ombra che la verifica empirica (la presenza di inefficienza) era in grado di proiettare sulla validità della teoria tradizionale (assenza di inefficienza).

Per inquadrare le diverse correnti di pensiero su questo tema e visto il carattere prevalentemente descrittivo di questa introduzione abbiamo idealmente suddiviso lo spazio della letteratura economica in tre grandi aree che, nonostante presentino confini sfumati o sovrapposti, identificano tre diversi modi di concepire l’inefficienza economica. A tale scopo, presentiamo una matrice di posizionamento (Figura 1.1) per illustrare schematicamente i diversi filoni del pensiero microeconomico che hanno contribuito a sviluppare il dibattito sull’efficienza produttiva.

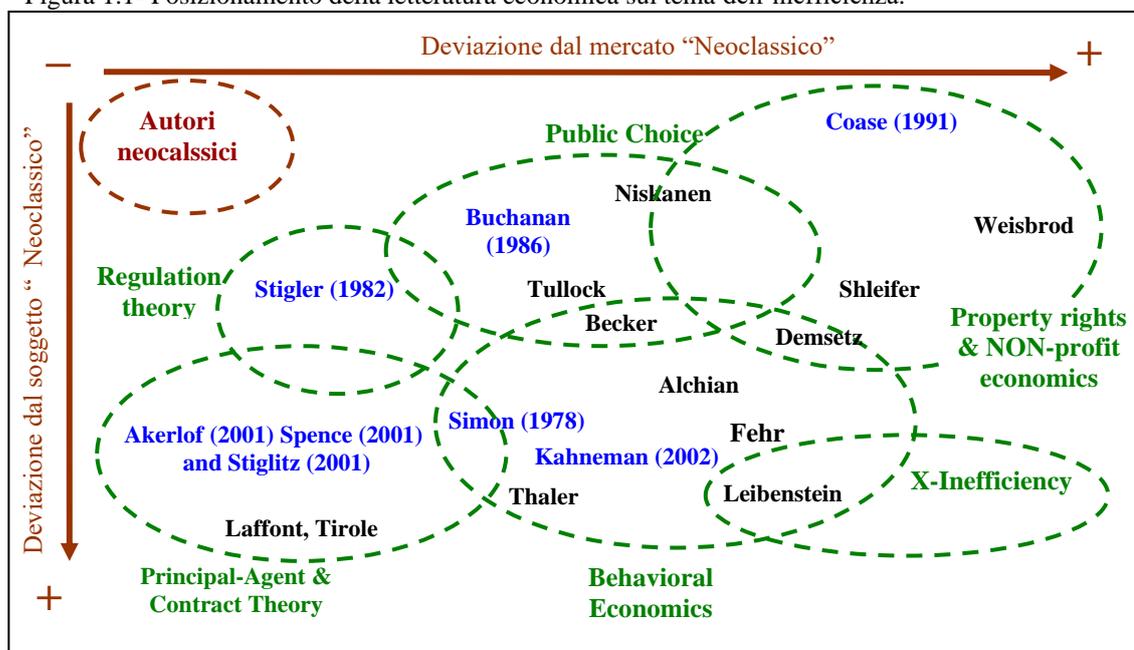
²⁷ Il passaggio è ripreso letteralmente dal trattato di sociologia generale del 1916.

²⁸ Traduzione personale tratta da Hicks (1935), *“The theory of monopoly: a survey”*.

Il diagramma raggruppa tutte le caratteristiche del modello neoclassico nel vertice nord-occidentale della matrice e da queste considera la possibilità di allontanarsene lungo due dimensioni: (1) il funzionamento e le caratteristiche del mercato (direttrice orizzontale) e/o (2) l'insieme di ipotesi relative al processo decisionale degli agenti economici (direttrice verticale).

La prima area, quella che include la teoria della regolamentazione, la teoria dei contratti e del principale agente, si trova in basso a sinistra. In questo gruppo abbiamo inserito alcuni autori che si sono concentrati maggiormente sul concetto di razionalità strumentale, alla base delle decisioni dell' homo oeconomicus, cercando di accomodare alcuni aspetti pratici molto importanti come il rapporto di agenzia, il rischio e l'incertezza, l'asimmetria informativa.

Figura 1.1- Posizionamento della letteratura economica sul tema dell'inefficienza.



Nota: Tra parentesi è indicato l'anno in cui l'autore è stato insignito del premio Nobel per l'economia.

In particolare, la teoria della regolamentazione (che ha in Stigler l'esponente più conosciuto) cerca di incorporare alcuni fenomeni politico-sociali (anche della peggior specie, come la corruzione e l'ingerenza del mondo imprenditoriale nella politica) in grado di spiegare "neoclassicamente" l'evidenza empirica sulla presenza di inefficienza misurata in molti settori. Secondo questa interpretazione, l'inefficienza altro non è che la nostra incapacità come studiosi e analisti di giudicare correttamente le azioni dei soggetti dei quali ignoriamo la reale funzione obiettivo e/o le risorse a disposizione. La teoria del Principale-Agente (*Principal-Agent Theory*) e la teoria dei contratti (*Contract Theory*) mettono invece l'accento sul ruolo che le asimmetrie informative (in senso

ampio) rivestono nelle interazioni economiche. Senza abbandonare l'ipotesi di razionalità degli individui, ed esaminando la struttura e la dinamica dei processi di decisione-esecuzione-controllo che dominano la vita economica, si comprende come, in assenza di perfetta informazione, i conflitti tra obiettivi divergenti non si possano conciliare con la piena efficienza.²⁹

La seconda area, nella quale troviamo la *Public Choice Theory*, la teoria dei diritti di proprietà e la teoria delle organizzazioni, si trova in alto a destra e include quegli economisti che hanno insistito su come il particolare profilo istituzionale (in senso lato) del mercato possa determinare uno scostamento dall'equilibrio postulato dalla teoria neoclassica. Ad esempio, il problema dei costi di transazione e della incompleta (o assente) definizione dei diritti di proprietà (Coase, 1937;1960), non riguarda la razionalità del comportamento dei soggetti, ma lo influenza e lo distorce facendolo sembrare, agli occhi di chi non considera questi aspetti, irrazionale o inefficiente. Inoltre, a partire da Buchanan and Tullock (1962) in avanti ("*Public Choice*"), viene teorizzata la possibilità che la ricerca di una rendita (*rent-seeking*) porti i soggetti economici a seguire comportamenti inefficienti, o apparentemente tali.³⁰ Alcuni autori si concentrano, invece, sul funzionamento interno dell'apparato burocratico (Niskanen, 1968; 1971) ipotizzando che la funzione obiettivo di chi prende le decisioni (i politici) e chi le attua concretamente (i funzionari dell'amministrazione), non coincidano. Questa deviazione sistematica nel comportamento individuale è la fonte principale dell'inefficienza, che spesso si osserva nel settore pubblico. Il paragone pubblico-privato si inserisce in un filone più ampio, quello dell'analisi dei diritti di proprietà, secondo il quale l'inefficienza osservata in alcuni settori dell'economia reale è spesso il riflesso dell'assetto proprietario e/o della forma istituzionale prevalente in quel mercato.³¹ In questo ambito desideriamo segnalare il contributo di quegli autori, in

²⁹ Il contributo della teoria dei contratti in materia di regolamentazione, in particolare l'opera di Laffont e Tirole (1993), è stato molto importante anche sul piano pratico, permettendo di analizzare gli effetti indesiderati di alcune forme di regolamentazione e fornendo alcuni strumenti economici per incentivare i soggetti controllati ad agire in modo efficiente.

³⁰ Alcuni autori, tra i quali Becker (1983), Demsetz (1988) e Posner (1974), mostrano come la competizione tra i soggetti nell'attività di *rent-seeking* porti il mercato verso un equilibrio molto vicino a quello neoclassico. In quello che può sembrare un paradosso, è l'inefficienza "competitiva" a garantire l'efficienza dei mercati.

³¹ In particolare, su questi temi si vedano Alchian (1965), Demsetz (1967), Alchian and Demsetz (1972) e Shleifer (1985).

modo particolare Weisbrod (1988, 1998, 2004), che si sono concentrati sulle particolarità del settore non-profit, fornendo valide argomentazioni (presenza di vincoli statutari, rilevanza della qualità, etc.) sul piano teorico per giustificare la presenza di inefficienza.

La terza area, che giace nell'angolo in basso a destra, è definita come *Behavioural Economics* ed è la più lontana rispetto alla teoria neoclassica. Questo filone, che abbraccia una vasta e ricca letteratura (*Behavioural Finance*, *X-inefficiency*, *Experimental Economics*, etc.), mostra un tratto comune che è quello di valorizzare la componente emotivo-psicologica che caratterizza il comportamento umano e quindi anche quello (semi-umano) previsto dalle teorie neoclassiche.³² Tra i molti economisti che si potrebbero includere in quest'area, citiamo il lavoro di Leibenstein (1966) che tratta in modo specifico il tema dell'efficienza produttiva e sviluppa un nuovo concetto di inefficienza che si discosta da quello neoclassico di inefficienza allocativa. L'attributo "X", scelto da Leibenstein per descrivere la natura di questa inefficienza, illustra bene l'idea che nel mondo reale numerose fonti di inefficienza si sovrappongono fondendosi in un'unica grandezza che, pur rimanendo difficile da scomporre e misurare, si può separare sul piano teorico.³³

Questa breve panoramica sui diversi modi di considerare l'efficienza produttiva nella letteratura ci permette, nonostante le semplificazioni adottate, di chiarire il quadro di riferimento teorico che fa da sfondo all'analisi empirica sull'efficienza delle case per anziani. L'inconciliabilità tra la teoria microeconomica neoclassica (che non contempla forme di inefficienza) e la misurazione empirica dell'inefficienza (che presuppone la sua esistenza) è un tema che resta aperto. Occorre ricordare, però, che l'obiettivo di questa tesi non è quello di discutere e verificare la validità delle teorie citate in precedenza. Tuttavia, siamo convinti che questa discussione possa contribuire a

³² A titolo di esempio citiamo il lavoro dello psicologo Kanheman che nel 2002 è stato insignito del premio Nobel all'economia "per aver integrato nella scienza economica le intuizioni della ricerca psicologica". Lo stesso Kanheman (coerentemente con i suoi studi sull'importanza della lealtà) riferendosi espressamente al processo di "integrazione", ha attribuito i meriti all'economista Thaler, co-autore di molte delle sue pubblicazioni economiche.

³³ Leibenstein sottolinea la contraddizione della teoria neoclassica che, mentre postula l'ottimizzazione delle risorse e l'efficienza delle decisioni aziendali, trascura il problema (a monte) dell'allocazione degli amministratori, ossia quella risorsa che indirizza quasi tutte le scelte operative dell'impresa. Nel settore pubblico si verifica invece il problema del "controllo dei controllori" che Alchian e Demsetz (1972) portano come giustificazione dell'inefficienza presente nel settore pubblico.

mitigare il (presunto) contrasto tra teoria ed empiria, permettendoci di sviluppare con più serenità il percorso di ricerca.

Per completare l'introduzione a questa tesi di dottorato, ci accingiamo a presentare la struttura del lavoro ed il contenuto dei capitoli che seguiranno, con la speranza di mantenere viva l'attenzione del lettore fino alle conclusioni della ricerca.

1.5 La struttura della tesi

Il desiderio di proporre un'analisi dell'efficienza del settore delle case per anziani svizzere, che fosse dettagliata nella trattazione dei singoli argomenti e, nel contempo, sufficientemente organica nell'impostazione generale, ci ha suggerito di organizzare i contenuti della ricerca in sette capitoli tematici.

Dopo questa introduzione, nel capitolo 2 dedicheremo ampio spazio alla presentazione del settore delle case per anziani in Svizzera, in modo da fornire tutti gli elementi di comprensione per le analisi empiriche che seguiranno. Inizialmente introdurremo una tassonomia delle case per anziani, mettendo in evidenza le tre principali tipologie di istituti che si sono diffusi in Svizzera. Il seguito del capitolo è dedicato all'analisi descrittiva di alcune dimensioni ritenute particolarmente interessanti: forma istituzionale, dimensione delle case per anziani, e ventaglio di prestazioni offerte. Un confronto intercantonale delle caratteristiche più significative, per gli scopi di quest'analisi, troverà spazio nella parte conclusiva del capitolo. L'analisi descrittiva delle caratteristiche di questo settore imprime un connotato empirico alla ricerca, fungendo da sfondo comune per gli argomenti trattati in seguito.

Il capitolo 3 si prefigge, invece, l'obiettivo di inquadrare teoricamente la ricerca. Partiremo quindi dalla teoria microeconomica (neoclassica) della produzione, presentando i concetti essenziali per lo sviluppo della parte empirica. L'accento verrà posto sulla teoria del duale sviluppatasi a partire dal lavoro di Shephard (1953) e ripresa più di recente da Diewert (1982) e Cornes (1992). Gli argomenti del capitolo sono organizzati in modo da discutere quasi esclusivamente quei concetti microeconomici funzionali alle analisi empiriche che seguiranno. In particolare, ci concentreremo sulla funzione di costo e sulle sue proprietà, evidenziando le differenze tra l'orizzonte temporale di breve e di lungo periodo. Inoltre, verrà presentato e discusso il concetto di economie di scala. Gli elementi essenziali della teoria neoclassica della produzione

serviranno per sviluppare il discorso nella direzione che più ci interessa in vista della parte empirica: la definizione di inefficienza produttiva. Prendendo spunto dal lavoro di Farrell (1957) esporremo i diversi concetti legati all'efficienza produttiva con particolare riguardo per l'inefficienza di costo.

Il capitolo 4 è dedicato ad approfondire le tecniche econometriche utilizzate in questa ricerca con l'obiettivo di chiarire: gli aspetti critici, i vantaggi e gli svantaggi connessi con le diverse procedure di stima. La comprensione della metodologia è fondamentale per affrontare l'analisi empirica con la prospettiva di generare risultati economicamente plausibili e di saperli interpretare correttamente. Più in dettaglio, discuteremo il problema di come scegliere adeguatamente i due elementi strutturali di una stima empirica: (1) il modello di costo e (2) la forma funzionale. La parte centrale del capitolo è dedicata ad una spiegazione della metodologia SFA seguita nella tesi, evidenziando i vantaggi connessi con il suo utilizzo. Una rassegna critica della recente letteratura econometrica sul tema della stima delle frontiere stocastiche occuperà il resto della sezione. In particolare, presenteremo le caratteristiche dei diversi filoni in cui la modellizzazione della SFA si è articolata, mettendo in risalto i 5 modelli concretamente impiegati nell'analisi empirica dell'inefficienza produttiva delle case per anziani in Svizzera. Con il capitolo 4 si chiude la parte dedicata alla metodologia, per lasciare spazio alla presentazione del modello di costo scelto per le stime e al commento dei risultati che ne scaturiscono.

I dettagli sulla definizione del modello di costo sono riportati nel capitolo 5, nel quale descriveremo anche i problemi legati ai dati a disposizione ed alla scelta delle variabili esplicative. Per contestualizzare le scelte fatte, inizieremo con una breve rassegna della letteratura che si è occupata di stime di funzioni di costo nel settore delle case per anziani. In questo modo potremo discutere la costruzione del modello di costo, alla luce dei problemi e delle possibili soluzioni già evidenziate da lavori precedenti. La rassegna copre 12 studi che abbracciano 18 anni di ricerca riferita a diversi paesi (in prevalenza Stati Uniti d'America e Svizzera). L'elemento che emerge con forza dal confronto dei diversi articoli è l'importanza di considerare la complessità del processo produttivo di una casa per anziani nella specificazione del modello di costo, poiché in ogni paese l'assistenza agli anziani assume forme molto differenziate ed è organizzata secondo modalità specifiche. Il resto del capitolo si concentra sulla spiegazione dettagliata delle variabili esplicative che costituiscono il modello di costo utilizzato

nelle stime. In questa tesi abbiamo utilizzato una funzione di costo totale Translog, specificata in modo neoclassico “ampio”, che affianca alle variabili strutturali (output e prezzi degli input) una serie di regressori in grado di cogliere fattori ambientali e caratteristiche specifiche. Le scelte compiute per definire il modello di costo impiegato in questa ricerca verranno motivate con riferimento alle discussioni metodologiche presentate in precedenza.

I risultati delle stime e le loro implicazioni economiche saranno discussi nel capitolo 6. Dopo una breve discussione sul campione di dati utilizzato nelle stime, presenteremo i risultati empirici generati dalla stima della funzione di costo, impiegando 5 diversi modelli econometrici. Con le stime dei parametri della funzione di costo è quindi possibile misurare l'efficienza di costo degli istituti per anziani, mettendo in luce il livello medio di inefficienza registrato nel settore, così come la distribuzione dei punteggi di inefficienza delle singole imprese. Il tema della stabilità dei risultati e dell'impatto del modello econometrico sulla misurazione dell'efficienza troveranno un adeguato spazio di discussione, mettendo in evidenza il valore informativo che può comunque emergere dal confronto dei diversi approcci. Il discorso si sposta quindi sul problema dell'eterogeneità, sia a livello cantonale sia rispetto ad altre dimensioni (forma istituzionale, grandezza degli istituti, etc.), sviluppato attraverso un confronto tra i punteggi di inefficienza dei rispettivi gruppi di case per anziani. Dopo aver discusso l'inefficienza di costo, discuteremo il valore delle economie di scala, stimato dai parametri della funzione di costo. Oltre alla misurazione del fenomeno verranno discusse le possibili implicazioni economiche ed eventuali raccomandazioni di politica sanitaria, che avranno carattere prevalentemente positivo.

Le considerazioni finali trovano spazio nel capitolo 7, con il quale termineremo il presente lavoro di dottorato. Dopo un riepilogo della ricerca, comprendente obiettivi, metodologia e risultati, saranno discussi alcuni aspetti di politica sanitaria, legati ai risultati empirici ed anche ai problemi metodologici emersi nel corso dell'analisi. Il capitolo si conclude con una riflessione orientata al futuro, evidenziando alcuni possibili sviluppi per proseguire e migliorare la ricerca in questo ambito.

“La nostra conoscenza può essere solo finita, mentre la nostra ignoranza deve essere necessariamente infinita.”

(Karl Raimund Popper, Filosofo)

2 Il settore delle case per anziani in Svizzera

L’attenzione delle autorità politiche verso un uso parsimonioso delle risorse si è manifestata nel corso degli ultimi decenni attraverso numerose misure di intervento in svariati campi dell’economia, tra i quali rientrano i settori sanitario e sociosanitario. La preoccupazione circa il continuo aumento dei costi ha spinto il legislatore ad introdurre nella Legge sull’assicurazione malattia (LAMal), elaborata nel 1994 alcuni articoli esplicitamente dedicati al tema del contenimento dei costi e della promozione dell’efficienza nel settore ospedaliero.³⁴

A riprova che l’impegno a contenere i costi nel settore delle case per anziani non è semplicemente una dichiarazione programmatica, possiamo osservare l’evoluzione dei sistemi cantonali di regolamentazione che nell’ultimo decennio hanno provveduto ad una continua riorganizzazione mirata ad ottimizzare le prestazioni e a contenere i costi. In ossequio ai principî del federalismo ogni cantone ha fatto le proprie scelte in autonomia e compatibilmente agli equilibri politici e finanziari di ciascuno Stato. In

³⁴ Nell’articolo 49, dove si definiscono le convenzioni tariffarie con gli ospedali, viene sancito il principio della comparabilità dei costi ospedalieri che devono essere rilevati in modo uniforme secondo lo strumento della contabilità analitica. Nel cpv. 7 viene precisato che il confronto tra i costi ospedalieri funge da strumento per verificare gli eccessi di costo non giustificabili il cui rimborso può essere ridimensionato.

molti cantoni si è deciso, ad esempio, di introdurre il contratto di prestazione come strumento di regolamentazione e di finanziamento del settore sociosanitario, con l'obiettivo di aumentare l'efficienza di questo settore; altri cantoni regolamentano il settore con strumenti più tradizionali come il controllo amministrativo e l'erogazione di sussidi alla gestione ed all'investimento. Nell'affrontare le scelte di regolazione, i singoli cantoni svizzeri sono confrontati con situazioni di partenza diverse (popolazione, territorio, risorse finanziarie, orientamenti politici, etc.) che portano alla fine a prediligere una determinata forma di intervento nel settore.

L'eterogeneità osservata nelle case per anziani scaturisce dall'intreccio delle diverse caratteristiche strutturali e delle politiche di regolamentazione differenziate da cantone a cantone. Per questo motivo sembra importante tracciare una descrizione del settore che non si limiti ai soli aspetti economici in modo da poter conoscere i diversi elementi del problema e concentrarsi, a ragion veduta, sugli aspetti legati all'efficienza economica degli istituti per anziani. Nel resto del capitolo forniremo, quindi, un quadro delle caratteristiche e delle peculiarità di questo settore in modo da preparare il terreno ai quesiti di carattere empirico che solleviamo in seguito.

2.1 Analisi descrittiva delle case per anziani in Svizzera

Si è soliti definire con il termine "casa per anziani" una struttura residenziale destinata, a titolo temporaneo o permanente, a persone che si ritrovano a dover fronteggiare, per l'avanzare dell'età, limitate capacità fisiche, relazionali o sociali. Un istituto per anziani offre di regola ai propri ospiti assistenza diurna e notturna, un servizio alberghiero, attività di cura, mobilitazione e ricreazione; è caratterizzata dall'apertura 24 ore su 24, per 365 giorni all'anno, e dal fatto di disporre di personale assistenziale e sanitario opportunamente qualificato.³⁵

A dispetto di questa definizione apparentemente precisa, in Svizzera l'universo delle case per anziani è piuttosto variegato. In questo capitolo cercheremo anzitutto di dar conto di questa varietà, presentando una nostra tassonomia delle case per anziani

³⁵ Una definizione analoga si trova in Travaglini (1998).

presenti in Svizzera, seguita da un'analisi descrittiva dei dati statistici in nostro possesso.

Le case per anziani attive sul territorio svizzero possono essere classificate in base ad almeno 3 criteri importanti: (1) il livello di complessità delle cure sanitarie offerte agli ospiti, (2) l'assetto proprietario dell'istituto e (3) la dimensione della casa in termini di posti letto. Per completare l'analisi descrittiva delle case per anziani desideriamo affiancare ai tre criteri menzionati una quarta dimensione valutativa: il ventaglio di servizi offerti da un istituto ai propri ospiti. La presentazione di questi aspetti distinti ci permette, da un lato di fornire un interessante sguardo d'insieme sul settore, e dall'altro di comprendere la sostanziale eterogeneità che lo contraddistingue. A livello cantonale emergono differenze importanti che presenteremo separatamente con lo scopo di preparare il terreno alla discussione sul ruolo del federalismo e della regolamentazione in questo settore.³⁶

2.1 Tassonomia: “Altersheime”, “Krankenheime” o “Pflegeheime”?

La situazione che riscontriamo oggi nel settore delle case per anziani in Svizzera è il frutto di scelte compiute nel passato e della naturale evoluzione della società. A suo tempo, nel progettare alcuni istituti per anziani non si era tenuto conto delle esigenze medico-infermieristiche di cui oggi mediamente necessitano gli ospiti. Le case per anziani erano concepite con lo scopo di offrire ai propri ospiti un servizio esclusivamente residenziale e ricreativo. Di conseguenza le infrastrutture non erano attrezzate per offrire ai residenti un'assistenza medica e infermieristica ed il personale non era necessariamente qualificato per gestire ospiti bisognosi di cure sanitarie.

Nonostante l'evoluzione del profilo medico degli anziani e i necessari adeguamenti delle strutture e del personale, la presenza di istituti concepiti come case di riposo non medicalizzate ha dato vita ad un processo spontaneo di scelta da parte degli ospiti. In queste case per anziani si stabiliscono preferibilmente persone che godono ancora di un discreto stato di salute e che sono in grado di gestire buona parte delle attività quotidiane in modo indipendente.

³⁶ Si veda Crivelli et al. (2002).

Altersheime. Nei cantoni di lingua tedesca questo tipo di istituti prende il nome di Altersheim (letteralmente: ospizio per la vecchiaia, casa per anziani) e si rivolge ad una cerchia di potenziali utenti che molto raramente si estende oltre i confini comunali. Conformemente alla suddivisione dei compiti tra cantoni e comuni, che nella maggior parte dei casi assegna all'istanza cantonale il ruolo trainante in materia sanitaria, lasciando ai comuni il compito di promuovere l'assistenza agli anziani, non è sorprendente il fatto che le competenze di sorveglianza e regolamentazione degli Altersheime, così come i maggiori oneri finanziari, spettino alle autorità comunali. Nei cantoni in cui esiste una legislazione specifica per gli Altersheime, il ruolo del cantone definito in questi testi di legge è limitato generalmente al versamento di un contributo finanziario (minoritario), in favore dei comuni o degli enti ai quali questi ultimi hanno delegato il compito dell'assistenza agli anziani, al fine di garantire un'equità di finanziamento tra comunità locali con una forza finanziaria differente.

Krankenheime. Di tutt'altro genere sono quegli istituti che, nei cantoni tedeschi, prendono il nome di Krankenheime (istituti di ricovero). Si tratta in questo caso di istituzioni che offrono in primo luogo cure medico-infermieristiche rivolte a pazienti per i quali si prevede una lunga degenza ospedaliera. Benché non si rivolgano specificamente alle persone anziane, la casistica di questi istituti è tuttavia tale per cui la maggior parte dei degenti (spesso più del 75% dei residenti) ha un'età superiore ai 65 anni. A differenza degli Altersheime, l'obiettivo dei Krankenheime non è quello di assicurare una forma di residenza definitiva ad una popolazione anziana, quanto piuttosto quello di offrire assistenza sanitaria. I Krankenheime hanno di regola un bacino di potenziali utenti che abbraccia una cintura di parecchi comuni, mentre i compiti di regolamentazione e di sussidiamento, così come è il caso per gli ospedali acuti, sono in prevalenza di competenza cantonale e vengono definiti dalle apposite leggi sanitarie.

Pflegeheime. Esiste infine una forma intermedia, i cosiddetti Pflegeheime (case di cura), i quali sono in sostanza istituti medicalizzati per persone anziane, in grado dunque di ospitare a titolo definitivo anche persone che necessitano in modo permanente di cure medico-sanitarie. Nel caso degli Pflegeheime, il compito primario della presa a carico della popolazione anziana e quello dell'assistenza sanitaria si sovrappongono, rendendo necessario il coinvolgimento contemporaneo di entrambi i livelli politici (ossia cantonale e comunale). All'atto pratico le differenze tra i diversi

concetti di casa per anziani sono sempre meno marcate e la nostra classificazione è pertinente in quei cantoni di lingua tedesca, dove la distinzione continua a sussistere anche sul piano legislativo.

Nei cantoni romandi e nel Ticino le differenze e la terminologia sono più sfumate dal momento che vengono semplicemente separati gli istituti che offrono prestazioni mediche e infermieristiche da quelli che si limitano a servizi residenziali. Nei cantoni latini si può parlare quindi di *établissement médico-social* (case per anziani socio-medicalizzate) e di *homes simples* (case per anziani), anche se non viene posta eccessiva enfasi in questa distinzione, soprattutto sul piano giuridico.

Se in passato le diverse tipologie di casa per anziani potevano trovare riscontro nella realtà, in questi ultimi anni si è assistito ad una riduzione delle differenze tra queste categorie di istituti, per almeno due ragioni. La prima è l'importante ampliamento dell'offerta di servizi di assistenza e cure a domicilio, che ha consentito a persone anziane che non soffrono di particolari problemi di salute di prolungare la permanenza al proprio domicilio, evitando così un trasferimento prematuro in una casa per anziani. La seconda è l'aumento della speranza di vita ed il conseguente invecchiamento della popolazione. Entrambi questi fattori hanno determinato un progressivo aumento dell'età media dei residenti in casa per anziani e nel contempo un incremento del grado di dipendenza e del fabbisogno di cure infermieristiche. Anche le persone che al momento di insediarsi in una casa per anziani godevano di buona salute, con l'invecchiamento necessitano di un'assistenza sanitaria. Per evitare che si debba sistematicamente ricorrere al trasferimento di questi casi in un istituto medicalizzato, molti *Altersheime* hanno finito per adattare le caratteristiche del proprio personale e l'infrastruttura a quelle di una struttura socio-medicalizzata. Si produce così una sorta di convergenza, determinata *in primis* dalla crescente omogeneità nei bisogni che caratterizzano gli ospiti residenti nei vari tipi di casa per anziani. Questa evoluzione è confermata dalla tendenza, nei cantoni svizzero-tedeschi, ad abolire l'attuale distinzione sul piano legislativo tra Alters-, Kranken- e Pflegeheime.

L'Ufficio Federale di Statistica (UFS), nel rilevare ed elaborare i dati sugli istituti non ospedalieri, ha adottato un sistema di classificazione che si basa sul profilo sanitario

degli ospiti.³⁷ Questo criterio, che richiama la tassonomia appena descritta, mira a raggruppare gli istituti secondo il grado di medicalizzazione espresso dall'insieme dei residenti. Se la differenza tra le tipologie di istituti per anziani è poco marcata (o è difficile trovare una linea di demarcazione), in termini pratici, e in particolar modo dal punto di vista statistico, il compito diventa ancora più arduo. Si deve, infatti, trovare un parametro quantitativo che permetta di attribuire i singoli istituti alle diverse categorie considerate. Anche le definizioni dell'UFS hanno subito diverse modifiche nel tempo, a dimostrazione della difficoltà di individuare, per questo settore, un criterio di classificazione che riuscisse a cogliere le reali differenze che distinguono le case per anziani svizzere, nel rispetto di alcune esigenze di carattere statistico e di alcuni vincoli di nomenclatura delle attività economiche (codice NOGA).³⁸ Gli aspetti salienti dell'attuale sistema di classificazione si possono riassumere in tre punti: (1) limitare la fluttuazione di unità statistiche da una categoria all'altra; (2) Individuare criteri che generino raggruppamenti relativamente uniformi e numericamente paragonabili; (3) consentire di individuare all'interno di ciascuna categoria eventuali differenze quantitative e qualitative tra le singole osservazioni. Il risultato finale di queste scelte è rappresentato nel diagramma seguente (Figura 2.1) che indica il processo di assegnazione delle unità statistiche, gli stabilimenti sociosanitari svizzeri, alle diverse categorie contemplate dall'UFS.

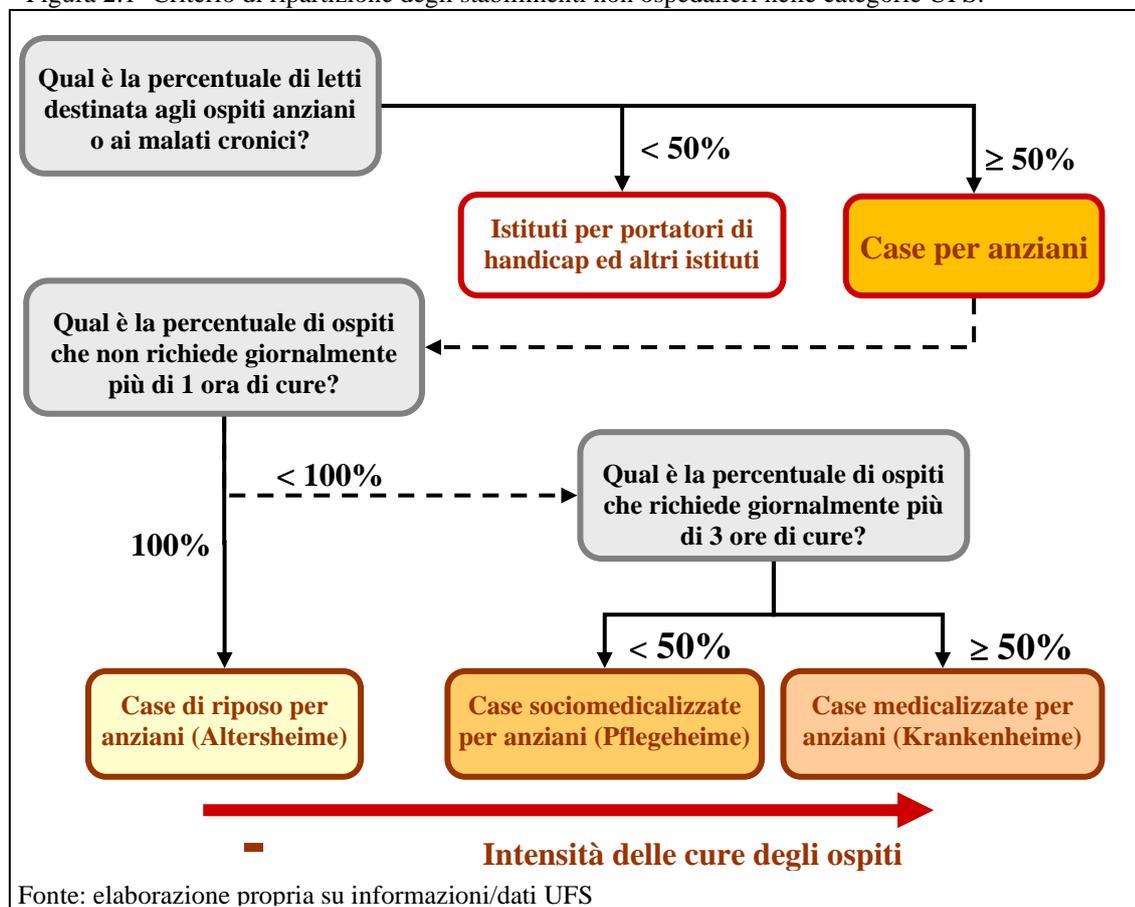
A scopo statistico sono definite case per anziani quegli istituti che riservano alle persone anziane e/o ai malati cronici una percentuale di letti superiore al 50%. L'insieme delle case per anziani viene poi ulteriormente suddiviso in tre gruppi che riflettono il diverso bisogno di cure da parte degli ospiti. L'istituto viene classificato come casa di riposo per anziani (*Altersheime*) se tutti gli ospiti richiedono, in media,

³⁷ Dal 1998 è disponibile la pubblicazione sugli istituti non ospedalieri Stat-Santé / Tableaux Standard che raccoglie ogni anno le informazioni individuali di quasi tutti gli istituti sociosanitari attivi in Svizzera.

³⁸ I criteri utilizzati nella redazione delle statistiche dell'UFS sono espressi in modo chiaro e sono contenuti in un'apposita pubblicazione denominata "Tipologia degli stabilimenti sanitari non ospedalieri". In questo lavoro faremo riferimento alla versione 4.1 (novembre 2001) che ha permesso di classificare gli istituti sociosanitari nel periodo coperto dalla presente analisi. Recentemente (novembre 2005) l'UFS ha modificato il sistema di classificazione degli istituti non ospedalieri ed un nuovo questionario di rilevazione statistica che sarà adottato nel 2006. Si veda l'indirizzo internet: http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen_quellen/blank/blank/ssmi/07.html

meno di un'ora di cure giornaliere, siano esse mediche (es. visite di controllo), infermieristiche (es. somministrazione di farmaci) o personali (es. aiuto nel vestirsi).

Figura 2.1- Criterio di ripartizione degli stabilimenti non ospedalieri nelle categorie UFS.



In caso contrario siamo in presenza di case per anziani con un certo grado di medicalizzazione che vengono ulteriormente suddivise basandosi sulla percentuale di ospiti che richiedono un'intensità di cure giornaliere superiore a tre ore: quando la percentuale non supera il 50% si rileva un istituto socio medicalizzato mentre negli altri casi l'istituto è definito casa per anziani medicalizzata.

La scelta di considerare un istituto come una casa per anziani è quindi basata sulla disponibilità di posti letto ad uso di persone anziane e/o cronicamente ammalate, posto che il numero di questa tipologia di letti costituisca la maggioranza. La linea di demarcazione tra ciò che si può definire casa per anziani dal punto di vista statistico, presta il fianco ad una critica importante, ossia il fatto di classificare in due categorie distinte alcuni istituti con una composizione di letti molto simile. Accade, infatti, che tutti gli istituti a cavallo della soglia scelta per discriminare tra case per anziani e istituti sociomedicalizzati finiscano in due gruppi diversi, pur trovandosi nelle immediate

vicinanze (sopra o sotto) dello spartiacque e presentando quote di letti riservati agli anziani che divergono solamente per pochi punti percentuali. Nel caso specifico, il problema non si pone visto che la distribuzione degli istituti non ospedalieri svizzeri è fortemente polarizzata rispetto alla composizione dei posti letto e solamente una piccola percentuale di stabilimenti si trova in una situazione di ambiguità.³⁹

All'interno della categoria delle case per anziani viene proposta un'ulteriore suddivisione che si basa sul profilo sanitario degli ospiti e sul loro livello di autonomia. Case per anziani propriamente dette (*Altersheime*) sono quegli istituti nei quali il 100% degli ospiti non necessita cure giornaliere che, in media, superano la durata di un'ora. Appare chiaro che questo criterio sia volutamente molto rigido per consentire di classificare con ragionevole sicurezza quegli istituti a carattere spiccatamente residenziale e ricreativo, nei quali si ammette (al limite per tutti gli ospiti) un minimo bisogno di cure di vario genere, quantificato mediamente in un'ora al giorno.

Tutti gli altri istituti sono raggruppati nella categoria delle case di cura per anziani che comprende una distinzione tra case medicalizzate e case socio-medicalizzate. Quest'ultima distinzione, che si basa sulla percentuale di ospiti che richiedono cure giornaliere superiori a 3 ore, è più sfumata e coglie solo in parte le reali differenze degli istituti che appartengono alle due categorie.

In conclusione, possiamo affermare che negli ultimi 20 anni, con il progressivo invecchiamento della popolazione, Alters-, Pflege- e Krankenheime hanno assunto delle caratteristiche produttive molto simili fra loro, rendendo inappropriata dal profilo economico una classificazione legata prettamente a ragioni storiche e giuridiche. Il criterio statistico utilizzato dall'UFS, che almeno a grandi linee rispecchia anche la tassonomia introdotta in precedenza, sembra preferibile per effettuare un'analisi descrittiva di questo settore distinguendo, se necessario, case per anziani a scopo residenziale ed altri istituti medicalizzati.

In base alle considerazioni fatte l'analisi statistica che segue verrà effettuata partendo da un campione di 11'637 istituti non ospedalieri su 5 anni (dal 1998 al 2002) per arrivare a 6'583 case per anziani, ulteriormente suddivise in due categorie: non

³⁹ Gli istituti non ospedalieri la cui quota di letti per anziani è compresa nell'intervallo 35-65%, sono molto pochi (inferiori al 3% di tutta la popolazione).

medicalizzate (74 unità) e medicalizzate (6'509).⁴⁰ La netta preponderanza di quest'ultimo raggruppamento fotografa in modo chiaro la progressiva scomparsa della casa di riposo propriamente detta, sostituita da istituti medicalizzati pronti ad accogliere una popolazione sempre più anziana e bisognosa di ricevere assistenza e cure mediche.

Nella parte empirica della tesi ci concentreremo sulle case per anziani che erogano, oltre ai servizi di assistenza, anche cure mediche, impiegando un criterio aggiuntivo rispetto alla classificazione dell'UFS.⁴¹ La tripartizione degli istituti per anziani descritta in precedenza (Alters-, Pflege- e Krankenhäuser) si può comunque ricostruire partendo dai risultati empirici, dal momento che il modello di costo impiegato include una variabile associata al numero medio di ore di assistenza per istituto.

2.1.1 Forme giuridiche e assetto proprietario

Un aspetto da approfondire per tracciare un quadro descrittivo completo delle case per anziani che operano in Svizzera è rappresentato dalla forma istituzionale, nella quale si riflettono da un lato l'assetto proprietario dell'istituto e dall'altro lo statuto giuridico.

Per classificare le case per anziani in base alla forma istituzionale occorre in primo luogo separare le organizzazioni pubbliche da quelle private, per poi suddividere ulteriormente queste ultime in non profit e for profit.⁴²

La prima distinzione non appare particolarmente problematica, poiché poggia sul criterio della proprietà del capitale che, nel settore delle case per anziani, è in genere completamente pubblico o integralmente privato.⁴³ La classificazione rispetto al criterio dell'attitudine a perseguire i profitti crea, invece, qualche problema. Ciò che caratterizza

⁴⁰ Il totale delle case per anziani è risultato di 6588 ma 5 unità sono state rimosse dal campione perché prive di valori nelle variabili chiave per definire le categorie menzionate (es. il numero di ore di cura).

⁴¹ Utilizzando i dati sui rimborsi ottenuti dagli istituti ai sensi della legge sull'assicurazione malattia di base (LAMal), è possibile individuare quegli istituti che non fatturano alcuna prestazione medica rimborsabile.

⁴² La letteratura economica sul settore non profit è molto ampia ed in continua evoluzione. Per una discussione generale si vedano ad esempio Weisbrod (1988, 2004), Zamagni (1998) e Powell and Steinberg (2006).

⁴³ La partecipazione congiunta di capitale pubblico e privato è del tutto possibile e legittima ed è una realtà in molti settori economici. Nel campo sociosanitario, tuttavia, le esperienze concrete sono molto limitate se escludiamo come forma di compartecipazione del capitale di rischio i sussidi pubblici all'investimento in istituti privati (ma regolamentati).

un'impresa for profit è il fatto di essere finalizzata al conseguimento di extra-profitti (quella parte di profitti che eccede una normale remunerazione del capitale investito e del rischio imprenditoriale), a vantaggio dei titolari del capitale aziendale. Al contrario le organizzazioni non profit sono imprese che non perseguono l'obiettivo di produrre extra-profitti, quanto piuttosto quello di generare un beneficio destinato ad una categoria di *stakeholders* diversa da quella degli *shareholders*. La distinzione non profit/for profit si gioca dunque sul piano degli obiettivi aziendali di un'istituzione, nella considerazione di una dimensione che è di natura squisitamente strategica e che quindi raramente si manifesta in modo chiaro verso l'esterno.⁴⁴

Vi sono pertanto due possibili vie d'uscita: (a) valutare le finalità di un'azienda sulla base di autodichiarazioni (rischiando però di commettere degli errori di classificazione, dal momento che non si può escludere a priori l'insorgere di incompatibilità tra gli obiettivi dichiarati ed il comportamento effettivo delle istituzioni); (b) considerare altre dimensioni che, nella maggior parte dei casi, appaiono verso l'esterno, quali l'esistenza di un utile di gestione, la forma giuridica scelta dall'azienda per raggiungere i propri obiettivi e la prassi di procedere alla remunerazione del capitale attraverso il versamento di dividendi.

L'esistenza di un profitto non è per nulla indicativo degli obiettivi principali dell'azienda, dal momento che non profit non significa necessariamente no profit (nel senso di zero profit). Un'impresa non profit può pertanto generare dei profitti, ma questi devono essere reinvestiti nelle attività dell'organizzazione, a favore degli utenti. È esclusa in altre parole la possibilità di distribuire gli utili.

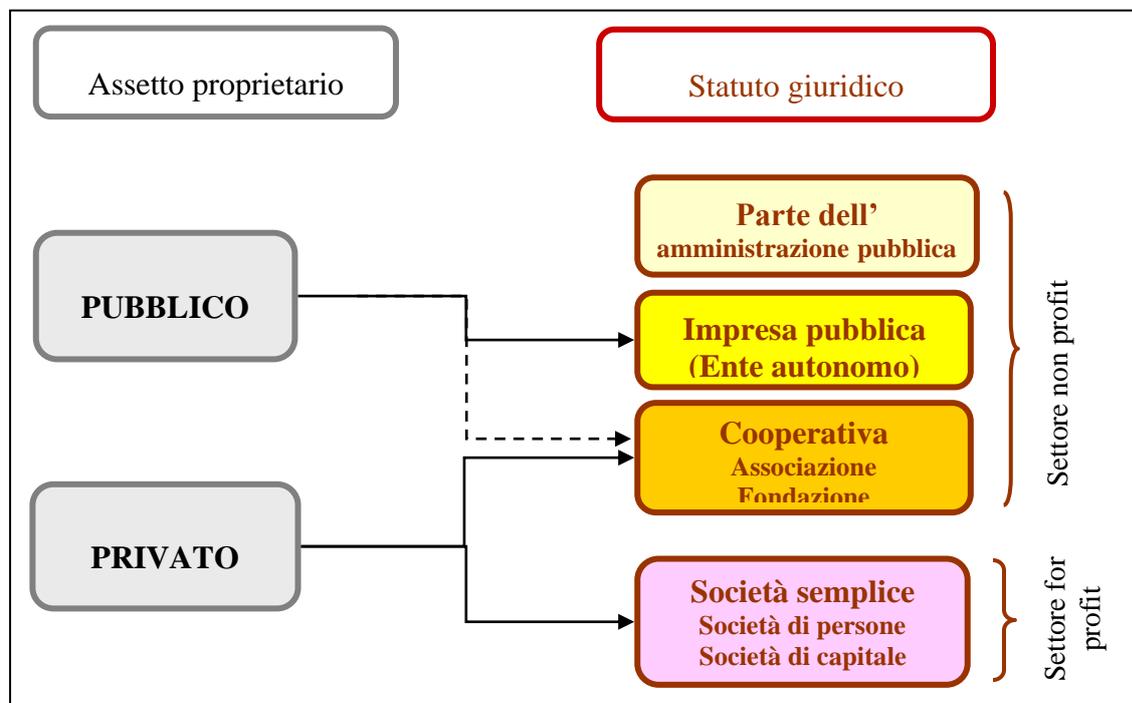
Nella letteratura anglosassone si è soliti ritenere il “divieto statutario di distribuire gli utili” un buon indicatore del fatto che l'azienda in questione sia non profit. Ciò non esclude tuttavia a priori che una società di capitale, nella quale i diritti di proprietà su eventuali extra-profitti sono per statuto allocati in corrispondenza delle quote di capitale investito dai soci, rinunci volontariamente alla distribuzione dei dividendi per reinvestire i profitti a vantaggio degli utenti.

Consapevoli che è sempre possibile avere una zona grigia nella quale è facile commettere errori di classificazione, abbiamo optato per una divisione delle case per

⁴⁴ Per un approfondimento sulla teoria economica del settore non profit si veda Borzaga (1998).

anziani secondo la forma istituzionale e in base al loro statuto giuridico (Figura 2.2). Sono state classificate come pubbliche le case per anziani che fanno parte dell'amministrazione federale, cantonale o comunale, oppure gli stabilimenti autonomi di diritto pubblico.

Figura 2.2 - Classificazione delle case per anziani secondo l'assetto proprietario e lo statuto giuridico.



Fra gli istituti privati sono state considerate non profit le associazioni, le fondazioni e le cooperative, e for profit le società semplici, le aziende individuali, le società in nome collettivo, in accomandita, a garanzia limitata e le società anonime.

Come visto in precedenza, esiste una sovrapposizione quasi perfetta tra assetto giuridico ed assetto istituzionale negli istituti pubblici, se si eccettua la possibilità di trasformare giuridicamente una casa per anziani (ad esempio scorporandola dall'amministrazione comunale per dar vita ad una fondazione a capitale pubblico). Per operare l'attribuzione delle società di diritto privato ad un determinato assetto istituzionale si è tenuto conto del fatto che nella letteratura sul "terzo settore" associazioni e cooperative sono ritenute le forme giuridiche più adatte per promuovere un beneficio mutualistico (*mutual benefit*), mentre la fondazione è la forma giuridica più indicata per il conseguimento di un beneficio pubblico (*public benefit*).

La distinzione spesso si basa sul fatto che la categoria dei beneficiari coincida o meno con la categoria dominante, per cui si parla di spirito mutualistico quando la cerchia dei beneficiari (stakeholders) è limitata al gruppo di persone alle quali spetta il

potere decisionale. Diversamente, quando la categoria dei beneficiari si estende a favore di interessi più ampi di carattere pubblico, si può parlare di *public benefit*.⁴⁵ Diversamente, quando la categoria dei beneficiari si estende a favore di interessi più ampi di carattere pubblico, si può parlare di *public benefit*. In questi casi solitamente viene costituita una fondazione il cui potere decisionale viene delegato ad un gruppo di amministratori che hanno il mandato di perseguire gli interessi sanciti dallo statuto. In ultima analisi risulta sempre difficile classificare con certezza un'organizzazione in un determinato gruppo conoscendo solo due aspetti quali la forma istituzionale e l'assetto proprietario che, sebbene molto indicativi dell'attitudine dell'impresa a realizzare profitti, lasciano spazio ad una zona di incertezza.

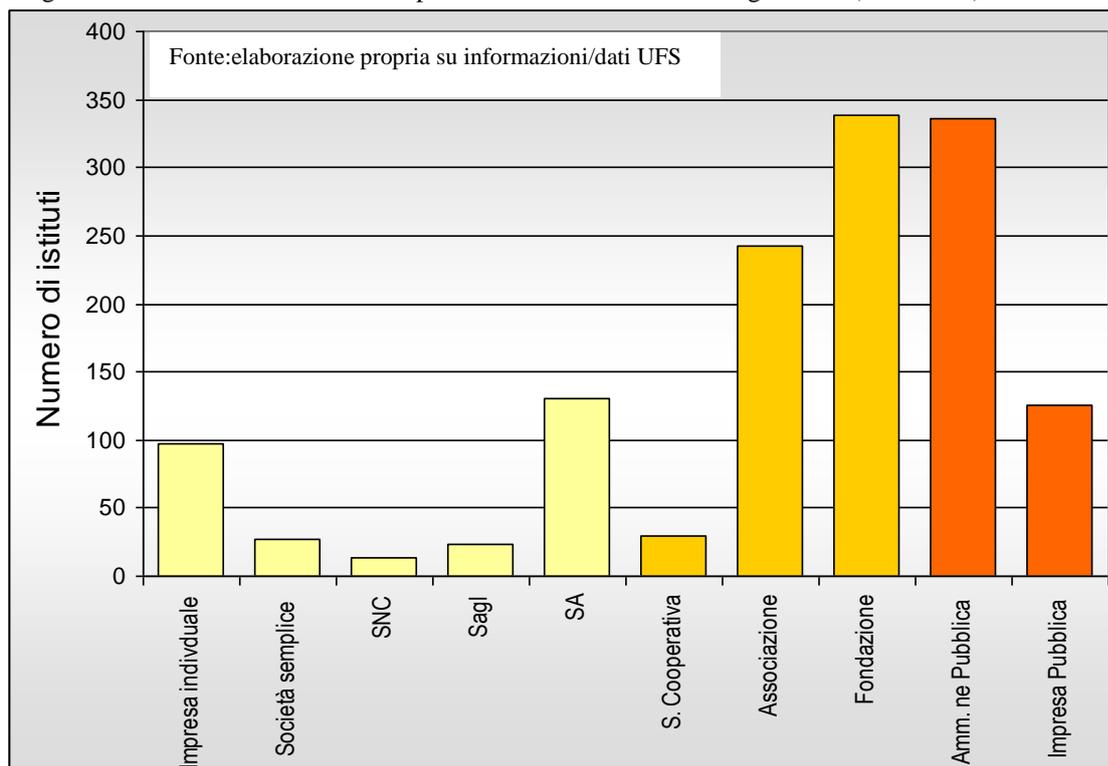
Nel diritto alcune forme istituzionali sono classificate come non idonee a perseguire uno scopo "economico pecuniario" (tipicamente le fondazioni, le associazioni e le cooperative) che è prerogativa delle società di persone o di capitali. Lo scopo di lucro (o la sua assenza) esprime, tuttavia, la natura e la filosofia di un'impresa, mentre la forma istituzionale riflette solamente una scelta strategica volta a perseguire nel modo migliore gli scopi dell'organizzazione. Anche da un punto di vista del diritto queste suddivisioni non sono così nette, poiché lo statuto giuridico costituisce soltanto un mezzo per raggiungere lo scopo statutario e se è vero che una certa forma giuridica può risultare più appropriata di un'altra, non vi è alcun obbligo di costituire l'organizzazione ricorrendo ad un determinato assetto giuridico.

L'esistenza di imprese che giuridicamente potrebbero realizzare e distribuire profitti (ad esempio le società di capitali) ma che, all'atto pratico, decidono di non perseguire uno scopo di lucro, ci ha suggerito di presentare (Figura 2.3) la composizione delle case per anziani svizzere rispetto alla loro forma istituzionale nell'anno 2002 per consentire al lettore di non perdere il dettaglio intorno al quale abbiamo costruito le tre tipologie di istituti (pubblici, privati non-profit e privati for-profit) viste in precedenza.

Come è facile notare osservando l'istogramma, gli istituti per anziani sono organizzati secondo un ampio spettro di forme giuridiche che confermano la complessa relazione che esiste tra scopo statutario e forma istituzionale scelta.

⁴⁵ Per un approfondimento sulla diversa natura delle organizzazioni non profit si veda Gui (1991).

Figura 2.3 - Distribuzione delle case per anziani in base allo statuto giuridico (anno 2002).



Nell'ambito delle forme giuridiche che si adattano meglio a perseguire uno scopo di lucro troviamo rappresentate tanto le società di persone quanto le società di capitali con una maggior numerosità all'interno di quest'ultima categoria delle Società Anonime. Il motivo potrebbe essere legato all'elevato ammontare di capitali che spesso è necessario investire in un istituto per anziani, per dotarlo di quelle infrastrutture ed attrezzature che il profilo sanitario degli ospiti richiede nella maggior parte dei casi, e che una SA è in grado di gestire più agevolmente.

Osservando il gruppo di istituti che si organizzano con uno statuto giuridico indicato a perseguire scopi diversi dal profitto economico, notiamo come le cooperative siano in netta minoranza a conferma del fatto che nel settore dell'assistenza agli anziani tende a prevalere la componente di public benefit su quella mutualistica. Questa constatazione risulta molto importante se consideriamo che la maggior parte degli istituti privati non profit riceve un sussidio da parte dello Stato, il quale riconosce, anche finanziariamente, l'esistenza e l'importanza dei benefici pubblici erogati dalle case per anziani private.

Dopo aver preso coscienza della composizione molto articolata di forme istituzionali che operano nel settore delle case per anziani in Svizzera, è possibile applicare il criterio visto in precedenza (assetto proprietario e statuto giuridico) per

suddividere gli istituti presenti nel campione in tre categorie: (1) pubblici, (2) privati non profit e (3) privati for profit. La consistenza numerica delle tre tipologie, che è indicata nella Tabella 2.1, si riferisce ai dati del 2002 e sottolinea in modo chiaro come la presenza delle imprese private a scopo di lucro (for profit) sia in netta minoranza rispetto alle altre due categorie.⁴⁶

Tabella 2.1 - Suddivisione degli istituti secondo la forma istituzionale (anno 2002).

Istituti pubblici	Istituti privati non profit	Istituti privati for profit
462	610	291

Fonte: elaborazione propria su informazioni/dati UFS

La definizione di queste tre categorie sarà ripresa anche nella parte empirica con l'obiettivo di verificare se il livello medio di inefficienza di costo misurato per le tre forme istituzionali risulti sensibilmente diverso.

2.1.2 Dimensione degli istituti

Il terzo criterio utilizzato per classificare le case per anziani è quello della dimensione, valutata in termini di posti letto. Dal punto di vista economico, si tratta di un criterio fondamentale, poiché si ricollega al discorso della dimensione ottimale che un istituto per anziani dovrebbe possedere per produrre in modo efficiente.

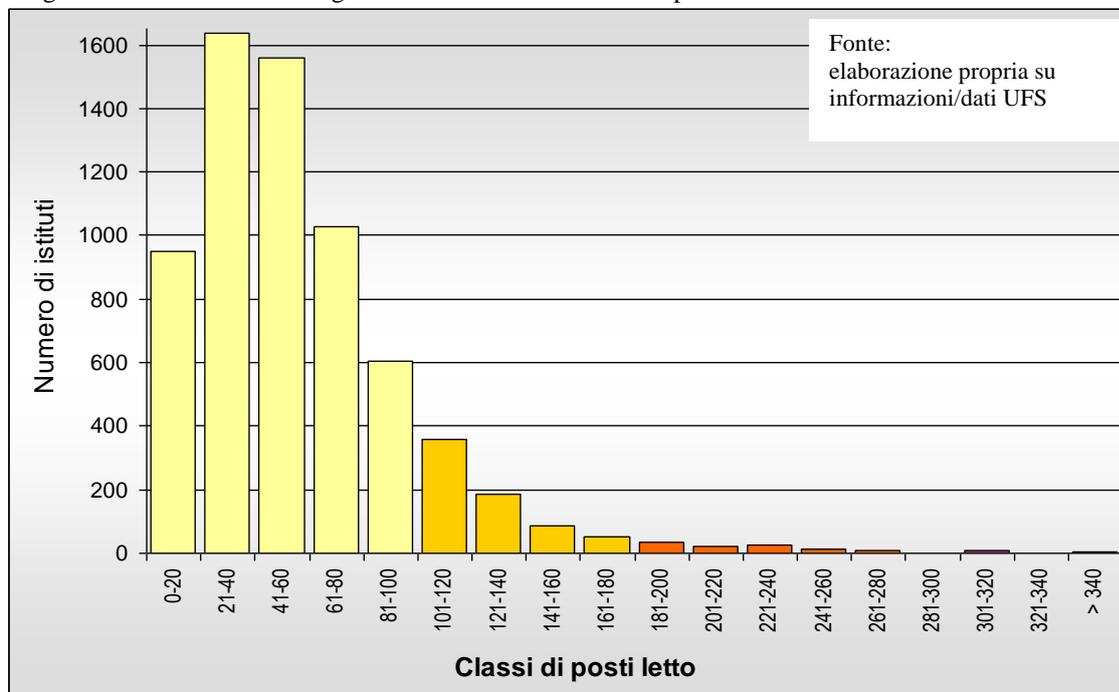
Il concetto di efficienza di scala, che costituisce uno degli obiettivi di ricerca principali di questo studio, è implicitamente un confronto tra istituti che operano con diversi livelli di capacità produttiva. Nel settore delle case per anziani possiamo utilizzare il numero di posti letto come misura della dimensione poiché, se considerato che il tasso di occupazione dei posti letto raggiunge mediamente il 96,6%, è legittimo guardare al numero dei letti come ad un indicatore sia della capacità produttiva sia del livello di attività di un istituto.

L'analisi che segue desidera fornire alcuni elementi descrittivi circa la distribuzione dei posti letto che osserviamo nel campione di istituti per anziani, con lo scopo di verificare innanzitutto se vi siano delle diversità strutturali nella dimensione tra le diverse forme istituzionali.

⁴⁶ Si noti che nella tabella 14 istituti non sono stati considerati dal momento che, secondo la codifica dell'UFS, risultavano chiese oppure organizzazioni straniere e/o internazionali.

La distribuzione dei posti letto si può osservare costruendo opportunamente un istogramma (Figura 2.4) che rappresenti la numerosità degli istituti per anziani nelle diverse classi di posti letto considerate.

Figura 2.4 - Distribuzione degli istituti secondo il numero di posti letto.



Una considerazione che emerge con forza è la forte asimmetria della distribuzione che si concentra prevalentemente in corrispondenza delle prime classi di posti letto, tanto che la soglia dei 100 letti comprende quasi il 90% delle osservazioni. La media campionaria risulta pari a 56.3 letti (deviazione standard pari a 41,67) a testimonianza del fatto che in Svizzera gli istituti per anziani presentano una dimensione contenuta.⁴⁷

Esplorando la parte della distribuzione riferita agli istituti con maggior numero di posti letto possiamo notare che, nonostante la casa per anziani più grande in assoluto conti 355 letti, solamente il 3% delle osservazioni superano la soglia dei 150 posti letto. Questa analisi, che suggerisce come in Svizzera le case per anziani tendano ad essere piccole nel senso assoluto del termine, non permette di capire se le diverse forme

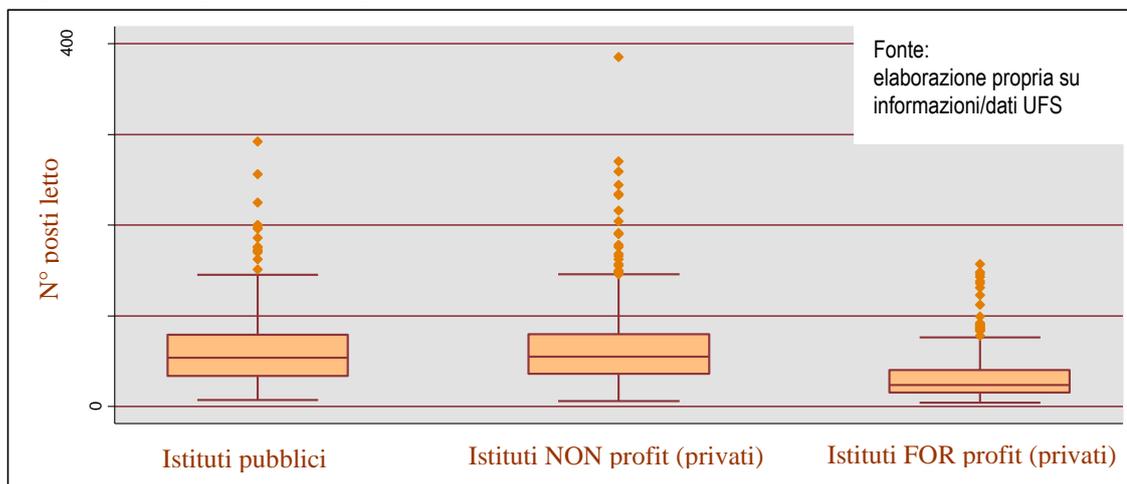
⁴⁷ A titolo di esempio riportiamo la dimensione media (letti) di una casa per anziani tratta da alcuni studi empirici basati su altre nazioni: Norvegia, 66 (Erlandsen and Førsund, 1999); Olanda, 157 (Kooreman, 1994); Italia, 71 (Rebba e Rizzi, 2002); USA (New York State), 201 (Vitaliano, 2003); Francia, 72 (Dervaux et al., 2006).

istituzionali definite in precedenza presentano differenze significative in termini di dimensione.

Uno strumento grafico molto utile per analizzare comparativamente le caratteristiche di una distribuzione è il cosiddetto box-plot. Sull'asse delle ordinate di un box-plot troviamo i valori della variabile di cui si considera la distribuzione (nel nostro caso specifico il numero dei posti letto), sulle ascisse le diverse tipologie di istituti che si desiderano confrontare. Il box-plot consente di fissare in un colpo d'occhio la posizione del valore mediano rispetto al primo ed al terzo quartile (i quali definiscono i lati del box), nonché tutte le osservazioni che distano rispetto al primo (verso il basso) e al terzo quartile (verso l'alto) di almeno 1,5 volte l'intervallo interquartile. Mediante questa rappresentazione grafica si rende visibile la densità di una distribuzione attorno alla mediana e l'eventuale simmetria delle osservazioni localizzate ai margini estremi della distribuzione.

I box-plot rappresentati nella Figura 2.5 si riferiscono ai dati relativi all'anno 2002 e ci consentono di cogliere immediatamente le analogie e le differenze nella distribuzione dei posti letto degli istituti svizzeri suddivisi nelle tre forme istituzionali: casa per anziani pubblica, privata non profit e privata for profit.

Figura 2.5 - Box-plot delle case per anziani, suddivise secondo la forma istituzionale.



La prima considerazione che emerge con forza è la somiglianza tra le distribuzioni degli istituti pubblici e privati non profit che esibiscono entrambi un valore mediano prossimo a 50 letti (54 per le case pubbliche e 55 per le private non profit) ed una forma dei box-plot pressoché identica per altezza e posizione. Notiamo infatti come la linea mediana separi in modo simmetrico i due rettangoli ad indicare che la parte centrale

della distribuzione (ricordiamo che il box contiene il 50% delle osservazioni essendo costruito a partire dal 25° per arrivare al 75° percentile) è piuttosto uniforme.

La presenza di alcuni istituti di grandi dimensioni, che ricadono ampiamente fuori dal limite superiore del box, è riscontrabile in entrambe le tipologie secondo un andamento molto simile a testimonianza del fatto che anche le “code” delle due distribuzioni si assomigliano. Il gruppo di case per anziani private for profit, al contrario, presenta una distribuzione diversa in almeno due aspetti: posizione e forma del box. Il valore mediano dei posti letto offerti dagli istituti privati for profit è pari a 24 ossia meno della metà del valore riscontrato per le altre forme istituzionali. Le case per anziani for profit, oltre ad essere più piccole, sono concentrate nella parte inferiore del box e questo si capisce dalla posizione della mediana che giace nella parte bassa del rettangolo. La distribuzione dei posti letto è quindi asimmetrica presentando un’elevata densità di osservazioni in corrispondenza degli istituti piccoli, contrapposta ad una “coda” rarefatta che giunge ad includere istituti di dimensioni relativamente grandi fino ad un massimo di 157 posti letto.

Il fatto che le case per anziani a scopo di lucro siano piccole in rapporto alle altre tipologie di istituti potrebbe indurre a sostenere un legame tra la dimensione e la capacità di realizzare profitti. Da un punto di vista descrittivo dobbiamo limitarci ad osservare che questa categoria di case per anziani offre un numero limitato di posti letto, anche se una spiegazione plausibile si può azzardare senza entrare nel dettaglio di questo fenomeno. Gli istituti che puntano alla realizzazione di profitti tendono ad orientare l’attenzione verso quelle nicchie di mercato maggiormente redditizie e si profilano, solitamente, come case per anziani con standard qualitativi elevati. Anche se la qualità è un attributo difficilmente osservabile, in questo settore la dimensione dell’istituto è percepita come un fattore che garantisce, se abbinato ad altre comodità (ad esempio i servizi alberghieri), un soggiorno particolarmente confortevole a fronte del quale l’utenza (o meglio, un certo tipo di utenza) è disposta a pagare rette di soggiorno elevate.

2.1.3 Ventaglio di prestazioni offerte

Dopo aver presentato le caratteristiche degli istituti per anziani rispetto a criteri tassonomici, giuridico-istituzionali e dimensionali, desideriamo completare la descrizione del settore con alcune semplici considerazioni sul ventaglio delle

prestazioni e dei servizi erogati dalle case per anziani. Tutte le strutture sono tenute ad offrire un servizio residenziale e assistenziale di base, che soddisfi le esigenze minime di un ospite in termini di assistenza e di cura. A questo servizio di base si aggiungono numerosi altri servizi ausiliari, la cui diffusione varia molto da un istituto all'altro. Le informazioni provenienti dall'UFS, che indicano la presenza o l'assenza dei servizi considerati dal questionario di rilevamento dati, consentono di tracciare una suddivisione dei servizi in 6 categorie principali: (1) cure mediche, (2) cure del corpo, (3) servizi sociali, (4) Terapie, (5) servizi tecnici, (6) Farmacia e laboratori.⁴⁸

Per avere un'indicazione generale sul numero e sulla composizione dei servizi offerti, abbiamo costruito la Tabella 2.2 che rileva il numero medio dei servizi erogati in ciascuna categoria dalle diverse tipologie di case per anziani.

Tabella 2.2 - Numero medio di servizi erogati, secondo la forma istituzionale

Forma Istituzionale	Categorie di servizi erogati						TOTALE SERVIZI
	Cure mediche	Cure del corpo	Servizi sociali	Terapie	Servizi tecnici	Farmacia e laboratori	
Istituto Pubblico	1,65	2,14	2,14	1,56	9,15	0,66	17,30
Istituto NON profit (privato)	1,84	2,23	2,35	1,83	9,28	0,93	18,45
Istituto FOR profit (privato)	2,18	2,92	2,96	2,35	8,54	1,19	20,14

Fonte: elaborazione propria su informazioni/dati UFS

Osservando la colonna con il totale dei servizi erogati si nota che in media gli istituti che offrono un ventaglio più ampio di prestazioni sono quelli a scopo di lucro seguiti dalle organizzazioni non profit. Questa constatazione deve tuttavia essere moderata considerando che il numero dei servizi attivi in una casa per anziani non è un corretto indicatore delle risorse assorbite e delle prestazioni erogate in quell'ambito, poiché non fornisce nessuna indicazione sull'importanza di quel servizio all'interno dell'istituto. Un'informazione che possiamo dedurre con relativa sicurezza è la complessità di funzionamento di una casa per anziani che utilizza i servizi tecnici

⁴⁸ La lista dei possibili servizi è molto dettagliata, come risulta dal questionario UFS consultabile al sito: http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen_quellen/blank/blank/ssmi/07.html.

(amministrazione, lavanderia, cucina, pulizia, etc.) per poter offrire ai propri ospiti servizi di cure in senso lato (cure medico-infermieristiche, servizi terapeutici, servizi medico-tecnici come l'odontoiatria e l'acustica,...) uniti a servizi sociali di varia natura (assistenza psicologica, servizi di animazione, formazione, assistenza religiosa, etc.). Dal confronto tra le diverse forme istituzionali emergono, tutto sommato, più analogie che differenze se eccettuiamo il fatto che gli istituti for profit, pur offrendo complessivamente un numero maggiore di servizi all'utenza in ogni categoria, utilizzano un minor numero di servizi tecnici.

Per cercare di capire meglio la composizione dei servizi erogati dalle case per anziani abbiamo costruito la Tabella 2.3 che propone, per una selezione di servizi generalmente offerti dagli istituti, il grado di diffusione di una particolare prestazione.

Tabella 2.3 - Grado di diffusione di particolari servizi erogati dalle case per anziani svizzere.

Servizio	Istituto Pubblico	Istituto Non profit	Istituto For profit	Servizio	Istituto Pubblico	Istituto Non profit	Istituto For profit
Servizi medici	40,3%	48,9%	68,7%	Economato	92,4%	90,5%	88,0%
Servizi psichiatrici	22,5%	29,2%	47,8%	Farmacia	44,6%	56,7%	67,0%
Cure medico infermier.	97,6%	98,5%	97,9%	Laboratorio di analisi	21,2%	36,1%	51,5%
Parrucchiere	73,8%	72,0%	85,9%	Fisioterapia	51,1%	58,4%	71,1%
Pédicure	74,9%	73,9%	86,6%	Ergoterapia	43,1%	43,8%	50,2%
Servizi odontotecnici	27,3%	31,3%	55,7%	Idroterapia	2,2%	3,6%	7,6%
Apparecchi acustiche	34,0%	38,7%	57,4%	Terapie particolari	10,6%	15,6%	21,6%
Servizio psicologico	18,8%	20,8%	38,8%	Dietologia	18,2%	25,4%	41,6%
Servizio sociale	31,4%	29,3%	41,9%	Psicoterapia	9,3%	14,8%	32,3%
Animazione e svago	75,5%	74,4%	84,2%	Formazione scolastica	2,4%	4,8%	7,2%
Manutenz.	96,5%	95,2%	91,1%	Formazione professionale	6,1%	7,2%	8,6%
Trasporti	76,6%	79,0%	86,6%	Laboratori (Ateliers)	6,5%	11,6%	13,7%

Fonte: elaborazione propria su informazioni/dati UFS

I servizi elencati sono la maggior parte di quelli che potenzialmente potrebbero essere erogati da un istituto per anziani. Si tratta di servizi che possono essere offerti internamente alla casa, oppure esternamente, vale a dire appoggiandosi a delle strutture esterne, ma facendo ricorso al proprio personale, oppure ancora essere acquistati da terzi, mediante un outsourcing completo del servizio in questione. Le cure medico-infermieristiche sono senza dubbio il servizio a più larga diffusione, in quanto vengono prestate da oltre il 95% degli istituti per anziani. Questo dato conferma che la maggior parte degli istituti per anziani in Svizzera offre servizi di assistenza e di cura e non solo servizi di tipo residenziale. La presenza (per almeno alcune ore la settimana) di un medico si riscontra in media in un istituto su due con maggior diffusione (70% circa) nelle strutture private for profit.

Altre tipologie di servizi a grande diffusione sono i servizi di cura del corpo, quali il parrucchiere o la pedicure. Non di rado le case per anziani dispongono di una farmacia o di laboratori di analisi al loro interno in particolar modo le case private for profit. Tra i servizi di terapia offerti sono molto diffusi la fisioterapia e l'ergoterapia, mentre altre forme terapeutiche (idroterapia, musicoterapica e altre) si trovano solo raramente. La psicoterapia è relativamente diffusa nelle case per anziani a scopo di lucro (30% circa) e poco frequente negli istituti pubblici (meno del 10%).

Il ventaglio dei servizi sociali erogati è molto ampio e tra questi i più rappresentati sono il servizio religioso, il servizio di animazione. Alcune prestazioni che richiedono una formazione specialistica (quali il sostegno psicologico, l'assistenza psichiatrica, i servizi sociali, etc.) sono ovviamente meno diffuse, mentre altri servizi, meno tipici del processo produttivo di una casa per anziani (laboratori, formazione professionale, etc.), sono presenti in via quasi eccezionale. Nella Tabella 2.3 non sono, invece, riportati i servizi tecnici imprescindibili (amministrazione, cucina, pulizia, lavanderia, ...) dal momento che la frequenza è molto vicina al 100% e non vi sono sostanziali differenze di importanza nelle tre tipologie di istituti.

In conclusione, l'analisi dello spettro di prestazioni erogate dagli istituti per anziani ci permette di capirne la complessità strutturale, ma anche le possibili differenze che si osservano nel panorama svizzero dell'assistenza agli anziani. In particolare, il confronto tra le caratteristiche degli istituti che appartengono a forme istituzionali diverse ci ha permesso di evidenziare alcune linee di comportamento generali che

indicano come le case per anziani a scopo di lucro siano più propense ad offrire una maggiore quantità di servizi e coprano con maggior frequenza anche alcune prestazioni particolari o specialistiche che difficilmente si trovano nelle altre strutture.

Per contro su alcuni servizi tecnici come l'economato e la manutenzione, gli istituti for profit sono meno attivi rispetto alle case per anziani pubbliche e a quelle private non profit. Senza entrare nel merito della questione, dall'analisi svolta appare chiaro che gli istituti pubblici e quelli privati non profit siano molto simili tra loro in quanto a prestazioni erogate e composizione delle stesse, mentre le case per anziani a scopo di lucro, pur restando in linea generale molto vicine, manifestano qualche tratto specifico.

2.2 Differenze cantonali nelle caratteristiche delle case per anziani

Una caratteristica distintiva del servizio di assistenza agli anziani è quella di essere legato intimamente con il territorio, rendendo necessario uno sguardo da una prospettiva meno generale, in grado di cogliere eventuali differenze di rilievo a livello regionale. Appare cioè molto difficile ignorare il ruolo che il federalismo svizzero gioca in questo settore, sia per quanto riguarda le caratteristiche degli istituti, sia in materia di regolamentazione del settore. Nonostante gli obiettivi di questo lavoro non siano centrati sulla teoria economica del federalismo né sulla verifica empirica dell'impatto della regolamentazione o delle sue determinanti, è necessario fornire un contesto chiaro e realistico della complessità del settore che si analizza. Per questa ragione desideriamo descrivere le differenze che emergono a livello cantonale con riferimento ad alcune caratteristiche importanti per la specificazione del modello di costo.

Come già discusso in precedenza, l'accento non verrà posto sui problemi legati alla regolamentazione cantonale e/o comunale per non introdurre una dimensione di complessità che rischia di compromettere gli obiettivi di ricerca di questa tesi. Con l'introduzione della LAMal nel 1994, infatti, l'attività regolatoria ha conosciuto una forte accelerazione, originando mutamenti anche profondi in brevi periodi di tempo. Il nostro studio è riferito agli anni dal 1998 al 2002 e durante questo periodo diversi cantoni hanno modificato i propri sistemi di regolamentazione del settore delle case per anziani.

Per poter entrare in materia, discutendo da un punto di vista teorico le diverse scelte cantonali e misurarne empiricamente gli effetti con l'analisi econometrica, bisognerebbe disporre di dati quantitativi e comparabili su tutti i 26 sistemi cantonali nel corso dei 5 anni coperti dalla banca dati. Vista l'impossibilità di disporre di queste informazioni, riteniamo più interessante considerare la regolamentazione come una caratteristica latente che genera eterogeneità nella banca dati a livello cantonale, riflettere poi, in sede di interpretazione dei risultati, sul possibile impatto della regolamentazione. Rimanendo volontariamente sul piano descrittivo, è comunque utile alla comprensione della banca dati illustrare le differenze che si riscontrano a livello cantonale nel settore delle case per anziani.

2.2.1 La capacità produttiva cantonale

La prima caratteristica che ci interessa analizzare riguarda la densità di posti letto in casa per anziani ogni 1000 abitanti. Nel 2002 la media svizzera di questo indicatore era pari a 10,44 letti con un minimo di 4,49 (registrato nel cantone Appenzello Interno) ed un massimo di 18,81 (nel cantone Appenzello Esterno). La distanza tra i due valori estremi è rilevante e contiene tutti gli altri cantoni che, in modo più o meno importante, si discostano dalla densità media nazionale.

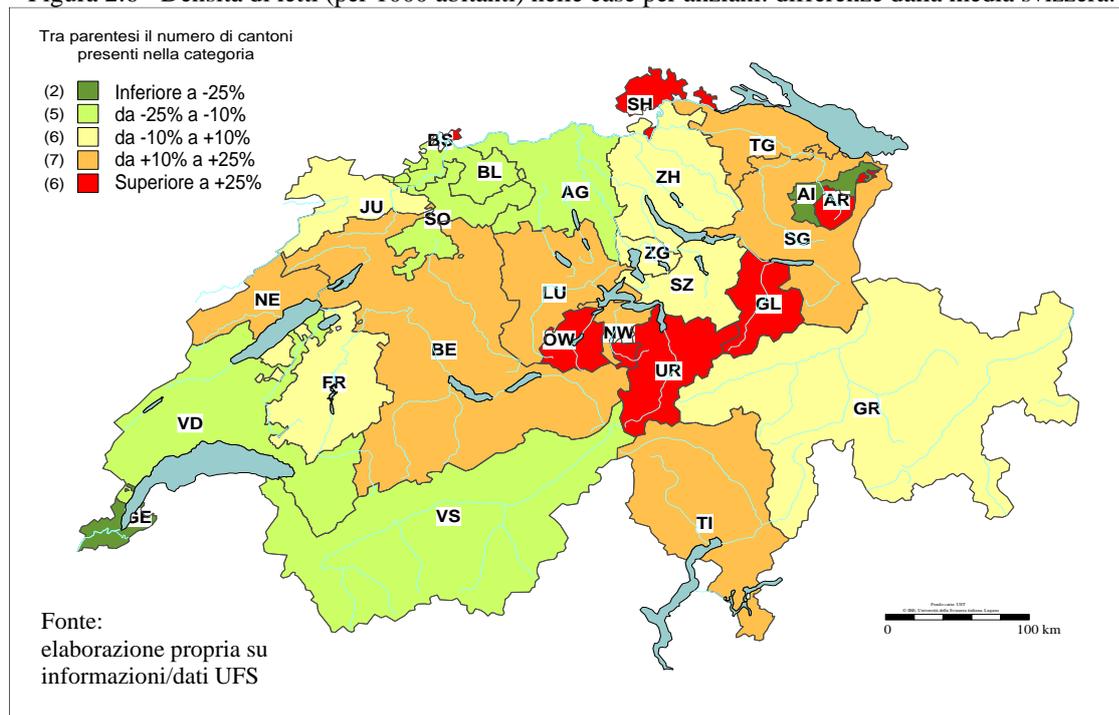
Per cogliere visivamente questa eterogeneità abbiamo costruito una carta tematica che rappresenta, attraverso 5 classi colorate, lo scostamento della densità cantonale dalla media svizzera in termini di posti letto offerti. Osservando la Figura 2.6, si può in qualche modo parlare di una differenziazione dei valori delle densità di letti secondo logiche regionali.

I cantoni romandi (Vaud, Vallese, Ginevra, Jura e Friburgo), con l'eccezione di Neuchâtel, ed i cantoni nord-occidentali ad eccezione di Basilea Città e di Sciaffusa sono contraddistinti da una densità di posti letto inferiore al valore medio.⁴⁹

Il canton Berna e tutti i cantoni situati nella Svizzera centrale (Lucerna, Obvaldo, Nidvaldo, Svitto e Uri) denotano una densità superiore alla media così come i cantoni della Svizzera orientale (Glarona, San Gallo e Appenzello Esterno). In Ticino e nei Grigioni si registra una densità superiore alla media svizzera.

⁴⁹ Si tratta dunque dei cantoni di Basilea Campagna, Soletta, Argovia, Turgovia, Zugo e Zurigo.

Figura 2.6 - Densità di letti (per 1000 abitanti) nelle case per anziani: differenze dalla media svizzera.



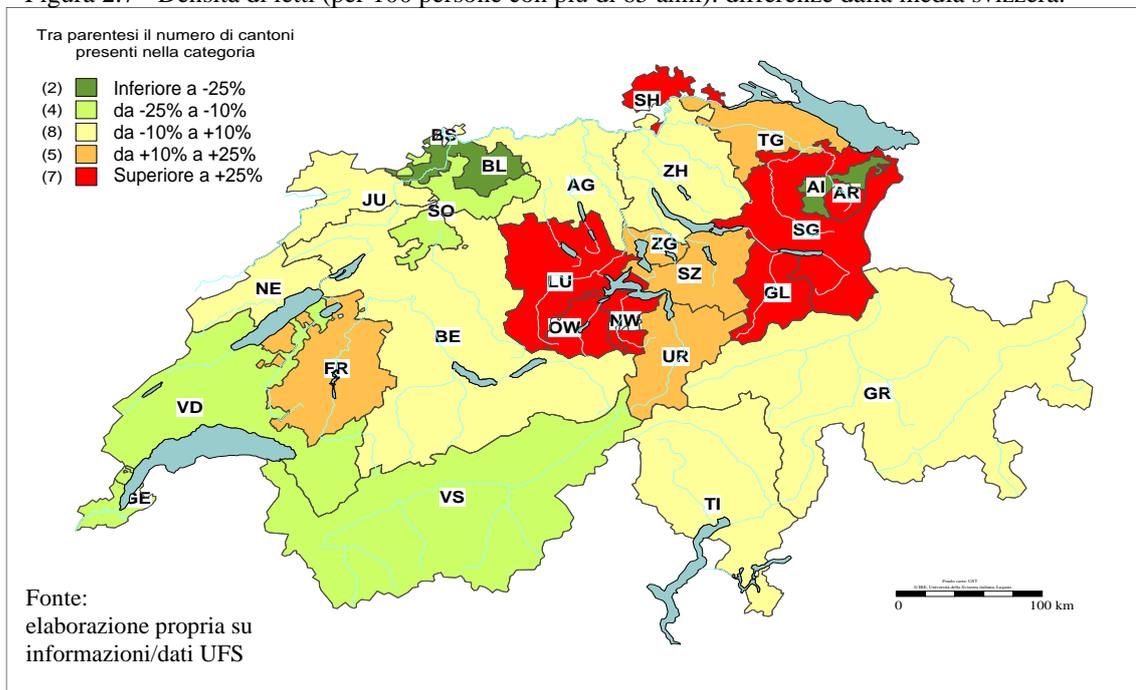
Queste differenze possono essere spiegate da numerosi fattori. In primo luogo, è ipotizzabile che, a seconda dei cantoni (rurali o urbani), le classi di età abbiano una prevalenza diversa e che dunque in alcuni cantoni la percentuale delle persone anziane sia maggiore (ed il fabbisogno di posti letto superiore), mentre in altri la popolazione sia tendenzialmente più giovane (e ciò legittimerebbe una densità dei letti inferiore alla media).

Per verificare questa ipotesi abbiamo calcolato la densità di posti letto ogni 100 persone anziane (persone con un'età superiore ai 65 anni) ed abbiamo riprodotto il risultato nella Figura 2.7. Se si tiene conto dell'anzianità della popolazione, l'immagine appare ancora più chiara con i cantoni romandi ed i cantoni nord-occidentali che tendenzialmente evidenziano una densità di letti inferiore alla media svizzera, pari a 6,71 letti ogni 100 anziani. Più densa di letti appare invece la Svizzera centrale e orientale (in particolar modo i cantoni Lucerna, Obwald, Nidwald, Glarona San Gallo, Appenzello Esterno e Sciaffusa).

L'eterogeneità nell'offerta di posti letto in casa per anziani non può essere pienamente spiegata dalla struttura demografica cantonale e chiama in causa altri fattori, tra i quali le specificità culturali delle diverse regioni, il ruolo assunto dalla famiglia nella società, aspetti di carattere socioeconomico e le priorità di politica regionale,

prima fra tutti l'importanza data allo sviluppo sul territorio di una rete di servizi di assistenza e cura a domicilio.

Figura 2.7 - Densità di letti (per 100 persone con più di 65 anni): differenze dalla media svizzera.



Non sorprende il fatto che siano proprio i cantoni della Svizzera tedesca, nei quali è maggiore l'autonomia dei comuni e la diffusione capillare di istituti non medicalizzati, ad evidenziare la più elevata percentuale di persone residenti in istituti per anziani. Analogamente non stupisce che siano proprio i cantoni della Svizzera romanda, nei quali lo Stato ha optato per una maggior centralizzazione dei poteri pianificatori e per la realizzazione di una rete di servizi medici ed infermieristici sul territorio, a presentare una più bassa percentuale di popolazione che vive in casa per anziani. L'accento diverso dato dalle comunità locali alle possibili strategie di presa a carico della popolazione anziana si riflette dunque nella densità degli istituti residenziali e in quella dei posti letto per numero di persone anziane.

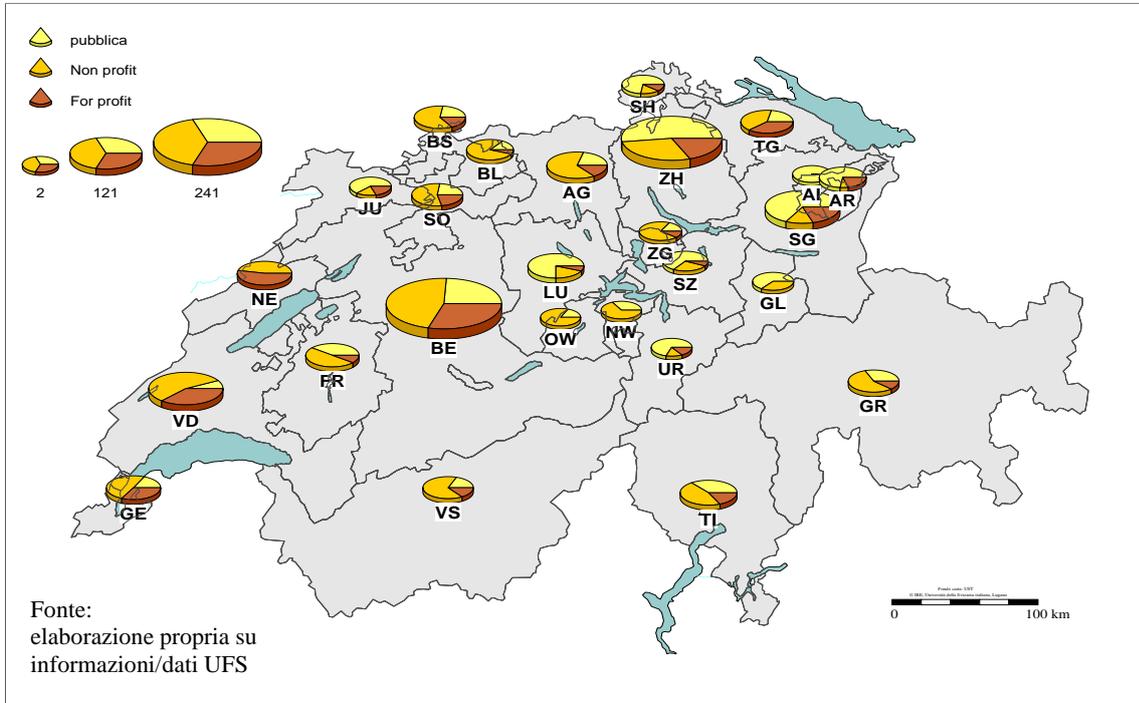
2.2.2 L'organizzazione istituzionale

Una seconda diversità che emerge da un confronto intercantonale è la frequenza con cui ricorrono le tre forme istituzionali identificate in precedenza. Nella Figura 2.8 abbiamo cercato di rappresentare graficamente la prevalenza di istituti pubblici, privati non profit e privati for profit nei diversi cantoni.

Entro i confini territoriali di ogni cantone, tracciati all'interno della cartina della Svizzera, compare un grafico a torta il cui volume corrisponde al numero complessivo

degli istituti, mentre la ripartizione della torta in 3 fette avviene conformemente al peso relativo (in termine di istituti) delle tre forme istituzionali: case per anziani pubbliche, private non profit e private for profit.

Figura 2.8 - Numero di istituti suddivisi secondo la forma istituzionale (anno 2002).



Fatta eccezione per il Canton Neuchâtel, in tutti gli altri cantoni gli istituti senza scopo di lucro (cioè le case per anziani pubbliche e quelle private non profit) sono in netta maggioranza. In alcuni cantoni sono gli istituti pubblici ad essere presenti con maggior frequenza (per esempio a Lucerna, a Zurigo, nella Svizzera orientale), mentre in altri sono più numerosi gli istituti privati non profit (per esempio a Berna, nel Canton Grigioni e nella Svizzera nord-occidentale). Nella Svizzera romanda (Vallese, Vaud, Ginevra e Neuchâtel), così come nella Svizzera nord-occidentale (Basilea Campagna e Basilea Città, Soletta e Argovia), a Berna, nel Canton Turgovia e nei semicantoni di Obwaldo e Nidwaldo gli istituti privati (non profit o for profit) rappresentano oltre il 75% degli istituti. Ancor più degno di nota è il fatto che nei cantoni di Ginevra, Neuchâtel e Vaud, così come nel Canton Berna e nel Canton Turgovia, la percentuale delle case per anziani for profit superi la soglia del 30 per cento, situandosi ad un livello nettamente superiore rispetto alla media svizzera (19,4%).

Le differenze nella prevalenza delle forme istituzionali si riflettono irrimediabilmente sul tipo di regolamentazione in atto nei vari cantoni. Senza procedere ad un'approfondita analisi storica del settore non è tuttavia possibile stabilire se vi sia

un nesso causale tra la diffusione delle diverse forme istituzionali e la tipologia di regolamentazione degli istituti adottata in un cantone. In altri termini non è dato di sapere se sia stato il ritardo con cui lo Stato ha riconosciuto la necessità di assumere un ruolo attivo nell'offerta di servizi residenziali per anziani ad aver in qualche modo incoraggiato la società civile a dare una risposta al problema mediante l'apertura di istituti privati non profit o se sia stata piuttosto la filosofia di regolamentazione adottata in un cantone ad aver favorito la diffusione delle forme istituzionali private.

A nostro avviso, con tutta probabilità, sono da ritenersi corrette entrambe le interpretazioni. La prevalenza di istituti privati non profit può essere probabilmente spiegata chiamando in causa la celerità con cui la società civile ha preso atto di determinati bisogni, anticipando la risposta istituzionale da parte dello Stato, ma l'argomentazione non tiene per quanto concerne gli istituti privati for profit. Non è pertanto possibile escludere a priori che la diffusione di case per anziani a scopo di lucro (per lo più case non medicalizzate) sia stata incoraggiata dal tipo di regolamentazione adottata.

Alla luce degli aspetti problematici che emergono dalla discussione, abbiamo deciso di non includere questa caratteristica del processo produttivo tra le variabili esplicative. La forma istituzionale viene comunque discussa nella parte empirica, dove metteremo a confronto i livelli medi di inefficienza misurati con riferimento alle diverse tipologie di istituti pubblici, privati non profit e privati for profit.

2.2.3 Il profilo dei residenti ed il ventaglio dei servizi erogati

Per arricchire il confronto intercantonale è utile considerare anche il profilo dell'utenza che potrebbe manifestare livelli di bisogno molto differenziati. A parità di sollecitazioni provenienti dal lato della domanda, ogni istituto potrebbe decidere con sensibilità diverse quali servizi attivare e secondo quali modalità. Per rendere conto di questa ulteriore dimensione dell'eterogeneità tra cantoni abbiamo costruito due diagrammi, nei quali rappresentiamo rispettivamente il fabbisogno medio di ore di cura e/o assistenza per ospite ed il numero di servizi medio erogato dalle case per anziani di ogni cantone svizzero.

Iniziamo con il profilo degli ospiti fornendo una breve spiegazione dell'indicatore del fabbisogno medio per residente per istituto, che ci accingiamo a costruire. Sulla base del numero di ore giornaliere che in media ogni singolo residente necessita per le cure

mediche o per l'assistenza nelle attività fondamentali (lavarsi, vestirsi, mangiare, etc.), i residenti sono attribuiti ad una categoria specifica tra le 4 previste.⁵⁰

Per associare ad ogni istituto un valore di sintesi che rifletta il fabbisogno di cure medio degli ospiti residenti abbiamo sostituito ad ogni intervallo categoriale un singolo valore. La Tabella 2.4 chiarisce il metodo seguito, simulando i dati di una casa per anziani e descrivendo la procedura di aggregazione.

Tabella 2.4 - Esempio di costruzione dell'indice del fabbisogno medio giornaliero di cure.

Classe di fabbisogno	N° residenti	Peso (quota %)	Valore attribuito alla classe di fabbisogno	Contributo indice di fabbisogno
0-1 ora al giorno	8	10%	0.5	0.05
1-3 ore al giorno	24	30%	2	0.6
3-5 ore al giorno	32	40%	4	1.6
> 5 ore al giorno	16	20%	6	1.2
TOTALI	80	100%	Indice del fabbisogno medio di cure	3.25

Il valore dell'indice del fabbisogno medio di cure è sostanzialmente una media ponderata delle ore di cura richieste dalle 4 tipologie di residenti, utilizzando come peso la quota percentuale di ciascuna classe di fabbisogno. L'indice che si ottiene permette di distinguere con una certa precisione le differenze nel profilo degli ospiti nei diversi istituti.

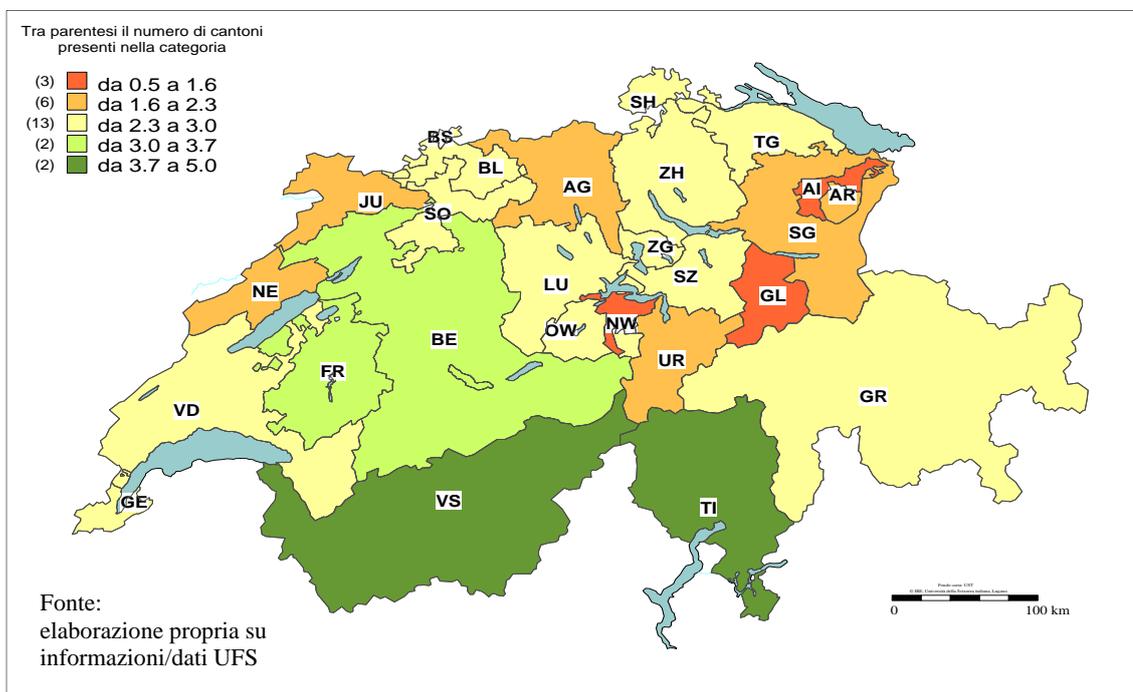
Dal punto di vista teorico, l'indicatore del fabbisogno medio di cure può variare tra il minimo di 0,5 (fabbisogno molto limitato) ed il massimo di 6 (fabbisogno molto elevato), anche se nel campione di case per anziani in nostro possesso i valori minimi e massimi riscontrati non raggiungono questi valori estremi.

Con riferimento all'anno 2002, possiamo osservare nella Figura 2.9 le differenze che esistono nel profilo medio degli ospiti che prevale nei diversi cantoni. Il quadro che emerge, pur prestandosi a diverse considerazioni di dettaglio, suggerisce che la maggior

⁵⁰ Le 4 categorie, basate sul numero di ore di cure/assistenza al giorno, sono: (1) meno di 1 ora, (2) 1-3 ore, (3) 3-5 ore e (4) più di 5 ore.

parte dei cantoni (13 su 26) è caratterizzata da un valore dell'indice del fabbisogno degli ospiti vicino alla media svizzera (2,78).⁵¹

Figura 2.9 - Livello del fabbisogno medio di cure e/o assistenza per istituto (anno 2002).



In alcuni cantoni (Vallese e Ticino su tutti) il valore medio del fabbisogno orario giornaliero è, tuttavia, molto elevato, mentre in altri cantoni (Appenzello interno, Nidwaldo e Glarona) è particolarmente basso. Sembra difficile, invece, raggruppare secondo un criterio geografico omogeneo i restanti 13 cantoni, i cui valori si allontanano dalla media svizzera.

Da questa semplice descrizione emerge l'importanza di cogliere e misurare l'influenza che le differenze nel fabbisogno medio degli ospiti possono esercitare nel processo produttivo di una casa per anziani. Inoltre, l'evidenza che in alcuni cantoni il livello medio del fabbisogno assume dei valori estremi (molto elevato o molto basso), obbliga a riflettere con una visione più ampia e dettagliata, considerando ad esempio il ruolo delle politiche pubbliche e dei servizi di cure a domicilio.

La dimensione del fabbisogno di cure dei residenti presenta elementi di eterogeneità che devono essere considerati nell'analisi, per questo motivo l'indice

⁵¹ La rappresentazione grafica è costruita in modo tale da coprire lo spettro dei valori rappresentati nel campione (minimo = 0,9 e massimo = 4,0) con 5 classi, facendo ricadere della classe centrale il valore medio (2,78) del campione.

medio del fabbisogno di cure sarà impiegato come regressore nella stima econometrica della funzione di costo. La discussione sul profilo della domanda si ferma (per ovvie ragioni editoriali) e lascia il posto all'analisi dei servizi erogati (lato dell'offerta) da parte degli istituti per anziani.

Prima di commentare le risultanze del confronto intercantonale, va ricordato che l'informazione sul numero complessivo di servizi attivati presso una casa per anziani non tiene conto della loro importanza relativa. Un uguale numero di servizi potrebbe quindi nascondere differenze profonde. La tipologia di servizi censiti dal questionario dell'UFS è ampia e copre tanto le attività imprescindibili per una casa per anziani (cure di vario genere, economato, mensa e caffetteria, lavanderia, animazione, sostegno psicologico, etc.), quanto quelle accessorie (farmacia, parrucchiere, servizi religiosi, terapie di vario genere, formazione, etc.). Il numero medio di servizi per istituto offerti in un cantone specifico può dunque variare da un minimo (teorico) di 0 ad un massimo di 30 servizi erogati.

Considerando i dati dell'anno 2002, presentiamo nella Tabella 2.5 i valori medi del numero di servizi erogati nei 26 cantoni svizzeri.

Tabella 2.5 - Numero medio di servizi erogati a livello cantonale (anno 2002).

Cantone	Media	Cantone	Media	Cantone	Media	Cantone	Media
OW	9.9	NW	13.0	TG	15.6	TI	18.2
FR	11.0	LU	13.1	AR	16.1	VD	21.0
BL	11.7	ZG	13.2	AG	16.2	NE	21.2
SZ	12.3	ZH	14.1	SO	16.3	JU	22.0
UR	12.4	SH	14.4	VS	16.8	GE	22.8
SG	12.5	AI	15.0	BE	16.9	Media	16.04
BS	12.7	GL	15.1	GR	17.8	Svizzera	

Fonte: elaborazione propria su informazioni/dati UFS

Organizzando i dati in ordine crescente si osserva che il cantone con il più basso numero medio di servizi è Obwaldo (9,9 servizi), mentre il cantone con il più alto numero di servizi è Ginevra (22,8 servizi), con la media svizzera pari a circa 16 servizi. Osservando la distribuzione dei valori estremi si nota come nella Svizzera "latina", vengano erogati in media molti più servizi. Per contro la zona dove si fornisce un ventaglio di prestazioni più limitato si concentra intorno al lago dei "quattro cantoni" e

comprende i cantoni della svizzera centrale (Lucerna, Obwaldo, Nidwaldo, Uri, Svitto e Zugo), oltre a Basilea e San Gallo.⁵²

Da sottolineare come l'eterogeneità cantonale, rilevata rispetto all'erogazione dei servizi, ignori sia la qualità e la composizione dei servizi, sia la presenza ed il peso di eventuali strategie di esternalizzazione delle attività. Si deve quindi riconoscere che, lavorando con un indicatore molto semplice (numero dei servizi), esistono alcune difficoltà nel catturare in modo compiuto l'eterogeneità nell'offerta rispetto a questa dimensione. Di conseguenza questo particolare aspetto del processo produttivo sarà incluso nella specificazione del modello di costo, ma sotto forma di variabile dummy, conferendogli così un peso maggiormente qualitativo.

Il processo produttivo di una casa per anziani si conferma, dunque, molto articolato e ricco di sfumature anche per quanto riguarda il profilo degli ospiti ed i servizi che gli istituti offrono loro. Anche in questo caso, però, le differenze emerse non si possono sempre cogliere con precisione e, soprattutto, identificare con i confini cantonali e/o linguistici.

2.2.4 Il costo medio per giornata di cura

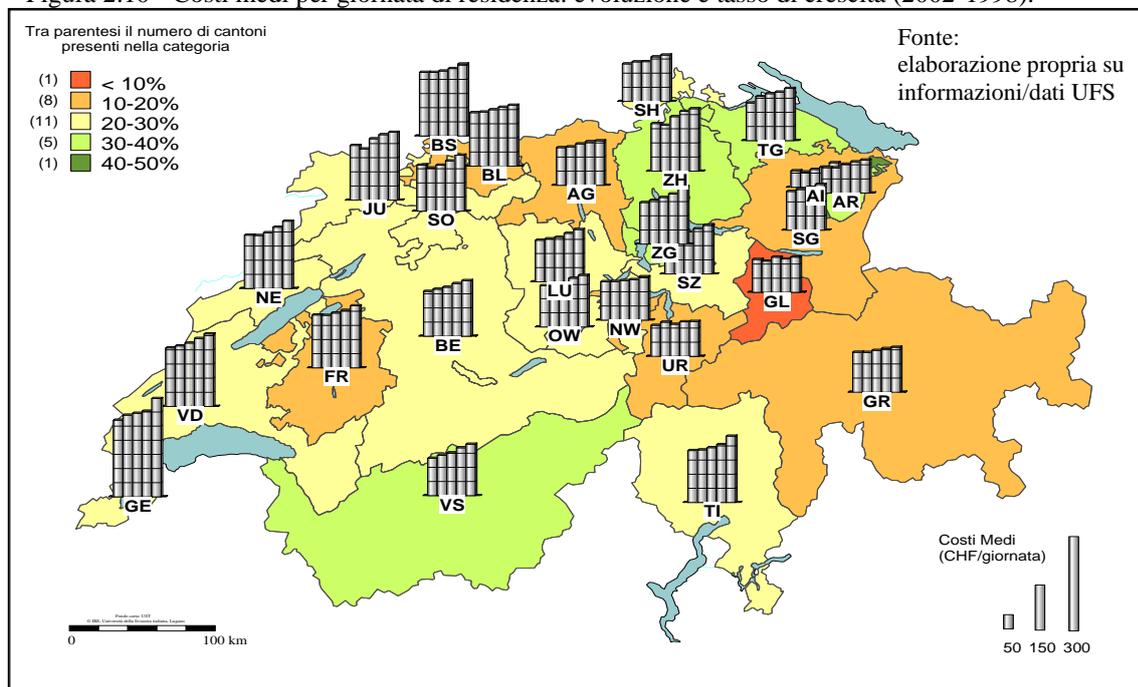
La panoramica dei confronti intercantionali si conclude con il dato più significativo per gli scopi della nostra analisi: il costo medio cantonale per giornata di residenza. Nella Figura 2.10 abbiamo riprodotto sia l'evoluzione dei costi medi nel corso del periodo considerato (1998-2002), sia il tasso di crescita percentuale dell'intero periodo rispetto al valore dell'anno iniziale.

In ogni periodo, l'indicatore del costo medio per giornata di cura a livello cantonale è costruito sommando i costi totali delle case anziani di ciascun cantone in quell'anno e dividendo il risultato per il numero complessivo di giornate di cura prestate durante l'anno in questione. L'evoluzione temporale dei costi medi cantonali è molto semplice da analizzare dal momento che in tutti i 26 cantoni svizzeri si assiste ad un aumento dei costi di gran lunga superiore alla crescita del livello dei prezzi. Se in tutta la Svizzera il costo medio per fornire una giornata di residenza in una casa per anziani è

⁵² Un'ipotesi abbastanza spontanea per spiegare il maggior numero di servizi erogati nei cantoni "latini" è quella di una minor diffusione in queste regioni del servizio di assistenza a domicilio ("Spitex"). Tuttavia, considerando le persone anziane (popolazione con più di 65 anni, anno 2002), l'utenza nei cantoni latini era pari al 22% della popolazione, contro il 16% del resto della Svizzera.

progressivamente più alto, quello che si differenzia tra cantone e cantone è la velocità con la quale i costi crescono.

Figura 2.10 - Costi medi per giornata di residenza: evoluzione e tasso di crescita (2002-1998).



A questo riguardo si nota come il cantone Glarona presenti il tasso di crescita del costo medio più contenuto (+ 7,1%), mentre il tasso di crescita più elevato va ad appannaggio del cantone Appenzello Interno (+ 49%). Un buon numero di cantoni (11) mostra un tasso di crescita compreso tra il 20% ed il 30%, mentre il numero di cantoni con tassi superiori a questa soglia sono relativamente pochi anche se alcuni (Ginevra, Zurigo, Turgovia e Vallese) di un certo peso in termini di giornate prodotte. A prima vista sembra corretto affermare che i cantoni con i tassi di crescita più moderati sono in prevalenza cantoni di piccole dimensioni e/o rurali.

Più che l'evoluzione dei costi ed il suo tasso di crescita è interessante paragonare i livelli assoluti raggiunti nei diversi cantoni in un singolo anno. La variabilità del costo medio fra un cantone e l'altro è assai rilevante. Considerando come riferimento l'anno 2002, ma il discorso vale per tutti gli altri anni, tra il cantone contraddistinto dal costo medio più basso (Appenzello Interno con 87,1 franchi) e quello con il costo medio più alto (Ginevra con 347,7 franchi), la differenza è ragguardevole (fattore 1:4 circa).

Il confronto suggerisce una tesi difficile da sostenere: lo stesso servizio può essere prodotto con un costo fino a 4 volte superiore a dipendenza del cantone nel quale viene erogato. La spiegazione delle differenze di costo medio tra i vari cantoni è un dato di

estrema importanza se consideriamo che l'obiettivo di questo lavoro è proprio la specificazione della struttura che genera tali costi. Ovviamente questa forte variabilità può avere molte cause (che bisognerà considerare nella stima della funzione di costo) e, per commentare nello specifico le differenze di costo medio suggerite dalla Figura 2.10, dobbiamo restare su un piano intuitivo.

Come possibili spiegazioni ci sembra utile menzionare perlomeno: (1) le diversità intercantionali nel costo della vita, che si riflettono in un livello medio dei salari ed un livello dei prezzi nel settore sanitario assai diverso, (2) la tipologia di casa anziani e (3) il numero di istituti (entrambi i fattori hanno delle implicazioni per quanto riguarda il case mix degli ospiti), (4) l'ampiezza dei servizi ed il livello di comfort mediamente offerto dagli istituti (quindi la qualità delle prestazioni erogate), (5) la dimensione delle case per anziani in termini di posti letto (con la conseguenza di incorrere in inefficienze di scala in cantoni piccoli e poco urbanizzati) e (6) l'esistenza di inefficienze di costo.

La panoramica sulle caratteristiche distintive di una generica casa per anziani ed il loro confronto intercantonale mette in evidenza come l'eterogeneità, certamente attesa a priori, insista su molte dimensioni. La semplice descrizione delle differenze cantonali nelle principali caratteristiche ha fatto emergere una complessità profonda ed un intreccio di peculiarità che non si riesce a ricondurre ad una chiave di lettura sistematica. In altre parole, non è possibile utilizzare criteri semplici (ad esempio la lingua, la dimensione, la posizione geografica, o altro ancora) per rendere conto a 360 gradi delle diverse situazioni cantonali.

Se da un lato questa considerazione conferisce grande peso alla dimensione cantonale dei dati quando si analizza una singola caratteristica, dall'altro lato evidenzia come l'eterogeneità complessiva trascenda le logiche cantonali. Dal punto di vista empirico sembra quasi un suggerimento a non insistere troppo sulla scomposizione cantonale, quanto più a cercare di cogliere i fattori che generano quell'eterogeneità rispetto alla quale le differenze cantonali sono la conseguenza e non la causa.⁵³

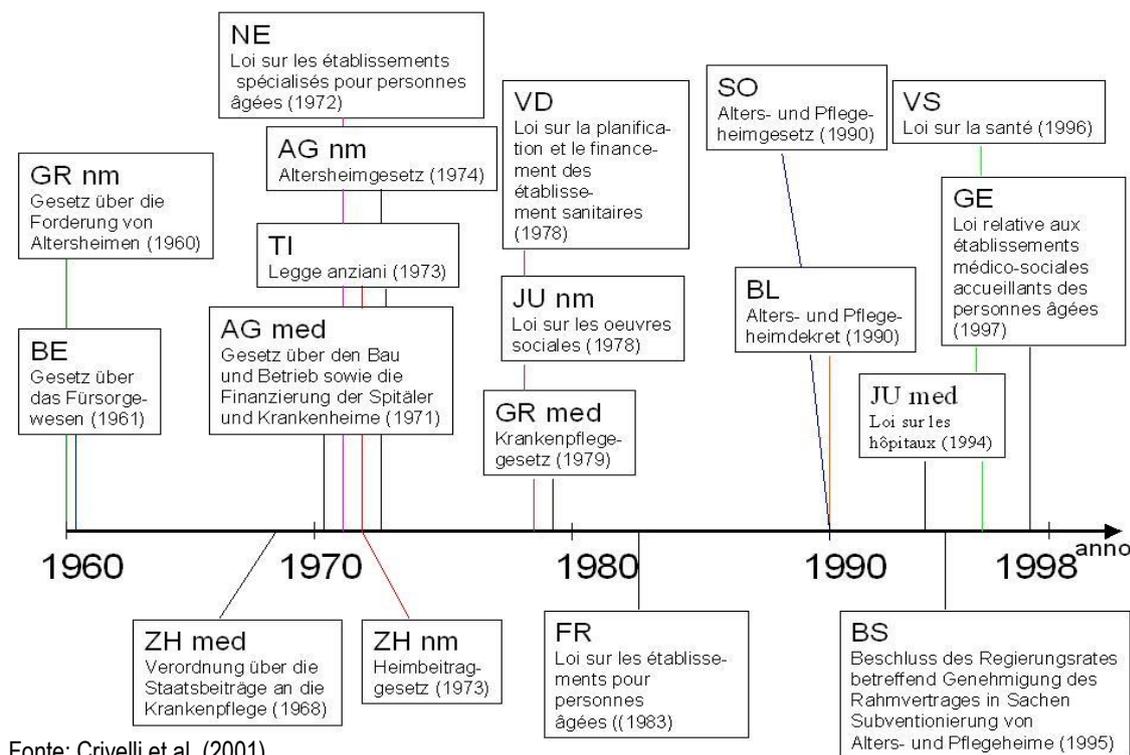
Prima di concludere la parte dedicata alle differenze tra i cantoni desideriamo ribadire nuovamente la scelta di non addentrarci nel tema della regolamentazione. Il

⁵³ L'importanza del livello cantonale e/o di altri criteri (regione, lingua, etc.), per aggregare tra loro i cantoni formando gruppi omogenei, è stata comunque oggetto di indagini econometriche che però non hanno dato buoni frutti, suggerendo strategie di stima diverse.

motivo principale è da ricondurre alla complessità del tema in rapporto alla carenza di informazioni dettagliate sulle scelte cantonali e comunali. Oltre al problema pratico del rilevamento primario dei dati, per il quale non si dispone delle risorse necessarie, esiste un altro problema legato al periodo particolare che il settore sta attraversando per quanto attiene alle scelte di regolamentazione.

Ogni singolo sistema cantonale andrebbe analizzato in una prospettiva storico-politica, attingendo alle basi legali e ai documenti dei dipartimenti sanitari dei 26 cantoni. A titolo di esempio, per dar conto dell'estrema complessità della problematica, desideriamo presentare una rappresentazione grafica (Figura 2.11) della cronologia delle leggi che hanno istituito la regolamentazione nella maggioranza dei cantoni svizzeri. La regolamentazione costituisce senza dubbio un vincolo importante, in grado di incidere sulle scelte delle case per anziani e riflettersi sulle osservazioni empiriche di una banca dati.

Figura 2.11 - Data di nascita delle leggi cantonali concernenti le case per anziani.



Da una prospettiva nazionale, quella adottata nel presente lavoro, è altrettanto vero che le diverse forme di regolamentazioni presenti a livello cantonale si configurano come un'altra forma di eterogeneità da considerare nelle stime.

Inoltre i benefici in termini di una miglior capacità descrittiva dei costi sarebbero probabilmente limitati, vista la difficoltà di conciliare i risultati di una simile mappatura con le esigenze quantitative di una stima econometrica.⁵⁴ L'insistenza nel sottolineare l'esclusione delle problematiche connesse alla regolamentazione non mira certo ad attenuare la responsabilità di questa omissione, quanto piuttosto a ricordare a se stessi ed al lettore che i risultati della ricerca andranno riletti e pesati anche alla luce di questa volontaria censura.

La descrizione delle caratteristiche del settore delle case per anziani permette di capire meglio quali siano gli aspetti più importanti da includere nella specificazione del modello di costo, mettendo comunque in guardia sulla difficoltà di considerarli tutti compiutamente (problema dell'eterogeneità non osservata). Al termine di questa carrellata di informazioni statistiche e di analisi descrittive della realtà svizzera nel settore dell'assistenza agli anziani desideriamo anticipare a grandi linee la struttura del modello di costo scelto per condurre le analisi empiriche.

Oltre all'output principale di una casa per anziani, misurato con il numero di giornate di residenza registrato in un anno, ed ai prezzi dei fattori produttivi, abbiamo considerato altre dimensioni che potrebbero contribuire a spiegare le differenze nei costi totali dei diversi istituti. In particolare, abbiamo considerato il fabbisogno di cure medio ed il livello di medicalizzazione delle singole case per anziani. A queste variabili, che caratterizzano l'output erogato, abbiamo affiancato altri elementi per cogliere alcune diversità strutturali ed organizzative, ad esempio, su: numero di servizi erogati, rapporto tra il personale medico e gli ospiti, presenza di ospiti con handicap fisici severi, intensità nell'impiego di personale a tempo parziale.

Senza entrare nel merito della discussione, si capisce come la definizione del modello di costo di una casa per anziani rappresentativa sia un'attività ricca di insidie. Nella scelta delle variabili esplicative, se da un lato sembra importante perseguire la precisione e la completezza, dall'altro le informazioni sono spesso limitate e, a volte, ridondanti tra loro. Inoltre, l'importanza ed il peso da attribuire ad ogni elemento sono

⁵⁴ Il problema della regolamentazione cantonale è stato affrontato in passato da Crivelli et al. (2002) limitatamente ad una selezione di 16 cantoni svizzeri. Nonostante il grande valore concettuale e di comprensione prodotto dall'indagine sulle forme di regolamentazione cantonale, l'impatto sui risultati econometrici e sulla capacità esplicativa dell'inefficienza di costo è risultato molto limitato.

diversi a seconda del tipo di analisi che si vuole compiere, conferendo maggiore libertà (con tutti i rischi annessi) al ricercatore.

Forniti alcuni elementi di comprensione circa l'oggetto di studio (le case per anziani svizzere) ed il contesto analitico (efficienza produttiva), restano da discutere la cornice teorica entro la quale costruiremo il lavoro e gli strumenti che permetteranno di eseguire materialmente le analisi. Ci riferiamo ai concetti economici che fungono da supporto a questa tesi, così come alla metodologia (e alle tecniche) che ci permetterà di produrre dei risultati empirici.

*“If you don't know where you are going,
any road will get you there.”*

(Lewis Carroll, Scrittore)

3 Elementi di teoria economica della produzione: la funzione di costo e l'efficienza produttiva.

La presente ricerca riguarda l'analisi dell'efficienza produttiva nel settore delle case per anziani in Svizzera e si propone di fornire una misurazione, in termini economici, delle prestazioni erogate dagli istituti di cura elvetici. Il concetto di efficienza produttiva è quindi riferito ai problemi economici che le imprese, in questo caso gli istituti per anziani, devono affrontare durante il processo di trasformazione dei fattori produttivi in beni e servizi da offrire sul mercato. Il processo produttivo deve infatti confrontarsi con il problema della scarsità delle risorse, operando delle scelte all'insegna dell'efficienza.

Nonostante la centralità che l'efficienza produttiva riveste nella teoria microeconomica dell'impresa, questo tema non rappresenta l'unico concetto di efficienza sviluppato dall'economia neoclassica. La teoria microeconomica identifica anche altri concetti molto importanti, quali l'efficienza nello scambio e l'efficienza dinamica. L'efficienza nello scambio sposta l'attenzione dalla singola unità produttiva all'intero mercato e mette in luce le virtù di un (adeguato) sistema di prezzi nel guidare la società verso una situazione dove i benefici (marginali) di chi consuma le risorse sono pari ai costi (marginali) di chi le produce. Ampliando ulteriormente l'orizzonte, l'efficienza dinamica si riferisce al processo di ottimizzazione

intertemporale, seguito dagli agenti economici confrontati con un determinato sistema di incentivi.

Nel corso del tempo, il concetto di efficienza produttiva è stato declinato dalla letteratura economica in diversi modi che, pur intrecciandosi tra loro o divergendo su alcuni punti, si basano (esplicitamente o implicitamente, seguendola o criticandola) sulla teoria neoclassica della produzione. Per questo motivo il capitolo è dedicato all'esposizione di alcuni concetti fondamentali della teoria microeconomica della produzione, selezionati *ad hoc* in base alla rilevanza per la parte empirica.

La prima parte del capitolo descrive e presenta le caratteristiche e le proprietà di una funzione di costo neoclassica regolare, mettendo in particolare l'accento su due aspetti: l'orizzonte temporale (breve o lungo periodo) e la definizione del concetto di economie di scala. La seconda parte del capitolo è invece dedicata alla presentazione dei diversi concetti di efficienza produttiva.

3.1 Elementi scelti di microeconomia della produzione

La teoria neoclassica della produzione assume che le imprese perseguano la minimizzazione dei costi attraverso una scelta razionale ed efficiente degli input. In pratica si ipotizza che l'impresa scelga una combinazione di input tecnicamente efficiente e che sia in grado di minimizzare i costi di produzione, dati i prezzi di mercato dei fattori produttivi. Le relazioni tra input-output e output-costi sono alla base della teoria economica del duale, che permette di derivare gli aspetti tecnologici-produttivi di un'impresa analizzando solamente il lato dei costi. La duplice natura della produzione (tecnica ed economica) dell'impresa si riflette anche sul concetto di efficienza che diventa una grandezza composita dal momento che incorpora due problemi distinti ma uniti (la scelta di una combinazione di input ottimale).

Il contenuto di questa sezione, che si basa principalmente sul lavoro di Shephard (1953), Diewert (1982), Chambers (1988) e Cornes (1992) sulla teoria della produzione duale, mira a presentare alcuni elementi di teoria economica indispensabili per preparare la discussione dei risultati empirici in merito alla struttura dei costi delle case per anziani in Svizzera.

Senza avanzare pretese di originalità, ci accingiamo ad esporre alcuni concetti relativi alla teoria microeconomica della produzione, sottolineando ancora una volta

come la scelta negli argomenti presentati sia dettata sostanzialmente dagli obiettivi di ricerca della tesi e dall'impostazione data all'analisi empirica. Per questa ragione, ad esempio, avendo scelto di stimare una funzione di costo, abbiamo deciso di presentare i concetti della teoria economica del duale partendo dal lato degli input ed enfatizzando maggiormente gli aspetti legati alla minimizzazione dei costi.⁵⁵

3.1.1 Tecnologia, funzione di produzione ed isoquanti

Per descrivere il processo produttivo di un'impresa neoclassica è necessario introdurre alcune definizioni che riguardano le possibilità di utilizzare risorse (input) per trasformarle in beni economici (output).

In termini formali possiamo considerare un vettore ($K \times 1$) non negativo di input $\mathbf{x} = (x_1, \dots, x_K)$ $\mathbf{x} \in \mathfrak{R}_+$ impiegato per ottenere un vettore ($M \times 1$) non negativo di output $\mathbf{y} = (y_1, \dots, y_M)$ $\mathbf{y} \in \mathfrak{R}_+$, grazie alla tecnologia di produzione T . L'insieme di tecnologie di produzione T rappresenta, quindi, tutte le alternative di produzione possibili, ossia quelle combinazioni input-output ottenibili con la tecnologia disponibile:⁵⁶

$$T = \{(x, y) : x \text{ può produrre } y\} \quad [3.1]$$

Per analizzare il problema produttivo di un'impresa è possibile, partendo dall'insieme di combinazioni descritto dalla [3.1], considerare solamente i processi tecnicamente efficienti, obiettivo implicito di un'unità produttiva. Se limitiamo la descrizione al caso di un singolo output, possiamo definire la funzione di produzione (FdP) come la relazione che lega un qualsiasi vettore di input $\mathbf{x} \in \mathfrak{R}_+$ alla massima quantità producibile di output $\mathbf{y} \in \mathfrak{R}_+$.⁵⁷

Considerando la produzione di un solo output (y è uno scalare), la FdP rappresenta lo strumento analitico sul quale costruire la teoria economica della produzione e può essere descritta in modo formale secondo la notazione utilizzata per l'insieme (\mathbf{x}, \mathbf{y}) delle alternative di produzione possibili:

⁵⁵ I concetti presentati in questo paragrafo si trovano con maggiori dettagli nei lavori di Kumbhakar and Lovell (2000), Tani (1986) e Filippini (1991), ai quali si rimanda per un approfondimento.

⁵⁶ Il concetto di tecnologia è definito in modo statico rispetto al tempo e allo spazio.

⁵⁷ La definizione è tratta da Tani (1986).

$$FdP: f(x) = \max \{y | x \in L(y)\}, \forall x \in \mathfrak{R}_+ \quad [3.2]$$

L'espressione [3.2] potrebbe descrivere il processo produttivo di una generica casa per anziani svizzera, nella quale l'output y è misurato dalle giornate di residenza degli ospiti e x è un vettore degli input impiegati nella produzione del servizio.

Per poter impiegare in modo costruttivo la FdP, questa dovrebbe garantire sempre una soluzione al problema di massimo presentato nell'espressione [3.2]. Un modo semplice per assicurare che la funzione di produzione risulti ben definita è quello di attribuire all'insieme T , descritto dalla [3.1], alcune proprietà desiderabili:

1. T non è l'insieme vuoto;
2. T è un insieme chiuso;
3. T è un insieme limitato superiormente per ogni \mathbf{x} finito;
4. Libera disponibilità degli input (\mathbf{x} weak disposability): se $(\mathbf{x}, \mathbf{y}) \in T \Rightarrow (\lambda \mathbf{x}, \mathbf{y}) \in T$ per $\lambda \geq 1$;
5. Possibilità di distruzione senza costo per l'output (\mathbf{y} weak disposability): se $(\mathbf{x}, \mathbf{y}) \in T \Rightarrow (\mathbf{x}, \square \lambda \mathbf{y}) \in T$ per $0 \leq \lambda \leq 1$;
6. Possibilità di inazione: $(\mathbf{x}, 0) \in T$ e se $(0, \mathbf{y}) \in T \Rightarrow \mathbf{y} = 0$;
7. T è un insieme convesso: se $\mathbf{y}, \mathbf{y}' \in T$ e $\theta \in [0, 1]$, allora $\theta \mathbf{y} + (1-\theta)\mathbf{y}' \in T$.

La proprietà 1 richiama l'attenzione sul fatto che nel processo produttivo analizzato gli input e gli output devono essere presenti (al limite in quantità pari a zero) poiché non è interessante parlare di un'impresa che utilizza una tecnologia inesistente.

La proprietà 2 prevede che il limite fissato dalla tecnologia sia effettivamente raggiungibile (ma non superabile). In parole più semplici significa che le imprese possono raggiungere la piena efficienza produttiva.

La proprietà 3 è un corollario del buon senso che stabilisce l'impossibilità di ottenere una quantità illimitata di output partendo da una quantità finita di input.

Le proprietà 4 e 5 sono conosciute con il nome di monotonicità debole e costituiscono, se considerate congiuntamente, la versione meno restrittiva della condizione di monotonicità forte, anche nota con il termine di *free disposability*. Il significato della condizione di monotonicità forte nella tecnologia è molto semplice

poiché prevede la possibilità di “sprecare” input e output durante il processo produttivo di trasformazione. Interpretata in altro modo significa che, partendo da una combinazione (\mathbf{x}, \mathbf{y}) , ogni combinazione $(\mathbf{x}', \mathbf{y}')$ che impieghi quantità non inferiori di input per ottenere quantità non superiori di output, deve essere producibile.⁵⁸ Vista in termini economici la condizione di monotonicità attribuisce all'impresa la flessibilità di espandere/contrarre il proprio livello di produzione in modo radiale (monotonicità forte) o in una delle due direzioni dello spazio input-output.

La proprietà 6 esprime in modo formale il fatto che l'origine degli assi, la combinazione $(0,0)$, appartenga alla tecnologia considerata. L'interpretazione economica di questa proprietà è duplice: è sempre possibile non produrre utilizzando delle risorse \mathbf{x} (*weak disposability*), ma non è possibile ottenere qualcosa dal nulla per cui se $\mathbf{x} = 0$, allora anche \mathbf{y} deve necessariamente essere 0 (*no free lunch*).

La proprietà 7 (convessità) non è sempre richiesta dalla teoria ed implica che una qualsiasi combinazione lineare di due elementi che appartengono all'insieme T deve appartenere al medesimo insieme. In altre parole, la convessità dell'insieme T, congiunta con la proprietà 6, sancisce l'impossibilità di avere rendimenti di scala crescenti nel processo produttivo.⁵⁹ Il significato economico di questa proprietà è abbastanza facile da comprendere poiché esprime da un punto di vista tecnico la difficoltà crescente ad ottenere output addizionale incrementando radialmente gli input utilizzati.

Nella realtà è difficile riscontrare in modo univoco tale comportamento produttivo, dal momento che processi produttivi diversi presentano caratteristiche estremamente variegate. La teoria economica e lo studio empirico di numerosi settori suggeriscono tuttavia tre conclusioni generali: (1) i rendimenti di scala variano molto a seconda delle caratteristiche del settore e, all'interno dello stesso settore, a dipendenza della scala di produzione considerata; (2) spesso in ogni settore si può individuare una determinata soglia a partire dalla quale i rendimenti di scala non sono più crescenti; (3) la scala di produzione espressa dalla maggioranza delle imprese appartiene a quel

⁵⁸ Formalmente la condizione di monotonicità forte prevede: $(\mathbf{x}, \mathbf{y}) \in T \Rightarrow (\mathbf{x}', \mathbf{y}') \in T, \forall (-\mathbf{x}', \mathbf{y}') \leq (-\mathbf{x}, \mathbf{y})$.

⁵⁹ Per una discussione sulle relazioni tra la convessità ed i rendimenti di scala si veda Tani (1986).

sottoinsieme di combinazioni produttive contraddistinte da rendimenti di scala non crescenti.

La teoria economica non può risolvere da sola l'indeterminazione del comportamento dell'output in risposta di variazioni scalari degli input e questo conferisce all'indagine empirica una valenza cognitiva molto importante. L'imposizione data a priori ad alcune restrizioni sulla tecnologia (come la convessità dell'insieme T) potrebbero influenzare i risultati ed è per questo motivo che alcune proprietà vengono utilizzate solo se strettamente necessarie.⁶⁰

Partendo dalla funzione di produzione è possibile spostare l'attenzione sul lato degli input e descrivere il problema produttivo in modo alternativo, introducendo il concetto di "insieme del fabbisogno di input".⁶¹ Questo insieme, denominato $L(y)$, esprime tutte le possibili combinazioni del vettore di input x in grado di produrre un dato vettore y di output:

$$L(y) = \{x : (x, y) \in T\}, \forall y \quad [3.3]$$

Se le proprietà viste in precedenza per l'insieme T sono rispettate, allora possiamo ricavare le seguenti proprietà per l'insieme $L(y)$:

1. $L(y)$ non è l'insieme vuoto per almeno un vettore di output;
2. $L(y)$ è un insieme chiuso
3. se x è finito e y è infinito $\Rightarrow x \notin L(y)$
4. se $x \in L(y) \Rightarrow \lambda x \in L(y)$ per $\lambda \geq 1$
5. $L(y) \subseteq L(\lambda y)$ per $0 \leq \lambda \leq 1$;
6. $0 \notin L(y)$ per $y \geq 0$; e $L(0) = \mathfrak{R}_+$;

⁶⁰ Spesso la scelta di considerare la convessità dell'insieme di produzione dipende dall'obiettivo dell'analisi. Ad esempio Mas Colell et al. (1995), nel presentare i problemi produttivi di un'impresa volta alla massimizzazione dei profitti, ne fanno uso esplicito per poter risolvere univocamente il problema di massimizzazione dei profitti. Cardani (1988), maggiormente interessato ai problemi produttivi dell'impresa (minimizzazione dei costi), utilizza il concetto di quasi-concavità della funzione di produzione, che garantisce la convessità degli isoquanti senza richiedere la convessità in senso stretto dell'insieme di produzione.

⁶¹ Espressione tradotta in modo libero dall'inglese "*Input requirement set*". La terminologia utilizzata dalla letteratura per descrivere lo stesso concetto non è sempre omogenea e si possono trovare anche espressioni come "input set" o "*input set of production technology*".

7. $L(y)$ è un insieme convesso.

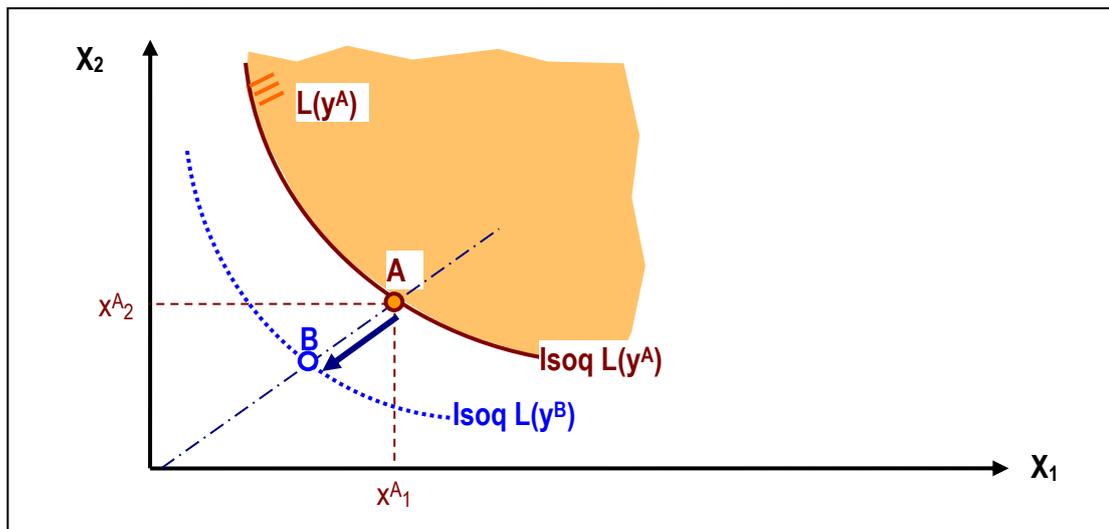
Le sette proprietà rispecchiano quanto già detto in precedenza e per questo le proprietà 4 e 5 si possono sostituire con la condizione più stringente di monotonicità forte (*strong monotonicity*).⁶²

Tra tutte le possibili combinazioni $L(y)$ concretamente realizzabili, le più interessanti dal punto di vista economico sono quelle che giacciono sul limite (*boundary*) dell'insieme $L(y)$, definendone la "frontiera tecnologica". In modo più formale possiamo introdurre il concetto di isoquante, che descrive questo sottoinsieme di $L(y)$:

$$\text{Isoq}L(y) = \{x: x \in L(y), \lambda x \notin L(y), \lambda < 1\} \quad [3.4]$$

Questa distinzione tra l'insieme $L(y)$ ed il suo confine $\text{Isoq}L(y)$, si può chiarire con l'ausilio di un semplice esempio (Figura 3.1) riferito ad una tecnologia produttiva che trasforma 2 input in un solo output.

Figura 3.1 - Rappresentazione grafica di un isoquante (2 input ed 1 output).



La definizione [3.4] indica le combinazioni di due input, x_1 e x_2 , che producono un dato output y senza sprecare (deliberatamente) gli input. In altre parole, significa che il vettore di input x , capace di produrre al massimo l'output y , non è più in grado di farlo quando viene contratto radialmente. Questo significa che, partendo ad esempio dal punto $A(x^A_1, x^A_2)$ sull'isoquante $\text{Isoq}L(y^A)$, una contrazione dei due input, lungo il

⁶² L'espressione formale è: $x \in L(y), x' \geq x \Rightarrow x' \in L(y)$ e $y' \leq y \Rightarrow L(y) \subseteq L(y')$.

raggio che parte dall'origine, individua un nuovo punto (B) sull'isoquante IsoqL(y^B), con la quantità y^B inferiore rispetto alla quantità y^A di partenza.

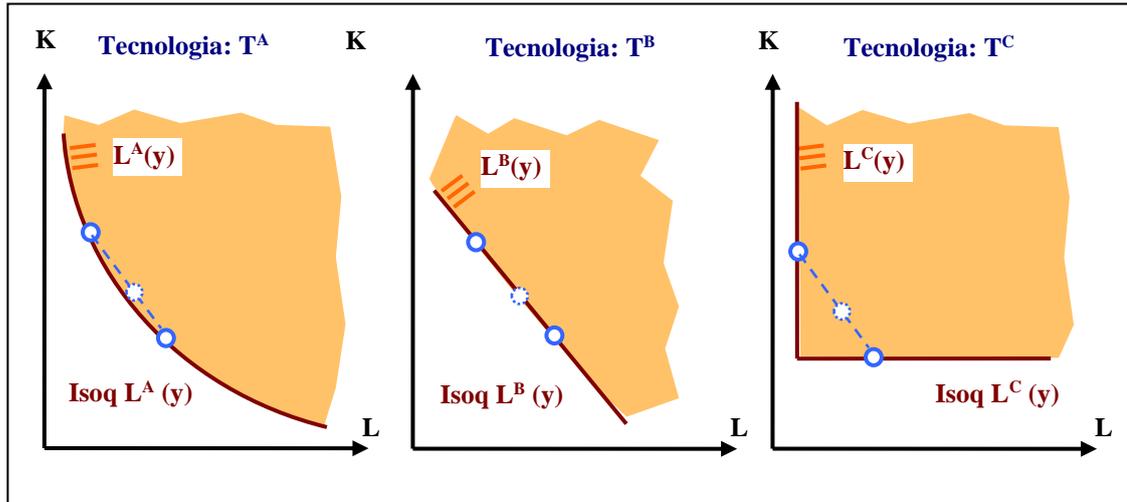
In parole più semplici, l'isoquante rappresenta l'insieme di combinazioni tecnicamente efficienti che appartengono a $L(y)$. Ovviamente il legame tra i due insiemi è molto stretto dal momento che l'isoquante rappresenta il contorno inferiore dell'insieme del fabbisogno di input.

Tra le proprietà dell'isoquante, che discendono direttamente da quelle dell'insieme $L(y)$, la convessità merita un accenno particolare poiché apre una discussione sulla forma che un isoquante può assumere e, di riflesso, sul grado di sostituibilità e/o complementarità dei fattori produttivi. Per definizione l'insieme $L(y)$ è convesso se, presi due punti qualsiasi appartenenti all'insieme, tutte le combinazioni lineari di tali punti appartengono anch'esse all'insieme. In termini economici questa proprietà ha un'implicazione molto semplice: se è possibile ottenere lo stesso output combinando in due modi diversi gli input, allora sarà possibile produrre il medesimo output utilizzando anche altri mix di input intermedi, rispetto alle due situazioni di partenza.

Immaginiamo che la casa per anziani A decida di offrire un certo numero di giornate di residenza in un anno (y_1^A), utilizzando solamente due input, lavoro (x_1) e capitale (x_2), le cui dotazioni sono rispettivamente pari a x_1^A e x_2^A . La casa per anziani B, invece, utilizza la medesima tecnologia produttiva, ma impiega una maggiore quantità di lavoro (x_1^B) ed una minore quantità di capitale (x_2^B) per generare il medesimo output (y_1^B). Se l'insieme del fabbisogno di input $L(y)$ è convesso, sarà possibile trovare un processo produttivo C che impiega un diverso mix di input, frutto di una combinazione lineare dei processi A e B, ma che genera lo stesso output. Se applichiamo la convessità ai punti che giacciono sul confine dell'insieme $L(y)$, discutendo dunque la convessità dell'isoquante, ci accorgiamo che tale proprietà non introduce a priori alcuna restrizione alla tecnologia produttiva per quanto attiene alla sostituibilità degli input.

In altre parole, è possibile descrivere processi produttivi con caratteristiche di sostituibilità degli input diametralmente opposte, ma che contemplano isoquanti convessi. A titolo di esempio, rappresentiamo (Figura 3.2) un processo produttivo caratterizzato da due input, lavoro (L) e capitale (K), ed un solo output (y).

Figura 3.2 - Forme di isoquanti convessi a confronto (2 input ed 1 output).



Definendo opportunamente la forma degli isoquanti è possibile rappresentare tre diverse tecnologie (T^A , T^B e T^C), che presentano caratteristiche molto diverse tra loro, pur soddisfacendo l'ipotesi di convessità degli isoquanti. La tecnologia A esprime una capacità di sostituzione tra i due fattori produttivi imperfetta e questo implica un certo grado di complementarietà tra gli input. Il caso B indica, invece, una perfetta sostituibilità dei fattori produttivi, mentre il caso C esprime la necessaria presenza di entrambi i fattori in una proporzione definita. Il rispetto della convessità, presente in tutti i 3 isoquanti, non impone alcuna restrizione sul grado di sostituibilità dei fattori produttivi.

Questa considerazione è importante, quando dalla teoria si passa alla verifica empirica della tecnologia presente in un determinato settore. Dall'osservazione della realtà è, infatti, possibile derivare alcune relazioni qualitative sul tipo di legame che caratterizza gli input all'interno di un particolare processo produttivo. Nell'ambito delle case per anziani, ad esempio, è difficile pensare che capitale e lavoro siano perfettamente sostituibili così come non sembra realistico considerarli perfetti complementi.

L'ipotesi di convessità dell'isoquante, pur richiamando alla mente quella sostituibilità imperfetta dei fattori di produzione che si riscontra nella maggior parte dei processi produttivi, non comporta alcuna restrizione in merito al grado di sostituibilità

tra i fattori. Per questa ragione la convessità degli isoquanti sarà adottata nel resto del lavoro.⁶³

3.1.2 La funzione di costo (totale) di lungo periodo

In questa sezione presentiamo alcuni semplici concetti riferiti alla teoria economica del duale con lo scopo di introdurre le caratteristiche principali della funzione di costo “neoclassica”.

Il termine “duale” si riferisce alla duplice dimensione, tecnica ed economica, che caratterizza il processo produttivo di un’impresa sia nello spazio delle grandezze fisiche sia in quello delle grandezze monetarie. L’impresa produce beni e servizi utilizzando fattori di produzione combinati tra loro grazie alla tecnologia. L’aspetto “primale” o primitivo analizza il legame tecnologico tra unità fisiche: quantità di output ottenute con una combinazione di quantità di input. Questo processo produttivo “primale” ha un riflesso ben definito in termini monetari che è il costo di produzione. L’aspetto “duale” della produzione consiste nell’osservare le grandezze economiche del processo produttivo (costi della produzione e prezzi dei fattori produttivi impiegati) e risalire da queste alla tecnologia che li ha generati.

In pratica la teoria del duale dimostra che le informazioni sulla tecnologia produttiva di un’impresa si possono ottenere dall’analisi dei costi, senza considerare direttamente la relazione primale tra quantità di input e di output. L’interesse per le funzioni di costo duali nell’ambito dell’economia applicata è motivato da due argomentazioni principali. La prima riguarda l’utilizzo di grandezze economiche, come il costo di produzione o il prezzo degli input, rispetto alla misurazione in unità fisiche. Il processo produttivo che trasforma gli input output è certamente importante da un punto di vista tecnico, in quanto permette di capire se e dove vi sono sprechi di risorse.

⁶³ Il problema della sostituzione dei fattori e le misure empiriche che permettono di rilevarla (elasticità di sostituzione) non sarà oggetto di una verifica empirica. Per un approfondimento su questi temi si veda ad esempio Filippini (1991).

Tuttavia, come avremo modo di precisare in seguito, l'efficienza tecnica è una condizione necessaria ma non sufficiente per garantire ciò che più interessa all'impresa (e all'economista): l'efficienza di costo. I fattori impiegati nella produzione di un determinato livello di output hanno un prezzo, che determina, data la combinazione tecnologica (mix di input) scelta per produrre, un dato livello di costo.

Risulta quindi preferibile analizzare la relazione tra l'output ed i costi di produzione in modo da poter spiegare in termini economici i fattori che determinano i costi di produzione. Il secondo motivo, di natura econometrica, è legato alla forte correlazione tra l'output e gli input (inevitabilmente alta trattandosi di una relazione tecnica) che rende meno efficienti le stime rispetto alla relazione economica della versione duale: costo/output/prezzo dei fattori.⁶⁴

Nel presente lavoro le analisi empiriche si basano sulla stima econometrica di una funzione di costo e questo motivo ci ha indotti a descrivere solamente le caratteristiche del lato duale (il costo) della produzione, rinviando alla vasta letteratura in materia il lettore interessato ad ulteriori approfondimenti.⁶⁵

Il legame che esiste tra produzione e costo di produzione è fondato sull'ipotesi che l'impresa segua un comportamento di minimizzazione del costo di produzione, in presenza di prezzi dei fattori fissati esogenamente.⁶⁶ In termini formali, se consideriamo un vettore strettamente positivo di prezzi $w = (w_1, w_2, \dots, w_N)$, il costo per produrre y sarà pari a $w^T x = \sum_n w_n x_n$. La funzione di costo è quindi rappresentata da:

$$C(y, w) = \min_x \{w^T x : x \in L(y)\} \quad [3.5]$$

⁶⁴ Esiste anche un ulteriore motivo di carattere empirico, che riguarda la possibilità di considerare processi produttivi input perfettamente complementari (tecnologia alla *Leontief*). In questi casi la funzione di produzione non è derivabile, ponendo alcuni problemi al ricercatore. Il problema scompare se consideriamo la corrispondente funzione di costo generata dal processo produttivo descritto.

⁶⁵ Per una discussione completa sugli aspetti della teoria del duale si veda ad esempio Cornes (1992), mentre un approfondimento sugli aspetti produttivi non presentati in questa tesi si vedano Chambers (1988) e Filippini (1991, 1997).

⁶⁶ In altre parole, una funzione di costo neoclassica che prevede la capacità di massimizzare la funzione obiettivo (minimo costo) in presenza di un vincolo esterno (i prezzi dei fattori fissati dal mercato).

Per ottenere informazioni complete sulla tecnologia di produzione, attraverso la sua rappresentazione duale, è tuttavia necessario che la funzione di costo $C(y,w)$ soddisfi le seguenti condizioni di regolarità, ricavate dalle proprietà di $L(y)$:

1. $C(y,w) > 0$ per $y \geq 0$ e $C(0,w) = 0$;
2. $C(y,\lambda w) = \lambda C(y,w)$ per $\lambda > 0$;
3. $C(y,w') \geq C(y,w)$ per $w' \geq w$;
4. $C(y,w)$ è una funzione CONTINUA e CONCAVA in w ;
5. $C(\lambda y,w) \geq C(y,w)$ per $\lambda \geq 1$;
6. se l'insieme T è convesso $\Rightarrow C(y,w)$ è una funzione convessa in y .

L'espressione [3.5] si riferisce ad una funzione di costo di lungo periodo (come si capisce dalla proprietà 1 che non ammette costi la presenza di costi fissi) non negativa (proprietà 1).

La proprietà 2 indica che la minimizzazione dei costi avviene sulla base dei prezzi relativi dei fattori e quindi una variazione scalare (λ) del prezzo degli input si trasferisce nella medesima proporzione al costo (omogeneità lineare).

La proprietà 3 stabilisce che la funzione di costo sia non decrescente nel vettore dei prezzi degli input e quindi se anche il prezzo di un solo fattore aumenta il costo di produzione non può diminuire.

La continuità è richiesta per poter differenziare la funzione rispetto al vettore dei prezzi degli input, mentre la concavità nei prezzi stabilisce che i costi totali di produzione reagiscano in modo meno che proporzionale (o al massimo proporzionale) alla variazione dei prezzi degli input. Queste due caratteristiche, sancite dalla proprietà 4, sono semplici ma molto importanti. Dal punto di vista economico, infatti, la ricerca del costo minimo spinge le imprese a sostituire i fattori produttivi, al variare dei loro prezzi e determina un impatto sui costi di produzione che dipende dal grado di sostituibilità degli input.

Se i fattori di produzione sono parzialmente sostituibili la variazione dei costi sarà meno che proporzionale rispetto alla variazione del prezzo degli input. I due casi estremi sono la tecnologia (alla *Leontief*) che non permette la sostituzione dei fattori di produzione e la tecnologia che permette una perfetta sostituzione tra i fattori. In

quest'ultimo caso i costi sono insensibili alla variazione dei prezzi relativi degli input, poiché è possibile rinunciare ai fattori il cui prezzo aumenta. Nel caso di perfetta complementarità invece, non è possibile modificare le proporzioni degli input e, applicando la proprietà 2, assisteremmo ad un trasferimento lineare della variazione dei prezzi degli input sul costo di produzione.

La Proprietà 5 esprime la monotonicità debole rispetto all'output y che si richiede ad una funzione di costo, mentre la proprietà 6 è necessaria solo se si ipotizza che la sottostante tecnologia T non sia un insieme convesso⁶⁷.

Un risultato notevole della teoria del duale è rappresentato dal lemma di Shephard che stabilisce una chiara relazione tra la funzione di costo [3.5] e le funzioni di domanda delle quantità fisiche degli input, condizionali ad un dato livello di output. Se la funzione di costo duale è differenziabile nei prezzi degli input, vale la seguente relazione:

$$x^*(y, w) = \nabla_w C(y, w) \quad [3.6]$$

La [3.6] ha una chiara interpretazione economica che attribuisce al vettore di input x^* la capacità di minimizzare i costi per produrre y , dato il vettore dei prezzi dei fattori w . In altre parole, per l'impresa che opera al costo minimo vale $w^T x^* = C(y, w)$, mentre per l'impresa che non minimizza i costi avremo $w^T x > C(y, w)$ e $x^* \neq x$.

Questa semplice considerazione indica che la stima empirica di una frontiera di costo permette di derivare le informazioni sulla tecnologia, dato che le proprietà di una funzione di costo si possono interpretare come il riflesso delle proprietà della funzione di produzione sottostante. Il lemma di Shephard è particolarmente interessante per analizzare il discorso della combinazione ottimale di input in risposta ai prezzi di acquisto dei fattori stessi e quindi della possibilità di sostituire tra loro i fattori di produzione, alla ricerca del minimo costo.

La domanda condizionale x_k dei singoli input, e quindi la combinazione dei fattori, è influenzata dal prezzo relativo degli input. In base alla [3.6] è possibile

⁶⁷ Anche nel caso della funzione di costo è possibile sostituire la monotonicità debole con la versione forte che formalmente recita: $C(y', w) \geq C(y, w)$ per $y' \geq y$.

descrivere l'andamento della domanda condizionale di un input partendo dalla derivata seconda della funzione di costo al prezzo w_j dei fattori:

$$\frac{\partial x_k(y, w)}{\partial w_j} = \frac{\partial^2 C(y, w)}{\partial w_k \partial w_j}, \quad k, j = 1, \dots, K \quad [3.7]$$

Tutte le domande condizionali dei fattori si possono quindi ricavare dalla matrice Hessiana della funzione di costo che, per rispettare le condizioni di regolarità, deve essere una matrice negativa semidefinita.⁶⁸

Il significato economico è molto semplice poiché indica la reazione negativa della quantità di un input alle variazioni del proprio prezzo: quando il prezzo relativo di un fattore aumenta, l'impresa cerca di utilizzarlo in misura minore. Da sottolineare che la combinazione ottimale di input risulta insensibile a cambiamenti equiproporzionali nel livello dei prezzi dei fattori, dato che in questo caso i prezzi relativi non si modificano.⁶⁹ Per contro, con l'aumentare dell'output è possibile che la condizione di costo minimo implichi un aggiustamento del mix ottimale di fattori produttivi, individuando quello che in economia si chiama "sentiero di espansione". La necessità di aggiustare la combinazione di input al crescere della produzione dipende dal tipo di tecnologia espressa dal processo produttivo.

Un caso particolare che merita un breve approfondimento è rappresentato dalla tecnologia omotetica, che implica un sentiero di espansione lineare. In altri termini le proporzioni tra gli input utilizzati per minimizzare i costi di produzione restano costanti per ogni livello di output, escludendo la possibilità di sostituzione dei fattori per cause diverse dalla modifica dei prezzi degli input. Da un punto di vista empirico questa distinzione ha un grande peso poiché, come verrà chiarito in seguito, nell'ambito di un'analisi econometrica la scelta della forma funzionale può imporre a priori alcune

⁶⁸ Nel caso di due input la matrice Hessiana ha dimensioni 2x2 ed è negativa semidefinita quando i due elementi della diagonale principale sono non positivi.

⁶⁹ Se la funzione di costo è omogenea di grado 1 nei prezzi dei fattori allora la sua derivata rispetto ai prezzi dei fattori sarà omogenea di grado 0. Dal lemma di Shephard sappiamo che le domande di input corrispondono al gradiente del primo ordine rispetto al prezzo dei fattori e quindi sono omogenee di grado 0 rispetto al vettore dei prezzi.

caratteristiche (ad esempio l'omoteticità) alla tecnologia produttiva, senza averne verificato la pertinenza con la realtà osservata.⁷⁰

Per fare un esempio concreto, il settore delle case per anziani in Svizzera si può sicuramente rappresentare in modo semplificato come un processo produttivo che trasforma due input, lavoro e capitale, in un solo output, le giornate di residenza degli ospiti. L'analisi descrittiva dei dati in nostro possesso evidenzia come l'intensità d'uso del capitale tenda a crescere con la dimensioni degli istituti suggerendo una certa cautela nell'imposizione di eventuali restrizioni come l'omoteticità alla tecnologia.⁷¹

Un'altra considerazione che emerge con forza dall'osservazione del settore dell'assistenza agli anziani, così come in ogni settore economico, è che la modifica dell'input-mix (ammesso che sia possibile) comporta sempre una riorganizzazione del processo produttivo e richiede tempo a sufficienza. Per questo motivo in economia si è soliti precisare quale orizzonte temporale si sta considerando nell'analisi e si distingue tra breve periodo e lungo periodo.

Finora tutti i ragionamenti presentati si riferivano implicitamente ad un'impresa che può disporre a piacimento di tutti gli input necessari, ivi compreso il tempo. In quest'ottica di lungo periodo, la minimizzazione dei costi (e quindi la scelta ottimale delle quantità dei fattori di produzione) non è soggetta ad alcun vincolo di capacità, poiché si considera la possibilità di modificare anche la scala di produzione. Nel breve periodo, invece, le scelte di minimizzazione dei costi devono necessariamente considerare la presenza di alcuni fattori produttivi che sono imm modificabili a breve termine.

L'istituto per anziani eroga servizi residenziali ed è evidente come la disponibilità dei posti letto sia soggetta a forti vincoli di capacità. Per questo motivo è bene riflettere sulle caratteristiche della funzione di costo quando alcuni input non sono modificabili, dato il breve orizzonte temporale (o la presenza di altri vincoli) considerato.

⁷⁰ Per un approfondimento si rimanda alla discussione sulle forme funzionali presentata nel capitolo 4.

⁷¹ Per una valutazione empirica dei problemi legati alla sostituzione dei fattori si veda Cawley et al. (2006).

3.1.3 La funzione di costo (variabile) di breve periodo

Il motivo che spinge gli economisti a distinguere tra breve e lungo periodo è legato al fatto di riconoscere che si tratta di due problemi di ottimizzazione delle risorse differenti. Nell'ottica di lungo termine, infatti, tutti gli input sono disponibili e si possono combinare secondo le possibilità tecnologiche, mentre nel breve termine esiste qualche tipo di vincolo sulle quantità di input.

La semplicità concettuale della distinzione tra breve e lungo periodo si scontra, però, con la difficoltà di individuare con precisione quali fattori siano variabili e quali fissi in un'ottica di breve periodo. In effetti, non è possibile classificare i diversi input in una lista chiusa dal momento che ogni fattore potrebbe, a seconda del processo produttivo, essere fisso o variabile nel breve periodo.⁷²

Da un punto di vista formale, possiamo trattare il breve periodo assegnando ciascun fattore di produzione del vettore \mathbf{x} degli input al sottoinsieme \mathbf{x}_1 degli input flessibili o a quello \mathbf{x}_2 degli input fissi. In questo modo l'insieme del fabbisogno di input di breve periodo è condizionato dal livello di output e dalle quantità dei fattori fissi:

$$L(y, \mathbf{x}_2) = \{ \mathbf{x}_1 : (\mathbf{x}_1, \mathbf{x}_2, y) \in T \} \quad [3.8]$$

La minimizzazione dei costi nel breve periodo che discende dalla [3.8] è legata ai soli fattori di variabili che definiscono la funzione di costo variabile di breve termine:

$$Cv(y, \mathbf{w}_1, \mathbf{x}_2) = \min_{\mathbf{x}_1} \{ \mathbf{w}_1^T \mathbf{x}_1 : \mathbf{x}_1 \in L(y, \mathbf{x}_2) \} \quad [3.9]$$

La [3.9] deve soddisfare le medesime caratteristiche della funzione di costo totale [3.5], che di seguito elenchiamo per semplicità di lettura:

1. $Cv(y, \mathbf{w}_1, \mathbf{x}_2) > 0$ per $y > 0$ e $Cv(0, \mathbf{w}_1, \mathbf{x}_2) = 0$;
2. $Cv(y, \lambda \mathbf{w}_1, \mathbf{x}_2) = \lambda Cv(y, \mathbf{w}_1, \mathbf{x}_2)$ per $\lambda > 0$;
3. $Cv(y, \mathbf{w}_1', \mathbf{x}_2) \geq Cv(y, \mathbf{w}_1, \mathbf{x}_2)$ per $\mathbf{w}_1' \geq \mathbf{w}_1$;
4. $Cv(y, \mathbf{w}_1, \mathbf{x}_2)$ è una funzione CONTINUA e CONCAVA in \mathbf{w}_1 ;
5. $Cv(\lambda y, \mathbf{w}_1, \mathbf{x}_2) \geq Cv(y, \mathbf{w}_1, \mathbf{x}_2)$ per $\lambda \geq 1$;

⁷² Si pensi ad esempio al fattore lavoro e alla facilità nel modificare in breve tempo la forza lavoro. Il tempo necessario per assumere o licenziare personale dipende dal tipo di contratto in vigore.

6. se l'insieme T è convesso $\Rightarrow C_v(y, w_1, x_2)$ è una funzione convessa in y .
7. $x_2 \geq x_2^* \Rightarrow C_v(y, w_1, x_2) \leq C_v(y, w_1, x_2^*)$.

Le 6 proprietà si interpretano in modo analogo e stabiliscono le caratteristiche della relazione tra l'output (y) ed i fattori variabili (x_1).

In aggiunta alle 6 proprietà già discusse per la funzione di costo totale è richiesto (proprietà 7) che i costi variabili non siano crescenti nel fattore fisso. Questo significa che se il vincolo costituito dal fattore fisso viene allentato (si può disporre di una quantità diversa di x_2), è possibile ri-ottimizzare anche i fattori variabili ottenendo un costo (variabile) uguale o inferiore a prima.

Nel breve periodo i costi totali si ottengono sommando il costo variabile relativo alla produzione di un certo output ai costi fissi generati dalla quantità (immodificabile nel breve periodo) di fattore fisso x_2 . Il costo totale di breve periodo, associato alla produzione di un dato livello di output, risulta solitamente più elevato del corrispondente costo totale di lungo periodo.

Il motivo è molto semplice ed è rappresentato dalla presenza del fattore fisso. Il legame tra il costo totale di lungo e di breve periodo si può formalizzare nel modo seguente: la minimizzazione dei costi variabili è una condizione irrinunciabile nel breve periodo e si affianca, attraverso la ricerca (e soluzione) della quantità ottimale di fattore fisso, alla minimizzazione dei costi totali nel lungo periodo. In pratica possiamo immaginare un'ottimizzazione in due fasi successive, ma intimamente legate, espresse da:

$$C(y, w) = \min_{x_2} C_v(y, w_1, x_2) + w_2^T x_2 \quad [3.10]$$

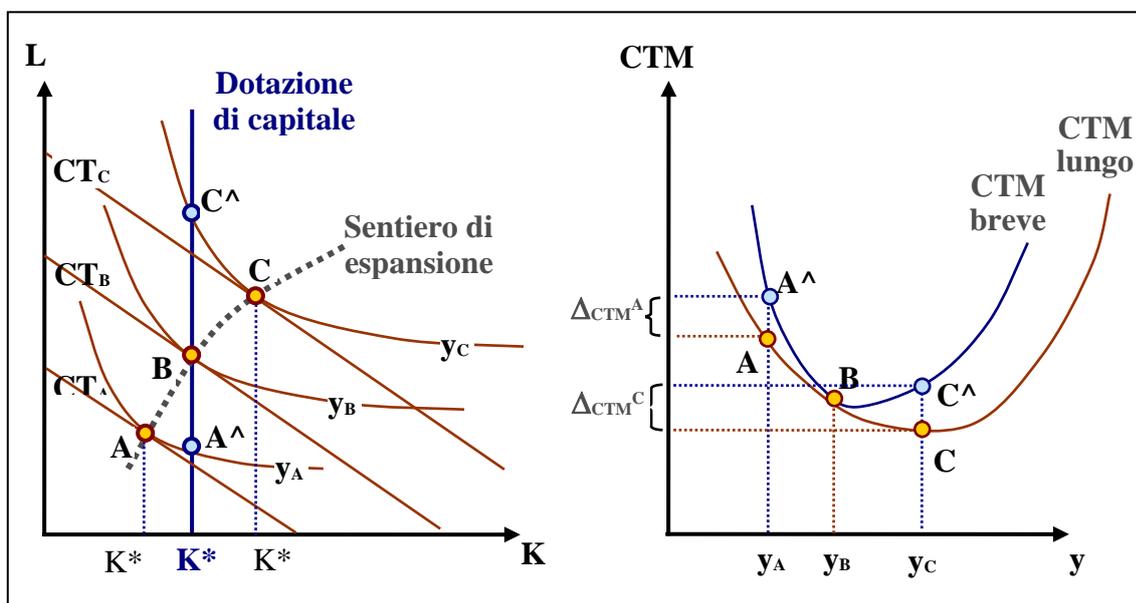
Sulla base della [3.10] la funzione di costo totale di lungo periodo $C(y, w)$ risulta l'involuppo inferiore della funzione di costo totale di breve periodo tanto nello spazio degli output quanto nello spazio dei prezzi dei fattori. Questo significa che i costi di breve periodo non sono mai inferiori a quelli di lungo periodo e che l'elasticità della domanda condizionale dei fattori rispetto al proprio prezzo è maggiore nel lungo periodo.

Osservando le condizioni del primo ordine della [3.10] possiamo trarre un importante insegnamento: il processo di aggiustamento nei fattori fissi va avanti finché il costo da pagare (w_2) è inferiore alla diminuzione dei costi variabili di breve periodo:

$$w_2 = -\nabla_{x_2} C_v(y, w_1, x_2) \quad [3.11]$$

Consideriamo, ora, un processo produttivo con un singolo output e due soli input e rappresentiamo (Figura 3.3) il legame tra i costi, totali e medi, di lungo e di breve periodo. Riferendoci al settore delle case per anziani possiamo esemplificare il legame tra i costi di breve e di lungo periodo, considerando il fattore lavoro come un input flessibile e il fattore capitale come un input fisso nel breve periodo.

Figura 3.3 - Condizioni di involuppo e costi medi di breve e di lungo periodo.



Il pannello di sinistra della Figura 3.3 rappresenta il sentiero di espansione nel lungo termine, periodo nel quale è possibile modificare anche le dotazioni di capitale. Nel breve periodo è possibile modificare solamente il fattore variabile (lavoro). I punti A, B e C si trovano sul sentiero di espansione ed indicano la combinazione di fattori produttivi che minimizza il costo totale per produrre rispettivamente le quantità di output y_A , y_B e y_C .

Nel breve periodo, tuttavia, il sentiero di espansione è determinato dalla quantità di capitale presente che è fissa e quindi per modificare l'output prodotto, minimizzando i costi nel breve periodo, si può agire solo sulla quantità di lavoro impiegata. Immaginiamo che la quantità di capitale corrisponda al livello k_B per cui la

combinazione ottimale dei fattori è rappresentata dal punto B. Nel breve periodo per diminuire la produzione fino al livello y_A , occorre ridurre la quantità di lavoro, mentre parte del capitale resta inutilizzata.

Come si vede dal grafico, la combinazione ottimale per produrre y_A nel breve periodo è indicata dal punto A^* , nel quale si registra una capacità produttiva sovradimensionata. Potendo aggiustare lo stock di capitale, infatti, ci si muoverebbe lungo l'isoquante fino al punto A (sostituendo l'esubero di capitale con l'impiego di forza lavoro), riducendo i costi totali di produzione rispetto al breve periodo. Il punto C rappresenta una situazione analoga nella quale l'impresa si trova con la necessità di aumentare la produzione senza poter aumentare lo stock di capitale. Per produrre l'output y_C si deve intensificare l'uso del fattore lavoro fino al punto C^* , quando la combinazione ottimale di lungo periodo prevede una combinazione (punto C) con più capitale e meno lavoro.

Nel pannello di destra della Figura 3.3 le considerazioni presentate sopra si possono tradurre in termini di Costi Totali Medi (CTM) di breve e lungo periodo, evidenziano bene il legame tra i costi nei due orizzonti temporali ed il concetto di inviluppo. Nel punto B del grafico la dotazione di capitale è ottimale e quindi il costo medio di breve e di lungo periodo coincidono. Nel breve periodo, se si desidera produrre di meno (esempio y_A) o di più (esempio y_C) rispetto all'output y_B , occorre muoversi lungo la CTM di breve periodo incorrendo in un eccesso di costo.

Nell'analisi empirica, le differenti condizioni di minimizzazione dei costi nel breve e nel lungo periodo si riflettono inevitabilmente nella scelta del tipo di funzione di costo da stimare. Una valutazione comparata dei vantaggi e svantaggi connessi ai due approcci, che sarà presentata in modo dettagliato nel capitolo 5, ci ha convinto ad optare per la funzione di costo totale. Nonostante la funzione dei costi variabili non sarà utilizzata nelle stime, la conoscenza delle sue proprietà si rivela utile per argomentare consapevolmente le proprie scelte, visto che l'approccio di lungo periodo scelto non comporta solo vantaggi.⁷³

⁷³ Per una discussione sui problemi legati alla misura delle economie di scala con una funzione di costo erroneamente specificata si veda Aletras (1999).

3.1.4 Economie di scala

Un aspetto importante dell'attività produttiva di un'impresa riguarda l'analisi dell'andamento dei costi di produzione in rapporto alla dimensione (scala) dell'impresa.

Seguendo Caves et al. (1984), il modo più diretto per misurare le economie di scala è quello di calcolare il reciproco dell'elasticità di costo rispetto all'output:

$$E_{Scala} = \frac{1}{\frac{\partial \ln C(y, w)}{\partial \ln y}} = \frac{1}{\varepsilon_y(y, w)} \quad [3.12]$$

La formulazione [3.12] si riferisce ad un generico processo produttivo ed accomoda anche il caso di un'impresa multiprodotto. Se, tuttavia, concentriamo l'interesse ad un solo output, la [3.12] si interpreta in modo molto semplice, poiché, seguendo Panzar and Willig (1977), l'elasticità di costo per un solo output si riduce al rapporto tra costo medio e costo marginale:

$$E_{Scala} = \frac{C(y, w)}{y[\partial C(y, w) / \partial y]} = \frac{AC(y, w)}{MC(y, w)} \quad [3.13]$$

L'indicatore espresso dalla [3.13], generalmente conosciuto con il termine di “*Economies of scale*” (in italiano: Economie di scala), può essere fonte di equivoci. Nell'ottica di lungo periodo la [3.13] indica la presenza o meno di economie di scala, poiché esprime la sensibilità dei costi (di lungo periodo) al variare dell'output.

La teoria economica, riconoscendo il fatto che l'output prodotto la dimensione dell'impresa non sono (soprattutto nel breve periodo) la stessa cosa, ha sviluppato e definito numerosi concetti quali: economie di densità/volume, economie di dimensione, economie di scala.⁷⁴ Date le caratteristiche produttive del settore dell'assistenza agli anziani, nel nostro lavoro la distinzione tra i vari concetti di economie di scala, economie di dimensione e di economie di densità non è di particolare rilevanza.⁷⁵

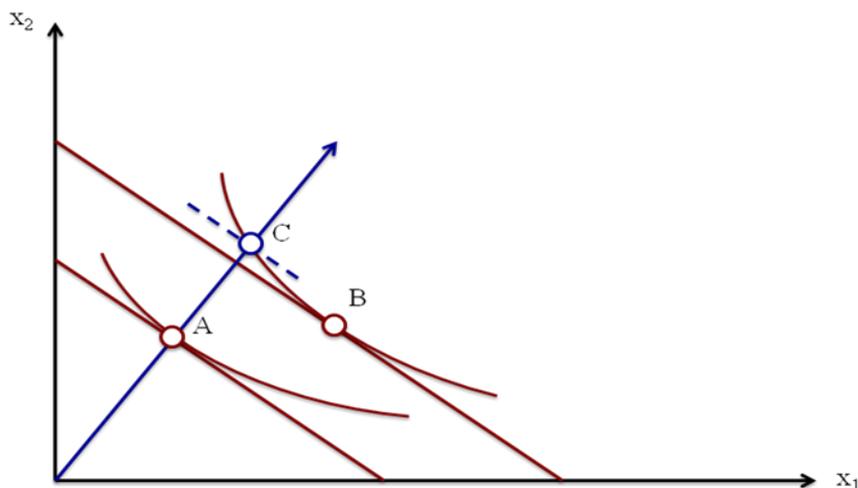
⁷⁴ Per un approfondimento su questi temi si rimanda a Chambers (1988) e Caves et al. (1984; 1985).

⁷⁵ La distinzione è molto importante in quei settori dove la produzione avviene nello spazio per mezzo di una rete (es. ferroviaria, elettrica, idrica, etc.). In questi casi, a parità di dimensione totale della rete è possibile aumentare l'output intensificando (densità) lo sfruttamento della rete.

Nel resto dell'analisi utilizzeremo per semplicità il termine economie di scala, non prima però di aver fatto una precisazione concettuale importante, che risale a Chambers (1988), e che riguarda la differenza tra economie di scala e rendimenti di scala. Per definizione, quest'ultimo concetto misura la variazione dell'output in rapporto ad una variazione equiproporzionale di tutti gli input (espansione radiale).

Nella Figura 3.4, partiamo dal punto A, nel quale l'impresa, dati i prezzi degli input, minimizza i costi di produzione per il livello di output che desidera produrre. Immaginiamo ora di espandere radialmente (per un fattore scalare) gli input, fino ad incontrare un nuovo isoquanto più esterno nel punto C. Il costo associato al punto C non necessariamente corrisponde al valore minimo possibile che, nel nostro esempio, si trova nel punto B. Modificando l'output, infatti, la condizione per produrre ai costi minimi potrebbe imporre una modifica nella combinazione degli input, a meno che la tecnologia non risulti omotetica.⁷⁶

Figura 3.4 - Elasticità di scala e di dimensione.



La precisazione sul concetto e sulla definizione di economie di scala è molto importanti, soprattutto quando si intende fornire una misura empirica del fenomeno, basata sui risultati della stima di una funzione di costo. Nel caso specifico del settore dell'assistenza agli anziani, il problema si riduce essenzialmente al diverso profilo temporale considerato. Sembra di poter dire che la dimensione di un'impresa, attiva in

⁷⁶ In assenza di questa condizione il termine corretto da adottare, secondo Chambers (1988), è quello di economie di dimensione e non di scala.

questo settore, risulta ben rappresentata dalla dimensione fisica della capacità produttiva (numero di letti a disposizione).

Le misure delle economie di scala date dalla [3.12] o dalla [3.13] forniscono un'indicazione diversa a seconda del tipo di funzione di costo scelta per la stima dei parametri. Per una funzione di costo totale di lungo periodo, la sensibilità dei costi rispetto all'output, e quindi la [3.12], indica sicuramente il concetto di economie di dimensione, secondo la definizione di Chambers. Nel caso di un orizzonte di breve periodo, invece, la scelta di utilizzare una funzione di costo variabile fornisce piuttosto una misura delle economie di volume (output), restando lo stock di capitale fisso. In quest'ultimo frangente è tuttavia possibile ricavare, dal parametro che indica le economie di volume, anche il valore delle economie di scala:

$$E_{Scale}^{Cv} = \frac{1 - \frac{\partial \ln Cv(y, w, K)}{\partial \ln K}}{\frac{\partial \ln Cv(y, w, K)}{\partial \ln y}} = \frac{1 - \varepsilon_K}{\varepsilon_y} \quad [3.14]$$

L'espressione [3.14], proposta da Caves et al. (1984), permette quindi di risalire alle economie di scala, partendo dalla stima di una funzione di costo variabile (nella quale lo stock di capitale è considerato fisso).

In conclusione, è bene ribadire che, da un punto di vista economico, la verifica e la misurazione della sensibilità dei costi al variare dell'output o della dimensione è di fondamentale importanza nei lavori empirici. La terminologia e l'opportunità di utilizzare alcune definizioni piuttosto che altre varia a dipendenza delle caratteristiche produttive del settore analizzato. Queste distinzioni e precisazioni sono molto importanti nella produzione di servizi a rete, meno per il settore delle case per anziani. L'importanza di precisare l'orizzonte temporale considerato nell'analisi empirica di questo settore impone, tuttavia, una chiara distinzione tra i due concetti esposti in precedenza: le economie di volume e le economie di scala.

3.2 L'(in)efficienza produttiva

Basandoci sugli elementi di teoria della produzione appena discussi, presenteremo in questa sezione la definizione ed i concetti di efficienza che si sono sviluppati in letteratura a partire dagli anni '50. Il paragrafo non vuole proporre una rassegna

completa di un filone di ricerca molto ricco ed articolato, quanto piuttosto una “guida rapida” per comprendere meglio la parte empirica.

Koopmans (1951), grande pioniera nella definizione e misurazione dell'efficienza, definiva inefficiente qualunque combinazione produttiva che, a parità di output, preveda un uso eccessivo di un qualunque input. In questa tesi consideriamo, invece, un concetto di efficienza radiale, sviluppato originariamente da Debreu (1951) e da Farrell (1957), che considera solamente le variazioni equiporzionali di tutti gli input.

L'inefficienza radiale deve parte del suo successo per la sua naturale compatibilità con il concetto di funzione di distanza, strumento proposto da Shephard (1953) per caratterizzare la tecnologia di processi produttivi multi-output sia nello spazio delle quantità (input-output) sia nello spazio monetario (costi-ricavi). Per definire la funzione di distanza, e quindi l'inefficienza, occorre specificare in quale direzione (ortogonale) stiamo orientando la misura. È infatti possibile esprimere una distanza nel senso degli output (*Output Oriented*) oppure nel senso degli input (*Input Oriented*). Misurare l'efficienza nel senso degli output significa stabilire il livello massimo di prodotto che una certa combinazione di input è in grado di generare, data la tecnologia. Orientare la misura nel senso degli input implica, invece, una definizione di efficienza come la quantità minima di fattori produttivi da impiegare per produrre un determinato livello di output, data la tecnologia.

Quest'ultima visione (*Input Oriented*) si presta meglio a descrivere gli sforzi produttivi delle case per anziani che, dato il cronico eccesso di domanda presente nel settore, devono erogare una quantità fissa di output, utilizzando meno input possibili. Di conseguenza in questa tesi adotteremo un concetto di efficienza produttiva che mira a minimizzare l'impiego di risorse per ottenere un livello di output prefissato, concentrando la nostra attenzione al lato degli input e dei costi di produzione che ne derivano (funzione di costo).

Il termine efficiente sarà quindi riferito a quelle unità produttive che soddisfano le definizioni date per la funzione di costo totale di breve [3.9] e di lungo [3.5] periodo. Per sua stessa natura, l'efficienza definisce un limite raggiungibile, ma invalicabile, che si può interpretare con il concetto di frontiera di produzione o di costo a dipendenza che si analizzi il problema nella sua versione primale o duale. La teoria economica della produzione e la misurazione delle performance di singole unità produttive danno luogo

a quattro concetti di efficienza che presenteremo separatamente: efficienza tecnica, efficienza di costo, efficienza allocativa ed efficienza di scala.

3.2.1 L'(in)efficienza tecnica

Il concetto di efficienza tecnica richiama alla mente la capacità di sfruttare al massimo consentito dalla tecnica le risorse a disposizione. In termini più formali, possiamo definire tecnicamente efficiente una combinazione produttiva nel seguente modo: $(x,y) \in T$ è tecnicamente efficiente se e solo se $(x',y') \notin T$ per $(-x',y') \geq (-x,y)$. Se ci riferiamo all'efficienza nel senso degli input possiamo, invece, dire che: il vettore di input x è tecnicamente efficiente se e solo se $x \in L(y)$.

In termini formali, seguendo la definizione data da Farrell (1957), possiamo esprimere una misura radiale dell'efficienza tecnica nel modo seguente:

$$ET^{Input}(x, y) = \min \{ \theta : \theta x \in L(y) \} \quad [3.15]$$

Questa definizione, che si basa sul concetto di funzione di distanza, è stata largamente impiegata in campo econometrico.⁷⁷ La [3.15] esibisce alcune proprietà desiderabili:

1. $ET^{Input}(x,y) \leq 1$;
2. $ET^{Input}(x,y) = 1 \Leftrightarrow x \in \text{Isoq } L(y)$;
3. $ET^{Input}(x,y)$ è non crescente in x ;
4. $ET^{Input}(x,y)$ è omogenea di grado -1 in x ;
5. $ET^{Input}(x,y)$ è invariante alle unità di misura di x e di y

Le prime due proprietà esprimono rispettivamente la normalizzazione scelta per indicare l'inefficienza e la scelta dell'isoquanto come frontiera di efficienza tecnica. La terza e la quarta proprietà si riferiscono a come la misura dell'efficienza reagisce ad un eventuale spreco di input. Se si utilizzano uno o più input per produrre la medesima quantità di output (proprietà 3), l'efficienza tecnica non può migliorare. La reazione di $ET^{Input}(x,y)$ allo spreco di input nelle medesime proporzioni, produce una modifica

⁷⁷ Per una presentazione dettagliata sul legame tra efficienza e funzioni di distanza si veda ad esempio Kumbhakar and Lovell (2000).

scalare, ma in senso opposto, alla misura di efficienza. In altre parole la proprietà 4 afferma che se, ad esempio, tutti gli input raddoppiano per produrre lo stesso output, l'efficienza tecnica viene dimezzata. La proprietà 5 rende la definizione dell'efficienza tecnica indipendente dalle grandezze utilizzate per misurare gli input e/o gli output e rende particolarmente interessanti i confronti tra unità e/o tecnologie produttive diverse.

L'efficienza tecnica esprime dunque una relazione quantitativa tra input ed output, che trova un immediato riscontro con la misura dell'(in)efficienza che ne deriva, riferita chiaramente all'aspetto primale della tecnologia. Per introdurre un concetto analogo, nella sua versione duale, occorre definire l'(in)efficienza di costo.

3.2.2 L'(in)efficienza di costo

L'attività produttiva di un'impresa implica l'impiego di risorse (input) e quindi determina un costo associato alla produzione dell'output. L'entità del costo di produzione dipende dalla combinazione dei fattori produttivi, dall'efficienza tecnica nel loro impiego e dal livello dei prezzi dei fattori stessi. Se la misura dell'efficienza di costo esprime un confronto tra la spesa effettiva e quella minima possibile per produrre lo stesso livello di output, allora è necessario disporre di un termine di paragone per valutare i costi di produzione.

Questo modo di ragionare è la conseguenza di una precisa ipotesi di comportamento attribuita alle imprese: la minimizzazione dei costi. La risposta al problema di produrre un determinato output al costo minimo è racchiusa dalla funzione di costo [3.5] che, di conseguenza, viene utilizzata come strumento di misura per rilevare eventuali eccessi di costo. L'indicatore di Efficienza di Costo (EC) è definito come:

$$EC(x, y, w) = C(y, w) / w^T x \quad [3.16]$$

Questa definizione esprime in modo chiaro l'efficienza di costo di una data combinazione produttiva e la lega alla funzione di costo che, fungendo da paragone per la piena efficienza, definisce la frontiera della tecnologia. La [3.16] è data dal rapporto tra il costo minimo e il costo osservato (per produrre un certo output), e misura la contrazione che si dovrebbe applicare al livello di costo osservato per poter ottenere il livello di costo minimo (efficiente).

Le proprietà espresse dalla misura di efficienza appena proposta sono:

1. $0 < EC(x, y, w) \leq 1$;
2. $EC(x, y, w) = 1 \Leftrightarrow x = x(y, w)$ e $w^T x = C(y, w)$;
3. $EC(x, y, w)$ è non decrescente in y ;
4. $EC(x, y, w)$ è omogenea di grado -1 in x ;
5. $EC(x, y, w)$ è omogenea di grado 0 in w .

La proprietà 1 indica lo spettro dei valori ammissibili secondo la definizione di efficienza data: il livello di efficienza raggiunto parte da 0 (escluso) ed è tanto più elevato quanto più ci si avvicina ad 1. La proprietà 2 stabilisce infatti che il limite superiore appartiene ai valori ammissibili e che l'efficienza di costo è pari ad 1 se e solo se il processo produttivo considerato minimizza i costi. La proprietà 3 indica che, se un processo produttivo è in grado di produrre un output maggiore (a parità di input e di prezzo dei fattori), l'efficienza di costo non deve diminuire.⁷⁸ La proprietà 4 rispecchia quanto detto per l'efficienza tecnica e cioè che se tutti gli input si modificano nella stessa proporzione, anche l'efficienza di costo subirà una variazione scalare inversa (se gli input raddoppiano, l'efficienza di costo dimezza). La proprietà 5 afferma che solo la variazione dei prezzi relativi dei fattori esercita un'influenza sull'efficienza di costo. Se tutti gli input modificano il loro prezzo nella medesima proporzione, non cambieranno le scelte dei fattori e quindi, nonostante la frontiera di costo si muova, l'effetto monetario dei prezzi si neutralizza nella misurazione dell'efficienza di costo.

L'efficienza di costo è idealmente il risultato di un successivo processo di ottimizzazione nel quale l'efficienza tecnica costituisce solamente il primo passo. Infatti non è sufficiente produrre un dato livello di output impiegando il minimo di risorse fisiche, poiché la combinazione di input scelta potrebbe, dati i prezzi dei fattori, generare un eccesso di costo.

⁷⁸ In questa proprietà si vede come il legame tra le diverse orientazioni della misura di efficienza non sia necessariamente lineare. L'aumento di output a parità di input è una misura del miglioramento dell'efficienza nel senso dell'output. Tuttavia, l'aumento dell'output (e quindi dell'efficienza in tal senso) non si deve necessariamente tradurre in un aumento proporzionale dell'efficienza di costo (che corrisponde ad un orientamento nel senso degli input), ma è sufficiente che questa non diminuisca.

Al contrario se un processo produttivo raggiunge l'efficienza di costo, questo implica necessariamente anche l'efficienza tecnica. Per poter distinguere tra uno spreco di risorse causato da una scelta non ottimale in senso tecnico o in senso economico, occorre specificare un nuovo concetto di efficienza: l'(in)efficienza allocativa.

3.2.3 L'(in)efficienza allocativa

Sulla scorta di quanto presentato, possiamo rappresentare l'efficienza di costo come il risultato di due componenti, l'efficienza tecnica e quella allocativa. Per questa ragione deriviamo la misura dell'efficienza allocativa, utilizzando la [3.15] e la [3.16].

La definizione di Efficienza Allocativa (EA) conseguente è espressa da:

$$EA = EC(x, y, w) / ET^{Input}(x, y) = C(y, w) / w^T \theta x \quad [3.17]$$

Anche l'efficienza allocativa presenta alcune proprietà che non presenteremo formalmente, poiché discendono per costruzione dalle proprietà già discusse per l'efficienza tecnica e di costo.⁷⁹ In estrema sintesi l'Efficienza Allocativa è misurata tra 0 e 1 ed è insensibile a variazioni equiproportionali degli input o dei loro prezzi.

Se riformuliamo la [3.17] mettendo in evidenza l'efficienza di costo, otteniamo un prodotto di 2 fattori: l'inefficienza tecnica e quella allocativa. Per un processo produttivo questo significa che l'inefficienza di costo (o complessiva) potrebbe dipendere: solo dall'inefficienza tecnica, solo dall'inefficienza allocativa o da entrambe. Per chiarire meglio queste differenze ricorriamo ad un esempio grafico (Figura 3.5) riferito ad una tecnologia produttiva che trasforma 2 input (x_1, x_2) in 1 solo output (y).

Nel pannello di sinistra è rappresentata una combinazione produttiva x^A che, dati i prezzi di mercato w^A , è in grado di produrre la quantità y^A di output. Da un punto di vista tecnico si registra un'inefficienza dal momento che la quantità y^A si può produrre utilizzando (nelle stesse proporzioni) minori quantità di input. La misura dell'(in)efficienza tecnica della combinazione x^A è data dalla [3.15] ed è pari a θ , ossia la contrazione radiale necessaria per riportare il punto considerato sull'isoquante $L(y^A)$.

La combinazione è tecnicamente efficiente ($\theta^A x^A$) ma non è in grado, tuttavia, di minimizzare i costi, poiché utilizza troppo intensamente il fattore di produzione x_1 il cui

⁷⁹ Per la formalizzazione delle proprietà della EA si rimanda a Kumbhakar e Lovell (2000).

La scomposizione dell'inefficienza di costo nelle componenti tecnica ed allocativa è molto importante da un punto di vista teorico ma richiede grande attenzione anche in sede di verifica empirica. La stima di una frontiera di costo mira infatti a determinare da un lato i parametri strutturali della tecnologia produttiva (sfruttando la teoria del duale) e, dall'altro, a misurare l'efficienza di costo delle singole imprese.

Queste informazioni possono essere estratte dai dati solo se il comportamento delle imprese persegue la minimizzazione dei costi che garantisce il rispetto teorico di tutte le proprietà descritte in precedenza.

Solo in questo caso è corretto utilizzare la frontiera di costo (teorica ed empirica) per definire i parametri della tecnologia sottostante e per misurare le performance economiche delle singole imprese. In assenza di un reale obiettivo di minimizzazione dei costi è comprensibile che alcune proprietà richieste (es. la concavità nel vettore dei prezzi degli input) non siano verificate empiricamente.

Inoltre, in questi casi, la stima della frontiera esprime una funzione di costo definita "comportamentale" (*behavioural cost function*) che approssima più o meno bene quella neoclassica. Questo aspetto è molto importante per le analisi empiriche e verrà ripreso durante la discussione del modello di costo scelto per rappresentare la tecnologia produttiva delle case per anziani in Svizzera.

3.2.4 L'(in)efficienza di scala

L'efficienza di scala è l'ultimo concetto che desideriamo definire per completare la descrizione degli strumenti adatti a descrivere, anche empiricamente, la tecnologia e le performance economiche delle imprese.

Abbiamo visto in precedenza che il legame tra dimensione ottimale, economie di scala e rendimenti di scala è molto stretto anche se i concetti sono differenziati. La minimizzazione dei costi (al variare dell'output) può richiedere una modifica del mix dei fattori produttivi per garantire la produzione al costo minimo. Il concetto di rendimenti di scala (riferito alla relazione tecnica tra input e output) e quello di economie di scala (riferito alla relazione economica tra output e costi) coincidono solo in presenza di una tecnologia omotetiche.⁸⁰ Nel caso di un singolo output l'aspetto

⁸⁰ Come discusso in precedenza (cfr. 3.1.4), quando la tecnologia non è omotetica si può introdurre (Chambers, 1988) la distinzione tra economie di dimensione ed economie di scala.

monetario (duale) dei rendimenti di scala si può cogliere osservando la sensibilità dei costi al variare dell'output, come già discusso in precedenza. Pertanto, utilizzando l'espressione [3.12], possiamo descrivere le economie di scala nel seguente modo:

$$E_{Scala} = \frac{1}{\varepsilon_y(y, w)} > 1 \quad \text{Economie di scala crescenti}$$

$$E_{Scala} = \frac{1}{\varepsilon_y(y, w)} = 1 \quad \text{Economie di scala costanti}$$

$$E_{Scala} = \frac{1}{\varepsilon_y(y, w)} < 1 \quad \text{Economie di scala decrescenti}$$

Di principio la teoria economica non pone alcuna restrizione ideologica sulle caratteristiche tecnologiche, considerando la possibilità che i rendimenti e le economie di scala varino a dipendenza dell'output. Il buon senso tuttavia suggerisce, in molti casi, un andamento dei rendimenti di scala dapprima crescente e poi decrescente, con la conseguenza che le economie di scala sono crescenti in una fase iniziale per poi decrescere.

Per misurare l'inefficienza di scala, tenendo conto dei due aspetti (fisico e monetario) che la determinano e del loro legame, occorre fissare un riferimento. Nel caso di un solo output, riferiremo l'inefficienza di scala alla regione dove i rendimenti di scala sono costanti, garantendo una corrispondenza tra output prodotto e minimizzazione dei costi unitari (concetto di dimensione ottimale) dell'impresa.⁸¹ Per spiegare come si costruisce la misura dell'inefficienza di scala, esemplifichiamo (Figura 3.6) un processo produttivo (1 input e 1 output) T^* che esprime rendimenti di scala costanti e uno T con rendimenti di scala variabili.⁸² Il motivo di confrontare due tecnologie diverse è quello di costruire una misura dell'efficienza di scala di una generica tecnologia T che ipotizziamo a rendimenti di scala variabili per contemplare tutte le possibilità.

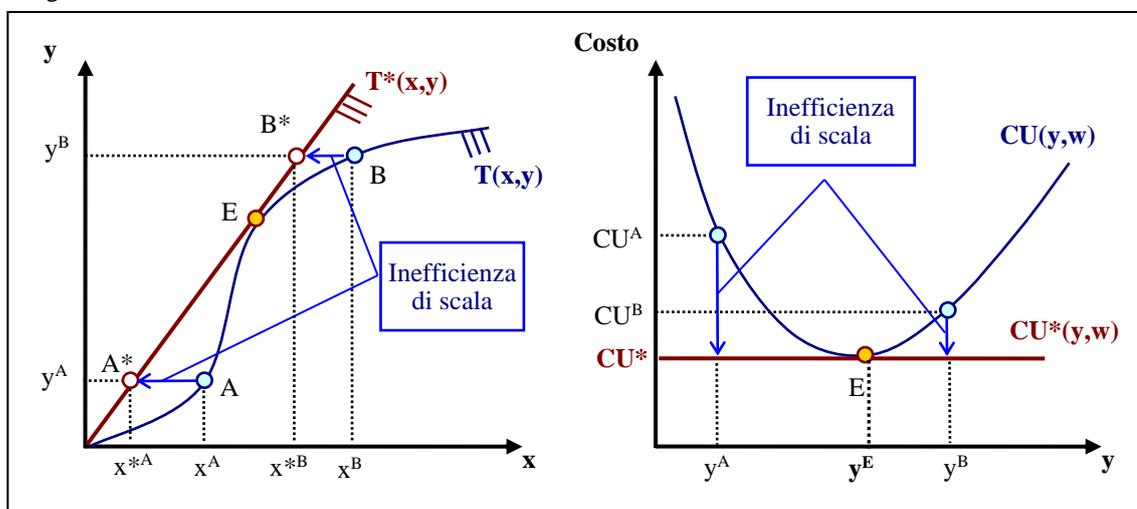
Se consideriamo la tecnologia T , possiamo giudicare efficienti dal punto di vista tecnico ed allocativo le combinazioni A e B . Lo stesso discorso vale per le

⁸¹ Per ulteriori approfondimenti sul tema si veda Färe et al. (1988).

⁸² La Tecnologia T^* (rendimenti di scala costanti) è scelta in modo tale da poter produrre tutte le combinazioni disponibili con la tecnologia T (rendimenti di scala variabili).

combinazioni A^* e B^* con riferimento alla tecnologia T^* che esprime rendimenti di scala costanti.

Figura 3.6 - L'inefficienza di scala.



Se, tuttavia, valutiamo l'efficienza tecnica ed economica dei punti A e B rispetto alla tecnologia T^* , notiamo la presenza di una forma di inefficienza legata alla scala produttiva dell'impresa. Se stiamo valutando il processo produttivo descritto dalla tecnologia T, questo apparente spreco di risorse non si può attribuire all'efficienza tecnica né a quella allocativa dato che T non può produrre A^* e B^* . Questa inefficienza è dovuta alla scala di produzione di T che risulta troppo piccola per produrre y^A (in presenza di rendimenti crescenti di scala) e troppo grande per produrre y^B (in presenza di rendimenti decrescenti di scala). Nel punto E la tecnologia T esprime rendimenti di scala costanti che sono presi come riferimento, utilizzando l'ipotetica tecnologia T^* , per valutare l'efficienza di scala di tutte le altre combinazioni produttive.

Anche dal punto di vista dei costi (duale alla produzione) possiamo ripetere il ragionamento e vedere come gli output y^A e y^B , nonostante minimizzino i rispettivi costi totali di produzione, corrispondano a costi unitari più elevati rispetto al punto E dove le economie di scala sono completamente sfruttate. Si potrebbe obiettare che l'inefficienza di scala è misurata con l'impiego di una tecnologia T^* che, essendo formulata teoricamente, potrebbe non trovare riscontro nella realtà. In verità T^* non è affatto arbitraria ed è costruita partendo da quella combinazione di T che è in grado di produrre l'output y^E al costo unitario minore possibile.

Questa osservazione è molto importante, poiché suggerisce la possibilità di produrre un qualsiasi livello di output y come multiplo o frazione del valore y^E in grado

di minimizzare i costi unitari. In termini concreti significa che in un mercato il cui output totale da produrre è pari a Y^A , si raggiunge l'efficienza di scala quando ogni processo produttivo ha la dimensione y^E da cui si ricava il numero di processi produttivi che è pari a $N^E = Y^A / y^E$.

Le caratteristiche del mercato concorrenziale garantiscono un aggiustamento spontaneo del numero di processi produttivi verso il valore N^E di equilibrio dove scompare l'inefficienza di scala. Nella realtà, ed in particolare nel settore delle case per anziani, non è detto che la scala di produzione sia quella ottimale dal punto di vista della società, né che tutti i processi produttivi abbiano uguale dimensione.⁸³

In questo capitolo abbiamo passato in rassegna alcuni elementi della teoria microeconomica della produzione, in modo da presentare con chiarezza ciò che dovremmo attenderci in un mondo perfetto senza sprechi. Sapendo che nella realtà difficilmente si realizza la piena efficienza produttiva, abbiamo discusso gli strumenti teorici per poter definire (e quindi misurare) le diverse forme di inefficienza produttiva.

Giunti a questo punto abbiamo lo strumentario teorico che occorre per affrontare l'analisi empirica, mentre non abbiamo discusso le modalità operative per poter misurare, partendo dai dati a nostra disposizione, l'inefficienza produttiva nel settore delle case per anziani in Svizzera. Nel prossimo capitolo affronteremo, quindi, la questione metodologica, discutendo in dettaglio il metodo econometrico scelto per questa ricerca: la frontiera di costo stocastica.

⁸³ Considerando esogeno l'output totale Y da produrre con la tecnologia T , la quantità y^i che ogni processo produttivo è chiamato a fornire dipende dal numero di processi produttivi N e dalla dimensione di ogni singolo processo produttivo. per tutti gli output y^j che non corrispondono alla quantità ottimale y^E si verifica un'inefficienza di scala.

*“La chance ne sourit qu'aux
esprits bien préparés.”*

(Louis Pasteur, Biologo)

4 La stima parametrica di una frontiera di costo stocastica: aspetti metodologici

Questo capitolo è dedicato all'esposizione dei problemi metodologici che si devono affrontare quando si vuole compiere un'indagine econometrica sull'efficienza produttiva delle case per anziani. L'obiettivo di una ricerca empirica è ovviamente quello di ottenere risultati credibili da un punto di vista economico (coerenti con la teoria e con il buon senso) ed affidabili sotto il profilo statistico.

Un'analisi dell'efficienza si può svolgere utilizzando numerosi strumenti metodologici, ognuno dei quali presenta caratteristiche peculiari che permettono di cogliere aspetti diversi di uno stesso problema. Nell'ambito degli studi empirici sull'efficienza produttiva in letteratura emergono due grandi famiglie metodologiche: i metodi non parametrici (es. Data Envelopment Analysis e Free Disposal Hull) ed i metodi parametrici (es. Stochastic Frontier Analysis). In questa tesi abbiamo deciso di utilizzare il metodo parametrico della frontiera di costo stocastica (SFA) quale strumento per analizzare la struttura dei costi e l'efficienza produttiva del settore delle case per anziani.

L'obiettivo di questo capitolo è in primo luogo quello di chiarire gli elementi che costituiscono la stima econometrica di una frontiera di costo stocastica e l'importanza che questi rivestono nel determinare i risultati empirici. Le tecniche di stima impiegate in questa tesi meritano un approfondimento per chiarire a quale filone econometrico

appartengono e, soprattutto, che vantaggi offrono rispetto ad altri approcci, in rapporto ai problemi che presentano. Affrontare questi argomenti, in particolare i vantaggi e gli svantaggi collegati con le diverse tecniche, consente di discutere consapevolmente i risultati economici che scaturiscono dall'analisi empirica, permettendo di ragionare in modo equilibrato sulle eventuali implicazioni di politica economica che discendono dalla ricerca.

Prima di analizzare in dettaglio le caratteristiche econometriche della SFA, desideriamo illustrare i tratti essenziali dei metodi non parametrici e confrontare criticamente i vantaggi comparativi nell'utilizzo dei metodi parametrici. Nell'ultima parte del capitolo, invece, ci sarà spazio per una rassegna mirata della letteratura econometrica del settore, con lo scopo di tracciare alcuni dei recenti sviluppi della metodologia delle frontiere stocastiche.

4.1 Metodi non parametrici e parametrici a confronto: DEA vs. SFA

Questo paragrafo si prefigge di presentare in modo estremamente semplificato le caratteristiche di due metodi molto impiegati dalla letteratura empirica sull'efficienza produttiva. Si tratta di confrontare la *Data Envelopment Analysis* (DEA) e la *Stochastic Frontier Analysis* (SFA), in rappresentanza delle rispettive famiglie di metodi non parametrici (DEA) e parametrici (SFA). Nonostante alcune importanti differenze di impostazione, entrambi i metodi seguono un approccio quantitativo per analizzare dati, con lo scopo di ricavarne informazioni numeriche riguardanti le performance economiche e/o le caratteristiche della tecnologia produttiva.

4.1.1 Metodi non parametrici: Data Envelopment Analysis (DEA)

La misura dell'efficienza si può ottenere dall'analisi di osservazioni empiriche utilizzando le tecniche non parametriche, lo strumento più conosciuto delle quali si chiama DEA, ossia l'analisi dell'involuppo dei dati.⁸⁴

Le tecniche non parametriche devono il proprio nome al fatto di non specificare la relazione che lega le variabili tra loro con una funzione parametrica. Input e output

⁸⁴ Per un'esposizione dei concetti teorici sulla DEA e la loro applicazione pratica con l'uso di software si consiglia la lettura di Cooper et al. (2007) e Zhu (2003). Per una lettura più avanzata si veda Daraio and Simar (2008).

osservati devono, però, rispettare alcuni vincoli e risolvere il problema di programmazione lineare che è alla base di queste metodologie.

La DEA si pone come un'applicazione empirica del concetto di funzione di distanza, che richiama le considerazioni fatte in precedenza sulla misurazione dell'efficienza di un processo produttivo. Ogni unità decisionale, che nel caso specifico può essere un'impresa che produce un vettore di output con un vettore di input, è confrontata con tutte le altre osservazioni sulla base di un problema di massimizzazione vincolata. La soluzione di questo problema determina l'espansione radiale dell'output (o la contrazione radiale dell'input) che porterebbe l'unità decisionale osservata sulla frontiera di produzione (o di costo).

Anche se non desideriamo trattare gli aspetti formali di questa tecnica, rimandando il lettore interessato alla produzione scientifica sull'argomento, occorre sottolineare come i punteggi di inefficienza di ogni unità produttiva siano il risultato di uno specifico problema di programmazione lineare, impostato dal ricercatore sulla base di alcune ipotesi di lavoro quali: l'orientamento dell'efficienza nel senso degli input o degli output, l'ipotesi sui rendimenti di scala, la presenza di alcuni vincoli, etc. Le stime dei valori di efficienza ottenuti con la tecnica DEA si prestano quindi ad effettuare un confronto quantitativo tra i processi produttivi osservati. Ogni performance è misurata e valutata rispetto a una tecnologia sconosciuta, ma linearmente derivata da processi produttivi esistenti che costituiscono i punti di riferimento (*peers*) per l'osservazione considerata.

Utilizzando i dati delle unità produttive secondo questa metodologia, è possibile effettuare un confronto tra le performance produttive di tutte le imprese presenti nel campione, definendo un sistema di *benchmarking* in un preciso momento temporale. Se i dati a disposizione hanno una struttura temporale (panel data) è possibile utilizzare la DEA per misurare, oltre che le performance delle singole imprese, il progresso tecnico.⁸⁵ La DEA è quindi una metodologia che si presta ad elaborare confronti tra

⁸⁵ Le stime delle funzioni di distanza ottenute con la DEA permettono di definire l'indice di Malmquist che è uno dei numerosi strumenti sviluppati dalla letteratura per condurre analisi sulla *Total Factor Productivity* (TFP) e sul progresso tecnico utilizzando i metodi non parametrici. Per un approfondimento sul tema si veda ad esempio Färe et al (1994).

unità produttive in modo abbastanza immediato e senza dover ricorrere a precise ipotesi circa la tecnologia del processo produttivo che si sta analizzando.

4.1.2 I metodi parametrici e la Stochastic Frontier Analysis (SFA)

I metodi parametrici si propongono di verificare empiricamente le caratteristiche della tecnologia produttiva attraverso l'analisi dei dati delle unità produttive con l'uso di tecniche econometriche. Il risultato di questo sforzo consiste nella stima dei parametri della funzione (di produzione o di costo) che descrive la tecnologia analizzata. Nelle tecniche non parametriche la relazione tra input e output è determinata da considerazioni di carattere economico-tecnico, che sfociano nella definizione di un modello di produzione o di costo. Queste relazioni, che discendono dall'analisi delle caratteristiche produttive del settore in esame, sono descritte dai parametri della funzione scelta come modello.

Lo scopo di questi metodi è quello di stabilire i parametri che caratterizzano la tecnologia in modo da poter analizzare le caratteristiche economiche del processo produttivo e, soprattutto, l'effetto quantitativo di ogni possibile cambiamento nelle variabili produttive. Con riferimento alla SFA, l'obiettivo è quello di stimare i parametri che descrivono la tecnologia efficiente nel senso discusso in precedenza, conciliando la natura deterministica della teoria economica con la natura stocastica delle osservazioni empiriche.

Gli aspetti formali e le procedure di stima utilizzate per questa tecnica saranno trattati in modo esaustivo nel prossimo capitolo, mentre di seguito desideriamo evidenziare come l'analisi parametrica della tecnologia e dell'efficienza sia una metodologia, solidamente strutturata da un punto di vista della teoria economica, che permette di misurare i parametri di una funzione di costo, considerando la natura stocastica del processo produttivo.

Anche per i metodi parametrici, qualora i dati presentino una struttura di tipo panel, è possibile costruire degli indici di progresso tecnico e misurare la produttività totale dei fattori o TFP (*Total Factor Productivity*).⁸⁶

⁸⁶ Per una discussione si veda ad esempio Coelli et al. (2003).

4.1.3 DEA vs. SFA: vantaggi e svantaggi nel misurare l'efficienza produttiva

In modo molto semplificato abbiamo descritto i due metodi principali per analizzare empiricamente l'efficienza produttiva. Ogni metodo presenta vantaggi e svantaggi, che devono essere ponderati anche sulla base degli obiettivi dell'analisi, della quantità e della qualità dei dati a disposizione.⁸⁷

Nell'ambito della ricerca economica applicata spesso si decide di utilizzare una sola metodologia. Questa constatazione, che trova molte buone motivazioni di ordine pratico, non implica tuttavia che gli strumenti descritti siano in competizione tra loro. Al contrario, è possibile utilizzare i metodi in modo complementare per verificare anche dal punto di vista dei risultati ottenuti con i diversi approcci gli effettivi vantaggi e svantaggi nel misurare le performance economiche delle imprese.⁸⁸

Il dibattito su quale famiglia (parametrica o non parametrica) di metodi sia preferibile è tuttora in corso e, poiché la letteratura progredisce continuamente, la discussione verte su nuovi argomenti. Per avere una visione aggiornata dello stato dell'arte dei principali filoni metodologici si suggerisce la lettura dell'opera curata da Fried et al. (2008). A distanza di 15 anni dall'uscita del loro libro, soprannominato "Yellow Book", gli autori discutono i risultati consolidati nel tempo nei diversi ambiti, presentano i più recenti sviluppi metodologici e tracciano le linee di sviluppo per risolvere i problemi ancora aperti.⁸⁹

Nella Tabella 4.1 presentiamo in modo schematico i principali vantaggi e svantaggi connessi con l'utilizzo dei due metodi in discussione. Dal paragone tra le due metodologie considerate emerge chiaramente come ogni approccio presenti punti di forza e di debolezza, rendendo impossibile una scelta aprioristica. Per questo motivo la produzione scientifica recente non cerca solo di migliorare le singole tecniche nei rispettivi punti di forza e/o di mitigare i punti di debolezza, ma anche di sviluppare approcci misti e di valutare comparativamente i risultati dei diversi metodi.⁹⁰

⁸⁷ Per un confronto sui rispettivi vantaggi e svantaggi delle due metodologie si veda Jacobs et al. (2006).

⁸⁸ Si veda ad esempio lo studio di Carrington et al. (2002) che applica diversi metodi al medesimo dataset di università australiane.

⁸⁹ Il soprannome "libro giallo" è dovuto al colore della copertina. Il titolo esatto del volume di Fried et al. (1993) è: *"The measurement of productivity efficiency. Technique and applications"*.

⁹⁰ Si veda ad esempio il lavoro di Sickles (2005).

Tabella 4.1 - Confronto tra i principali metodi per la misura delle performance.

	Data Envelopment Analysis (DEA)	Stochastic Frontier Analysis (SFA)
Vantaggi	<ul style="list-style-type: none"> ▪ Si presta ad analizzare i processi multiprodotto ▪ Fornisce l'indicazione dei <i>peers</i> (imprese efficienti con input-output mix simili) ▪ Non assume a priori una specifica forma funzionale per la frontiera e per la distribuzione del termine di inefficienza 	<ul style="list-style-type: none"> ▪ Tiene conto del disturbo stocastico e identifica gli outliers ▪ Permette l'uso di variabili ambientali in modo semplice ▪ Possibilità di testare i parametri e di formulare previsioni
Svantaggi	<ul style="list-style-type: none"> ▪ Sensibile al disturbo statistico dei dati ▪ Richiede dataset ampi per garantire la stabilità dei risultati ▪ Non permette l'uso di test di ipotesi (parzialmente risolvibile con le tecniche bootstrapping) 	<ul style="list-style-type: none"> ▪ I risultati sono sensibili alla specificazione della forma funzionale e del termine di errore ▪ Richiede dataset con molte osservazioni ▪ Problemi nella separazione dell'inefficienza nelle sue componenti

Rispetto agli scopi della nostra analisi abbiamo preferito indirizzarci verso i metodi parametrici, ed in modo particolare verso le frontiere stocastiche, per tre ordini di ragioni: (1) interesse metodologico verso le recenti tecniche panel (Greene, 2005) nell'ambito della SFA, (2) importanza attribuita alla misurazione delle economie di scala e (3) caratteristiche e qualità dei dati a disposizione.

Trattandosi di una scelta discrezionale, non commenteremo oltre il primo motivo, mentre il problema della misurazione delle economie di scala merita una breve spiegazione. Con la tecnica DEA è possibile individuare l'inefficienza di scala confrontando i risultati del modello a rendimenti di scala variabili (RSV) con quelli della specificazione a rendimenti di scala costanti (RSC). Questo modo di procedere, oltre ad essere molto sensibile alla presenza di osservazioni estreme, si rivela meno accurato nel misurare il valore dell'elasticità di costo rispetto all'output. Visto l'interesse empirico a determinare l'importanza delle economie di dimensione per il settore delle case per anziani, preferiamo affidarci ai metodi parametrici che sono in grado di fornire indicazioni più precise su questo aspetto.

La motivazione legata ai dati riguarda principalmente la struttura e la qualità delle informazioni contenute nella banca dati. Trattandosi di un rilevamento ad ampio spettro sulla realtà degli istituti non ospedalieri, nei dati è presente una forte eterogeneità. Se aggiungiamo le difficoltà del rilevamento statistico a considerare tutte le caratteristiche

del settore dell'assistenza agli anziani (problema dell'eterogeneità non osservata), i dati conterrebbero verosimilmente un elevato "rumore stocastico di fondo" (*white noise*). La metodologia SFA, che rispetto alla DEA permette di accomodare direttamente nel modello di costo anche le variabili *dummies* ed è in grado di filtrare parte del disturbo presente nei dati, sembra quindi più adatta per il nostro caso.⁹¹

Dopo aver presentato i motivi che ci hanno spinto a prediligere l'approccio parametrico e in particolare la frontiera di costo stocastica, è tempo di discutere a fondo i problemi e le caratteristiche associati a questa metodologia.

4.2 Elementi per la stima di una frontiera di costo stocastica

Il primo punto che desideriamo affrontare è quello di chiarire come mai, per analizzare l'efficienza produttiva del settore delle case per anziani, abbiamo scelto di analizzare una funzione di costo, invece che una funzione di profitto, di ricavo o di produzione.⁹² Date le caratteristiche del settore delle case per anziani in Svizzera, possiamo escludere le frontiere di profitto e di ricavo, che si adattano meglio all'analisi di mercati maggiormente competitivi (frontiera di profitto) e con processi produttivi multiple-output (frontiera di ricavo). Inoltre, l'obiettivo della ricerca è quello di descrivere e parametrizzare le caratteristiche tecnologiche del settore dell'assistenza agli anziani, concentrandoci maggiormente sugli aspetti produttivi. La scelta si riduce quindi a due opzioni: la frontiera di produzione e la frontiera di costo. I dati in nostro possesso avrebbero consentito di utilizzare entrambe le specificazioni, ma, per le ragioni che a seguire presenteremo, abbiamo deciso di stimare una frontiera di costo.

Il primo motivo riguarda l'impossibilità delle frontiere di produzione di accomodare una specificazione multiprodotto. La variabile dipendente è infatti l'unico output che deve essere spiegato dalle quantità dei fattori produttivi strutturali ed eventualmente da alcune caratteristiche "ambientali" (fattori produttivi *sui generis*). Questo problema non si pone nella frontiera di costo che, potendo includere tra i

⁹¹ Alcune recenti sviluppi delle tecniche non parametriche permettono di considerare e filtrare il disturbo stocastico presente nei dati. Per un approfondimento si veda Daraio e Simar (2008).

⁹² In letteratura si trovano applicazioni per ciascuna delle tipologie di frontiera citate. Per una discussione di queste tipologie di frontiera stocastica si veda Kumbhakar and Lovell (2000).

regressori più di un output, consente di specificare meglio le caratteristiche qualitative del servizio erogato dalle case per anziani.⁹³

Una seconda importante ragione è legata all'interesse della ricerca a determinare l'efficienza di costo (efficienza complessiva) delle case per anziani in Svizzera. La frontiera di produzione, non considerando le grandezze monetarie (i prezzi degli input), permette di determinare solamente l'efficienza tecnica, condizione necessaria ma non sufficiente per l'efficienza complessiva. Inoltre, l'impiego di una frontiera di costo riduce i problemi econometrici (multicollinearità) dovuti alla forte correlazione, che di norma si osserva, tra le quantità dei diversi input.⁹⁴

Stabilito che l'analisi verterà sulla stima di una frontiera di costo, occorre illustrare gli elementi fondamentali che concorrono ad una verifica empirica dell'efficienza basata sulla stima econometrica di una frontiera di costo.

Questi elementi essenziali per svolgere un'analisi di frontiera stocastica sono tre: (1) il modello di costo, (2) la forma funzionale e (3) il modello econometrico utilizzato per la stima di una frontiera di costo. Di seguito presentiamo i punti critici dei primi due elementi, dedicando agli aspetti econometrici il paragrafo successivo.

4.2.1 Modello di costo

In termini molto generali, definire il modello di costo significa rappresentare la realtà produttiva del settore che si osserva attraverso le variabili e le relazioni che il modello di costo prevede. In altre parole, significa dichiarare quali sono gli elementi in grado di influenzare i costi, precisando bene di quali costi si sta parlando e a che tipo di processo produttivo vengono riferiti.

La teoria economica (neoclassica) definisce il costo totale come la spesa minima che è necessario sostenere per produrre, con una data tecnologia, una certa quantità di prodotto. Questa definizione, che non lascia spazio all'inefficienza (di costo), indica che lo sforzo produttivo di un'impresa sarà sempre orientato a combinare gli input secondo il loro miglior uso possibile in relazione al prezzo del fattore stesso. Le differenze nei costi di produzione tra un'impresa e l'altra devono necessariamente riflettere (avendo

⁹³ Seguendo Baumol et al. (1982), anche in un contesto single-output è possibile equiparare le dimensioni qualitative del prodotto a dei veri e propri output (accessori).

⁹⁴ Per una discussione su questi temi si veda ad esempio Jacobs et al. (2006).

escluso la possibilità di comportamenti inefficienti) il prezzo pagato per i fattori e la quantità di output prodotta.

Nella realtà, tuttavia, sappiamo che il bene prodotto da imprese diverse non è mai perfettamente omogeneo, così come le caratteristiche del processo produttivo variano da impresa a impresa. Il costo di produzione dipende anche da queste caratteristiche specifiche dell'impresa. Il modello teorico di costo che nasce da questo ragionamento, pur condividendo l'approccio previsto dalla teoria neoclassica, se ne discosta per considerare anche le particolarità del processo produttivo e le variabili di tipo ambientale.⁹⁵

Richiamando la definizione [3.5], possiamo specificare la generica funzione di costo totale di lungo periodo di un qualunque settore economico come $C(y,w)$. I costi totali C di una generica impresa dipendono dal vettore degli output (y) e dal vettore dei prezzi degli input (w) necessari alla produzione. La definizione di un processo produttivo multiprodotto permette infatti di considerare, in aggiunta alla produzione fisica (output propriamente detto), anche le variabili ambientali e le caratteristiche specifiche dell'output.

Quando si passa da una definizione generale ad una più dettagliata e riferita ad un contesto preciso, occorre prestare attenzione a due aspetti: il primo riguarda il tipo di costo che si intende misurare, mentre il secondo si riferisce alla scelta delle variabili che spiegano le caratteristiche del processo produttivo. I costi di un'impresa si possono infatti scomporre in costi variabili e costi fissi, riflettendo quest'ultima categoria di costi l'impossibilità di modificare alcuni fattori produttivi in breve tempo. Analizzare una funzione dei soli costi variabili, riconoscendo quindi la presenza di alcuni vincoli decisionali, implica una prospettiva di breve periodo nella quale si considerano fissi i costi derivanti da alcuni fattori produttivi che non è possibile modificare anche volendo.

Per semplificare, nell'ottica di breve periodo si considera come data la dimensione dell'impresa; pertanto, lo stock di capitale utilizzato dall'impresa per produrre non può essere modificato. Se, al contrario, desideriamo analizzare l'impresa nel lungo periodo,

⁹⁵ Per una discussione approfondita sui problemi metodologici legati alla specificazione di un modello di costo in ambito ospedaliero si veda Breyer (1987), mentre per una trattazione più generale su come includere le caratteristiche dell'output e le variabili ambientali in una funzione di costo neoclassica si vedano Spady and Friedlaender (1978), Baumol et al. (1982) e Grannemann et al. (1986).

è necessario considerare tutte le voci di costo, comprese quelle che sono fisse nel breve periodo, ma che diventano variabili nel lungo periodo. Il motivo risiede nel fatto che avendo a disposizione un lasso di tempo sufficientemente lungo è possibile modificare anche la dimensione dell'impresa, aggiustandola secondo necessità.⁹⁶

Dopo aver stabilito che tipo di costi si analizzeranno, occorre elencare le variabili che definiscono concretamente il modello. Queste variabili devono essere in grado di cogliere gli aspetti più importanti del processo produttivo e devono riuscire a descriverlo in modo preciso. Quando si compie un'analisi empirica, volta quindi a confrontare i costi previsti dal modello con quelli osservati nella realtà, la specificazione del modello di costo è molto importante.⁹⁷

Il problema della specificazione del modello è senza dubbio un punto chiave della ricerca empirico-econometrica che, tuttavia, non trova (e non può trovare) una risposta univoca. Il motivo è facile da intuire e risiede nella difficoltà di conciliare, da un lato, una formulazione generale che permetta il confronto tra i dati di molte unità produttive e, dall'altro lato, una descrizione precisa del processo produttivo che sappia cogliere le particolarità di ogni impresa.

La specificazione del modello di costo è quindi una scelta discrezionale degli autori che deve essere esposta in modo chiaro per legittimare i risultati della propria ricerca e, non meno importante, per tracciarne gli eventuali limiti, consentendo una corretta interpretazione dei risultati dello studio.

4.2.2 Forma funzionale

Il secondo elemento importante per la stima di una frontiera di costo è la scelta della forma funzionale da attribuire al modello. La caratteristica principale dei metodi parametrici è quella di individuare una relazione matematica che lega tra loro la variabile che si vuole spiegare (nel nostro caso i costi delle case per anziani) e tutti quei fattori che la possono influenzare. Il modello di costo è quindi definito da una funzione per la quale esistono, dato un modello di costo, molte possibili specificazioni riguardo alla forma funzionale impiegata. Nell'ambito dell'economia applicata, la discussione

⁹⁶ Per una trattazione formale delle funzioni di costo di breve e di lungo periodo si rimanda al Cap. 3.

⁹⁷ In ambito sanitario la letteratura economica applicata ha dedicato grande attenzione a questo problema. Alcuni esempi si trovano in Breyer (1987) e Dor (1989).

sulle caratteristiche e sulle proprietà della forma funzionale di un modello econometrico è un tema che suscita grande interesse e che, con il passare del tempo, ha visto emergere ed affermarsi alcune forme funzionali, in modo prevalente rispetto ad altre.

Al ricercatore interessa poter scegliere consapevolmente la forma funzionale che egli impone al modello economico senza possibilità di sapere a priori la vera forma funzionale del fenomeno che si analizza. Per questo motivo è molto importante chiarire le caratteristiche che una determinata forma funzionale può offrire a priori in modo da poter scegliere quella maggiormente adeguata agli scopi della ricerca. Dopo il processo di stima è possibile trovare delle regole decisionali che aiutano a selezionare ulteriormente la forma funzionale. Queste “regole” sono rappresentate dai test statistici e dal ragionamento economico, che consentono di definire (anche se in modo parzialmente discrezionale) quale sia la forma funzionale che meglio si adatta al campione di dati a disposizione. Il problema della scelta della forma funzionale è stato trattato con particolare attenzione da Lau (1986) il quale ha proposto una serie di criteri per poter scegliere (o quantomeno scartare) una o più forme funzionali: Consistenza teorica, Dominio di applicabilità, Flessibilità, Facilità computazionale, Conformità empirica.⁹⁸

La **consistenza teorica** significa che una forma funzionale deve essere in grado di soddisfare i requisiti e le proprietà previste dalla teoria economica sottostante. Nello specifico, dovendo stimare una funzione di costo, è importante che la forma funzionale sia compatibile con le principali caratteristiche di una funzione di costo. In pratica la funzione scelta deve essere omogenea di 1°, non decrescente e concava nel vettore dei prezzi e non deve essere decrescente nell’output. Alcune di queste proprietà si possono imporre a priori o, in alternativa, si possono verificare in seguito con opportuni test sui risultati delle stime. Il vantaggio di imporre una data condizione è la certezza che sia rispettata, ma il prezzo da pagare è spesso una diminuzione della flessibilità, altro requisito desiderabile.

Il **dominio di applicabilità** riguarda l’interazione tra i parametri stimati empiricamente, la teoria economica ed i valori delle variabili (dipendente ed indipendenti) del modello. In pratica si tratta di stabilire l’insieme di tutti i valori

⁹⁸ Per una discussione ben argomentata delle 5 proprietà elencate si veda Kuenzle (2005).

ammissibili per le variabili indipendenti che, dati i parametri stimati, garantiscono il rispetto delle proprietà economiche (consistenza teorica) da parte della forma funzionale considerata⁹⁹.

La **flessibilità** è un attributo molto importante che può essere descritto in termini generali come una forma funzionale vincolata solamente dal rispetto della consistenza teorica. In altri termini è possibile scegliere un insieme di parametri tali per cui qualsiasi parametro di interesse da essi derivato può assumere, ad un determinato punto di approssimazione nello spazio della variabile, un valore arbitrario e teoricamente consistente¹⁰⁰. Nell'economia applicata la forma funzionale flessibile si traduce spesso nel concetto di flessibilità del secondo ordine o Diewert-Flexibility che limita i parametri di interesse alla seconda derivata della vera (ma sconosciuta) funzione che si cerca di approssimare.¹⁰¹

La **facilità computazionale** è un fattore altrettanto importante che non si può definire in modo univoco ma che si ricollega ad una serie di criteri che presentiamo brevemente. I parametri da stimare dovrebbero essere facilmente ottenibili dai dati e questo nella pratica significa che la forma funzionale scelta è lineare (al limite dopo una trasformazione conosciuta, ad esempio log-log). Da un punto di vista matematico, la forma funzionale dovrebbe essere espressa in forma chiusa.¹⁰² Qualora si stimi un sistema di equazioni (es. una funzione di costo e le relative domande condizionali dei fattori), tutte le funzioni dovrebbero avere la stessa forma algebrica e differire solamente nei parametri. Da ultimo la forma funzionale dovrebbe possedere il minor numero possibile di parametri, per un dato livello di flessibilità, in modo da preservare

⁹⁹ Per quanto riguarda la stima della funzione di costo, il dominio di applicabilità ideale dovrebbe estendersi a tutti i prezzi positivi degli inputs e a tutti gli outputs non negativi. Nella pratica è possibile garantire che questo accada solo sacrificando la flessibilità. In alternativa si può verificare quale sia il sottoinsieme di prezzi e output che garantiscono la consistenza teorica mantenendo la flessibilità.

¹⁰⁰ Per una discussione sulle forme funzionali flessibili e le loro proprietà si veda Feger (2000).

¹⁰¹ La scelta di limitare la flessibilità al secondo ordine è giustificata quando lo scopo della stima è quello di determinare l'elasticità e l'elasticità incrociata (ad esempio dei prezzi degli input), per le quali è sufficiente derivare due volte la funzione primitiva stimata (ad esempio una funzione di costo).

¹⁰² Nelle applicazioni empiriche il rispetto di questa condizione, necessario per riuscire a produrre le stime, è sempre verificato. Per una discussione teorica si veda Lau (1986).

preziosi gradi di libertà per i test statistici che si basano sulla teoria dei grandi campioni.¹⁰³

La **conformità empirica** è l'ultimo criterio da considerare nella scelta della forma funzionale ed è anche il più aleatorio da enunciare. In teoria questa proprietà indica che la forma funzionale dovrebbe essere consistente con la realtà empirica che si sta analizzando. In pratica le evidenze empiriche cambiano da caso a caso ed è molto difficile individuare qualche principio generale. Per portare un esempio abbastanza istruttivo, se consideriamo una funzione di produzione potremmo escludere a priori (o non desiderare) che le elasticità di sostituzione tra tutte le coppie di input siano identiche quando vi sono tre o più fattori produttivi. In un simile caso si potrebbero sostituire gli input (ad un medesimo fattore di sostituzione costante) indipendentemente dal livello di produzione e di impiego degli input stessi. Le forme funzionali (ad esempio la Cobb-Douglas) che impongono questa restrizione non supportata dall'evidenza dei fatti, andrebbero impiegate con grande circospezione (o meglio ancora non andrebbero utilizzate) qualora l'intento del ricercatore fosse indirizzato alla stima delle elasticità di sostituzione tra i fattori produttivi.

Il problema di fondo nella scelta di una forma funzionale risiede nell'impossibilità di soddisfare contemporaneamente tre requisiti importantissimi quali: la coerenza intrinseca con le proprietà economiche previste dalla teoria, l'assenza di restrizioni aprioristiche alla tecnologia, la facilità e l'efficienza computazionali. Alcune forme funzionali impiegate in letteratura come la Cobb-Douglas, le funzioni CES (elasticità di sostituzione costante), e le funzioni Leontief sono interessanti perché soddisfano tutte le proprietà economiche discusse in precedenza, tuttavia non sono funzioni flessibili. Se si desidera che i parametri non siano costretti dalla forma funzionale stessa ad assumere determinati valori, si preferirà certamente una forma funzionale flessibile come la Translog, le funzioni quadratiche o la Generalised Leontief.¹⁰⁴

¹⁰³ Da notare che anche in presenza di sufficienti gradi di libertà è preferibile limitare il numero dei parametri per non accentuare il possibile problema della multicollinearità, tipico di alcune forme funzionali flessibili.

¹⁰⁴ Questa forma funzionale è stata proposta originariamente da Diewert (1971) per un processo produttivo monoprodotta e successivamente da Hall (1973) per il caso multiprodotta. Diewert e Wales (1987) hanno proposto una variante per approssimare una versione che fosse flessibile anche quando la funzione approssimata è sconosciuta.

Queste forme funzionali sono definite flessibili perché in grado di approssimare localmente una qualunque funzione continua differenziabile per due volte. Per loro natura, le forme flessibili descritte non sono in grado di approssimare globalmente la funzione sconosciuta e offrono il fianco a questa critica che ha spinto alcuni autori (Gallant, 1981) a sviluppare forme globalmente flessibili come l'*Augmented Fourier*. In questo caso il tentativo di correggere le limitazioni delle forme flessibili tradizionali si paga con l'elevato numero di parametri da stimare e la difficoltà di applicare praticamente queste forme funzionali nei lavori empirici. Il discorso sulle forme funzionali si complica ulteriormente se consideriamo un processo produttivo multiprodotto con la possibilità di considerare anche elementi nulli nel vettore degli output che, nel caso delle trasformazioni logaritmiche (es. Translog) pone dei seri problemi.

Box and Cox (1964) hanno proposto una trasformazione dell'output non logaritmica che ammetta la presenza di output pari a zero.¹⁰⁵ Da un punto di vista pratico, tuttavia, la trasformazione Box-Cox presenta lo svantaggio di non essere lineare nei parametri, limitando la scelta della tecnica econometrica e dei modelli di frontiera disponibili. La discussione sulle proprietà di tutte le famiglie di forme funzionali elaborate dalla letteratura potrebbe proseguire ma, un'analisi di dettaglio sulle caratteristiche e le limitazioni delle forme funzionali va oltre gli scopi di questa ricerca, che si è limitata a sottolineare l'importanza di scegliere la forma funzionale in modo consapevole.¹⁰⁶

Di seguito presentiamo le caratteristiche delle due forme funzionali maggiormente impiegate negli studi empirici relativi alle frontiere stocastiche di costo: la funzione di costo Cobb-Douglas e la funzione di costo Translog.

4.2.2.1 Funzione di costo Cobb-Douglas

La forma funzionale Cobb-Douglas deve il suo nome agli autori che, nel 1928, la proposero come una forma funzionale idonea per descrivere la tecnologia di produzione. La principale caratteristica della Cobb-Douglas è quella di essere duale a sé

¹⁰⁵ Caves et al. (1980) hanno proposto, partendo dalla trasformazione Box-Cox, un modello generalizzato per la Translog.

¹⁰⁶ Una descrizione dettagliata delle forme funzionali maggiormente impiegate nella letteratura empirica si trova in Zoric (2006) e Kuenzle (2005).

stessa (*self dual*) nel senso che data una funzione di produzione Cobb-Douglas la funzione di costo corrispondente è anch'essa Cobb-Douglas. Quando un'impresa utilizza K input per produrre M output secondo la tecnologia Cobb-Douglas, la funzione di costo corrispondente si può esprimere come:

$$C = \alpha \prod_{i=1}^M y_i^{\beta_i} \prod_{j=1}^K w_j^{\gamma_j} \quad \text{con } i = 1, 2, \dots, M \text{ e } j = 1, 2, \dots, K \quad [4.1]$$

Nella [4.1] i costi C sono determinati dal valore y_i assunto dagli M output prodotti (o dalle loro caratteristiche qualitative) e dal livello w_j dei prezzi dei K fattori produttivi. Si nota come la funzione Cobb-Douglas, altrimenti detta log-log, non sia lineare nelle variabili che la costituiscono. Tuttavia, con una semplice trasformazione logaritmica la [4.1] genera una nuova espressione:

$$\ln C = \ln \beta_0 + \sum_{i=1}^M \beta_i \ln y_i + \sum_{j=1}^K \gamma_j \ln w_j \quad [4.2]$$

L'equazione [4.2] è lineare nei parametri ed è semplice sia da stimare econometricamente sia da interpretare economicamente.

Da un punto di vista della teoria economica i parametri β_i e γ_j devono essere sempre positivi per garantire che i costi di produzione non decrescano all'aumentare dell'output e dei prezzi degli input. Inoltre, una funzione di costo dovrebbe essere omogenea di 1° rispetto al prezzo dei fattori. La funzione Cobb-Douglas rispetta questa prescrizione teorica solo quando $\sum_{j=1}^K \gamma_j = 1$. Nel processo di stima è possibile imporre tale restrizione ai parametri o, in alternativa, normalizzare la [4.2] rispetto ad uno dei prezzi dei fattori. Nel primo caso è possibile condurre un test diagnostico sulla restrizione imposta, mentre nel secondo caso l'imposizione è strutturalmente imposta dividendo i costi e i prezzi degli input per uno dei prezzi degli input preso a riferimento.¹⁰⁷

¹⁰⁷ L'ipotesi di omogeneità lineare nei prezzi degli input si può testare utilizzando ad esempio il Wald test tra la stima del modello ristretto e di quello libero.

Applicando il lemma di Shephard alla [4.2] possiamo determinare le domande condizionali dei fattori produttivi e determinare le quote di costo che risultano espresse dai parametri γ_j , fornendo una chiave interpretativa molto immediata. Di facile derivazione è anche l'elasticità di costo che, per ciascuna delle variabili esplicative, è espressa dal parametro corrispondente.

La tecnologia descritta da una funzione Cobb-Douglas impone alcune restrizioni che è bene sottolineare dal momento che riguardano la sostituibilità dei fattori produttivi e il valore delle economie di scala. La Cobb-Douglas è una funzione che impone un'elasticità di sostituzione dei fattori costante e pari a 1, escludendo la possibilità di analizzare empiricamente questo aspetto. La distinzione tra rendimenti ed economie di scala non assume rilevanza dal punto di vista empirico, dal momento che la funzione Cobb-Douglas è omotetica.

Tuttavia, la misura delle economie di scala, che è determinata dalla somma delle elasticità di costo rispetto agli output, risulta costante per tutto lo spettro di produzione. A volte, questa limitazione è in contrasto con la realtà osservata di alcuni settori nei quali, a causa delle differenti economie di scala realizzate, una dimensione ottimale emerge con chiarezza.¹⁰⁸

Il grande pregio della funzione Cobb-Douglas è quello di possedere le caratteristiche economiche della funzione di costo elencate per la [3.5] e di essere molto facile da stimare empiricamente. La relazione lineare tra i suoi parametri e il numero contenuto degli stessi le hanno conferito grande risalto ed interesse. La semplicità computazionale della Cobb-Douglas si ottiene pagando un prezzo in termini di flessibilità. Molte caratteristiche tecnologiche sono imposte a priori o non si possono stimare, rendendo poco attraente questa forma funzionale per l'indagine empirica di alcuni aspetti della tecnologia produttiva, in particolare la sostituibilità dei fattori e la misurazione delle economie di scala.

¹⁰⁸ Senza abbandonare del tutto la Cobb-Douglas è possibile rendere variabili con la scala le elasticità di costo aggiungendo i termini del secondo ordine per il vettore degli output.

4.2.2.2 Funzione di costo Translog

La funzione Translog, nome breve per *TRANScendental LOGarithmic function*, fu proposta all'inizio degli anni '70 da Christensen, Jorgenson e Lau (1971, 1973) per estendere le possibilità di indagine empirica con uno strumento più flessibile. In effetti la translog si propone di approssimare in un punto specifico la vera funzione sconosciuta che rappresenta la tecnologia. Utilizzando l'espansione di Taylor del secondo ordine nella sua forma logaritmica si può approssimare la funzione di costo sconosciuta $C^*(y,w)$ con la seguente espressione:

$$\begin{aligned} \ln C = & \ln \alpha + \sum_{i=1}^M \beta_i \ln y_i + \sum_{j=1}^K \gamma_j \ln w_j + \\ & + \frac{1}{2} \sum_{i=1}^M \beta_{ii} \ln y_i \ln y_i + \frac{1}{2} \sum_{j=1}^K \gamma_{jj} \ln w_j \ln w_j + \\ & + \sum_{i=1}^M \sum_{l=1}^M \beta_{il} \ln y_i \ln y_l + \sum_{j=1}^K \sum_{k=1}^K \gamma_{jk} \ln w_j \ln w_k + \sum_{i=1}^M \sum_{j=1}^K \delta_{ij} \ln y_i \ln w_j \end{aligned} \quad [4.3]$$

con $i = 1, 2, \dots, M$ e $j = 1, 2, \dots, K$

Pur non conoscendo le vere caratteristiche del processo produttivo osservato, nel punto di approssimazione scelto, la [4.3] è in grado di rappresentare in modo adeguato la funzione di costo multiprodotto che ne deriva.¹⁰⁹ La scelta del punto di approssimazione è discrezionale, ma comporta una conseguenza molto importante: l'affidabilità dei risultati delle stime perde progressivamente forza più ci si allontana dal punto di riferimento scelto. Per questo motivo, da un punto di vista economico è interessante approssimare la funzione sconosciuta in corrispondenza di un indicatore empiricamente significativo come la media o la mediana.¹¹⁰ Dividendo tutte le variabili esplicative per il punto di approssimazione scelto (ad esempio la mediana) si ottiene una normalizzazione dei dati che permette di interpretare in modo locale le stime ottenute a partire dalla [4.3].

¹⁰⁹ Per una presentazione dell'espansione di Taylor applicata ad una funzione di costo Translog si veda, ad esempio, Filippini (1991).

¹¹⁰ Solitamente si preferisce l'uso della mediana che rispetto alla media è meno sensibile ai valori estremi presenti nel campione e, in chiave distributiva, si interpreta in modo più trasparente. Il problema della scelta del punto di approssimazione, con riferimento alla stima di una funzione di costo per le case per anziani, è discusso da Christensen (2004).

I parametri della translog si possono suddividere in parametri del 1° ordine, del 2° ordine e parametri dei prodotti incrociati. Questa suddivisione è evidenziata nella [4.3] dalla disposizione su tre righe dei vari termini. Questi gruppi di parametri esprimono complessivamente quella flessibilità della tecnologia che non troviamo nella Cobb-Douglas. A ben vedere la Cobb-Douglas si potrebbe definire una translog con i parametri del 2° ordine e dei prodotti incrociati (seconda e terza riga della [4.3]), ristretti forzatamente a 0. In assenza di tali restrizioni, i parametri da stimare risultano molto più numerosi e precisamente sono pari a $\frac{1}{2}(n+1)(n+2)$, quando $n = M + K$ è il numero di regressori. Rispetto alla Cobb-Douglas si nota un aumento notevole del numero di parametri da stimare che cresce alla velocità $n/2$ in relazione al numero n di regressori utilizzati. Questo aggravio di parametri da stimare presenta due conseguenze negative: la riduzione del numero di gradi di libertà per la stima e (2) il possibile insorgere del fenomeno della multicollinearità.¹¹¹

La forma funzionale translog non garantisce a priori le condizioni di regolarità necessarie per rappresentare in modo adeguato una funzione di costo e quindi occorre verificare se, e sotto quali condizioni, i parametri stimati rispecchiano la teoria economica. Tra le diverse proprietà elencate in precedenza, ve ne sono tre essenziali per definire la regolarità di una funzione di costo: (1) la monotonicità della funzione di costo rispetto agli output e ai prezzi degli input, (2) L'omogeneità lineare della funzione di costo rispetto ai prezzi dei fattori produttivi e (3) la convessità della funzione di costo rispetto ai prezzi dei fattori produttivi. La prima condizione (monotonicità) è rispettata quando l'elasticità di costo risulta maggiore o uguale a zero. Applicando la [3.12] alla [4.3] otteniamo quindi:

$$\begin{aligned} \varepsilon_{y_i} &= \frac{\partial \ln C}{\partial \ln y_i} = \beta_i + \sum_{l=1}^M \beta_{il} \ln y_l + \sum_{j=1}^K \delta_{ij} \ln w_j \geq 0 \\ \varepsilon_{w_j} &= \frac{\partial \ln C}{\partial \ln w_j} = \gamma_j + \sum_{k=1}^K \gamma_{jk} \ln w_k + \sum_{i=1}^M \delta_{ik} \ln y_i \geq 0 \end{aligned} \tag{4.4}$$

con $i \in 1, 2, \dots, M$ e $j \in k = 1, 2, \dots, K$

¹¹¹ In presenza di campioni di dati sufficientemente grandi, il problema dei gradi di libertà è trascurabile, mentre non lo è la presenza di multicollinearità che potrebbe portare alla stima di parametri con il segno errato o con valori di significatività statistica molto bassi. Per una discussione sul problema si veda ad esempio Filippini (1997).

Nell'espressione dell'elasticità di costo appena derivata, si nota un primo vantaggio della flessibilità della translog rispetto alla Cobb-Douglas, osservando come la [4.4] esprima valori di ε_{yi} che dipendono dall'output, originando economie di scala e di dimensione variabili.¹¹²

La funzione di costo deve essere linearmente omogenea nei prezzi degli input. Questa condizione si può soddisfare in diversi modi: imponendo alcune restrizioni ai parametri della [4.3], verificando a posteriori (con opportuni test econometrici) la condizione, oppure normalizzando il costo ed i prezzi dei fattori per il prezzo di un input scelto come numerario.

Le restrizioni da imporre ai parametri per garantire l'omogeneità lineare sono:

$$\sum_{j=1}^K \gamma_j = 1, \quad \sum_{j=1}^K \gamma_{jk} = 0, \forall k, \quad \sum_{j=1}^K \delta_{jk} = 0, \forall k \quad [4.5]$$

Un'alternativa all'imposizione dell'omogeneità lineare attraverso i parametri è quella di normalizzare i prezzi degli input e i costi totali per il prezzo di un fattore scelto come numerario. Da un punto di vista economico questa operazione porta allo stesso risultato, ma lascia i parametri liberi da restrizioni durante il procedimento di stima.

Il lemma di Shephard, che, come visto in precedenza, consente di derivare dalla funzione di costo le quantità ottimali dei fattori, si può applicare nella forma logaritmica specificata dalla [4.3] ottenendo le quote di costo S_j di ciascuno dei K input:

$$S_j \equiv \frac{\partial \ln C(y, w)}{\partial \ln w_j} \equiv \frac{\partial C(y, w)}{\partial w_j} \frac{w_j}{C(y, w)} = \gamma_j + \sum_{k=1}^K \gamma_{jk} \ln w_k + \sum_{i=1}^M \delta_{ij} \ln y_i \quad [4.6]$$

Il valore delle quote di costo, derivate dall'applicazione del lemma di Shephard, permette di verificare se la funzione di costo è non decrescente nel vettore dei prezzi degli input, il che si verifica quando la [4.6] è maggiore o uguale a zero. Da notare che, per un processo produttivo monoprodotto, nel punto di approssimazione il valore delle quote di costo S_j è semplicemente dato dal parametro del primo ordine γ_j . Inoltre, dal

¹¹² Nel caso della translog la tecnologia non è necessariamente omotetica e quindi i concetti di economie di scala e di dimensione vanno distinti.

momento che l'omogeneità lineare è imposta (direttamente o tramite la normalizzazione), la somma delle quote di costo S_j deve essere uguale a uno.¹¹³

La concavità della funzione di costo rispetto al vettore dei prezzi degli input è l'ultima condizione richiesta per la regolarità della funzione di costo che discuteremo. Questa condizione si verifica quando la matrice Hessiana corrispondente, ottenuta derivando per due volte la funzione di costo rispetto ai prezzi dei fattori, è negativa semi-definita. Diewert e Wales (1987) dimostrano che la condizione di concavità rispetto ai prezzi degli input si verifica quando la seguente matrice è negativa semidefinita:

$$\Gamma(y, w) = \begin{vmatrix} \gamma_{11} - S_1(1 - S_1) & \gamma_{12} + S_1S_2 & \cdots & \gamma_{1K} + S_1S_K \\ \gamma_{21} + S_2S_1 & \gamma_{22} - S_2(1 - S_2) & \cdots & \gamma_{2K} + S_2S_K \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{K1} + S_KS_1 & \gamma_{K2} + S_KS_2 & \cdots & \gamma_{KK} - S_K(1 - S_K) \end{vmatrix} \quad [4.7]$$

Occorre sottolineare che gli elementi della matrice non sono determinati solamente dai valori dei parametri γ_{jk} , ma anche dalle quote di costo S_j .

Pertanto, la matrice [4.7] $\Gamma(y, w)$ dipende dall'output e dai prezzi degli input, visto che le quote di costo sono una funzione dell'output y e dei prezzi degli input w , secondo la relazione [4.6] definita in precedenza. Da notare che la simmetria dei parametri $\gamma_{jk} = \gamma_{kj}$ non implica la simmetria degli elementi jk della matrice $\Gamma(y, w)$ poiché la simmetria delle quote di costo non è affatto garantita.

A titolo di esempio consideriamo una funzione di costo monoprodotto con due soli input, lavoro (L) e capitale (K), che potrebbe descrivere in modo generale il processo produttivo di una casa per anziani.

Nel punto di approssimazione scelto (normalmente il valore mediano delle osservazioni) la [4.7] si riduce ad una matrice 2x2 i cui elementi jk sono simmetrici e caratterizzati nel modo seguente:

¹¹³ Normalizzando i prezzi dei fattori per imporre strutturalmente l'omogeneità lineare, si perde come regressore l'input numerario i cui parametri si possono ricostruire utilizzando la condizione di omogeneità lineare imposta dalla [4.5]. Il coefficiente del primo ordine dell'input numerario, ad esempio, si ottiene per complemento all'unità della somma di tutti i K-1 coefficienti γ_j del primo ordine.

$$\Gamma_{L,K} = \begin{vmatrix} \gamma_{LL} - \gamma_L \gamma_K & \gamma_{LK} + \gamma_L \gamma_K \\ \gamma_{LK} + \gamma_L \gamma_K & \gamma_{KK} - \gamma_L \gamma_K \end{vmatrix} \quad [4.8]$$

In questo caso la verifica della concavità della funzione di costo rispetto ai prezzi dei fattori produttivi è molto semplice dato che tutti gli elementi della matrice [4.8] $\Gamma_{L,K}$ si ricavano dalle stime dei parametri della funzione.¹¹⁴ Occorre sottolineare che la verifica empirica (ex-post) della concavità nei prezzi degli input è necessaria per coniugare il rigore della teoria economica con la flessibilità della Translog.¹¹⁵ Imporre la concavità è in effetti possibile ma al costo di distruggere la flessibilità di questa forma funzionale.

L'utilizzo della Translog, in conclusione, permette di soddisfare in modo automatico alcune proprietà della funzione di costo, mentre altre proprietà vanno imposte o verificate a posteriori. Trattandosi di un'approssimazione locale intorno ad un punto di riferimento ricavato dai dati (la media, la mediana o un altro indice di posizione), i risultati delle stime di una Translog richiedono grande attenzione interpretativa. In particolare, quando ci si allontana dal punto di approssimazione scelto, le stime dei parametri possono soffrire di importanti distorsioni. Questa criticità impone l'uso di una certa cautela nell'interpretazione economica dei risultati di una stima empirica che si dimostra econometricamente affidabile solo nei pressi del punto di approssimazione dei dati.

Scelta la forma funzionale e definito il modello di costo, tutto è pronto per estrarre dai dati empirici le informazioni di interesse per il ricercatore. Questa ultima fase, detta in breve stima econometrica, comporta numerose problematiche in relazione alle differenti tecniche disponibili per la stima empirica dei parametri di una funzione di costo. I modelli parametrici per l'analisi delle performance sono accomunati dal fatto di generare delle stime empiriche per i parametri del modello utilizzato.

¹¹⁴ Nel caso specifico la matrice [4.8] è negativa semi-definita se tutti gli elementi della diagonale principale sono negativi e cioè se $\gamma_{LL} < \gamma_L \gamma_K$ e $\gamma_{KK} < \gamma_L \gamma_K$.

¹¹⁵ La concavità nei prezzi degli input è una proprietà che dovrebbe valere in tutto il dominio della funzione. La verifica della proprietà nel solo punto di approssimazione scelto obbliga ad interpretare i risultati delle stime in modo locale. Ryan and Wales (2000) e Chew et al. (2005) hanno sviluppato alcune tecniche che permettono di risolvere parzialmente il problema, imponendo la concavità in modo indiretto (tramite la scelta del punto di approssimazione).

Ciò che invece li distingue è la tecnica econometrica con la quale si arriva a determinare il valore di tali parametri: accanto alle semplici regressioni lineari, come il metodo dei minimi quadrati ordinari OLS (*Ordinary Least Square*) possiamo infatti trovare stime di massima verosimiglianza o avanzate tecniche di simulazione (massima verosimiglianza simulata).

Gli aspetti econometrici delle frontiere stocastiche sono alla base delle proprietà e delle possibilità che i differenti modelli di frontiera di costo offrono al ricercatore ed è per questo motivo che dedicheremo il prossimo paragrafo ad una discussione delle caratteristiche dei diversi approcci.

4.3 Stima econometrica di una frontiera di costo

Per determinare empiricamente i parametri di una funzione di costo (e di una qualunque altra funzione) si possono utilizzare gli strumenti teorici dell'econometria ed i software appositamente sviluppati per le applicazioni empiriche. I problemi legati alla procedura di stima che caratterizzano i lavori di ricerca applicata sono presenti anche quando si desidera determinare il valore dei parametri di una frontiera di costo.

In questo caso, tuttavia, il concetto stesso di frontiera (inteso come limite invalicabile) di costo pone un altro problema, poiché le procedure econometriche devono permettere di accomodare i dati empirici in modo tale che nessuna osservazione risulti oltre la frontiera stimata. I costi osservati attraverso i dati empirici, possono scostarsi dal minimo teorico stimato dalla frontiera generando dei residui econometrici, i quali inevitabilmente si interpretano come inefficienza.

Il cuore della discussione sui modelli di frontiera verte proprio sul modo di specificare il termine di errore, rappresentato dai residui della procedura di stima. Ad esempio, le ipotesi di partenza circa i residui della regressione permettono di distinguere tra l'approccio deterministico e quello stocastico (SFA) nell'analizzare ed interpretare le stime di una funzione frontiera di costo.

Nel filone della SFA è poi rilevante la distinzione tra quei metodi che attribuiscono al termine non casuale di errore una precisa distribuzione di probabilità e quelli che invece non fanno alcuna assunzione. Di particolare interesse è poi la struttura temporale delle osservazioni empiriche che apre uno spazio di discussione in merito alle tecniche di stima panel che permettono di separare meglio l'eventuale eterogeneità non

osservata dal modello dalla misura dell'inefficienza. In presenza di dati panel è anche opportuno distinguere tra quei modelli che considerano l'inefficienza come costante nel tempo ed altri che la reputano variabile.

La formulazione più o meno articolata del modello di errore pone anche dei quesiti di natura econometrica nel senso che alcune specificazioni richiedono l'uso di tecniche più sofisticate di stima ed il ricorso, a volte, a metodi di iterazione numerica come la stima di massima verosimiglianza simulata o stimatori non lineari molto complessi. Queste considerazioni meritano un approfondimento specifico per comprendere le caratteristiche economiche dei diversi modelli disponibili e le relative difficoltà econometriche che emergono nel processo di stima.

Dopo una presentazione dei metodi più diffusi nella letteratura empirica sulle frontiere di costo, esporremo alcune considerazioni comparative per poter decidere quale modello si addice maggiormente al nostro obiettivo di ricerca.

4.3.1 L'approccio deterministico all'analisi delle frontiere

La stima di una funzione di costo determina i parametri delle variabili del modello e genera, confrontando le osservazioni empiriche dei costi con i valori previsti, i residui della regressione. Il termine di errore di una stima OLS, distribuito normalmente intorno allo zero, non si può interpretare come inefficienza di costo, dal momento che quest'ultima può assumere solo valori positivi.

Il modo più semplice per ottenere una frontiera deterministica di costo, partendo dai risultati di una stima OLS della funzione di costo, è noto con il termine di COLS o "*Corrected Ordinary Least Squares*".¹¹⁶ In pratica la funzione di costo stimata viene traslata verso il basso in modo che tutti i residui della regressione giacciono sopra o al limite sulla funzione di costo traslata che diventa una frontiera di costo. Partendo da una stima lineare nei parametri è quindi facile ottenere una frontiera di costo poiché è sufficiente modificare un solo parametro, la costante della regressione.

¹¹⁶ Questo metodo fu proposto originariamente da Winsten (1957). Per una discussione approfondita riguardante gli approcci econometrici applicati all'analisi dell'efficienza, si vedano ad esempio Kumbhakar and Lovell (2000) e Greene (2008).

Se consideriamo ad esempio la forma funzionale Cobb-Douglas [4.2], la stima OLS di una generica funzione di costo monoprodotto, espressa in forma log-log, si può rappresentare nel seguente modo:

$$\text{Ln}C_i = \beta_0 + \beta_y \ln y + \sum_{j=1}^K \beta_j \ln w_{ji} + \varepsilon_i \quad [4.9]$$

Nella [4.9] il logaritmo dei costi di ogni i -esima osservazione ($i = 1, 2, \dots, N$) è determinato dalla funzione di costo specificata, i cui veri parametri sono rappresentati dalle stime dei coefficienti.¹¹⁷ I residui della regressione $\hat{\varepsilon}_i$ rappresentano l'errore casuale (ε_i) che è distribuito normalmente intorno allo zero, potendo assumere valori positivi e negativi. Partendo dai risultati della [4.9] possiamo ricavare la seguente espressione che definisce la frontiera deterministica di costo secondo il metodo COLS:

$$\text{Ln}C_i = \beta_0^* + \beta_y \ln y + \sum_{j=1}^K \beta_j \ln w_{ji} + u_i^* \quad [4.10]$$

Il termine costante, stimato seguendo la [4.10], è per costruzione pari a $\hat{\beta}_0^* = \hat{\beta}_0 - |\min\{\hat{\varepsilon}_i\}|$, mentre i residui sono definiti dal termine $\hat{u}_i^* = \hat{\varepsilon}_i + |\min\{\hat{\varepsilon}_i\}|$.

La traslazione della costante, che garantisce un valore dei residui maggiore o uguale a zero, permette di interpretare la componente u_i^* in modo deterministico come uno scostamento dai costi minimi espressi dalla frontiera COLS. I parametri stimati seguendo l'approccio COLS sono consistenti da un punto di vista econometrico e permettono di soddisfare empiricamente anche i precetti teorici di una frontiera di costo.¹¹⁸ Il metodo COLS permette di determinare empiricamente l'inefficienza di costo e misurarla con un indicatore derivato dalle stime dei parametri della frontiera.

¹¹⁷ I veri parametri della funzione di costo non sono conosciuti. I valori ottenuti attraverso la procedura di stima OLS $\hat{\beta}$ sono consistenti (anche se non efficienti) con il vettore β dei veri parametri della funzione.

¹¹⁸ Per una prova della consistenza dello stimatore COLS si veda Greene (1980). Lo stesso Greene (1997) mostra come i parametri della funzione [4.10] si possano stimare in modo più efficiente con la tecnica di massima verosimiglianza. Questa procedura, tuttavia, richiede la conoscenza (o l'imposizione) della distribuzione del termine di errore u_i^* .

Seguendo la definizione [3.16] data in precedenza, l'inefficienza di costo di ogni singola impresa si può misurare come il rapporto tra il costo minimo determinato dalla frontiera C_i^* ed il costo effettivamente osservato C_i nei dati:

$$EC_i = \frac{C_i^*}{C_i} = \frac{\exp\left[\hat{\beta}_0^* + \hat{\beta}_y \ln y + \sum_{j=1}^K \hat{\beta}_j \ln w_{ji}\right]}{\exp\left[\hat{\beta}_0^* + \hat{\beta}_y \ln y + \sum_{j=1}^K \hat{\beta}_j \ln w_{ji} + \hat{u}_i^*\right]} = \exp\left[-\hat{u}_i^*\right] \quad [4.11]$$

L'inefficienza di costo è quindi definita nello spazio $[0,1]$ in corrispondenza delle due situazioni estreme di piena efficienza ($EC = 1$) e di completa inefficienza ($EC = 0$). Un modo alternativo per esprimere la misura dell'inefficienza di costo è quello di definire l'eccesso di costo in termini percentuali rispetto al costo minimo teorico rappresentato dalla frontiera. Tale misura, che varia da 0 ad infinito e si ottiene sottraendo 1 al reciproco della [4.11], risulta più intuitiva da interpretare e sarà impiegata (e ridiscussa) nella parte dedicata alle analisi empiriche.

Un secondo modello di frontiera deterministica impiegato nella ricerca empirica è il “*Modified Ordinary Least Squares*” o MOLS, il quale segue sostanzialmente l'approccio appena discusso per il COLS.¹¹⁹ In questo caso, però, l'intercetta stimata con l'OLS viene modificata utilizzando il valore atteso della componente di inefficienza u_i e non con il valore assoluto del minor residuo ε_i della regressione. L'unica differenza rispetto al modello [4.11] consiste nella definizione di $\hat{\beta}_0^* = \hat{\beta}_0 - E[u_i^*]$ e di $\hat{u}_i^* = \hat{\varepsilon}_i + E[u_i^*]$.

Rispetto al COLS questo approccio è meno sensibile ai valori estremi dei residui, ma obbliga anche ad attribuire alla componente u_i una specifica distribuzione. Inoltre, con il metodo MOLS non vi è garanzia che tutti i costi osservati siano maggiori (o al limite uguali) rispetto ai costi teorici indicati dalla frontiera. Questo perché i residui negativi della stima OLS potrebbero (e di solito è così nella pratica) risultare maggiori del valore atteso $E[u_i]$ determinando un valore negativo per la componente di inefficienza difficile da interpretare economicamente.

¹¹⁹ Questo approccio è stato proposto da Afriat (1972) e ripreso da Richmond (1974).

La definizione di un modello deterministico di frontiera di costo si basa sull'ipotesi che ogni eccesso di costo misurato dai parametri della frontiera (attraverso una manipolazione algebrica dei residui della regressione) sia da attribuire a comportamenti non efficienti da parte dell'impresa. In altri termini non si concede che la realtà produttiva osservata, il modello che la descrive ed i dati che la rappresentano, possano avere delle ripercussioni sull'inefficienza non attribuibili alle imprese.

Se, invece, si considerano i processi produttivi come fenomeni stocastici, le informazioni contenute nei residui sono molteplici. Da un punto di vista econometrico, infatti, il termine di errore cattura almeno tre distinte componenti: (1) l'omissione nel modello di alcune variabili esplicative, (2) gli errori di misurazione e (3) la componente di casualità nei comportamenti dei soggetti economici.¹²⁰ Alla luce di questo ragionamento, interpretare i residui della regressione come inefficienza delle imprese appare poco soddisfacente, vista la presenza di questi tre elementi di disturbo.

Considerare il processo produttivo perfettamente determinato dalla funzione di costo implica il conseguente problema di basare la valutazione dell'efficienza su quelle poche (di solito una sola) unità statistiche che determinano la traslazione della funzione di costo, trasformandola in una frontiera. I residui ottenuti dalla stima della funzione di costo, che determinano lo spostamento del termine di paragone (*yardstick*), non considerano la possibilità che una parte dell'errore sia dovuta a fenomeni casuali fuori dal controllo delle imprese. Questa componente, che non dovrebbe entrare nella misura empirica dell'inefficienza dell'impresa, non esprime l'errata scelta nel processo di minimizzazione dei costi (inefficienza), quanto piuttosto la fisiologica variabilità di alcune decisioni (natura stocastica del processo produttivo) o l'imprecisione nel modello e nei dati.

In contrapposizione ai modelli di frontiera deterministici troviamo l'approccio della "*Stochastic Frontier Analysis*" o SFA che si propone di stimare una frontiera di costo che consideri la componente casuale e che riesca a filtrare gli effetti di questo termine nella misurazione dell'inefficienza.

¹²⁰ Questa scomposizione del residuo di una regressione, tratto da Salvatore (1985), mette in luce la sottile differenza che esiste tra l'ipotesi di perfetta razionalità, postulata dalla teoria economica e la constatazione empirica di razionalità limitata o comportamento razionale stocastico.

4.3.2 L'approccio stocastico all'analisi delle frontiere (SFA)

L'analisi empirica dell'inefficienza delle imprese si può condurre utilizzando le frontiere stocastiche di costo. Questo approccio si pone come obiettivo quello di conciliare l'impostazione neoclassica (minimizzazione dei costi) teorica con una verifica (e relativa misurazione) empirica delle performance delle imprese. In un contesto stocastico, la struttura del residuo si può scomporre in due parti che descrivono effetti molto diversi: gli effetti casuali e l'inefficienza delle imprese.

Entrambi gli effetti sono normalmente presenti e, da un punto di vista empirico, si presentano inevitabilmente confusi dal momento che confluiscono nei residui della regressione. La letteratura sulla SFA ha sviluppato numerosi modelli che si differenziano nel tipo di dati a disposizione, nella tecnica econometrica usata per le stime, ma soprattutto nel modo con il quale viene specificato il termine di errore.

La struttura dei dati è una caratteristica molto importante dal punto di vista econometrico che distingue in modo netto i modelli Cross-Section da quelli Panel-Data. I primi riguardano una serie di osservazioni riferite a diverse unità statistiche valutate in un punto del tempo, mentre i dati panel sono un insieme di dati Cross-Section osservati su più periodi temporali. L'approccio econometrico ai dati panel permette di sfruttare la struttura temporale dei dati per ottenere informazioni più specifiche sulle caratteristiche delle singole imprese (mitigando il problema dell'eterogeneità presente nei dati) o per migliorare la precisione delle stime.¹²¹

Da un punto di vista dell'inefficienza e della sua misurazione l'utilizzo dei modelli panel apre un altro tipo di discussione che verte sul comportamento dell'inefficienza nel corso del tempo. Rispetto a questa dimensione si distinguono quei modelli che considerano l'inefficienza costante nel tempo dai modelli che permettono all'inefficienza di modificarsi da un periodo all'altro.

Un altro problema da considerare è quello relativo all'eterogeneità dei processi produttivi non catturata dal modello. Come accennato in precedenza, se ciò che finisce nei residui della regressione si trasforma in una misura dell'inefficienza, è opportuno discutere, da un punto di vista econometrico, il problema dell'eventuale traccia di

¹²¹ I dati panel si prestano anche all'uso delle tecniche di stime Cross-Section. In questo caso le informazioni temporali vengono ignorate e si parla di modelli *pooled*.

eterogeneità non osservata presente nei residui. Questa discussione ha dato origine ad una serie di modelli che cercano di trattare il problema dell'eterogeneità per mitigare le conseguenze che può avere sulla misurazione empirica dell'efficienza e sulla sua interpretazione economica. Nell'ambito della SFA sono stati seguiti numerosi approcci che si differenziano rispetto alle tre dimensioni menzionate: tipo di dati considerati per le stime (cross-sectional vs. panel data), comportamento dell'inefficienza rispetto al tempo (costante vs. variabile) e modellizzazione dell'eterogeneità non osservata (approcci che non la considerano vs. approcci che la considerano).

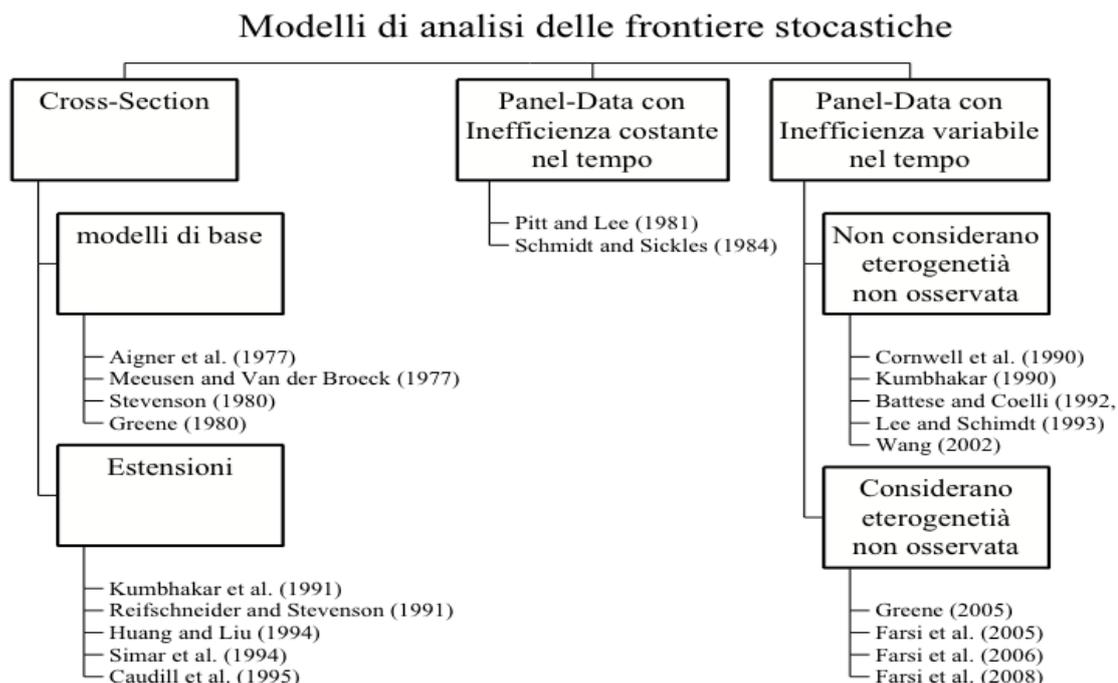
Quando una banca dati contiene informazioni ripetute nel tempo e riferite alle medesime unità statistiche siamo in presenza di una base statistica che esprime due dimensioni: la variabilità dei dati tra le N diverse unità (*Cross-Section*) e la variabilità nel corso dei T periodi di tempo di ogni osservazione (*Time-Series*). In questi casi è comunque possibile ignorare la struttura temporale dei dati e raggruppare (*pool*) le $N \times T$ osservazioni, considerandole in modo distinto. I modelli che non hanno (o non considerano) la dimensione temporale sono conosciuti come modelli *Cross-Section*, mentre i modelli che sfruttano le informazioni sulla struttura temporale della banca dati si definiscono modelli *Panel-Data*.

Riprendendo i tre criteri appena discussi, presentiamo in forma schematica (Figura 4.1) alcuni contributi sul tema delle frontiere stocastiche, riconosciuti dalla letteratura come modelli di riferimento per la stima dell'inefficienza produttiva.

Nella famiglia dei modelli Cross-Section (o Pool) troviamo il lavoro di Aigner Lovell and Schmidt (1977) e quello di Meeusen and Van der Broeck (1977) considerati i “capi stipite” della SFA. Le differenze tra questi contributi e gli altri modelli che abbiamo denominato di base, consistono nel tipo di distribuzione assegnato alla componente asimmetrica del termine di errore. All'inizio degli anni '90 il modello originale ha conosciuto alcune interessanti estensioni che hanno considerato il ruolo delle variabili esogene sul termine di inefficienza.

Nel secondo gruppo, modelli panel con inefficienza costante nel tempo, abbiamo inserito solamente i due modelli più conosciuti: Pitt and Lee (1981) e Schmidt and Sickles (1984). Il primo modello estende il lavoro di Aigner et al. (1977) al caso dei panel data ed utilizza la tecnica della massima verosimiglianza, specificando un modello ad effetti casuali o *Random Effect Model* (REM).

Figura 4.1 - Famiglie di modelli per l'analisi delle Frontiere Stocastiche.



Il lavoro di Schmidt and Sickles (1984) sfrutta, invece, le proprietà degli stimatori panel per determinare l'inefficienza, senza bisogno di definire a priori la distribuzione della componente asimmetrica del termine di errore. Entrambi i modelli immaginano che l'inefficienza di ogni individuo (impresa) sia persistente, causata quindi da fattori strutturali. La famiglia dei modelli panel data, che ammettono la possibilità che l'inefficienza vari nel tempo (oltre che tra gli individui), è molto ricca e si è sviluppata a partire dagli anni '90.

La possibilità di specificare l'evoluzione nel tempo dell'inefficienza in molti modi diversi ha contribuito alla fioritura di parecchi lavori in questa area, anche se una grande distinzione di fondo era già chiara con i due contributi più datati, i lavori di Kumbhakar (1990) e di Cornwell, Schmidt and Sickles (1990). Per lavorare con approcci basati sulla massima verosimiglianza (è il caso di Kumbhakar, 1990) occorre specificare la distribuzione delle componenti del termine di errore, mentre per lavorare senza ipotesi a priori sulle distribuzioni (Cornwell et al., 1990) sono richiesti molti più parametri da stimare. Senza entrare nel dettaglio di tutti i modelli elencati, occorre sottolineare che nessun modello, tra quelli discussi finora, si preoccupa di filtrare dalla misura dell'inefficienza l'eterogeneità non osservata, eventualmente presente nei dati.

Recentemente, il lavoro di Greene (2005) ha arricchito la famiglia dei modelli panel ad inefficienza variabile nel tempo considerando, nel procedimento di stima

dell'inefficienza, il ruolo dell'eterogeneità non osservata. Sfruttando la potenza di calcolo dei moderni processori è infatti rivisitare i modelli ad effetti fissi (*Fixed Effect*, FE) e casuali (*Random Effect*, RE) depurando dalla misura dell'inefficienza il disturbo causato dall'eterogeneità non osservata.

Seguendo la ripartizione appena discussa, nel prossimo paragrafo esporremo con maggiori dettagli alcuni dei modelli menzionati nella Figura 4.1, in modo da poter argomentare, in chiusura di capitolo, la scelta dei modelli utilizzati in questa tesi.

4.4 Principali modelli econometrici in uso nella SFA

In questa sezione desideriamo presentare con maggiori dettagli le diverse famiglie di modelli di frontiera stocastica delineate in precedenza. L'esposizione degli aspetti econometrici principali dei modelli Cross-Section servirà come base di partenza per estendere l'analisi alle diverse ramificazioni della famiglia dei modelli Panel. Per questi ultimi modelli cercheremo di evidenziare tanto l'aspetto della variabilità nel tempo dell'inefficienza quanto il ruolo giocato dall'eterogeneità presente nei dati.

Prima di passare in rassegna le diverse tipologie di modelli panel è utile ricordare come la definizione e la specificazione dell'inefficienza (e quindi la sua misurazione) siano in parte il frutto di una scelta arbitraria del ricercatore. Questa considerazione, evidenziata chiaramente da Deprins and Simar (1989), ci ricorda che l'inefficienza si nasconde tra i residui della regressione, che sono la sola grandezza osservabile. La misura dell'inefficienza dipende dal modello utilizzato per stimare la frontiera di costo e dalla struttura che immaginiamo abbia l'inefficienza stessa.

La specificazione del modello riflette quindi alcune convinzioni del ricercatore sul fatto che in un determinato settore, ad esempio, l'inefficienza di un'impresa sia costante nel tempo o che alcuni fattori siano da considerare esogeni piuttosto che una caratteristica del processo produttivo o, ancora, che l'inefficienza si mescoli con l'eterogeneità causata da fattori non osservabili.

4.4.1 Modelli Cross-Section

Il primo modello di frontiera stocastica di produzione fu proposto in letteratura da Aigner et al (1977) e da Meeusen et al. (1977) in modo indipendente e nello stesso anno. Il modello proposto prevedeva una funzione di produzione logaritmica e un

termine di errore composto da due termini in grado di catturare una componente casuale dovuta alla natura stocastica della realtà osservata e una componente sistematica in grado di catturare l'inefficienza produttiva delle singole osservazioni.

Riprendendo l'esempio della log-log [4.9] è possibile adattare il modello di base per descrivere una frontiera stocastica di costo:

$$\text{Ln}C_i = \beta_0 + \beta_y \ln y + \sum_{j=1}^K \beta_j \ln w_{ji} + v_i + u_i \quad [4.12]$$

Il termine di errore (ε_i) presente nella [4.9] è la parte dei costi osservati che un modello deterministico non riesce a spiegare ed è stato ridefinito come la somma di due fattori: una componente stocastica (v_i) ed una componente sistematica (u_i). Il termine v_i è distribuito in modo simmetrico per catturare l'effetto del disturbo statistico o di fenomeni casuali, mentre il termine u_i assume solo valori non negativi ($u_i \geq 0$) per descrivere e considerare gli effetti dell'inefficienza di costo. Quando si stima una frontiera di costo, il termine di errore corrisponde ai residui della regressione che contengono entrambe le componenti originando una casistica variegata di situazioni, riassunta come segue:

- Il residuo è positivo ($\varepsilon_i > 0$), e l'osservazione è efficiente ($u_i = 0$);
- Il residuo è positivo ($\varepsilon_i > 0$), e l'osservazione è inefficiente ($u_i > 0$);
- Il residuo è nullo o negativo ($\varepsilon_i \leq 0$), e l'osservazione è efficiente ($u_i = 0$);
- Il residuo è nullo o negativo ($\varepsilon_i \leq 0$), e l'osservazione è inefficiente ($u_i > 0$).

Per misurare correttamente l'inefficienza (u_i) è quindi necessario riuscire a scomporre i residui della regressione (ε_i) nelle due componenti, ma questa operazione richiede la conoscenza (o la supposizione) delle funzioni di probabilità della componente stocastica (white noise) e del termine di inefficienza.

La frontiera stocastica di costo si può stimare con la tecnica della Massima Verosimiglianza (ML), ipotizzando le seguenti distribuzioni:

1. $v_i \sim \text{iid } N(0, \sigma_v^2)$;
2. $u_i \sim \text{iid } N^+(0, \sigma_u^2)$;
3. v_i e u_i sono distribuite in modo indipendente tra loro e rispetto ai regressori.

Questo modello, che viene definito dalla letteratura come *Normal-Half Normal*, ricalca la formulazione originariamente proposta da Aigner, Lovell and Schmidt (1977) che si può stimare empiricamente massimizzando la seguente funzione di verosimiglianza:¹²²

$$\ln L = K - N \ln \sigma + \sum_i \ln \Phi\left(\frac{\varepsilon_i \lambda}{\sigma}\right) - \frac{1}{2\sigma^2} \sum_i \varepsilon_i^2 \quad [4.13]$$

Dove:

- K è una costante,
- N è il numero di osservazioni,
- $\Phi(\bullet)$ è la funzione di distribuzione della normale standard,
- ε_i è il termine di errore (o residuo della regressione),
- $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$
- $\lambda = \sigma_u / \sigma_v$.

Oltre al vettore β dei parametri della funzione di costo, si determina una stima delle deviazioni standard (σ_u e σ_v) di entrambe le componenti del termine di errore. Questa informazione è espressa in modo sintetico dal parametro λ , che indica il contributo relativo delle componenti di inefficienza (u_i) e stocastiche (v_i) rispetto al termine di errore. Quando λ è molto vicino a 0, il contributo del termine di inefficienza (la parte asimmetrica del termine di errore) è molto piccolo e questo riconduce la stima di massima verosimiglianza ai risultati della stima dei minimi quadrati ordinari.¹²³

Un aspetto interessante della SFA è la possibilità di misurare i valori delle inefficienze di costo delle singole unità, utilizzando i risultati della regressione. Le ipotesi sulle distribuzioni delle componenti del termine di errore permettono infatti di ricavare un'informazione sulla componente di inefficienza (u_i) partendo dall'osservazione del residuo (ε_i) della regressione. In pratica, si applica la formula di

¹²² Per una derivazione formale della funzione di verosimiglianza si veda Kumbhakar and Lovell (2000).

¹²³ Coelli (1995) ha proposto un test statistico per verificare se l'ipotesi $\lambda = 0$ si può rifiutare. In questo caso anche in presenza di valori molto piccoli di λ , la struttura a due componenti del modello è giustificabile.

Jondrow et al (1982) che definisce, utilizzando la media della distribuzione di u condizionata dall'osservazione di ε , uno stimatore puntuale della componente di inefficienza (u_i) nel seguente modo:

$$E(u_i | \varepsilon_i) = \sigma_* \left[\frac{\phi(\varepsilon_i \lambda / \sigma)}{1 - \Phi(-\varepsilon_i \lambda / \sigma)} + \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right] \quad [4.14]$$

Dove:

- $\sigma_* = \sigma_u \sigma_v / \sigma$

Determinati i valori individuali della componente di inefficienza (u_i) è possibile applicare la [4.12] e calcolare l'inefficienza di costo di tutte le osservazioni. Da un punto di vista econometrico, la stima delle inefficienze di costo dipende dalle assunzioni fatte circa la distribuzione della componente u_i per la quale la letteratura ha proposto altre distribuzioni oltre a quella "half-normal".¹²⁴ La scelta della distribuzione della componente u_i solleva problemi di carattere econometrico che non intendiamo discutere in questa sede ponendo invece l'attenzione sul significato di inefficienza che si attribuisce alla componente asimmetrica dell'errore. Per questo motivo è indispensabile che la distribuzione di u_i possa assumere solamente valori non negativi, mentre la forma specifica della distribuzione passa in secondo piano dal momento che la vera distribuzione dell'inefficienza è comunque sconosciuta.

Nella valutazione dell'inefficienza di singole unità produttive è importante riuscire a considerare il ruolo di determinati fattori esogeni o di caratteristiche specifiche delle singole osservazioni che potrebbero influenzare la misura dell'inefficienza. Nel settore delle case per anziani, così come altri settori, il processo produttivo è influenzato da molti fattori che sono al di fuori del controllo dell'impresa. Il profilo sanitario degli ospiti, la localizzazione geografica, la forma istituzionale, le norme di regolamentazione, sono alcuni esempi di fattori che sono difficilmente controllabili dal management dell'impresa ma che possono incidere, in modo positivo o negativo, sulle performance di costo delle singole case per anziani. Il problema di

¹²⁴ In particolare, ricordiamo la distribuzione esponenziale proposta da Aigner et al. (1977) e da Meeusen and Van der Broeck (1977), la distribuzione normale troncata proposta da Stevenson (1980) e la distribuzione gamma proposta da Greene (1980a, 1980b).

considerare in modo adeguato i fattori esogeni si può in parte risolvere includendo nel modello un vettore di regressori che sia in grado di catturare l'effetto delle variabili ambientali e, conseguentemente, di depurare la componente di inefficienza (u_i) da questo disturbo. Da un punto di vista econometrico, introdurre un set di variabili per considerare i fattori esogeni, non comporta alcun cambiamento, anche se i valori stimati di u_i dovrebbero essere più precisi.¹²⁵

Un modo alternativo per considerare gli effetti di fattori esogeni sull'inefficienza di costo delle singole osservazioni è quello di individuare i fattori esogeni che potrebbero incidere sul livello di inefficienza e ricorrere ad una procedura di stima a due stadi. Nel primo stadio si ignorano i fattori esogeni e si procede ad una normale stima della frontiera stocastica in modo da ottenere una stima della componente di inefficienza (u_i). Nel secondo stadio, i punteggi di inefficienza vengono spiegati dai fattori esogeni attraverso una regressione.

La procedura solleva numerosi problemi dal punto di vista econometrico. Un primo punto delicato riguarda l'esclusione volontaria delle variabili ambientali nel primo stadio che se correlate con i regressori introdurrebbero una distorsione nelle stime dei parametri. Tuttavia, nel secondo stadio è desiderabile la correlazione tra queste variabili e la componente di inefficienza (u_i), che si cerca di spiegare. Questo approccio esprime una contraddizione intrinseca, dato che, in presenza di risultati molto significativi nella regressione del secondo stadio, la distorsione delle stime (dovuta all'omissione delle variabili ambientali nel primo stadio) sarebbe molto probabile.¹²⁶

Un certo scetticismo sulla correttezza metodologica del metodo a due stadi è riconosciuto in letteratura e sono state proposte alcune soluzioni interessanti. Kumbhakar et al. (1991) hanno specificato un modello di frontiera stocastica che incorpori i fattori esogeni direttamente nel termine di inefficienza (u_i):

$$u_i = \gamma' z_i + e_i \quad [4.15]$$

$$\text{con } u_i \sim N^+(\gamma' z_i, \sigma_u^2)$$

¹²⁵ Come per gli altri regressori, è necessario che le variabili ambientali non siano correlate con le due componenti (u_i e v_i) del termine di errore.

¹²⁶ Si veda ad esempio Deprins and Simar (1989).

Questa assunzione permette di cogliere le differenze sistematiche di inefficienza causate dai fattori esogeni (z_i) permettendo di utilizzare una procedura di stima in un singolo stadio utilizzando la tecnica di massima verosimiglianza.¹²⁷ Altri lavori hanno approfondito ed esteso l'idea di inserire direttamente i fattori esogeni come una componente del fattore u_i di inefficienza. Huang and Liu (1994) hanno considerato anche le interazioni tra i regressori e le variabili ambientali, mentre Simar et al. (1994) e Caudill et al. (1995) hanno considerato la possibilità che i fattori esogeni (z_i) influenzino anche la varianza della componente di inefficienza.

4.4.2 Modelli Panel-Data

I problemi relativi alla specificazione del termine di inefficienza e del ruolo dei fattori esogeni è ancora più interessante se si abbandona il terreno dei modelli *Cross-Section* per esplorare il mondo dei *Panel-Data*, caratterizzati da una struttura temporale più complessa. In presenza di dati panel, la specificazione generale di un modello di frontiera di costo viene leggermente modificata rispetto alla [4.12] per considerare la presenza di osservazioni ripetute nel tempo:

$$\ln C_{it} = \beta_0 + \beta_y \ln y_{it} + \sum_{j=1}^K \beta_j \ln w_{jit} + v_{it} + u_{it} \quad [4.16]$$

dove $i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$; $j = 1, 2, \dots, K$.

Anche in questo caso, il termine di errore (ε_{it}) è definito dalla somma di due fattori: una componente stocastica (v_{it}) ed una componente sistematica (u_{it}) che cattura l'inefficienza dell'unità osservata. La componente di inefficienza può essere una caratteristica che varia nel tempo o che resta costante, originando due famiglie distinte di modelli che ci accingiamo a presentare.

4.4.2.1 Modelli Panel con inefficienza costante nel tempo

Possiamo immaginare che il livello di inefficienza di ogni osservazione dipenda dalle caratteristiche della singola unità, ma non dal periodo temporale di osservazione. Ciò che determina l'inefficienza di ogni unità i è quindi costante nel tempo. In termini

¹²⁷ La funzione di verosimiglianza risulta una generalizzazione del modello di normale troncata, proposto da Stevenson (1980).

pratici questo significa sostituire il termine u_{it} dell'equazione [4.16] con il termine u_i che denota l'assenza di variabilità nel tempo dell'inefficienza.

Con questa specificazione del termine di errore è possibile adattare i classici modelli sviluppati dall'econometria dei panel-data (*Fixed Effect* e *Random-Effect*) al contesto delle frontiere stocastiche, come suggerito da Schmidt and Sickles (1984). Un approccio alternativo consiste nella stima di massima verosimiglianza estendendo il modello cross-section, come proposto da Pitt and Lee (1981).

Il modello *Fixed Effect* (FE), proposto da Schmidt and Sickles (1984), presuppone una struttura ad effetti fissi, riferita ad ogni osservazione i , che cattura il termine di inefficienza. La specificazione del modello, riferito al solito esempio log-log dell'espressione [4.9], risulta quindi:

$$\ln C_{it} = \alpha_i + \beta_y \ln y_{it} + \sum_{j=1}^K \beta_j \ln w_{jit} + v_{it} \quad [4.17]$$

dove $\alpha_i = \beta_0 + u_i$

Questo modello, che non richiede alcuna assunzione sulla distribuzione della componente u_i , ha il pregio di generare stime consistenti dei parametri. Nel caso di una stima lineare nei parametri, quando $N \rightarrow \infty$, lo stimatore $\hat{\beta}$ è in grado di stimare in modo consistente il vettore β sconosciuto.

Un secondo vantaggio di questa specificazione è che la proprietà della consistenza viene mantenuta anche in presenza di una correlazione tra gli effetti individuali u_i ed il termine casuale v_{it} o i regressori. Inoltre, la componente casuale non deve necessariamente seguire la distribuzione normale, ma si ipotizza una distribuzione iid $(0, \sigma_v^2)$ non correlata con i regressori della funzione di costo.

Il modello FE, che viene stimato da una regressione con N dummy per catturare gli effetti specifici di ogni singola unità, è noto come LSDV (Least Square Dummy Variables). La stima delle componenti individuali di inefficienza (u_i) si ottiene a partire dalle intercette specifiche $\hat{\alpha}_i$ con una procedura piuttosto semplice: si individua $\hat{\alpha}_i = \min\{\hat{\alpha}_i\}$ e si determinano le componenti individuali come $\hat{u}_i = \hat{\alpha}_i - \hat{\alpha}$. Per costruzione tutti i valori della componente di inefficienza sono definiti in modo tale che $\hat{u}_i \geq 0$. Utilizzando la [4.11] si possono calcolare i punteggi di inefficienza, che saranno espressi in modo relativo alla miglior performance presente nel campione.

Molto spesso nella ricerca empirica la numerosità Cross-Section del campione è sufficientemente grande (ad esempio il nostro database ha $N > 350$) da assicurare che le stime dei parametri β con il modello ad effetti fissi siano consistenti. Come detto questa proprietà non viene compromessa nel caso gli effetti individuali u_i fossero correlati con la componente casuale v_i ed i regressori.

A ben vedere in settori come quello delle case per anziani è facile ipotizzare che una certa eterogeneità (riferita alle singole osservazioni) non osservabile, ma correlata con i regressori venga catturata dalla componente individuale e rilevata come inefficienza. In questi casi è importante sapere che il modello ad effetti fissi permette di ottenere stime meno distorte rispetto ad altri modelli.

Il modello FE presenta, tuttavia, una seria limitazione che riguarda la dimensione temporale dei dati ed è noto in letteratura come “*incidental parameter problem*”. La stima degli effetti individuali α_i , che si ottiene con lo stimatore within sulla base delle T osservazioni di ogni i individuo, risulta inconsistente quando il numero di periodi T è limitato. Il problema della stima delle inefficienze individuali tende a svanire quando T cresce e scompare (Kumbahakar and Lovell, 2000) se $T \rightarrow \infty$, evento molto improbabile vista la normale dimensione temporale dei dati panel disponibili per la ricerca applicata.

Il numero di periodi T a disposizione (lunghezza del panel) non è un elemento sufficiente a garantire l’accuratezza delle stime, dal momento che la trasformazione delle variabili si basa sulla differenza dalla media temporale delle osservazioni. Anche un panel “lungo” (es. 10 anni) potrebbe dare problemi di stima qualora la variabilità all’interno dei gruppi (within variation) fosse molto limitata, poiché in questi casi l’elevato numero di periodi T sarebbe associato ad un basso valore informativo, replicando di fatto T volte lo stesso valore.¹²⁸

A conclusione di questi ragionamenti è utile sottolineare come nel modello ad effetti fissi sia importante chiarire se un panel è corto o lungo, ma soprattutto come la brevità e/o la scarsa variazione nel tempo dei regressori distorcano le stime delle componenti individuali.

¹²⁸ Il caso estremo è rappresentato da un panel con T molto grande ma con nessuna variabilità nel tempo. Dal punto di vista econometrico si tratta della medesima serie Cross-Section ripetuta per T volte. Gli aspetti problematici dello stimatore within groups sono evidenziati ad es. in Cameron and Trivedi (2005).

La capacità del modello FE di catturare in modo veritiero la componente di inefficienza delle singole osservazioni dipende anche dal tipo di eterogeneità presente nel campione. In presenza di caratteristiche individuali (non osservate) che restano costanti nel tempo la capacità del modello FE di rilevare con precisione l'inefficienza è compromessa poiché queste variabili, non possono essere separate dalla componente di inefficienza.

La probabilità di trovarsi in una simile situazione è piuttosto elevata nel settore delle case per anziani, dove esistono numerose differenze tra gli istituti che potrebbero influenzare i costi di produzione (posizionamento geografico, qualità delle cure offerte, età delle infrastrutture, forma giuridica e assetto proprietario, etc.) delle diverse imprese. Molte di queste caratteristiche, pur variando tra i diversi istituti, restano invariate nel tempo. L'effetto di tali variabili (osservate o meno) sui costi, non essendo spiegata dal modello, sarebbe catturato dagli effetti fissi $\hat{\alpha}_i$, determinando una stima della componente di inefficienza individuale \hat{u}_i distorta dalla sovrapposizione dell'eterogeneità. In altre parole, le differenze di costo, che si potrebbero spiegare considerando questa eterogeneità cross-section (ma time-invariant), si mescolano ai maggiori costi legati all'inefficienza vera e propria.

I punteggi di inefficienza sono costruiti, in modo relativo, per comparazione con la miglior performance $\hat{\alpha}_i$, il cui livello di costo potrebbe essere molto basso, sia per l'assenza (ipotizzata) di inefficienza, sia a causa di particolari caratteristiche che involontariamente il parametro $\hat{\alpha}_i$ cattura. Di conseguenza, il campo di variazione dei punteggi di inefficienza calcolati con il modello FE risulta spesso molto ampio e caratterizzato da un'inefficienza media piuttosto elevata.

In un contesto caratterizzato dalla presenza di eterogeneità non osservata, come il settore delle case per anziani, questo aspetto metodologico deve essere preso in debita considerazione, se si utilizza il modello FE con l'obiettivo di confrontare il livello dell'inefficienza di costo delle singole osservazioni. In conclusione, il modello ad effetti fissi, pur presentando indubbi vantaggi nella corretta stima dei parametri della regressione, presenta alcune serie limitazioni nel determinare con precisione il valore dei punteggi di inefficienza.

Un'alternativa interessante al modello FE è rappresentata da una specificazione che considera gli effetti individuali come casuali con media e varianza costanti. Questo modello, sviluppato da Schmidt and Sickles (1984), si definisce ad effetti casuali, o modello *Random Effect* (RE), e non attribuisce una distribuzione precisa alla componente individuale u_i , mentre ipotizza che il termine u_i non sia correlato né con il termine casuale v_{it} né con i regressori. Con riferimento all'esempio log-log dell'espressione [4.9], il modello RE risulta pertanto:

$$\begin{aligned} \ln C_{it} &= \alpha^* + \beta_y \ln y_{it} + \sum_{j=1}^K \beta_j \ln w_{jit} + v_{it} + u_i^* \\ \text{dove } \alpha^* &= \beta_0 + E(u_i), \\ u_i^* &= u_i - E(u_i) \quad \text{da cui } E(u_i^*) = 0 \end{aligned} \quad [4.18]$$

Questo modello si può stimare con una procedura in due stadi, nota con il termine di *feasible GLS* (*Generalized Least Square*), nella quale si ottengono le stime di tutti i parametri tramite l'OLS. Dopo aver determinato le due componenti della varianza dell'errore (utilizzando uno dei diversi metodi proposti dalla letteratura) si ri-stima il modello con il *feasible GLS* e si ottengono le stime finali dei parametri.

Una volta determinati empiricamente i valori dei parametri è possibile risalire alla stima delle componenti individuali di inefficienza \hat{u}_i tramite una serie di passaggi algebrici a partire dai residui della regressione. L'effetto che la componente casuale v_{it} esercita su ciascun individuo i ha un valore atteso pari a zero e quindi, se consideriamo un numero elevato (al crescere di T) di periodi tende a scomparire.

La media temporale di ciascun individuo i (*group means*) rappresenta quindi una possibile stima \hat{u}_i^* della componente u_i :

$$\hat{u}_i^* = \frac{1}{T} \sum_{t=1}^T \left[\ln C_{it} - \alpha^* - \hat{\beta}_y \ln y_{it} - \sum_{j=1}^K \hat{\beta}_j \ln w_{jit} \right] \quad [4.19]$$

Come nel caso del modello ad effetti fissi, si determinano le componenti individuali in modo relativo alla miglior performance presente nel campione, utilizzando $\hat{u}_i = \hat{u}_i^* - \min_i \{\hat{u}_i^*\}$. Tutti i valori della componente di inefficienza sono

definiti in modo tale che $\hat{u}_i \geq 0$ e, utilizzando la [4.11], si possono calcolare i punteggi di inefficienza nel modo usuale.

Rispetto al modello FE, il grande vantaggio del modello RE (stimato con la procedura GLS) è quello di consentire l'utilizzo di regressori costanti nel tempo. In questo modo l'effetto di caratteristiche osservate, specifiche degli individui, non viene confuso con l'effetto dell'inefficienza individuale.¹²⁹ Questa proprietà è certamente desiderabile, ma si ottiene al costo di assumere l'indipendenza degli effetti individuali dai regressori. Se questa ipotesi, per la quale Hausman e Taylor (1981) hanno sviluppato un apposito test statistico, non fosse rispettata, la stima GLS produrrebbe valori distorti per i parametri del modello.

Nel settore delle case per anziani l'eterogeneità dell'output è in parte dovuta alla presenza di caratteristiche specifiche degli istituti (variabili ambientali) che non variano nel tempo. Il fatto di poter includere nel modello anche variabili ambientali, permette di filtrare dall'inefficienza almeno la parte di eterogeneità osservabile (costante nel tempo), rendendo la specificazione ad effetti casuali più interessante rispetto al modello ad effetti fissi.

Se le caratteristiche dell'output o l'eterogeneità del settore suggeriscono questi argomenti come giustificazione del modello RE, occorre tenere presente che la contropartita riguarda la consistenza delle stime dei parametri della frontiera che potrebbero risultare distorti a causa della violazione delle ipotesi necessarie per la stima GLS utilizzata da questo modello. È infatti possibile che, non potendo osservare ed includere nel modello tutte le caratteristiche time-invariant, una parte dell'eterogeneità finisca per sovrapporsi alla componente u_i , indebolendo l'ipotesi di non correlazione tra gli effetti individuali ed i regressori richiesta dal modello ad effetti casuali.

In sostanza il grande vantaggio dei modelli FE e RE consiste nel non imporre a priori alcuna distribuzione alla componente individuale di inefficienza e riuscire, tuttavia, a determinare empiricamente una misura delle singole inefficienze. Inoltre, la plausibilità di assumere l'indipendenza tra i regressori e la componente u_i (per poter

¹²⁹ Da notare che l'effetto di caratteristiche individuali costanti nel tempo ma non osservabili (e quindi escluse dal modello) continua ad essere catturato dai residui, mescolandosi con la componente individuale di inefficienza delle singole osservazioni.

utilizzare il RE) si può verificare empiricamente con il test di Hausman-Taylor che suggerisce l'una o l'altra specificazione.

La semplicità econometrica di questi modelli è un punto di forza innegabile così come l'assenza di imposizioni sulla distribuzione della componente di inefficienza che è uno degli elementi al centro dell'interesse dell'analisi delle frontiere stocastiche. Un approccio alternativo ai metodi LSDV e GLS è rappresentato dalle stime di massima verosimiglianza che, però, richiedono la conoscenza (o la scelta discrezionale) della distribuzione della componente u_i , senza la quale è impossibile ricavare la funzione di verosimiglianza.

La stima di un modello ad effetti casuali si può ottenere anche utilizzando la tecnica di massima verosimiglianza. Questo approccio è stato proposto da Pitt and Lee (1981), che estendono il modello di Aigner et al. (1977) al contesto dei panel data. Le assunzioni di base sono identiche ipotizzando una distribuzione delle due componenti del termine di errore v_{it} e u_i rispettivamente normal e half-normal. La componente di inefficienza è considerata costante nel tempo e quindi la funzione di densità di $u \geq 0$ risulta indipendente dal tempo ed assume la forma:

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \quad [4.20]$$

La componente stocastica v_{it} del termine di errore dipende dal tempo e la corrispondente funzione di densità di $v = (v_1, \dots, v_T)'$ risulta:

$$f(v) = \frac{1}{2\pi^{T/2}\sigma_v^T} \exp\left\{-\frac{v'v}{2\sigma_v^2}\right\} \quad [4.21]$$

Le equazioni [4.20] e [4.21], unitamente alle ipotesi di indipendenza delle due componenti del termine di errore, permettono di derivare la funzione (il logaritmo della) di verosimiglianza da massimizzare rispetto ai parametri dei regressori della frontiera di costo ed alle deviazioni standard delle due componenti del termine di errore.¹³⁰

¹³⁰ Per una presentazione completa dei passaggi che portano all'espressione della funzione di verosimiglianza si veda Kumbhakar and Lovell (2000).

Come per il modello Cross-Section è possibile ricavare una stima dei valori individuali di inefficienza partendo dai residui della regressione, utilizzando la formula di Jondrow et al. (1982) che nel caso in esame risulta:

$$E(u_i | \varepsilon_i) = \mu_{*i} + \sigma_* \left[\frac{\phi(-\mu_{*i} / \sigma_*)}{1 - \Phi(-\mu_{*i} / \sigma_*)} \right] \quad [4.22]$$

Dove:

- $\mu_{*i} = \frac{T\sigma_u^2 \bar{\varepsilon}_i}{\sigma_v^2 + T\sigma_u^2} \quad \text{con } \bar{\varepsilon}_i = \frac{1}{T} \sum_t \varepsilon_{it}$
- $\sigma_*^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + T\sigma_u^2}$

Un potenziale vantaggio derivante dall'uso della stima di massima verosimiglianza è quello di ottenere stime più accurate dei parametri della frontiera stocastica, a patto che non siano violate le assunzioni fatte circa: (1) le distribuzioni delle componenti di errore, (2) la loro reciproca indipendenza e (3) la non correlazione con i regressori. Il modello RE-ML è stato utilizzato molto in letteratura e una parte del suo successo è sicuramente spiegata dalla facilità con la quale si possono ricavare i punteggi di inefficienza grazie alla formula di Jondrow et al. (1982) che si può costruire in tutti i pacchetti software econometrici con alcuni comandi di algebra lineare.¹³¹

Come per il modello Cross-Section, anche nel caso dei Panel-data sono state considerate diverse alternative alla distribuzione half-normal per la componente individuale di inefficienza. Dal punto di vista metodologico non cambia nulla se non la derivazione della funzione di verosimiglianza da massimizzare e la relativa distribuzione condizionale u_i/ε_{it} che si utilizza per stimare i valori di inefficienza.

In conclusione, i modelli panel che considerano l'inefficienza costante nel tempo si dividono nei modelli che considerano gli effetti individuali come fissi (FE-LSDV) o

¹³¹ Le versioni più recenti di alcuni software econometrici (es. Limdep 8.0 e Stata 10) prevedono una routine di calcolo per i valori condizionali di u_i data l'osservazione dei residui ε_{it} della regressione.

come casuali. La famiglia di questi ultimi si divide ulteriormente, a seconda dell'approccio econometrico scelto per la stima, in RE-GLS e RE-ML.

L'ipotesi che nel corso del tempo l'inefficienza di ogni singola impresa resti costante merita un'adeguata argomentazione economica, soprattutto se si considerano periodi di tempo abbastanza lunghi. In un contesto di competizione (anche se imperfetta) è logico immaginare che tutte le imprese cercheranno di correggere quei comportamenti che hanno generato inefficienze nel passato.

La scelta di utilizzare un modello ad inefficienza costante nel tempo significa ipotizzare, almeno nel periodo osservato, che le imprese non riescano a far fronte a decisioni errate o a ridurre eventuali sprechi. Nel settore delle case per anziani, caratterizzato da molti elementi aleatori e da lenti ma costanti cambiamenti, appare difficile escludere a priori che le imprese restino immobili, ciascuna con il proprio problema, senza riconoscerlo e/o risolverlo.

Per superare questo limite, e in alternativa ai modelli presentati, la letteratura ha elaborato approcci simili dal punto di vista econometrico, ma che considerano l'inefficienza come una grandezza libera di variare sia rispetto agli individui sia nei confronti del tempo. L'elemento di novità di questi modelli consiste, dunque, nel trattare l'inefficienza come una grandezza non strutturale (e quindi fissa nel tempo) che ad ogni periodo si manifesta in modo diverso anche se riferita al medesimo individuo.

4.4.2.2 *Modelli Panel con inefficienza variabile nel tempo*

La disponibilità di osservazioni ripetute nel tempo riferite alla stessa impresa, permette di cogliere (in modo parziale) la natura dinamica dell'attività economica. Il fatto di ipotizzare che ogni impresa esprima un proprio livello di inefficienza, che riflette le diverse decisioni, situazioni, fattori contingenti, è in linea con il buon senso. La natura dinamica degli eventi, unita alla continua modifica del contesto competitivo, rende molto difficile la piena ottimizzazione delle risorse e giustificano la presenza di inefficienze che, secondo questo ragionamento, potrebbero assumere livelli diversi alla fine di ogni periodo. Pertanto, questi modelli assumono che l'inefficienza possa variare sia rispetto agli individui sia rispetto ai periodi di tempo presenti nel campione di dati.

In questa sezione analizzeremo più in dettaglio alcuni modelli ad inefficienza variabile nel tempo iniziando con il lavoro di Cornwell Schmidt and Sickles (1990).

La specificazione del modello segue la strategia di stima già analizzata per i modelli FE e RE con inefficienza costante nel tempo, ma in ogni periodo t di osservazione esiste un valore comune a tutti gli individui (α_i) che si somma alla componente individuale di inefficienza (u_{it}) di quel periodo determinando un'intercetta che dipende tanto dagli individui quanto dal periodo di osservazione. Dal punto di vista empirico il modello richiede, in aggiunta ai parametri β e σ_v , la stima di $N \times T$ intercette, cosa impossibile visto che i parametri superano i dati a disposizione.

Per risolvere il problema gli autori hanno espresso l'intercetta α_{it} in funzione del tempo secondo la formulazione seguente:

$$\alpha_{it} = \theta_{1i} + \theta_{2i}t + \theta_{3i}t^2 \quad [4.23]$$

In questo modo si riducono i parametri relativi alle intercette a $3 \times N$, numero comunque molto elevato e che non permette di stimare il modello se T non è almeno pari a 4 periodi.

Questa specificazione permette di stimare il modello utilizzando sia l'approccio ad effetti fissi sia quello ad effetti casuali a dipendenza delle restrizioni imposte ai parametri θ . Il modello degenera al caso con inefficienza costante nel tempo se θ_{2i} e $\theta_{3i} = 0 \forall i$. Se ipotizziamo $\theta_{2i} = \theta_2$ e $\theta_{3i} = \theta_3$ il modello, invece, si trasforma in un FE dal momento che si aggiunge un termine temporale comune ($\theta_2 t + \theta_3 t^2$) alla componente θ_{1i} specifica per individuo. Una possibile lettura di questa struttura del termine di inefficienza è che ogni individuo presenta la propria inefficienza specifica, mentre l'evoluzione temporale segue un percorso identico per tutte le N osservazioni. In alternativa, si può considerare il termine temporale come la modifica della frontiera dovuta dal progresso, che lascia tuttavia inalterata (e quindi costante e pari a θ_{1i}) l'inefficienza delle singole osservazioni.

Le strategie di stima proposte cambiano leggermente a seconda della specificazione ad effetti fissi o casuali, e si basano su stime a più stadi per riuscire a determinare tutti i parametri θ relativi alle componenti individuali.¹³² Anche nel caso dei modelli ad inefficienza variabile nel tempo la specificazione FE non permette di

¹³² Maggiori dettagli sulle strategie di stima dei modelli FE e RE si veda Kumbhakar and Lovell (2000).

includere tra i regressori, variabili caratteristiche costanti nel tempo. Questo problema si può superare con l'impiego della specificazione RE ad effetti casuali, ma in questo caso il problema si sposta sulla consistenza delle stime GLS che potrebbero risultare distorte se l'inefficienza di costo è correlata con i regressori.¹³³

Lee and Schmidt (1993) hanno proposto una versione alternativa nella quale gli effetti individuali di inefficienza in ogni periodo di tempo (u_{it}) sono definiti dal prodotto tra gli effetti individuali (u_i) ed i parametri di un set di dummy relative ad ogni periodo di tempo. Questo modello, che si può stimare utilizzando un approccio ad effetti fissi o ad effetti casuali, è stato ripreso da Ahn et al. (1994) utilizzando la tecnica del metodo dei momenti generalizzato.

Partendo dal modello generale descritto dalla [4.16], Kumbahakar (1990) ha espresso la componente di inefficienza u_{it} secondo la seguente parametrizzazione temporale:

$$u_{it} = [1 + \exp(\gamma t + \delta t^2)]^{-1} \cdot u_i \quad [4.24]$$

Il modello si può stimare con la tecnica di massima verosimiglianza in modo analogo al modello di Pitt and Lee (1981) con effetti casuali e inefficienza costante nel tempo. L'aggiunta dei parametri γ e δ permette, tuttavia, di modellare (monotonicamente) la pendenza e la concavità della traiettoria evolutiva dell'inefficienza dei singoli individui. In altre parole, il valore di questi due parametri aggiuntivi determina l'andamento temporale dell'inefficienza che può solo aumentare o solo diminuire ad un ritmo crescente o decrescente.

Questo modello è molto parsimonioso nei parametri, ma, come contropartita, introduce una certa rigidità nel comportamento temporale dell'inefficienza di ogni singolo individuo. La [4.24] impedisce ad esempio che nel corso del tempo l'inefficienza diminuisca per un certo periodo per poi aumentare in seguito.

¹³³ Gli autori, riconoscendo il problema, hanno proposto uno stimatore con variabili strumentali che produce stime consistenti.

Seguendo l'approccio di Kumbhakar (1990), Battese and Coelli (1992) propongono una specificazione alternativa per la componente di inefficienza u_{it} che viene definita come una funzione esponenziale del tempo:

$$u_{it} = \exp[-\eta(t - T)] \cdot u_i \quad [4.25]$$

Il modello richiede la stima di un solo parametro aggiuntivo, ma impone all'inefficienza una struttura evolutiva ancora più rigida: a dipendenza del segno di η l'inefficienza continua a crescere ($\eta < 0$), a decrescere ($\eta > 0$) o resta costante ($\eta = 0$).

Da un punto di vista economico, imporre una dinamica temporale identica a tutte le osservazioni significa che i ranghi di inefficienza tra le diverse imprese, rilevati nel primo periodo di osservazione, si mantengono inalterati in tutti i periodi successivi. La variabilità dell'inefficienza nel tempo è quindi riferita alla sua misurazione in senso assoluto, mentre dal punto di vista relativo non cambiano i rapporti tra le diverse imprese.

Questa mancanza di flessibilità presenta molte affinità con un concetto di inefficienza persistente, poiché causata da fenomeni radicati nelle singole imprese, mentre appare in contrasto con l'idea che l'inefficienza si "rinnovi" nel tempo in modo imprevedibile, frutto di nuovi shock, di errori o di situazioni particolari. Il modello di Cornwell et al. (1990) è certamente più adatto per rappresentare questo tipo di inefficienza, ma richiede la stima di un numero elevatissimo di parametri ed è virtualmente impraticabile con panel di lunghezza non estesa (4-7 anni), ossia una buona parte dei panel che sono concretamente disponibili per la ricerca empirica.

La letteratura sui panel data nella SFA ha indagato numerosi altri aspetti come il problema dell'eteroschedasticità o l'importanza del progresso tecnologico, particolarmente significativo nei panel lunghi.

Ad esempio, nel modello di Battese and Coelli (1995) si propone un modello ad inefficienza variabile nel tempo che segue la distribuzione normale troncata $u_{it} \sim N^+(\mu_{it}, \sigma_{it}^2)$ nella quale si ipotizza che $\mu_{it} = \gamma' z_{it}$ e $\sigma_{it}^2 = \exp(\gamma' z_{it})$. In questo modo le caratteristiche del fattore di inefficienza (u_{it}) sono determinate dal vettore z di fattori esogeni e il modello può essere stimato simultaneamente con la tecnica della massima

verosimiglianza.¹³⁴ Il vantaggio di questo modello è quello di considerare (almeno in parte) l'influenza che l'eterogeneità esercita sulle stime dell'inefficienza. Il modello di Battese and Coelli (1995), tuttavia, non risolve completamente il problema, ignorando i possibili effetti causati dall'eterogeneità non osservata. Per questo motivo nella prossima sezione presenteremo alcuni modelli (recenti) che cercano di considerare in modo adeguato il fenomeno dell'eterogeneità non osservata.

4.4.2.3 *Modelli Panel che considerano l'eterogeneità non osservata*

La panoramica sui principali modelli panel-data ha evidenziato come i diversi approcci siano in grado di descrivere e misurare l'inefficienza di costo. La stima dell'inefficienza ottenuta con i modelli analizzati in precedenza non è sempre precisa, dal momento che si mescola con le differenze di costo causate da fattori esogeni. Nel modello ad effetti fissi, dove questo problema è particolarmente marcato, le stime di inefficienza incorporano quelle differenze di costo che si potrebbero giustificare con fattori esogeni costanti nel tempo, ma non osservati.¹³⁵

Il problema è parzialmente risolto dal modello ad effetti casuali che permette di inserire tra i regressori eventuali fattori esogeni costanti nel tempo. Tuttavia, il modello RE, che ipotizza l'inefficienza costante nel tempo, non è in grado di filtrare quella parte di eterogeneità non osservata che non varia nel tempo. Anche i modelli che non considerano costante nel tempo l'inefficienza, ipotesi poco realistica in alcuni settori e con panel molto lunghi, non sono in grado di separare in modo soddisfacente l'eterogeneità non osservata dall'inefficienza.

Una soluzione concettualmente semplice, ma empiricamente impraticabile, è quella di specificare in modo più dettagliato il modello e disporre di tutte le variabili che sono in grado di spiegare le differenze di costo tra le osservazioni. Purtroppo, a volte, anche quando è facile individuare un fattore esplicativo, la misurazione delle variabili (es. qualità delle cure, motivazione del personale, qualità delle dotazioni tecniche) non

¹³⁴ Per stimare empiricamente, sulla base del modello di Battese and Coelli (1995), i parametri di una frontiera stocastica e i punteggi di inefficienza è disponibile un apposito software gratuito sviluppato da Tim Coelli e denominato Frontier.

¹³⁵ La specificazione econometrica del modello FE impedisce, infatti, di avere regressori che, pur variando da individuo a individuo, non variano nel tempo.

si presta ad essere quantificata in modo soddisfacente. In altri casi, le informazioni che si cercano non sono disponibili a causa della mancanza dei dati statistici.

Questa banale considerazione trova la sua dignità nel sollevare con forza il problema dell'eterogeneità non osservata che, piaccia o meno, è sempre presente nelle analisi empiriche. Per questo motivo una possibile soluzione è quella di considerare questa presenza nella specificazione econometrica in modo da riuscire ad isolare l'effetto che interessa e cioè l'inefficienza.

L'effetto dell'eterogeneità sulle stime di inefficienza è un problema che affligge i modelli ad effetti fissi e che può essere mitigato (ma non completamente risolto) da una specificazione ad effetti casuali che sia in grado di catturare, oltre all'eterogeneità osservata con le variabili del modello, la parte di eterogeneità non osservabile costante nel tempo. Questo modello viene definito da Greene (2005) come “*True Random Effect*” (TREM), la cui specificazione, riadattata secondo la [3.16], risulta:

$$\text{Ln}C_{it} = \hat{\beta}_0 + \hat{\beta}_y \ln y_{it} + \sum_{j=1}^K \hat{\beta}_j \ln w_{jit} + \omega_i + v_{it} + u_{it} \quad [4.26]$$

L'elemento di novità è rappresentato dal termine costante nel tempo ω_i , che costituisce l'effetto casuale specifico di ogni individuo. Questo parametro, che non deve essere correlato con tutti gli altri elementi del modello, si prefigge di catturare l'eventuale presenza di eterogeneità (costante nel tempo) tra individui. A prima vista, il termine di errore del modello [4.26] sembra composto da tre elementi (ω_i , v_{it} e u_{it}), il che pone un problema di identificazione.

Se classifichiamo il disturbo secondo la tipologia delle fonti di provenienza, possiamo evidenziare, da una parte, l'eterogeneità intesa come differenza nelle caratteristiche specifiche dei diversi individui e, dall'altra, un termine di errore ε_{it} che racchiude sia la componente sistematica di inefficienza (u_{it}), sia il disturbo casuale (v_{it}). Definito in questo modo il modello risulta un normale RE con il termine di errore variabile nel tempo ($\varepsilon_{it} = v_{it} + u_{it}$) che segue una distribuzione asimmetrica, data dalla somma di una normale (v_{it}) e di una mezza-normale (u_{it}).

Per ogni individuo i , la funzione di densità del termine di errore ε_{it} (condizionata da ω_i) risulta:

$$f(\varepsilon_{i1}, \dots, \varepsilon_{iT} | \omega_i) = \prod_{t=1}^T \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon_{it}}{\sigma}\right) \cdot \Phi\left(\frac{\varepsilon_{it}\lambda}{\sigma}\right) \quad [4.27]$$

Per definire una funzione di verosimiglianza che si possa massimizzare nei parametri di interesse ($\alpha, \beta, \sigma, \lambda$) è necessario integrare fuori dalla [4.27] il parametro ω_i che cattura l'eterogeneità, per ottenere l'espressione della densità incondizionata del termine ε_{it} :

$$f(\varepsilon_{i1}, \dots, \varepsilon_{iT}) = \int_{\omega_i} \prod_{t=1}^T \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon_{it}}{\sigma}\right) \cdot \Phi\left(\frac{\varepsilon_{it}\lambda}{\sigma}\right) g(\omega_i) d\omega_i \quad [4.28]$$

Purtroppo, l'integrale della [4.28] non ha una forma chiusa ed è quindi impossibile da risolvere analiticamente. Interpretando l'integrale come valore atteso rispetto a ω_i , è possibile sostituire all'integrale analitico una versione simulata. La simulazione consiste nel generare casualmente una serie di valori ω_i basandosi sulla distribuzione (reale o presunta di ω_i). Se il numero totale di ripetizioni R è sufficientemente grande, è possibile arrivare ad una buona approssimazione dell'integrale che permette di stimare tutti i parametri del modello.

La (log)verosimiglianza simulata risulta quindi:

$$\text{Log } L_s(\alpha, \beta, \sigma, \lambda, \theta) = \sum_{i=1}^N \log \frac{1}{R} \sum_{r=1}^R \left[\prod_{t=1}^T \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon_{it} | \omega_{ir}}{\sigma}\right) \cdot \Phi\left(\frac{\varepsilon_{it}\lambda | \omega_{ir}}{\sigma}\right) \right] \quad [4.29]$$

Per poter generare casualmente un valore ω_{ir} in ognuna delle R ripetizioni è necessario attribuire a ω_i una distribuzione specifica dalla quale estrarre i valori. L'utilizzo di una distribuzione normale con deviazione standard pari a θ , permette di ottenere una ragionevole approssimazione della vera funzione di (log)verosimiglianza a patto di generare un numero elevatissimo di estrazioni casuali. La procedura si può accelerare utilizzando la sequenza di Halton che permette di coprire in modo uniforme

tutto il dominio dell'integrale con un numero inferiore (da cinque a dieci volte) di ripetizioni.¹³⁶

Il comportamento del modello "True" RE nei confronti della misurazione dell'inefficienza di costo potrebbe discostarsi in modo anche marcato rispetto ad altri modelli ad effetti casuali come il modello RE di Pitt and Lee (1981) analizzato in precedenza. Per verificare questa possibilità Greene (2005) ha in effetti comparato le stime di inefficienza generate dal modello "True" RE con altre specificazioni ad effetti casuali e con modelli Cross-Section.

Lo studio, basato su un campione di banche commerciali americane nel periodo 1996-2000, ha evidenziato come le differenze di comportamento tra i vari modelli siano più grandi di quanto le statistiche descrittive non lascino capire e, soprattutto, come sia difficile trovare una sistematica a tali differenze. La bassa correlazione che si rileva tra le stime di inefficienza generate con i vari modelli, è un segnale chiaro di come la scelta della specificazione econometrica possa incidere sui risultati.

Un punto delicato dei modelli RE è rappresentato dalla possibile violazione dell'ipotesi che il termine di inefficienza non sia correlato con i regressori. Questo problema, noto come "*Heterogeneity bias*", si riferisce al fatto che una parte dell'eterogeneità non osservabile, correlata con i regressori, è catturata dal termine di inefficienza, introducendo una distorsione nelle stime.

Come discusso in precedenza, il modello ad effetti fissi (FEM) risente in misura minima di questo problema, motivo per il quale è ampiamente usato in letteratura. Per l'analisi delle frontiere stocastiche, tuttavia, la forma tradizionale (regressione lineare LSDV) del modello FEM non riesce a separare l'eterogeneità dall'inefficienza.

Per risolvere (almeno in parte) il problema, Greene (2005) ha proposto un modello denominato "True Fixed Effect Model" (TFEM) e specificato nel modo seguente:

$$\ln C_{it} = \hat{\beta}_0 + \hat{\beta}_y \ln y_{it} + \sum_{j=1}^K \hat{\beta}_j \ln w_{jit} + \alpha_i + v_{it} + u_{it} \quad [4.30]$$

¹³⁶ Invece di utilizzare la tecnica Monte Carlo con estrazioni (pseudo)casuali, il metodo di Halton si basa su una scelta selettiva dei valori (tecnica Quasi Monte Carlo) che, pur non essendo casuale, garantisce una convergenza molto più rapida verso il valore atteso dell'integrale simulato.

Apparentemente la [4.30] corrisponde ad una specificazione tradizionale di frontiera stocastica, con la differenza che si tratta di un modello non lineare a causa della specificazione della funzione di densità del termine di errore ($v_{it} + u_{it}$).

Il modello [4.30] implica, oltre alla stima dei parametri strutturali (β , σ e λ), la stima simultanea di una grande quantità di parametri accessori (*nuisance parameters*), contemplati dalla verosimiglianza condizionata. La funzione di verosimiglianza si può massimizzare in modo algoritmico (“*brute force*”) con il metodo di Newton, ricavando poi i parametri della funzione in modo iterativo.¹³⁷

Dal punto di vista pratico, si pone un problema che riguarda la potenza di calcolo richiesta per ottenere le stime nel caso di modelli molto ricchi di parametri e panel molto estesi. Sfruttando le proprietà delle matrici coinvolte nella procedura di iterazione è possibile semplificare il calcolo che rientra, anche in presenza di panel molto grandi (dell’ordine di grandezza di 100’000 osservazioni), tranquillamente nelle capacità esecutive di un comune PC domestico.

Questo modello, così come il modello ad effetti fissi tradizionale, si rivela tuttavia poco affidabile quando i dati hanno una dimensione temporale limitata a pochi periodi (nel nostro caso 5 anni) e la variabilità temporale (*within variation*) dei regressori è contenuta. Inoltre, la procedura di stima richiede che si arrivi ad una convergenza dell’iterazione, cosa che non sempre accade.¹³⁸

Il vero problema del modello “*True*” FE, come sottolineato dallo stesso Greene (2005), riguarda la rilevanza empirica di un problema econometrico, noto in letteratura con il nome di “*incidental parameters problem*”. Le considerazioni fatte da Greene (2005) evidenziano come la distorsione sui parametri della regressione sia potenzialmente lieve. Tuttavia, il problema si manifesta nella stima dei parametri σ e λ che determinano i valori dell’inefficienza, causandone una sovrastima. Per le ragioni

¹³⁷ Per i dettagli sulle funzioni di verosimiglianza e sulla struttura dell’iterazione si veda il paper di Greene (2002a).

¹³⁸ Il problema della convergenza nelle procedure iterative è un elemento importante da considerare, poiché permette, in caso di convergenza, di corroborare i risultati delle stime. In caso di mancata convergenza dell’algoritmo (entro il numero di iterazioni prefissate) non sempre si ottengono delle stime e, quando ciò accade, occorre cautela nell’interpretazione dei risultati.

discusse, e visto l'interesse a misurare in modo accurato l'inefficienza di costo, il modello TFEM sarà escluso nella parte empirica.

Per correggere il problema del “*heterogeneity bias*”, Farsi, Filippini e Kuenzle (2005) hanno proposto alcuni modelli basati sulla formulazione RE di Mundlak (1978). La specificazione degli effetti specifici di ogni individuo, riadattata alla notazione della [4.17], risulta:

$$\alpha_i = \alpha_0 + \gamma_y \bar{y}_i + \gamma_j \bar{w}_{ji} + \delta_i \quad [4.31]$$

dove:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T [\beta_y \ln y_{it}] \quad e \quad \bar{w}_{ji} = \frac{1}{T} \sum_{t=1}^T [\beta_j \ln w_{jit}] \quad \forall j$$

$$\delta_i \sim iid N(0, \sigma_\delta^2)$$

In questa specificazione l'eterogeneità non osservata (α_i) è una funzione delle medie temporali (*within groups* o *group means*) delle variabili esplicative (y_{it} e w_{it}) del modello di costo. Il parametro γ cattura l'effetto della correlazione tra eterogeneità ed i regressori, mentre la componente δ_i esprime l'inefficienza, costante nel tempo, ipotizzata ortogonale ai regressori. La correzione del *Heterogeneity Bias* può essere applicata sia al modello con effetti casuali TREM (proposto da Greene, 2005), sia al modello tradizionale (RE). Con riferimento a quest'ultimo modello, l'espressione estesa della specificazione “Mundlak”, ottenuta aggiungendo al modello [4.18] il regressore \bar{y}_i (per l'unico output) ed il set di regressori \bar{w}_{ji} (per i K input), risulta quindi:

$$\ln C_{it} = \beta_0 + \beta_y \ln y_{it} + \sum_{j=1}^K \beta_j \ln w_{jit} + \gamma_y \bar{y}_i + \sum_{j=1}^K \gamma_j \bar{w}_{ji} + \delta_i + \varepsilon_{it} \quad [4.32]$$

Il modello [4.32] si può stimare con la procedura GLS e presenta il vantaggio di corrispondere allo stimatore *within* del modello ad effetti fissi, risultando quindi non distorto.¹³⁹ Anche il modello TREM, stimato impiegando la massima verosimiglianza (simulata), si può correggere con la specificazione “Mundlak”, semplicemente esprimendo il termine di errore (ε_{it}) nei due elementi che lo compongono: parte stocastica (u_{it}) e parte di inefficienza (v_{it}). Rispetto al modello RE, tuttavia, per il

¹³⁹ Una prova della corrispondenza tra lo stimatore RE “Mundlak” e lo stimatore FE si trova in Hsiao (2003)

modello TREM la correzione del problema della “*Heterogeneity bias*” non è completa.¹⁴⁰

Nell’applicazione al settore delle case per anziani in Ticino (un cantone della Svizzera) gli autori confrontano le stime generate da sei diverse specificazioni econometriche, evidenziando marcate differenze tra i diversi modelli, tanto nelle statistiche descrittive, quanto nelle correlazioni dell’inefficienza.¹⁴¹ Le stesse considerazioni si possono trovare, ad esempio, in Farsi, Filippini e Kuenzle (2005) e Farsi, Filippini e Greene (2005) che hanno applicato alcuni tra i modelli di frontiera stocastica descritti in precedenza a dati svizzeri riferiti rispettivamente al settore ferroviario e al settore dei trasporti locali (bus). Gli autori dimostrano che la correlazione tra i punteggi di inefficienza di costo misurata con le diverse metodologie è, in generale, moderata. Appare significativo che questa regolarità empirica della bassa correlazione tra i diversi modelli, sia un risultato che si ripete in modo analogo, nonostante le grandi differenze che esistono nei settori considerati.

I modelli proposti da Greene (2005), denominati “True” FE e “True” RE, sono indubbiamente interessanti poiché riescono a catturare (attraverso i parametri α_i e ω_i) quella parte di eterogeneità non osservata che resta costante nel tempo e che finirebbe inevitabilmente mischiata alle stime di inefficienza. Le ipotesi di base di questi modelli, tuttavia, non prevedono la possibilità che l’inefficienza sia costante nel tempo ed attribuiscono la sua eventuale presenza all’effetto di caratteristiche non osservate. Per questo motivo è possibile che, se da un lato i modelli FE e RE tradizionali tendano a sovrastimare l’inefficienza, dall’altro, i modelli “True” FE e RE tendano a sottostimare l’inefficienza. La conclusione che si potrebbe trarre, applicando il buon senso, è che la verità stia a metà strada o, meglio, che l’effettivo valore dell’inefficienza non sia espresso da nessuno dei due modelli che, però, identificano con buona precisione il campo di variazione della reale misura di inefficienza.

¹⁴⁰ La stima di massima verosimiglianza si basa infatti sull’ipotesi di asimmetria del termine di errore che rende il modello non lineare. Di conseguenza l’interpretazione di stimatore *within* viene a cadere.

¹⁴¹ In controtendenza all’andamento generale, la correlazione (nei ranghi) tra le specificazioni RE-GLS e “True” RE risulta molto elevata sia per i modelli “originali” (superiore a 0,98) sia per la versione “Mundlak” (superiore a 0,95).

La discussione sulla scelta dell'approccio econometrico e le sue conseguenze è un tema importante che, sebbene già parzialmente indagato durante la descrizione dei singoli modelli, trova uno spazio di approfondimento, volto a sottolineare pregi e difetti dei diversi approcci, nella prossima sezione.

4.4.3 Pro e contro nella scelta del modello econometrico

Nell'ambito delle frontiere stocastiche (SFA), la letteratura ha sviluppato una gamma molto vasta di modelli econometrici per dare risposta ai diversi problemi che contraddistinguono l'analisi empirica dell'efficienza produttiva. Alcuni problemi riguardano soprattutto la sfera econometrica (le ipotesi sulle distribuzioni delle componenti di errore o sulla correlazione di questi ultimi con i regressori), mentre altri sono di natura più economica (ad esempio la distinzione tra eterogeneità ed efficienza o la variabilità di quest'ultima nel corso del tempo).

I lavori di economia applicata, pur guardando con interesse agli sviluppi metodologici più recenti, sono sempre confrontati con il problema concreto di scegliere l'approccio econometrico adatto agli scopi della ricerca. Pertanto, nel processo di selezione del modello econometrico troviamo almeno tre dimensioni distinte che concorrono a determinare le scelte del ricercatore: (1) gli obiettivi specifici dell'analisi, (2) le caratteristiche del settore considerato e (3) il tipo di banca dati a disposizione. I tre elementi citati si possono combinare in molti modi, rendendo impossibile qualsiasi confronto astratto tra i modelli presentati in precedenza. Per questo motivo è importante analizzare il contesto di ricerca nel quale si opera e presentare una comparazione critica dei modelli econometrici discussi, in relazione agli obiettivi del lavoro, alle caratteristiche del settore delle case per anziani in Svizzera ed alla specificità della banca dati disponibile.

La prima constatazione è relativa ai dati che, contenendo informazioni sulle case per anziani svizzere per gli anni 1998-2002, presentano una struttura panel, permettendo, sia dal punto di vista econometrico che da quello economico, di specificare modelli più articolati. Poiché i dati lo consentono, sembra preferibile adottare una metodologia come quella dei Panel-data, in grado di sfruttare le informazioni sull'andamento temporale delle variabili. Nel nostro caso, tuttavia, la

durata del panel è limitata a soli 5 anni, riducendo in parte il vantaggio offerto dai modelli Panel rispetto alle specificazioni “*Pooled*” Cross-Section.¹⁴²

Se, comunque, concentriamo le nostre preferenze ai modelli panel, emerge con forza una seconda considerazione, riguardo alla scelta dell’approccio ad effetti fissi o ad effetti casuali. Hsiao (2003) sottolinea come la scelta, che in qualche modo riflette alcune considerazioni arbitrarie del ricercatore, sia irrilevante quando il numero di periodi temporali T è molto grande.¹⁴³ In presenza di un panel di 5 anni (breve), invece, diventa molto importante l’entità della variazione “*within*” (intra gruppo), poiché modifica l’importanza relativa di pregi e difetti dei due diversi approcci econometrici. Un limitato numero di periodi e/o una bassa variazione temporale dei principali regressori, accentuano i problemi connessi con l’uso degli effetti fissi.¹⁴⁴

Per indagare la natura e l’importanza delle distorsioni provocate da un panel corto (e comunque con T fisso), Greene (2005) ha condotto un’analisi Monte Carlo su dati provenienti da 500 banche commerciali americane nel periodo 1996-2000.¹⁴⁵ I risultati principali evidenziano che nel caso delle frontiere stocastiche, al contrario dei modelli binari, la distorsione sui parametri della regressione è lieve anche nel caso di panel brevi. Tuttavia, il problema si manifesta nella stima dei parametri (λ e σ) che determinano i valori dell’inefficienza. L’errore di stima introdotto dall’“*incidental parameters problem*”, risulta in media circa il 25% ad indicare una chiara sovrastima dei punteggi di inefficienza.

Se i problemi sollevati dalle caratteristiche dei dati a disposizione, che rappresentano un vincolo insormontabile, indicano una ragionevole preferenza per una specificazione (Panel) ad effetti casuali, all’interno di questa famiglia rimangono numerosi nodi da sciogliere. Primo fra tutti riguarda le diverse ipotesi previste dal

¹⁴² Baltagi and Griffin (1984), evidenziano come gli stimatori within e between deteriorino le loro performance in modo marcato quando la dimensione temporale del panel scende sotto i 7 anni.

¹⁴³ Hsiao (2003), fa notare come l’“*incidental parameter problem*”, si manifesti anche nell’inferenza dei parametri strutturali (β) del modello che sono condizionati dalla stima del parametro incidentale (α).

¹⁴⁴ Su questo punto Cameron and Trivedi (2005) sono chiari nell’indicare (p. 715) due debolezze pratiche del modello FE che: (1) non può includere regressori *time invariant*, (2) anche le stime dei parametri strutturali (β) possono risultare molto imprecise se la variabilità temporale è bassa.

¹⁴⁵ La banca dati, proveniente da un database curato dalla Federal Reserve Bank di Chicago, consisteva di 5'000 unità per 5 anni. Un campione casuale di 500 banche è stato impiegato per stimare una frontiera stocastica di costo, con una forma funzionale Cobb-Douglas.

modello in merito alla distribuzione della componente di inefficienza. Con riferimento a tutti quei modelli che richiedono una scelta specifica per la distribuzione del termine di inefficienza, si può evidenziare come i risultati dell'inefficienza siano sensibili alla distribuzione utilizzata.

Il problema della scelta della distribuzione dell'inefficienza, ampiamente riconosciuto in letteratura, è senza soluzione dal momento che il ricercatore non può conoscere la vera distribuzione dell'inefficienza. Nonostante tutti i modelli siano plausibili a priori, la stima dell'inefficienza ottenuta con distribuzioni diverse potrebbe evidenziare forti differenze, soprattutto per quanto riguarda le prestazioni relative (ranking) delle imprese.

Se l'interesse del ricercatore è focalizzato alla definizione di una graduatoria, ad esempio nel contesto di un'analisi di *benchmarking*, diventa indispensabile verificare la stabilità dei risultati rispetto alle diverse distribuzioni del termine di inefficienza.¹⁴⁶ Nel nostro caso, l'analisi dei punteggi di inefficienza delle singole case per anziani non scende nel dettaglio delle singole prestazioni e, di conseguenza, è meno interessata dal problema della scelta di una distribuzione specifica per la componente di inefficienza.

Da sottolineare, invece, il diverso modo con il quale i punteggi di inefficienza si determinano nei modelli RE stimati con la massima verosimiglianza, attraverso la formula di Jondrow et al. (1982), rispetto ai modelli RE basati sulla regressione GLS. Questi ultimi, non ipotizzando alcuna distribuzione per la componente individuale, ricavano le stime di inefficienza da una trasformazione algebrica dei residui senza imporre ulteriori condizioni. D'altra parte, i modelli GLS contrappongono al pregio della maggior flessibilità, il difetto di esasperare i valori estremi del campo di variazione dell'inefficienza, specialmente se la lunghezza del Panel è limitata.

Il contesto empirico nel quale si svolge la nostra analisi non fornisce elementi decisivi di sostegno per uno o per l'altro approccio, anche se la durata di soli 5 anni del periodo di osservazione potrebbe disturbare maggiormente i punteggi di inefficienza

¹⁴⁶ Anche le stime dei parametri del modello di costo sono influenzate dalla scelta della distribuzione della componente di inefficienza. Tuttavia, l'evidenza empirica dimostra come le stime dei coefficienti siano molto stabili, spostando l'importanza del problema sulla stima dei punteggi di inefficienza.

ricavati dall'approccio di stima GLS rispetto a quello basato sulla massima verosimiglianza.

L'aspetto che forse merita una discussione più estesa e contestuale al settore delle case per anziani è quello della separazione dell'inefficienza dall'eterogeneità. La difficoltà pratica di individuare uno strumento econometrico in grado di depurare le stime di inefficienza dalle differenze individuali tra le imprese è preceduta dal problema di specificare in modo corretto la natura stessa dell'inefficienza.

L'inefficienza di costo è una grandezza molto semplice da definire a livello concettuale. Essa misura il rapporto tra i costi effettivamente sostenuti da un'impresa e quelli che, secondo il miglior utilizzo delle risorse e della tecnologia disponibile, un'ipotetica unità produttiva efficiente avrebbe realizzato. In altre parole, si tratta di misurare l'eccesso di costo rispetto ad un termine di paragone, rappresentato, nel caso della SFA, dalla frontiera di costo stocastica.

Ancora più semplice è immaginare che, in un contesto competitivo, ogni singola impresa si sforzi di eliminare o di ridurre al minimo ogni costo superfluo e quindi ogni forma di inefficienza. Ciò nonostante, l'evidenza empirica (sostenuta dal buon senso) indica l'esistenza di forti differenze nei costi di imprese con caratteristiche simili (ad esempio le case per anziani), segnalando la probabile presenza di forme di inefficienza.

La vera complicazione risiede, dunque, nel comprendere quali possono essere le cause che generano l'inefficienza di costo nel settore delle case per anziani. Come già osservato in precedenza, la convinzione che i servizi residenziali per anziani siano erogati in un mercato perfettamente concorrenziale escluderebbe la rilevanza empirica di un'analisi di inefficienza, poiché tutto ciò che non riusciremmo a spiegare sarebbe attribuito al disturbo statistico.

Il mercato delle case per anziani in Svizzera non sembra presentare tutte le caratteristiche del modello neoclassico, soprattutto per la presenza di monopoli naturali locali nelle zone poco popolate, per l'eterogeneità della domanda e dell'offerta e per la forte regolamentazione che caratterizza il settore frammentandolo di fatto in un insieme di mercati regionali.

Abbandonando l'ipotesi di perfetta competizione in questo settore ed accettando come strutturali le caratteristiche menzionate, si producono di colpo due effetti: la possibile (anche se indesiderata) presenza di inefficienza e l'importanza di non

confondere quest'ultima con l'eterogeneità che caratterizza il settore. Stabilito che dal punto di vista teorico il settore delle case per anziani possa esprimere un elevato grado di eterogeneità ed un possibile livello di inefficienza, il passo successivo è quello di capire come queste due dimensioni si comportino rispetto al passare del tempo.

Alcune caratteristiche che differenziano tra loro le case per anziani sono permanenti o comunque relativamente costanti nel tempo (es. posizione geografica, il numero di letti, la forma istituzionale, regolamentazioni cantonali e comunali, etc.), mentre altre sono soggette a variazioni temporali più o meno ampie (es. profilo sanitario degli ospiti, composizione del personale, dotazioni tecniche e apparecchiature mediche, ventaglio di servizi offerti all'interno della struttura, etc.).

Dal punto di vista empirico, in effetti, quando parliamo di eterogeneità dobbiamo precisare se ci riferiamo a quelle caratteristiche che possiamo osservare dai dati o se, invece, si tratta di differenze ragionevolmente presumibili ma non direttamente osservabili. La distinzione tra eterogeneità osservata e non osservata è rilevante dal momento che, mentre quella osservata si può parametrizzare all'interno del modello con opportune variabili, quella non osservabile viene catturata dai residui della regressione.¹⁴⁷

Nei modelli econometrici analizzati non si arriva mai ad una separazione completa dell'eterogeneità non osservata dall'inefficienza dal momento che entrambe possono avere una componente costante nel tempo ed una variabile, per un totale di 4 fattori. Le quattro componenti, che si sommano al disturbo statistico, finiscono tutte nei residui della regressione, rendendo impossibile di fatto una completa separazione dei termini.

Ad esempio, nel "True" RE, dove l'inefficienza è variabile nel tempo (u_{it}) e l'eterogeneità inosservata è costante (ω_i) nel tempo, la presenza di un eventuale parte di inefficienza costante nel tempo (strutturale) non verrebbe rilevata come tale, nascondendosi sotto forma di eterogeneità. Per contro in questo stesso modello la misura dell'inefficienza potrebbe includere anche quella componente di eterogeneità dovuta a cause contingenti e quindi variabili nel tempo.

¹⁴⁷ Su questo punto si veda ad esempio Jacobs et al. (2006) che, oltre a discutere gli aspetti teorici, propongono un'analisi di sensitività applicata ad un panel di 740 osservazioni composte da 185 ospedali acuti inglesi nel periodo 1994-1997.

Appare chiaro come l'adeguatezza del modello econometrico dipenda in buona parte dalle caratteristiche del settore in questione e delle informazioni statistiche in possesso del ricercatore. Nel settore delle case per anziani, le differenze che riguardano il lato della domanda sono sostanzialmente esogene e misurabili con sufficiente precisione.¹⁴⁸ Il rilevamento dei bisogni dell'utente (pratica obbligatoria negli istituti elvetici) permette quindi di avere informazioni sul profilo dell'utenza nelle diverse case per anziani osservate, riuscendo a controllare una parte dell'eterogeneità presente nel campione. Per contro, molte dimensioni del processo produttivo (legate prevalentemente al lato dell'offerta) non si possono osservare nei dati statistici. La maggior parte di queste differenze nelle caratteristiche degli istituti, che non si possono osservare nei dati (eterogeneità non osservata), è di tipo strutturale nel senso che una modifica in breve tempo di questi fattori è molto difficile, se non impossibile. Questo, però, non significa che le differenze tra le case per anziani siano sempre esogene e che non si modifichino su un periodo di tempo sufficientemente lungo.

Alcune decisioni (es. ampliamento della struttura, modifica dell'organizzazione o dell'assetto istituzionale, livello di qualità nelle cure erogate, etc.) rientrano sicuramente nella sfera di influenza del management dell'impresa, che può decidere di intervenire. Altri elementi sono quasi certamente al di fuori del controllo dell'impresa (regolamentazione, densità della domanda, case-mix degli ospiti, prezzi di alcuni input, etc.), ma anche questi fattori possono evolvere. Analogamente possiamo discutere se l'inefficienza sia in prevalenza un fenomeno specifico dell'impresa (e quindi costante nel tempo) o se, invece, rifletta anche elementi che variano nel tempo, modificando il valore dell'inefficienza da periodo a periodo. In questo settore possiamo trovare alcuni validi argomenti a sostegno dell'ipotesi che, su un arco di tempo di 5 anni, l'inefficienza delle singole case per anziani sia in prevalenza costante nel tempo.

La spiegazione principale si riferisce alla lentezza con la quale avvengono i cambiamenti in questo settore e al basso livello di competizione presente nel mercato. In tale situazione è lecito attendersi l'insorgere di comportamenti opportunistici che potrebbero tradursi in forme di inefficienza strutturale. Come suggerito da Eraclito (*Panta rei*), tutto scorre (e cambia) anche ciò che a prima vista, come alcune

¹⁴⁸ In Svizzera le case per anziani, a causa delle regole di gestione delle liste d'attesa, sono fortemente limitate nella "scelta dei clienti" (cream skimming), legittimando l'ipotesi di esogenità dell'output.

caratteristiche delle case per anziani, sembra immutabile. Il periodo di tempo considerato dall'analisi (1998-2002) è certamente breve, ma caratterizzato da una certa trasformazione sia spontanea, per rispondere ai cambiamenti del livello e delle caratteristiche della domanda, sia indotta dalle numerose modifiche nella regolamentazione che il settore ha conosciuto a partire dagli anni '90 (in particolare dal 1994 con l'entrata in vigore della LAMal, la legge federale sull'assicurazione malattia).

A questi elementi potremmo unire la convinzione che il processo produttivo di tutte le imprese sia costellato (soprattutto nel settore dei servizi sociosanitari) da una serie di eventi non sistematici, ma comunque imputabili alla gestione di impresa e non al caso. La figura di inefficienza che si compone è piuttosto quella di un elemento che varia certamente da impresa a impresa, ma anche nel tempo. In conclusione, per analizzare l'efficienza produttiva nel settore delle case per anziani in Svizzera nell'arco del periodo 1998-2002, è difficile poter scegliere un unico approccio econometrico. A ben vedere vi sono buone argomentazioni a sostegno dei diversi modelli e, come detto, i diversi modelli potrebbero generare risultati molto diversi, sia per quanto concerne i parametri della funzione di costo, sia per l'analisi dell'inefficienza.

Nella Tabella 4.2, desideriamo presentare un quadro sintetico delle caratteristiche dei 5 modelli che utilizzeremo per stimare una frontiera stocastica di costo nel settore delle case per anziani in Svizzera. Il confronto delle diverse caratteristiche dei cinque modelli econometrici impiegati nella parte empirica, legittima a nostro avviso la decisione di non scegliere un singolo approccio, consentendoci di verificare la stabilità dei risultati delle stime ottenute con i 5 modelli e di ricavarne alcune indicazioni critiche.

Ricordiamo, infatti, che uno degli obiettivi di questa tesi, sebbene non il principale, è la discussione di quale sia la metodologia appropriata per la stima di una frontiera di costo nel settore delle case per anziani. Procedendo alla stima dei parametri e dell'inefficienza con più modelli, simili nell'approccio ma diversi nelle caratteristiche econometriche, sarà possibile arricchire la presente discussione, finora condotta sul piano teorico, con alcuni elementi suggeriti dall'evidenza empirica per questo settore.

Dopo la discussione delle tecniche econometriche da impiegare, rivolgiamo la nostra attenzione ai problemi legati alla specificazione del modello di costo, che saranno discussi nel prossimo capitolo.

Tabella 4.2 - Riepilogo dei modelli econometrici da considerare per le stime.

Modelli con inefficienza costante nel tempo			
Autore e Sigla	Specificazione dell'errore tecnica di stima	Componente individuale di inefficienza	Osservazioni
Schmidt and Sickles (1984) FEM	$\varepsilon_{it} = U_i + V_{it}$ LSDV	$\hat{u}_i = \hat{\alpha}_i - \min_i \{\hat{\alpha}_i\}$	problemi se T è ridotto e/o se within variation contenuta Non richiede ipotesi sulla distribuzione di u_i Stime di inefficienza inattendibili
Schmidt and Sickles (1984) REM	$\varepsilon_{it} = U_i + V_{it}$ Feasible GLS	$\hat{u}_i = \hat{u}_i^* - \min_i \{\hat{u}_i^*\}$ $\hat{u}_i^* = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}$	Non richiede ipotesi sulla distribuzione di u_i Si richiedono indipendenza reciproca tra u_i , v_{it} ed i regressori Sovrastima dell'inefficienza (Heterogeneity bias)
Pitt and Lee (1981) ML-REM	$\varepsilon_{it} = U_i + V_{it}$ ML $u_i \sim iid N^+(0, \sigma_u^2)$	$\hat{u}_i = E[u_i \hat{\varepsilon}_{it}]$	Stime di inefficienza sensibili alla distribuzione di u_i Sovrastima dell'inefficienza (Heterogeneity bias)
Modelli con inefficienza variabile nel tempo			
Autore e Sigla	Specificazione dell'errore tecnica di stima	Componente individuale di inefficienza	Osservazioni
Aigner, Lovell and Schmidt (1977) POOL	$\varepsilon_{it} = u_{it} + V_{it}$ ML estimation $u_{it} \sim iid N^+(0, \sigma_u^2)$	$\hat{u}_{it} = E[u_{it} \hat{\varepsilon}_{it}]$	Non sfrutta le informazioni Panel (problemi di specificazione) Meno sensibile al problema dell'Heterogeneity bias nei punteggi di inefficienza
Greene (2005) TREM	$e_{it} = \omega_i + u_{it} + v_{it}$ $= \omega_i + \varepsilon_{it}$ ML estimation (simulata) $u_{it} \sim iid N^+(0, \sigma_u^2)$ $\omega_i \sim iid N(0, \sigma_\omega^2)$	$\hat{u}_{it} = E[u_{it} \hat{\varepsilon}_{it}]$	Permette una separazione parziale tra inefficienza ed eterogeneità Sottostima dell'inefficienza

“As soon as you can say what you think, and not what some other person has thought for you, you are on your way to being a remarkable man.

(James Matthew Barrie, Scrittore)

5 Definizione del modello di costo per il settore delle case per anziani in Svizzera

Le premesse fatte fino a questo punto ci permettono di rivolgere, con il giusto spirito ed un adeguato strumentario, la nostra attenzione sul tema centrale di questo lavoro: l'analisi dell'efficienza produttiva nel settore delle case per anziani in Svizzera. Il capitolo 4 ha fornito tutti i chiarimenti metodologici per specificare e stimare una frontiera stocastica di costo ed interpretarne i risultati. A questo punto occorre definire un modello di costo che rappresenti in modo adeguato il processo produttivo delle case per anziani e che permetta di confrontare i dati provenienti dagli istituti di tutta la Svizzera. Per inquadrare meglio le problematiche empiriche connesse alla misurazione dell'inefficienza produttiva, il capitolo si apre con una rassegna di alcuni lavori sulla stima di una funzione di costo per le case per anziani. In seguito, illustreremo le difficoltà teoriche e pratiche che normalmente si riscontrano nel definire un modello di costo rappresentativo per il settore delle case per anziani, concludendo il capitolo con una descrizione di dettaglio della specificazione che abbiamo utilizzato per le stime.

L'interesse delle scienze economiche verso l'analisi empirica della struttura dei costi nel settore delle cure agli anziani ha origini lontane ed ha prodotto una vasta letteratura, focalizzata su molteplici aspetti della produzione ed erogazione dei servizi residenziali agli anziani. In questo ambito, a partire dagli anni '80, si è sviluppato e

consolidato un filone di ricerca applicata, basato sulle tecniche statistico-econometriche. L'obiettivo comune a questa parte di letteratura è la spiegazione del processo produttivo di una casa per anziani generica e la stima econometrica dei parametri di un modello di costo, con lo scopo di verificare empiricamente la validità di alcune ipotesi specifiche.

In ambito sanitario (ospedali e case per anziani), la scelta della forma funzionale e delle variabili esplicative da utilizzare nella regressione hanno dato luogo ad un interessante dibattito sui possibili approcci da seguire nella specificazione del modello di costo. Il problema nasce dalla difficoltà di considerare in modo soddisfacente la natura multiprodotto dei servizi ospedalieri e di assistenza agli anziani con un numero gestibile di variabili e rimanendo coerenti con la teoria microeconomica della produzione.

La letteratura ha evidenziato uno spettro di possibili approcci i cui estremi opposti sono rappresentati, da una parte, dalle specificazioni *ad hoc* e, dall'altra, dalle funzioni di costo neoclassiche. La funzione di costo *ad hoc*, largamente impiegata negli studi sui costi del settore sanitario fino all'inizio degli anni '80, mira a spiegare la variazione statistica osservata in alcuni indicatori molto significativi (costo per paziente, costo per giornata, ecc.) tramite una regressione multipla. Per questo motivo, nel lato destro dell'equazione vengono incluse in modo piuttosto acritico quelle variabili, per le quali si ipotizza un nesso di causalità e per le quali si abbiano informazioni disponibili.¹⁴⁹ Questo approccio, non rispettando le caratteristiche economiche prescritte per una funzione di costo, è legittimato a scegliere forme funzionali molto semplici che permettono di accomodare un grande numero di regressori, descrivendo in modo ricco il processo "processo produttivo".

Il prezzo da pagare è, tuttavia, la parziale incoerenza con la teoria economica (neoclassica) della produzione. Per questo motivo alcuni autori insistono sull'importanza di una specificazione teoricamente fondata, che preveda come variabile dipendente il costo (totale o variabile) e come regressori gli output ed il vettore dei prezzi degli input. A partire dai lavori di Bishop (1980), Conrad and Strauss (1983) e Cowing and Holtmann (1983), si forma un consenso intorno all'importanza di specificare in modo strutturato la funzione di costo, accettando le conseguenze

¹⁴⁹ Per un'esemplificazione di alcuni regressori impiegati in questi studi si veda Breyer (1987).

(maggiore complessità della forma funzionale e aumento del numero di parametri da stimare) in termini econometrici.

Cercando di avvicinare i due estremi, costituiti dalle funzioni di costo *ad hoc* e da quelle neoclassiche, alcuni autori come Grannemann et al. (1986) propongono un approccio, denominando dagli autori “*hybrid cost function*”, che permette di arricchire la specificazione neoclassica (senza comprometterne le proprietà) con variabili esplicative “ambientali”, che descrivono cioè alcune caratteristiche del processo produttivo, diverse da output e prezzi degli input. La specificazione di un modello di costo “misto” o “neoclassico in senso ampio” ha trovato molti seguaci che, riconoscendo il bisogno di un compromesso tra i due approcci estremi, hanno utilizzato diverse modalità e terminologie per descrivere la funzione di costo impiegata.¹⁵⁰

I tre approcci discussi (ad hoc, neoclassico e neoclassico in senso ampio) non esauriscono il dibattito sulla specificazione del modello di costo che deve anche precisare il tipo di obiettivo che si attribuisce alle imprese osservate. In questo senso è possibile distinguere chiaramente l’approccio neoclassico tradizionale, che presuppone la piena minimizzazione dei costi (di breve o di lungo periodo), dall’approccio “*behavioural*”.¹⁵¹ Questo secondo approccio considera in modo più pragmatico l’attività produttiva dell’impresa, osservando come spesso non si verificano le condizioni, postulate dalla teoria, per garantire una piena ottimizzazione delle risorse. Le imprese, quindi, vuoi per vincoli regolatori, vuoi per decisioni opportunistiche del management o per caratteristiche strutturali del mercato, potrebbero non seguire con rigore lo scopo della minimizzazione dei costi.

Come sottolineato da Bös (1986), anche in settori dove la minimizzazione dei costi potrebbe non essere rispettata, si può comunque utilizzare l’approccio neoclassico per determinare empiricamente le relazioni tra i costi e gli elementi strutturali della funzione di costo. Tuttavia, la relazione tra le variabili non sarà di tipo esclusivamente tecnico (funzione di costo neoclassica), ma piuttosto di tipo comportamentale (funzione

¹⁵⁰ La stessa terminologia è ripresa, per le case per anziani, da Nyman (1988), Dor (1989) e Christensen (2004). Vita (1990) invece, stimando i costi nel settore ospedaliero, denomina l’approccio seguito come “*Eclectic cost function*”. Nell’ambito dei trasporti la terminologia è ancora diversa con Spady and Friedlaender (1978) che propongono una funzione di costo, definita dagli autori edonica (*Hedonic Cost Function*), in grado di considerare gli attributi qualitativi in una specificazione neoclassica.

¹⁵¹ In ambito sanitario il primo ad adottare questo termine per una funzione di costo è stato Evans (1971).

di costo neoclassica *behavioural*). Questa osservazione è molto importante poiché permette di lavorare con l'impianto della teoria microeconomica della produzione, pur in contesti come quello delle case per anziani con una forte presenza di imprese ed istituzioni che potrebbero non rispettare pienamente i postulati della teoria neoclassica della produzione.¹⁵²

In ambito sanitario, i lavori empirici basati sulla definizione e stima di funzioni di costo hanno evidenziato molti aspetti, che in letteratura sono controversi e lunghi dall'essere risolti. Alcuni di questi dibattiti sulla corretta specificazione del modello riguardano: la scelta della forma funzionale, la specificazione dei costi (totali o variabili), l'inclusione e la misurazione degli aspetti qualitativi, l'endogenità delle variabili.¹⁵³ La discussione di questi aspetti sarà svolta al termine della rassegna di studi sul settore delle case per anziani, mentre è utile anticipare la precisazione concettuale e terminologica sul tipo di modello di costo (neoclassico, *ad hoc* o strutturale in senso ampio). È quindi giunto il momento di analizzare alcuni studi sulla struttura dei costi delle case per anziani che, per analogia o per differenza, aiutano a comprendere le scelte fatte nella parte empirica della tesi.

5.1 Rassegna della letteratura sulla struttura dei costi nel settore delle case per anziani

La stima di una funzione di produzione o di costo è lo strumento di base attraverso il quale tutti gli autori hanno perseguito gli obiettivi delle rispettive ricerche empiriche. Alcuni studi si sono concentrati maggiormente sulla misurazione delle economie di scala, di scopo e dell'inefficienza, mentre altri lavori hanno posto l'accento sul ruolo della qualità nella funzione di costo. Un altro tema di indagine proposto dalla letteratura riguarda l'impatto sull'efficienza di costo delle varie forme istituzionali presenti nel settore delle case per anziani (pubbliche, private non profit, private for profit).

¹⁵² Per una discussione sulla specificazione "*behavioural*" al settore delle case per anziani si veda ad esempio Farsi and Filippini (2004).

¹⁵³ Per una discussione generale sui problemi di specificazione del modello di costo in ambito ospedaliero si veda Smet (2002).

La variabilità che si registra negli obiettivi delle analisi, si ritrova anche nella scelta del tipo di modello (funzioni strutturali, ibride e ad hoc), nella definizione del processo produttivo (single o multiple output) e nella lista delle variabili. Altre differenze riguardano la definizione dell'orizzonte temporale, che si riflette inevitabilmente sull'impiego di funzioni di costo di breve o di lungo periodo.

A dispetto di una forte eterogeneità presente in letteratura, due elementi sono facilmente comparabili negli studi di costo per il settore delle case per anziani: (1) la specificazione delle variabili esplicative del modello e (2) i risultati sulle economie di scala. A seguire presentiamo una selezione di articoli, che ripercorrono in ordine cronologico alcune applicazioni relative al settore delle case per anziani, mettendo in evidenza gli obiettivi di ricerca, le scelte metodologiche effettuate, i dati a disposizione ed i principali risultati conseguiti.

McKay (1988) stima una funzione di costo totale per un campione di 82 case per anziani attive in Texas nel 1983. I dati provengono dal dipartimento della sanità e si riferiscono ad un campione stratificato rispetto alla dimensione degli istituti. Partendo dal totale di 679 case per anziani l'obiettivo del paper è quello di determinare le economie di scala, considerando nella specificazione del modello gli effetti della qualità delle cure. Per ridurre il problema dell'eterogeneità vengono considerate nell'analisi solamente le case per anziani for profit che offrono esclusivamente cure intermedie.¹⁵⁴

Per la stima empirica viene specificata, in due varianti alternative, una funzione di costo totale translog per considerare l'impatto della qualità. Il modello include le giornate di residenza come misura dell'output ed i prezzi di 4 inputs: infermieri, aiutanti, infrastrutture e dotazioni tecniche.¹⁵⁵ La specificazione "neoclassica" di partenza non riesce a cogliere le differenze di costo dovute a diversi livelli qualitativi, per catturare le quali si introduce una variabile "qualità" nel modello. Non disponendo di valutazioni specifiche sulla qualità delle cure, la misura proposta è il rapporto tra il

¹⁵⁴ Le 1001 case per anziani censite in Texas nel 1983 sono classificate, a dipendenza del livello di cure erogate, in: case medicalizzate (skilled) e case intermedie (intermediate). L'80% circa degli istituti offre esclusivamente cure intermedie e di questo sottoinsieme la maggior parte (679 case per anziani) persegue scopi di lucro (for profit).

¹⁵⁵ I prezzi degli input di ogni istituto sono calcolati come il rapporto tra le spese imputate a tale categoria diviso una misura della quantità del fattore considerato. Il personale è misurato in ore effettive, mentre per il capitale si utilizzano i m² di superficie dell'istituto. Gli equipaggiamenti sono considerati un elemento proporzionale al numero di giornate di residenza che viene usato come proxy.

numero di ore (infermieristiche) di cura per giornata, argomentando la scelta sulla scorta di alcuni studi sulle determinanti degli outcome sanitari del processo di cura.¹⁵⁶ Il modello viene leggermente riformulato, considerando le “ore infermieristiche” alla stregua di un fattore quasi-fisso, determinato dal livello dell’output e della qualità. I parametri delle funzioni di costo (con e senza correzione per la qualità) sono stimati con la tecnica del sistema di equazioni e la procedura SUR (Seemingly Unrelated Regression), proposta da Zellner (1962), che garantisce una maggior efficienza (sicuramente auspicata visto il limitato numero di gradi di libertà).

I risultati empirici forniscono indicazioni in controtendenza rispetto agli studi dell’epoca, che non rilevavano la presenza di importanti economie di scala. Alla media dei dati (25'000 giornate in una casa per anziani da 100 letti circa), entrambe le specificazioni rilevano la presenza di deboli economie di scala. Il paper sottolinea l’importanza del rigore metodologico (economico ed empirico) e traccia due implicazioni di politica sanitaria. La prima è che in Texas il tasso di occupazione è relativamente basso (80%), ma questo non è dovuto alla dimensione media degli istituti, quanto piuttosto al loro numero elevato. La dimensione ottimale di una casa è stimata in circa 133 letti (contro una dimensione media di 100 letti). La seconda, invece, indica che il risparmio complessivo, teoricamente ottenibile sfruttando al meglio la capacità produttiva, raggiunge dimensioni non trascurabili rispetto alle risorse assorbite dal settore.

Nyman (1988), riferendosi ai dati delle case per anziani dello stato di New York per l’anno 1983, stima una funzione di costo con lo scopo di determinare il livello dei costi marginali per diverse tipologie di utenti. L’obiettivo del paper è quello di smentire empiricamente l’ipotesi che il fenomeno del cream-skimming, nei confronti degli ospiti “Medicaid”, sia causato da un tasso di rimborso inferiore ai costi marginali di produzione.¹⁵⁷

¹⁵⁶ Il riferimento è allo studio di Linn et al. (1977) dove si evidenzia che la quantità di ore di cura per paziente incide positivamente sugli esiti delle cure.

¹⁵⁷ Medicaid è un programma pubblico (a livello statale) di sostegno finanziario per i bassi redditi, che si sostituisce al cliente della casa per anziani, riconoscendo un rimborso di tipo cost-plus. Per questa ragione gli istituti sono interessati ad escludere questa tipologia di clienti se, al margine, il rimborso non copre le spese di una giornata di cura.

L'autore sostiene che, in un contesto di eccesso di domanda indotto dalla regolamentazione CON (Certificate of Need), la selezione dei pazienti sarebbe presente anche se il rimborso per i clienti "Medicaid" fosse superiore ai costi.¹⁵⁸ L'impresa che massimizza i profitti con un vincolo di capacità produttiva, seleziona i clienti sulla base dei margini di profitto associati alle diverse tipologie (in questo caso "Medicaid" e "Privati") di residenti.

Il dataset utilizzato è composto da 419 case per anziani (sulle circa 600 attive nel 1983 nello stato di New York) suddivise in: (1) istituti specializzati nella ricezione di clienti con un profilo sanitario severo (definiti SNF "Skilled Nursing Facility"), (2) istituti che trattano solo clienti con profilo sanitario intermedio (definiti INF "Intermediate Nursing Facility") e (3) istituti che accettano entrambe le tipologie di residenti.¹⁵⁹

Dal punto di vista metodologico viene specificata, seguendo la proposta di Grannenmann et al. (1986), una funzione di costo "ibrida" che permette maggior flessibilità nella definizione del modello e consente (al contrario della translog) la stima di un processo multiprodotto, nel quale alcuni output sono assenti. Il modello di costo prevede infatti, 4 output per le cure residenziali (giornate di residenza suddivise per profilo sanitario: severo, intermedio, lieve e autosufficiente) 2 output per le cure offerte ai non residenti (visite ambulatoriali e giornaliere) e le visite domiciliari. La specificazione del modello di costo è molto ricca e considera, oltre ai prezzi di 4 categorie di lavoratori (Infermieri qualificati, non qualificati, aiutanti e personale amministrativo): il case-mix dei pazienti, il livello di concorrenza, il mix di clienti "medicaid-Privati", il tipo di localizzazione (urbana o meno) e di istituzione (for profit o non profit).¹⁶⁰

¹⁵⁸ Il CON, introdotto per la prima volta nello stato di New York nel 1966, è l'autorizzazione che permette l'espansione della capacità produttiva di un istituto esistente (o l'apertura di uno nuovo). Il certificato è rilasciato dal regolatore sulla base di un reale bisogno e costituisce di fatto un limite alla capacità totale dei posti letto disponibili.

¹⁵⁹ La classificazione basata sul livello del profilo sanitario prevede anche la categoria lieve (Domiciliary Care) e la categoria autosufficiente (Independent Living).

¹⁶⁰ L'autore solleva anche il problema di scegliere tra una funzione di costo di breve periodo, escludendo gli input fissi (capitale e spese amministrative), e di lungo periodo, optando per quest'ultimo approccio.

I risultati della stima OLS permettono di misurare il costo marginale di una giornata di cura e il livello delle economie di scala. Alla media dei dati, l'autore rileva economie di scala specifiche sia per i pazienti SNF (1,06) che per i pazienti INF (1,23) rendendo attenti alla necessità di non estendere questi valori all'intero spazio produttivo.¹⁶¹ Le stime dei parametri della funzione di costo vengono poi utilizzati per discutere il tema centrale del paper, ossia la definizione di un rimborso per i clienti "Medicaid" con profilo sanitario severo in grado di scoraggiare l'esclusione di questi ultimi in favore dei clienti privati o comunque con profilo sanitario intermedio.

Gertler and Waldman (1992) stimano una funzione di costo per un campione di 279 case per anziani riferite allo Stato di New York nell'anno 1980, con lo scopo di considerare l'impatto della qualità sui parametri strutturali del modello di costo. Gli autori propongono una specificazione dove la qualità è esogena ed uguale per tutti gli istituti, contrapposta ad un modello nel quale il livello di qualità è scelto dall'impresa con l'obiettivo di massimizzare i profitti. Le case per anziani, selezionate per l'analisi, sono tutte a scopo di lucro e competono per attirare due tipologie di ospiti: privati o "Medicaid", in un contesto di eccesso di domanda per questi ultimi.¹⁶² Per questo motivo gli autori sostengono che l'output è esogeno (come i prezzi degli input e le caratteristiche della domanda), mentre gli istituti fissano il prezzo del servizio ed il livello di qualità.

L'obiettivo del paper è quello di sottolineare l'importanza di una corretta specificazione dei parametri e quali possono essere le conseguenze negative di tali errori sulle indicazioni di politica sanitaria. Il modello teorico, con qualità endogena e non osservabile, viene specificato, seguendo l'approccio neoclassico in senso ampio, con una funzione di costo translog nella quale si considerano i costi totali per giornata in funzione dell'output, dei prezzi degli input e delle caratteristiche della domanda.¹⁶³ L'uso delle medie giornaliere per tutte le variabili si può spiegare con la decisione di

¹⁶¹ Trattandosi di una specificazione multiprodotto, le economie di scala sono definite, seguendo Baumol et al. (1982), dal rapporto tra il costo medio incrementale di un output ed il rispettivo costo marginale.

¹⁶² I clienti Medicaid, non dovendo pagare personalmente, sono interessati solamente alla qualità delle cure, mentre i clienti privati, che pagano di tasca propria, guardano sia la qualità che il prezzo.

¹⁶³ Nel modello empirico la qualità non viene inclusa, poiché non è osservata. Specificando il sistema composto dalla funzione di costo totale e dalle "share equations" (Lemma di Shephard) si riesce tramite sostituzione ad esprimere il parametro della qualità (endogena) come una funzione dei parametri dei regressori.

scegliere gli ospiti come output (invece che le giornate) e l'elevato tasso di occupazione delle strutture. Gli input considerati sono il lavoro qualificato, il lavoro non qualificato e le forniture (pasti, materiale di consumo, etc.), mentre sono escluse le apparecchiature e gli stabili.¹⁶⁴ Per l'applicazione empirica gli autori seguono una procedura, proposta da Chamberlain (1982), basata sulla stima di minima distanza che permette di stimare il modello in modo consistente, recuperando i parametri di impatto della qualità, specificati nella forma ridotta del modello.¹⁶⁵

Il risultato empirico più importante è la differenza dei parametri stimati con i due approcci che indicano, nel caso di qualità esogena, la presenza di economie di scala (1.06), mentre non si hanno economie di scala significative quando la qualità viene considerata endogena.¹⁶⁶ Con riferimento alla media dei dati (una casa per anziani rappresentativa con 120 ospiti al giorno), gli autori si avvalgono dei parametri delle stime per formulare due considerazioni di politica sanitaria: (1) l'effetto della competizione sulla qualità è positivo, tuttavia (2) questo significa che ridurre il livello di qualità, abbassando la competizione, comporta un notevole risparmio di costo.¹⁶⁷

Il paper di **Vitaliano and Toren (1994)**, che rappresenta una delle prime applicazioni delle frontiere stocastiche al settore delle case per anziani, si propone di misurare e confrontare l'inefficienza di costo delle case per anziani dello stato di New York relative a due anni distinti. Distaccandosi in parte dalla letteratura dell'epoca, concentrata principalmente sul ruolo della qualità nella specificazione della funzione di costo, gli autori si concentrano sulla misurazione dell'inefficienza a livello di settore e di singola casa per anziani, con l'obiettivo di determinare empiricamente i fattori esplicativi dell'inefficienza. I dati a disposizione riguardano 164 case per anziani per

¹⁶⁴ Le infrastrutture ed i costi di capitale connessi sono considerati come un fattore quasi fisso e non considerati nell'analisi che quindi si profila (anche se non esplicitamente dichiarato) come una specificazione di breve periodo.

¹⁶⁵ La spiegazione della strategia di stima si trova in appendice al paper, mentre le prove della consistenza dello stimatore e della sua normalità asintotica si trovano nel paper originale di Chamberlain (1982).

¹⁶⁶ Nel paper vengono presentate le economie di scala per tre diversi livelli di qualità (minimo medio e massimo), ottenuti come previsione del modello sulla base dei dati osservati per il campione. Solo per un livello minimo di qualità si registrano economie di scala.

¹⁶⁷ Una deviazione standard in meno nel livello di competizione (misurato con l'indice di Herfindahl) comporta una riduzione del 20% nel livello dei costi e solamente del 2,5% nel livello di qualità.

ospiti con profilo sanitario severo (“skilled nursing facility” o SNF) riferite all’anno 1987 e 1990.

Il modello considera una funzione di costo totale con una specificazione neoclassica in senso ampio, affiancando cioè alle variabili strutturali (output e prezzi dei fattori) una variabile di case mix ed una proxy per la qualità, con lo scopo di considerare almeno in parte l’eterogeneità dell’output (misurato con le giornate di residenza). La variabile che coglie le differenze nel profilo sanitario degli ospiti di una struttura è rappresentata dalla quota di residenti associati alla classe RUG (Resources Utilization Groups) meno costosa.¹⁶⁸ La qualità dell’output viene misurata sulla base di un protocollo di qualità a più dimensioni e con numerosi items, rispetto al quale si contano le manchevolezze riscontrate.¹⁶⁹ La specificazione degli input ricalca più o meno quanto già osservato negli altri lavori (lavoro qualificato, lavoro non qualificato spese di capitale).

Per verificare l’impatto della forma istituzionale e dell’assetto proprietario sull’inefficienza degli istituti, vengono inserite 4 *dummies* per distinguere le case private for profit (in tre diversi regimi di proprietà) e pubbliche da quelle private non profit, utilizzate come riferimento di base.¹⁷⁰ La specificazione econometrica è definita da una funzione Cobb-Douglas, stimata con la tecnica di massima verosimiglianza seguendo il modello di Aigner et al. (1977).¹⁷¹ Dai risultati empirici emergono alcuni spunti interessanti. il livello di inefficienza medio è passato dal 31% nel 1987 al 24% nel 1990 anche se la differenza non è statisticamente significativa. Nonostante i costi di produzione delle diverse forme istituzionali siano diversi (le relative *dummies* risultano statisticamente significative) le differenze di inefficienza tra le diverse tipologie sono trascurabili. Le economie di scala, la cui misura non costituiva un obiettivo del paper, si

¹⁶⁸ Gli ospiti sono suddivisi in 5 classi, basate sull’utilizzo delle risorse che provocano. La classe nella quale i residenti necessitano cure meno intense o impegnative determina anche un costo minore.

¹⁶⁹ La costruzione di questa variabile, che è descritta in ogni dettaglio nell’appendice del paper, non cattura certamente la complessa natura della qualità nelle case per anziani, ma si propone di filtrare almeno una parte dell’eterogeneità dovuta a livelli qualitativi molto diversi.

¹⁷⁰ La scelta delle Non-profit come riferimento è spiegata dalla maggior rappresentazione (67) di questa forma organizzativa nel campione (164). Le case per anziani pubbliche sono 16, mentre le restanti strutture for profit sono suddivise in “corporate owner” (25), “proprietorship” (14) e “partnership” (42).

¹⁷¹ La scelta della Cobb-Douglas è giustificata dagli autori con la scarsa significatività dei parametri di ordine superiore al primo nella specificazione translog.

possono ricavare dal coefficiente dell'output e risultano pari a 1.10, evidenziando la possibilità di ridurre i costi medi all'aumentare dell'output.¹⁷²

Blank and Eggink (2001) analizzano il settore delle case per anziani in Olanda con lo scopo di determinare la struttura dei costi ed il livello medio di inefficienza. Il campione di dati è un panel bilanciato, composto da 110 istituti osservati su un periodo di 8 anni, di cui 7 consecutivi (1984, 1987-1993). L'approccio seguito dagli autori è sostanzialmente quello di Gertler and Waldman (1992) con le necessarie modifiche per adattare la specificazione teorica al contesto olandese, diverso da quello americano. In Olanda, infatti, il settore dell'assistenza agli anziani avviene in un contesto fortemente regolamentato: gli istituti sono non-profit, i budget sono fissati ex-ante dall'autorità, esiste una pianificazione centralizzata della capacità produttiva, gli ospiti sono assegnati in modo esogeno, le prestazioni sono rimborsate, il livello dei salari è soggetto a vincoli, etc. Gli autori vedono nella qualità l'unico strumento per competere sul mercato e, in prospettiva dinamica, garantire la sopravvivenza dell'impresa.

Nel paper troviamo una distinzione tra qualità esogena (dovuta principalmente al profilo sanitario degli ospiti e alle caratteristiche della domanda) e qualità endogena. La prima è misurabile ed approssimata nella stima da tre variabili (indice di dipendenza, l'indice di Herfindal e le giornate di cura diurne), mentre la qualità endogena (servizi offerti, comfort della struttura, qualità dello staff, etc.) non si può osservare. Entrambe le componenti della qualità, insieme alla misura quantitativa dell'output (giornate di cura), costituiscono le variabili esplicative dell'equazione denominata "output corretto per la qualità". Il problema di identificazione è risolto per sostituzione, utilizzando l'equazione (Cobb-Douglas) che specifica la qualità, nel sistema di equazioni composto da una funzione di costo variabile (translog) e dalle relative equazioni delle quote di costo.¹⁷³

I parametri del modello ridotto, così ottenuto, sono stimati seguendo la procedura SUR (Seemingly Unrelated Regression) proposta da Zellner (1962), ignorando la struttura panel dei dati. Imponendo alcune restrizioni ai parametri relativi alla qualità, è

¹⁷² La Cobb-Douglas non è una forma funzionale flessibile. Le stesse economie di scala sono, quindi, imposte sia agli istituti molto piccoli sia a quelli molto grandi.

¹⁷³ Per poter identificare il modello, l'equazione della qualità endogena viene sostituita nella funzione di costo, imponendo alcune restrizioni ai parametri. Inoltre, l'equazione di una delle quote di costo deve essere omessa per evitare che la matrice di varianza-covarianza dell'errore risulti singolare.

possibile confrontare i risultati del modello che specifica entrambe le componenti (endogena ed esogena), con i risultati di un modello dove la qualità è esclusivamente esogena. Nonostante il test sul rapporto di log-verosimiglianza favorisca chiaramente il modello completo (con qualità endogena ed esogena), i parametri delle due specificazioni risultano sostanzialmente uguali.¹⁷⁴ In entrambi i modelli le economie di scala non risultano apprezzabili mentre il livello dell'inefficienza di costo, ottenuta dai residui della regressione sfruttando la struttura panel dei dati, è leggermente più elevato nel modello con qualità esogena (11%), rispetto alla specificazione con qualità endogena (7%), risultato in linea con le argomentazioni teoriche degli autori.¹⁷⁵

Il paper di **Filippini (2001)** presenta la struttura dei costi del settore delle case per anziani nel Cantone Ticino, uno dei 26 stati della Confederazione Helvetica. L'obiettivo di questo lavoro è quello di offrire un contributo empirico alla discussione sulle economie di scala del settore che, se presenti, potrebbero influenzare le scelte di politica sanitaria e di regolamentazione. In Svizzera l'erogazione dei servizi agli anziani è responsabilità dei singoli cantoni che delegano, a loro volta, il compito operativo ai comuni e/o alla società civile, riservandosi il compito di finanziatore e di regolatore.¹⁷⁶ In Ticino le case per anziani sono confrontate con un eccesso di domanda cronico ed una scarsa mobilità dei richiedenti, due condizioni che legittimano la scelta di considerare esogeno l'output.

Per descrivere la tecnologia di produzione di questo settore, l'autore specifica e stima una funzione di costo totale translog per un campione di 36 case per anziani attive in Ticino nel periodo 1993-1995. I dati, che provengono dal dipartimento di sanità del Cantone, riguardano solamente le case per anziani (pubbliche e private non profit) che ricevono un finanziamento pubblico e sono, per contro, soggette a regolamentazione. L'output è misurato in giornate di residenza, mentre gli input considerati sono il lavoro ed il capitale. Entrambi sono calcolati dividendo le spese, indicate nel conto economico,

¹⁷⁴ Gli autori attribuiscono questo inatteso risultato (che suggerisce come la distorsione dei parametri introdotta da un modello con qualità esogena sia limitata o trascurabile) alla relativa uniformità nel livello qualitativo delle case per anziani.

¹⁷⁵ I dettagli della procedura, basata sul lavoro di Berger et al. (1993), per ricavare i punteggi di inefficienza sono descritti in appendice.

¹⁷⁶ In Ticino la concessione di sussidi alle case per anziani pubbliche e private non profit è subordinata al rispetto di numerose leggi e requisiti in materia di qualità delle strutture, delle dotazioni e delle competenze del personale.

per un indicatore della quantità: il numero di addetti (equivalenti a tempo pieno) per il fattore lavoro ed il numero di letti come misura dello stock di capitale.¹⁷⁷ Oltre agli elementi della funzione di costo neoclassica (output e prezzi degli input), il modello include una variabile che misura le differenze nel profilo sanitario degli ospiti (indice di dipendenza) e una variabile che approssima il livello di qualità, misurato dal rapporto tra infermieri effettivi e prescritti dal cantone.¹⁷⁸ Il modello è completato dall'aggiunta di una *dummy* (per considerare gli istituti che offrono anche appartamenti protetti) e di un trend lineare che cattura l'effetto del cambiamento tecnologico. I risultati empirici sono ottenuti specificando un sistema di equazioni composto dalla funzione di costo e dalle quote relative agli input.

Le stime dei parametri, generate con la procedura SUR (Zellner, 1962), indicano un effetto positivo della qualità sui costi e la presenza di economie di scala. Il valore delle economie di scala, che nel punto di approssimazione dei dati (capacità produttiva pari a 57 letti) è pari a 1.18, diminuisce all'aumentare della dimensione degli istituti e scompare, quando si raggiunge una capacità produttiva pari a 120 letti.

Crivelli, Filippini e Lunati (2002) analizzano la struttura dei costi e l'inefficienza delle case per anziani svizzere, considerando un campione di dati riferito all'intera nazione per l'anno 1998. Il paper rappresenta la prima applicazione nel settore dell'assistenza agli anziani in Svizzera della tecnica delle frontiere stocastiche. L'obiettivo principale è la stima di una funzione di costo totale, che descriva una tecnologia produttiva comune a livello nazionale in grado di misurare la presenza di economie di scala ed il livello di inefficienza raggiunto dalle singole case per anziani. Il settore dell'assistenza agli anziani, in Svizzera, risulta fortemente eterogeneo a livello cantonale per quanto riguarda le forme di regolamentazione, l'offerta complessiva di letti, l'assetto istituzionale e il livello dei costi medi per giornata di cura. Per contro vi sono tratti che accomunano gli istituti nei diversi cantoni svizzeri come: la prevalenza

¹⁷⁷ La banca dati non permetteva la separazione delle principali categorie professionali che operano in una casa per anziani. Per questo motivo il prezzo del lavoro si può considerare come una media ponderata, i cui valori e pesi sono sconosciuti. Per il prezzo del capitale l'autore ha utilizzato, seguendo Friedländer and Wang Chiang (1983) e Filippini and Maggi (1993), l'approccio residuale.

¹⁷⁸ Gli ospiti sono classificati in tre classi di dipendenza, distinte a seconda del fabbisogno di cure e della capacità di svolgere le attività quotidiane (lavarsi, vestirsi, mangiare, etc.) e numerate da 1 a 3 in ordine crescente di gravità. L'indice di dipendenza, che è la media dei punteggi riportati dagli ospiti, approssima il profilo sanitario dei residenti. Le case per anziani non sono obbligate a rispettare questa indicazione sull'impiego di personale, volta a garantire un livello ideale di qualità.

delle istituzioni non profit (pubbliche e private), un cronico eccesso di domanda, una limitata mobilità intercomunale degli ospiti, una pressione da parte delle autorità finanziatrici (cantoni e comuni) al contenimento dei costi.¹⁷⁹

Dal punto di vista metodologico gli autori specificano una frontiera di costo stocastica, nella quale la componente asimmetrica dell'errore (il termine che cattura l'inefficienza) segue una distribuzione esponenziale. La forma funzionale utilizzata è la translog, che permette di misurare le economie di scala in diversi punti di approssimazione, mentre la specificazione del modello è di tipo "behavioural". I costi totali sono descritti dall'output e dai prezzi dei fattori, come in un modello strutturale, ai quali si aggiungono una serie di regressori che descrivono il comportamento ("behaviour") produttivo delle imprese di questo settore. Gli input considerati sono il lavoro ed il capitale, misurati dividendo le rispettive spese per un indicatore di quantità.¹⁸⁰ L'output è espresso dalle giornate di residenza ed è integrato da una variabile che approssima il case mix degli ospiti e la qualità delle cure.¹⁸¹

Seguendo McKay (1988), per catturare il livello di qualità offerto dalle singole case per anziani viene definita una variabile che misura l'ammontare di ore di assistenza e cura ricevute in media da un singolo ospite. Il modello considera anche due variabili che misurano le differenze nella composizione dello staff (percentuale di personale medico per paziente) e nel ventaglio dei servizi offerti.¹⁸² Per completare il modello sono inserite le *dummies* cantonali ed una *dummy* per quegli istituti che offrono posti letto in appartamenti protetti.¹⁸³

¹⁷⁹ Per una descrizione del settore delle case per anziani in Svizzera si rimanda al capitolo 2.

¹⁸⁰ La disponibilità limitata delle informazioni finanziarie non ha permesso di distinguere tra diverse categorie professionali il fattore lavoro. Il prezzo del lavoro è quindi definito come il totale delle spese del personale diviso il numero di addetti (equivalenti al tempo pieno), mentre le spese per il capitale (ammortamenti e interessi) sono divise per il numero di letti che approssima lo stock di capitale.

¹⁸¹ Non disponendo di misure più precise gli autori hanno utilizzato, come indicatore del profilo sanitario dei residenti, il rimborso medio per ospite che la legge federale sull'assicurazione malattia riconosce per le spese sanitarie. Esiste infatti una forte correlazione positiva tra il livello di dipendenza (e in generale il degrado dello stato di salute) e il livello di spese mediche (rimborsabili) sostenute.

¹⁸² Le informazioni nella banca dati permettevano di capire solo la presenza di un determinato servizio e quindi l'aggregazione lineare delle attività presenti in un istituto potrebbe non riflettere l'effettiva complessità ed importanza dei servizi erogati.

¹⁸³ La possibilità di effettuare le stime a livello cantonale è preclusa dalla frammentazione dei dati. La maggior parte dei cantoni contano meno di 40 case per anziani rendendo impossibile una specificazione translog di un modello con più di tre o quattro variabili.

Per la stima dei coefficienti, ottenuta con la tecnica di massima verosimiglianza, gli autori hanno utilizzato solamente 886 delle circa 1'300 case per anziani attive in Svizzera, a causa di dati incompleti o di valori poco plausibili. I valori empirici dei parametri sono economicamente sensati e la funzione di costo rispetta nel punto di approssimazione dei dati (la mediana) le condizioni di regolarità.¹⁸⁴

Le stime rilevano la presenza di economie di scala che tendono a diminuire con il crescere della dimensione. Calcolate nel punto mediano, che corrisponde ad una capacità di 48 letti, il valore delle economie di scala è pari a 1,04 e scende a 1 intorno a 90 letti, che rappresentano la dimensione ottimale. I punteggi di inefficienza di costo delle singole case per anziani sono calcolati, a partire dai residui della regressione, utilizzando la formula di Jondrow et al. (1982). Il risultato generale rileva un livello medio di inefficienza delle case per anziani relativamente contenuto (13% circa), mentre non si registrano differenze significative (da un punto di vista statistico) nel livello medio di inefficienza delle diverse forme istituzionali e di regolamentazione.

Rebba e Rizzi (2002) analizzano la struttura dei costi delle case per anziani in una regione italiana (Veneto), stimando una funzione di costo per un campione di 61 istituti con riferimento all'anno 2000. Il servizio di assistenza agli anziani è regolamentato in diversi aspetti (fissazione dei contributi per le cure mediche, pianificazione dell'offerta, assegnazione dei posti ai richiedenti, dotazioni di personale, etc.) ed opera in un contesto di eccesso di domanda. Gli autori, dopo una dettagliata descrizione delle caratteristiche specifiche del settore delle case per anziani in Veneto, presentano la specificazione generica di un modello di costo adatto a descrivere la realtà esaminata.

La scelta metodologica si orienta ad una funzione di costo totale translog, con lo scopo di misurare il valore delle economie di scala e dell'inefficienza di costo.¹⁸⁵ Per considerare l'impatto della qualità e di altri fattori (forma istituzionale, tipologia di ospiti, intensità di assistenza erogata, etc.) che incidono sui costi, la misura quantitativa dell'output viene corretta da un vettore di caratteristiche, che interagiscono linearmente

¹⁸⁴ L'omogeneità lineare nei prezzi degli input è stata imposta, mentre dai coefficienti stimati della funzione di costo si è verificato che la matrice Hessiana fosse negativa semidefinita.

¹⁸⁵ L'approccio seguito è simile ad una specificazione endogena della qualità, come visto in Gertler and Waldman (1992) e Blank and Eggink (2001), anche se in modo più semplificato. La qualità (e gli altri fattori) sono inclusi nell'output e non interagiscono con i parametri strutturali della funzione di costo, determinando semplicemente una traslazione verticale del livello dei costi.

(nei logaritmi) con l'output. La misura dell'output considerata è il numero di giornate di residenza, mentre gli input sono il lavoro, il capitale e le altre spese. I prezzi sono definiti utilizzando le informazioni contabili sulle rispettive quote di costo divise per una misura fisica della quantità del fattore produttivo.¹⁸⁶ Gli autori lamentano la mancanza di una misura attendibile del profilo sanitario degli ospiti (case-mix) e di una misura diretta della qualità delle cure (approssimata da alcuni indicatori spesso utilizzati come le ore di cura per giornata o superficie disponibile per ogni letto). Il vettore dei fattori qualitativi ed ambientali è completato da una serie di *dummies* relative alla forma organizzativa ed istituzionale dell'istituto.

I dati a disposizione per le stime sono solamente 50 (in rappresentanza di 61 istituti sugli oltre 300 censiti in Veneto nel 2000), a causa della presenza di alcune reti locali di strutture che consolidano i dati contabili. Questa limitazione nei gradi di libertà suggerisce agli autori di non utilizzare la tecnica delle frontiere stocastiche e di determinare i punteggi di inefficienza delle singole case per anziani con una frontiera deterministica. I parametri della funzione di costo (e delle domande condizionali degli input) sono stimati con la tecnica SUR proposta da Zellner (1962) permettono di verificare la struttura dei costi delle case per anziani venete e la loro inefficienza. L'elasticità di costo rispetto all'output è pari a 0,91 ed indica che all'aumentare delle giornate di residenza i costi crescono in modo meno che proporzionale.

Gli autori, tuttavia, non enfatizzano la presenza di economie di scala considerando il parametro come non conclusivo vista l'elevata probabilità (circa il 18%) che sia uguale ad 1. L'analisi dell'inefficienza di costo esprime la presenza di un grande potenziale di miglioramento nelle performance degli istituti visto che il livello medio di inefficienza è in media pari a 40% (eccesso di costo rispetto al valore efficiente). Come rilevato dagli autori, la natura deterministica del metodo COLS contribuisce ad esasperare i valori estremi e la dispersione dei punteggi di inefficienza che, nel caso delle case per anziani in Veneto, sembrano troppo elevati per essere pienamente credibili.

¹⁸⁶ Per misurare il lavoro ed il capitale, seguendo la letteratura prevalente, gli autori utilizzano lo staff (unità a tempo pieno) ed il numero di letti. Le altre spese sono normalizzate con il totale dei costi, causando qualche problema di interpretazione e violando l'ipotesi di non correlazione dei regressori.

Vitaliano (2003), a distanza di 10 anni dal lavoro con Toren (Vitaliano and Toren, 1994), analizza la situazione delle case per anziani dello stato di New York, considerando 228 istituti non profit per l'anno 1993, per stabilire se l'obiettivo della massimizzazione dei profitti è realistico per tutte le strutture. Come discusso in precedenza, il settore delle case per anziani nello stato di New York è caratterizzato da un forte eccesso di domanda ed una regolamentazione nell'offerta aggregata (CON) e nelle tariffe per i pazienti "Medicaid".

L'autore formalizza un modello di comportamento analogo a Nyman (1988), nel quale la qualità, che entra come argomento sia nella funzione di costo sia nelle funzioni di domanda dei clienti privati (insieme al prezzo) e dei clienti "Medicaid", è lo strumento utilizzato per selezionare il mix che massimizza i profitti.¹⁸⁷ Il vincolo di non distribuzione degli utili, che caratterizza le istituzioni non profit, potrebbe indurre le case per anziani a massimizzare, al posto dei profitti, una funzione di utilità (dei managers) le cui variabili esplicative sono: i profitti, la qualità e gli eccessi di input.¹⁸⁸ Dalla discussione delle condizioni del primo ordine del problema emerge una strategia semplice per verificare empiricamente il tipo di comportamento delle singole imprese, che consiste nel confronto tra il costo marginale con il rimborso previsto. Un costo marginale significativamente maggiore del rimborso indica che l'impresa non sta massimizzando i profitti.¹⁸⁹

La stima dei costi marginali per giornata di cura implica la definizione di un modello di costo ed una strategia econometrica. La necessità di confrontare le tariffe di rimborso delle 16 categorie RUG alla quale gli ospiti "Medicaid" sono attribuiti comporta la specificazione di un processo produttivo multiprodotto. L'assenza di alcune tipologie di paziente in una determinata casa per anziani significa rilevare un vettore di output con molti elementi pari a 0 e questo costituisce un grave ostacolo all'uso di

¹⁸⁷ Come predetto da Nyman (1988), l'introduzione di un rimborso basato sui RUG, differenziato per livello di severità del profilo sanitario ha determinato un profondo cambiamento nel mix di ospiti Medicaid". In particolare, il numero di ospiti con profili più severi (e costosi), che in passato venivano esclusi perché poco remunerativi, è aumentato sensibilmente.

¹⁸⁸ Seguendo Weisbrod (1998), in una casa per anziani non profit i profitti non devono essere distribuiti agli azionisti, mentre il management alla ricerca di rendite potrebbe appropriarsene in diversi modi. Inoltre, i profitti sono un mezzo necessario per aumentare il livello di qualità o per acquistare input non necessari che conferiscono utilità diretta o indiretta al management.

¹⁸⁹ Per semplificare l'esercizio econometrico gli autori si sono concentrati solo sui pazienti "Medicaid".

forme funzionali come la Cobb-Douglas o la Translog. L'autore utilizza la trasformazione Box-Cox per ciascuno dei 16 output e propone una funzione di costo totale neoclassica con due input (lavoro e capitale) e corretta da: un indice di qualità delle cure, dal numero di servizi offerti e dalla percentuale di ricavi da ospiti "Medicaid".¹⁹⁰ L'indice di qualità è una misura mutuata da una guida alle case per anziani (Bua, 1997), che valuta le manchevolezze dell'istituto rispetto a numerosi parametri medici e strutturali.¹⁹¹ Il numero di servizi attivi nelle strutture è inserito per cogliere quelle differenze nei costi che non sono attribuibili al case-mix (i gruppi RUG) o alla qualità delle cure.¹⁹²

I risultati empirici della stima si concentrano sul confronto tra i costi marginali di ogni output ed il relativo rimborso.¹⁹³ Dal test su tutti i 228 istituti non profit considerati gli autori concludono che 134 case per anziani hanno un comportamento compatibile con la massimizzazione dei profitti, 49 istituti massimizzano qualcosa di diverso dai profitti e per 45 strutture non si può stabilire con sufficiente confidenza statistica.

Farsi and Filippini (2004) stimano una frontiera stocastica di costo per un campione di 36 case per anziani attive in Ticino, un cantone della Svizzera, nel periodo compreso dal 1993 al 2001. Lo scopo del paper è quello di misurare empiricamente il livello di inefficienza di costo delle case per anziani, per verificare l'ipotesi che la forma istituzionale o l'assetto proprietario possano indurre delle differenze sistematiche di inefficienza. In Ticino la forte regolamentazione sulla qualità, sull'offerta e sull'assegnazione dei pazienti alle strutture lascia poco spazio alla competizione di mercato volta alla massimizzazione dei profitti. Gli autori ipotizzano quindi che l'obiettivo di minimizzazione dei costi possa essere indebolito a causa dei problemi di

¹⁹⁰ Rispetto ad altri studi, il costo del capitale viene calcolato utilizzando i valori previsti dalla regressione del tasso di interesse nominale del debito nei confronti (1) della lunghezza del periodo di rimborso e (2) della percentuale di indebitamento.

¹⁹¹ L'indice, che si basa sui punteggi misurati nel periodo 1993-1996, assume valore 1 quando la qualità è molto bassa e valore 6 quando la qualità è molto alta.

¹⁹² Il numero di servizi considerati arriva ad un massimo di 20 e copre gli aspetti strutturali e funzionali della casa per anziani che costituiscono un elemento della qualità complessiva dell'istituto.

¹⁹³ Le economie di scala non vengono discusse in dettaglio. Tuttavia, dai parametri della funzione di costo si possono calcolare le economie di scala radiali alla media dei dati e le economie di scala per ogni singolo output. Nella nota 19 del paper si rileva l'assenza di economie di scala.

agenzia che sono presenti nella catena decisionale di queste organizzazioni.¹⁹⁴ Dal punto di vista empirico l'obiettivo è quello di stimare una frontiera di costo in modo da ottenere i valori di inefficienza delle singole case per anziani per verificare l'effetto delle diverse forme istituzionali e dell'assetto proprietario. Il campione è infatti costituito da 14 istituti privati non profit (organizzate come fondazioni) e 22 istituti pubblici. La gestione di questi ultimi è a carico delle amministrazioni comunali per la maggior parte dei casi (17), mentre 5 istituti sono gestiti in modo (relativamente) autonomo da fondazioni a capitale pubblico. La struttura panel dei dati permette la specificazione di un modello di frontiera stocastica di costo ad effetti casuali e gli autori utilizzano il metodo proposto da Schmidt and Sickles (1984).

Il modello descrive una funzione di costo totale translog di tipo “*behavioural*”, poiché si ammette la possibilità che, a causa dei vincoli imposti dalla regolamentazione e dal ruolo giocato dalla qualità, il comportamento dell'impresa non sia descritto da una minimizzazione incondizionata dei costi come previsto dalla teoria neoclassica. La specificazione dei regressori è analoga al lavoro di Filippini (2001) e prevede: un solo output (giornate di residenza), due input (lavoro e capitale), una proxy per il case-mix dei residenti (indice di dipendenza), una proxy per la qualità delle cure (rapporto staff effettivo rispetto allo staff indicato dal cantone) ed un trend temporale. Visto che la qualità non può essere usata per selezionare gli ospiti, che devono fare riferimento al proprio comune di domicilio, gli autori considerano la variabile come esogena e la misurano con l'indice di dipendenza elaborato dal dipartimento di sanità del cantone.¹⁹⁵ Nel paper si ammette la possibilità che le case per anziani rilevino in modo eterogeneo la dipendenza degli ospiti, introducendo un errore sistematico nella variabile e una conseguente distorsione dei parametri stimati. Per verificare questo potenziale problema gli autori specificano un modello alternativo senza considerare la qualità tra i regressori.

I risultati delle stime del modello inclusivo della qualità indicano che il livello di inefficienza medio è contenuto per tutte le tipologie di istituzione e varia dal 7% per le

¹⁹⁴ Le ipotesi del paper sono in linea con la letteratura sui diritti di proprietà (es. Alchian, 1965 e Demsetz, 1967), sulla teoria della burocrazia (Niskanen, 1968 e 1971) e sul settore non profit (Weisbrod, 1998).

¹⁹⁵ L'indice è costruito assegnando una misura da 1 a 3 ad ogni ospite sulla base del livello di autonomia (dove 3 rappresenta la massima dipendenza) del singolo residente rispetto ad attività quotidiane. Il valore medio di dipendenza degli ospiti di una casa anziani costituisce l'indice di dipendenza.

fondazioni private (8% per le 5 fondazioni pubbliche) al 10% per gli istituti pubblici. Nel modello senza qualità il livello medio dell'inefficienza è più elevato, ma l'ordinamento delle istituzioni non cambia. Le differenze tra istituzioni sono corroborate dal test di Kruskal-Wallis che (al 5% di confidenza) rileva sia una maggior efficienza delle case per anziani private rispetto a quelle pubbliche sia una maggior efficienza delle fondazioni rispetto alla gestione comunale.¹⁹⁶ I parametri della funzione di costo, che permettono di calcolare le economie di scala in diversi punti di approssimazione dei dati, indicano che nel settore delle case per anziani in Ticino la maggioranza degli istituti potrebbe produrre a costi medi inferiori se operasse su scala più grande. Le economie di scala sono sostanzialmente stabili nell'intervallo di produzione prevalente e passano da 1,13 nel 1° quartile a 1,12 nel punto mediano per scendere a 1,11 nel 3° quartile.¹⁹⁷ Da un punto di vista di politica sanitaria questi risultati, che indicano la presenza di un'offerta frazionata in tante case per anziani di piccole dimensioni, suggeriscono di considerare l'importanza economica della capacità per la pianificazione degli istituti di nuova e futura costruzione.

Christensen (2004) propone un'analisi delle economie di scala e di scopo delle case per anziani dello stato dell'Illinois (USA) basata sulla tecnica di regressione per quantili. L'autore sostiene che nel dibattito sulla corretta definizione e specificazione di una funzione di costo per questo settore poco peso sia dato alle differenze strutturali che esistono nei diversi punti (ad esempio i decili) della distribuzione dei costi. Una casa per anziani con 10 posti letto difficilmente produce un servizio organizzandosi in modo analogo ad un istituto dotato di 500 posti letto. Il campo di variazione della capacità produttiva di una casa per anziani, misurato dal rapporto tra la dimensione massima e quella minima è quasi sempre molto ampio e non lontano dall'esempio citato.¹⁹⁸

Su queste premesse, e in linea con la letteratura già presentata, il paper considera un modello di costo che include: una variabile "qualità delle cure erogate" (ore di cura di infermieri registrati per giornata), una variabile "qualità dei servizi offerti" (superficie struttura per letto) e un controllo per fattori quali la localizzazione urbana (Chicago e St

¹⁹⁶ Il Kruskal-Wallis (1952) test è un test non parametrico che si basa sul confronto dei ranghi di due sottocampioni per stabilire se gli stessi appartengono o meno alla stessa popolazione.

¹⁹⁷ In termini di posti letto, l'approssimazione al 1° quartile, alla mediana ed al 3° quartile dei dati, corrisponde rispettivamente ad una casa per anziani con una dotazione di 47, 61 ed 80 posti letto.

¹⁹⁸ In Illinois nel 1997 la casa per anziani più piccola era dotata di 20 letti, mentre la più grande di 508.

Louis) o la forma istituzionale (Pubblica, privata non profit, privata for profit). La parte neoclassica del modello, vista la presenza di un vincolo importante (CON) nell'espansione della capacità produttiva, è definita con un'ottica di breve periodo che porta alla specificazione di una funzione di costo variabile. La definizione di un processo multiprodotto che eroga servizi di cure intermedie, specialistiche ("skilled") o di entrambi i tipi rende difficile l'impiego di una Cobb-Douglas o di una Translog tipica dei modelli strutturati "*behavioural*".

Per garantire un certo rigore nella struttura dei costi, senza perdere la flessibilità delle funzioni ad hoc, gli autori specificano una funzione ibrida, sostanzialmente analoga a quella di Grannemann et al. (1986). La tecnica di regressione usata permette di verificare il comportamento dei parametri della funzione in punti diversi della distribuzione dei costi. In pratica, invece che derivare un solo vettore di parametri da approssimare in differenti punti (es. tecnica OLS) si ottengono tanti vettori quanti sono i punti (quantili) della distribuzione considerati. Gli autori, seguendo Koenker and Bassett (1978), utilizzano la routine di STATA per ottenere i parametri di ogni decile della distribuzione dei costi. I dati a disposizione degli autori si riferiscono a 600 istituti per anziani attivi nel 1997 nello Stato dell'Illinois, dove le case per anziani offrono servizi di cura a pazienti "Skilled Care" (SC) e "Intermediate Care" (IC). La maggior parte delle strutture (411) è attrezzata per ospitare entrambe le tipologie, mentre 121 istituti offrono solamente letti a pazienti IC e le restanti 68 case per anziani sono esclusivamente per ospiti SC. Questa ripartizione del campione permette di condurre alcuni confronti sulla presenza di economie di scopo, uno degli obiettivi del paper.

Per quanto riguarda le economie di scala, gli autori rilevano un chiaro vantaggio ad espandere la dimensione solamente nei primi decili, ossia per case per anziani con capacità inferiori a 60-70 posti letto. A prescindere dai risultati empirici sulle economie di scala e di scopo, non sempre in linea con la letteratura passata, il messaggio centrale del lavoro è l'attenzione che occorre prestare all'interpretazione delle stime.¹⁹⁹ Se ci si allontana dal punto di approssimazione scelto (ad esempio la media) per generare i

¹⁹⁹ Nel paper viene precisato l'interesse a determinare le economie di scala (concetto di lungo periodo) partendo dalla stima dei parametri di una funzione di costo variabile. Nel capitolo 3 abbiamo già discusso la possibilità di definire, attraverso la formula [3.20], le economie di scala per un'impresa che opera in equilibrio di breve periodo.

parametri, occorre cautela nel formulare giudizi per realtà che appartengono ad una posizione molto diversa nella distribuzione dei dati.

Il paper di **Farsi Filippini and Kuenzle (2005)** ha un taglio marcatamente metodologico e confronta il comportamento di diverse tecniche econometriche rispetto al problema di separare l'eterogeneità non osservata dalla misurazione dell'inefficienza di costo. Gli autori si propongono di applicare diverse specificazioni di modelli panel ad un campione di 36 case per anziani attive in Ticino (CH) nel periodo 1993-2001.

Il modello di costo (ed il dataset) è specificato in modo analogo al lavoro di Farsi and Filippini (2004) e, come già discusso in precedenza, offre una buona rappresentazione della struttura dei costi del settore delle case per anziani, anche se alcuni fattori esogeni o caratteristiche specifiche dei singoli istituti non sono considerati dal modello di costo. Nel processo empirico di stima dei parametri, questa eterogeneità non osservata, che sarebbe in grado di spiegare una parte della variabilità dei costi, si confonde con i residui della regressione e quindi con la misura dell'inefficienza.

L'accesso a dati panel consente agli autori di applicare una serie di modelli panel con proprietà diverse, alcuni dei quali possono ridurre il problema dell'eterogeneità non osservata. In particolare, si considerano 6 diversi modelli: “*Fixed and Random Effect*” (Schmidt and Sickles, 1984), “*Pooled Stochastic Frontier*” (Aigner et al., 1977), “*True Random Effect*” (Greene, 2005), più una specificazione in linea con Mundlak (1978) del “*Random Effect*” e del “*True Random Effect*”. Gli obiettivi del paper sono quelli di verificare quanto la stima dei parametri della frontiera di costo translog risulti sensibile alle diverse specificazioni e, soprattutto, come la diversa capacità dei modelli di considerare l'eterogeneità non osservata, influenzi le misure di inefficienza.²⁰⁰

I risultati delle stime dei coefficienti del primo ordine sono, in tutte le 6 specificazioni, significativi e coerenti con le aspettative: la funzione di costo è crescente nell'output, nei prezzi dei fattori, nel profilo sanitario degli ospiti e nella qualità delle cure erogate.²⁰¹ Le differenze tra i parametri nei diversi modelli, che esprimono una

²⁰⁰ Per una presentazione dei diversi modelli utilizzati si rimanda al capitolo 4. Nel paper gli autori prefigurano la direzione e l'importanza delle distorsioni provocate dall'eterogeneità in rapporto alla specificazione econometrica scelta, lasciando alla misurazione empirica la conferma delle ipotesi.

²⁰¹ Le condizioni di regolarità nel punto di approssimazione delle variabili non sono soddisfatte. Gli autori giustificano questo risultato con la presenza di vincoli regolatori sul mercato del lavoro che impediscono una libera minimizzazione dei costi, come richiesto dalla funzione di costo totale.

certa convergenza tendenziale, sono indagate in dettaglio per verificare se esiste una sistematica nella distorsione delle stime introdotta dall'eterogeneità non osservata.

Preso come riferimento il modello ad effetti fissi (FE), che non è afflitto dal problema, si constata come i modelli che maggiormente ignorano il fenomeno dell'eterogeneità non osservata (*Pooled* e REM) si discostino maggiormente dai valori stimati in modo consistente dal modello FE. Inoltre, la correzione "Mundlak" proposta dagli autori per il modello "*True Random Effect*" esprime parametri quasi identici rispetto al modello ad effetti fissi, preso come riferimento. L'elasticità di costo rispetto all'output varia da un minimo di 0,75 ad un massimo di 0,92 e da questi risultati si può desumere che, in corrispondenza di una capacità produttiva mediana (circa 59 letti), esistono economie di scala non pienamente sfruttate. Rispetto ai parametri della funzione di costo, la stima dell'inefficienza presenta un campo di variazione molto più ampio, rilevando un'inefficienza media del 20% nel modello FE paragonato ad un valore inferiore al 5% nel caso del "*True RE-Mundlak*". Il risultato, atteso dagli autori, dimostra che il fenomeno dell'eterogeneità non osservata è molto marcato nella stima dell'inefficienza più che nei parametri della frontiera di costo.

La rassegna esposta copre circa 20 anni di lavori empirici, dedicati all'analisi (parametrica) della struttura dei costi ed alla misurazione delle economie di scala e dell'efficienza di costo nel settore delle case per anziani. Un elemento che emerge con forza dal confronto dei diversi lavori è l'importanza di considerare la complessità del processo produttivo di una casa per anziani nella specificazione del modello di costo. A prescindere dagli obiettivi di ricerca specifici dei singoli paper discussi, un elemento importante nell'analisi empirica di questo settore è la comprensione delle caratteristiche strutturali degli elementi costitutivi del mercato: domanda, offerta e regolamentazione. In ogni paese l'assistenza agli anziani assume forme molto differenziate ed è organizzata secondo modalità specifiche che rendono difficile un'analisi comparata di lavori empirici basati su realtà profondamente diverse. Per questo motivo, nella Tabella 5.1, presentiamo un riepilogo di sette modelli in rappresentanza di tutti i diversi mercati analizzati in questi dodici papers recensiti.²⁰²

²⁰² Dei 4 articoli riferiti alle case per anziani nello stato di New York abbiamo selezionato l'unico che utilizza la metodologia delle frontiere stocastiche. Dei 3 articoli riferiti al cantone Ticino abbiamo considerato quello più recente e con un taglio più metodologico.

Le diverse realtà analizzate, unite a diversi obiettivi di ricerca e disponibilità di informazioni statistiche, dipingono un quadro piuttosto eterogeneo di specificazioni sia per quanto riguarda le variabili utilizzate nei modelli sia per le tecniche di stima.

Tabella 5.1 - Riepilogo dei modelli di costo specificati per diversi paesi.

Autore	Modello di costo e forma funzionale	Dati	Tecnica di stima	Economie di Scala	Inefficienza di costo (media)
McKay (1988)	Modello strutturale Costi totali Translog	82 CPA 1983 Texas-USA	Sistema equazioni SUR	1,09 -1,12 ^a alla media dei dati (100 letti)	Non rilevata
Vitaliano & Toren (1994)	Modello Behavioural Costi totali Cobb-Douglas	164 CPA 1987 e 1990 New York-USA	SFA: Aigner et al. (1977) ^b Cross-S., ML	1,10 alla media dei dati (146 letti)	Inefficienza: 29% Nessun effetto forma istituzionale
Blank & Eggink (2001)	Modello Behavioural con qualità endogena ^c Costi variabili Cobb-D. (qualità) / Translog (Costi)	110 CPA 1984,1987-93 Olanda	Sistema equazioni SUR	Assenti	Modello con qualità endogena 7% Modello con qualità esogena 11%
Crivelli, Filippini & Lunati (2002)	Modello Behavioural Costi totali Translog	886 CPA 1998 Svizzera	SFA: Aigner et al. (1977) ^d Cross-S., ML	1,04 alla mediana dei dati (48 letti)	Inefficienza: 13% Nessun effetto forma istituzionale e regolamentazione
Rebba e Rizzi (2002)	Modello Behavioural ^e Costi totali Translog ^f	61 CPA 2000 Veneto-IT	Sistema equazioni SUR Frontiera COLS	1,09 alla media dati (68 letti)	Inefficienza: 40%
Christensen (2004)	Modello Ibrido con qualità esogena Costi variabili	600 CPA ^g 1997 Illinois-USA	Regressione a quantili (10 decili)	1,28 solamente nel 1° decile (63 letti)	Non rilevata
Farsi, Filippini & Kuenzle (2005)	Modello Behavioural Costi totali Translog	36 CPA 1993-2001 Ticino-CH	6 Frontiere Stocastiche Panel Data ^h	1,08-1,33 ⁱ alla mediana dei dati (59 letti)	Inefficienza: FE 20% RE 9% TRE-Mundlak 4%

^a i valori si riferiscono a due specificazioni diverse (senza qualità e con qualità)

^b la distribuzione del termine asimmetrico è una normale troncata

^c un modello con qualità esogena è usato per confrontare le stime

^d la distribuzione del termine asimmetrico è esponenziale

^e l'output è una misura composita corretta con un vettore di caratteristiche qualitativo-ambientali

^f un modello translog non ristretto e due specificazioni Cobb-Douglas sono confrontate al modello principale

^g il modello è stimato anche per 3 sottogruppi di output: solo SC (68) solo IC (121) ed entrambe (411)

^h i modelli (FE, RE, RE-Mundlak, Aigner et al. (1977), True RE e True RE-Mundlak) sono stimati con diverse tecniche

ⁱ valori riferiti al modello che esprime le minori (Pooled SF) e le maggiori (FE e RE-Mundlak) economie di scala

Un elemento che accomuna gli studi empirici nel settore dell'assistenza agli anziani è la difficoltà a considerare in modo soddisfacente l'eterogeneità che caratterizza, seppur in modo diverso da paese a paese, il processo produttivo di una casa per anziani, come pure l'eterogeneità nel profilo degli ospiti.

Le differenze tra gli istituti si riferiscono soprattutto al profilo sanitario degli ospiti, al livello qualitativo delle cure erogate ed al ventaglio di servizi presenti nella struttura. La presenza di questi fattori, che neppure il rilevamento di statistiche sempre più dettagliate permette di cogliere pienamente, costituisce un problema dal punto di vista econometrico. L'eterogeneità del settore si riflette in una grande variabilità dei costi che, però, il modello non è in grado di rilevare correttamente se non si dispone di un'adeguata misura numerica di questi elementi qualitativo-ambientali.

In letteratura per contenere gli effetti di questo problema si possono individuare sostanzialmente due approcci: (1) definire delle variabili quantitative che approssimino in modo ragionevole una parte di questa eterogeneità, (2) ammettere la presenza di eterogeneità non osservabile e considerarla nella specificazione econometrica. Un secondo problema comune da considerare è se le variabili incluse nella regressione si possono considerare esogene, come richiesto per la corretta stima di una funzione di costo.

Nella rassegna abbiamo evidenziato come il problema di una possibile endogeneità di alcuni regressori, in particolare l'output e la qualità, vada affrontato in modo pragmatico, analizzando il contesto empirico in cui ci si muove. In particolare, la presenza di una regolamentazione marcata nel settore da analizzare può avvalorare l'ipotesi di esogenità dell'output e della qualità. La conoscenza delle caratteristiche specifiche del mercato dell'assistenza agli anziani permette quindi di definire nel migliore dei modi anche la strategia econometrica da seguire. Un altro nodo da sciogliere è quello della funzione obiettivo da attribuire alle case per anziani che operano in un settore caratterizzato dalla compresenza di imprese a scopo di lucro, imprese pubbliche e istituzioni non profit. In termini pratici questo significa dover decidere se la minimizzazione dei costi è un comportamento realistico e se considerare un orizzonte temporale di breve o di lungo periodo, decidendo rispettivamente tra la stima di una funzione di costo variabile o di costo totale.

Un punto, invece, sul quale l'evidenza empirica è contrastante riguarda la determinazione delle economie di scala in questo settore. Tutti gli studi, riconoscendo l'elevata intensità di lavoro che caratterizza l'erogazione del servizio di assistenza agli anziani, ipotizzano una scarsa rilevanza delle economie di scala a partire da una certa dimensione, oltre la quale il beneficio di condividere alcuni servizi strutturali imprescindibili (lavanderia, cucina, presidi medici e attrezzature, spazi comuni, etc.)

scompare. Dal punto di vista empirico la stima di una funzione di costo permette di stimare quale sia tale soglia nel campione di dati analizzato e, da un confronto dei diversi studi, se esista un'indicazione generale. Nel paragone tra i diversi lavori occorre, tuttavia, considerare tre elementi: (1) il calcolo delle economie di scala dipende dalla correttezza delle stime che possono essere sensibili alla tecnica econometrica considerata, (2) la dimensione media delle case per anziani può essere molto diversa e (3) le indicazioni fornite da una funzione di costo approssimata localmente potrebbero risultare poco attendibili lontano dal punto di approssimazione. Queste considerazioni suggeriscono cautela nel confronto dei risultati empirici di studi condotti sul medesimo settore, ma relativi a situazioni diverse, analizzate con metodologie simili (ma non uguali) e con obiettivi scientifici non sempre coincidenti.

Alla luce di questa panoramica sugli studi empirici, risulta ora più semplice analizzare le scelte compiute in questo lavoro empirico in merito alla specificazione del modello di costo per una generica casa per anziani svizzera.

5.2 Il processo produttivo di una generica casa per anziani

Secondo una rappresentazione molto generale e sintetica del settore delle case per anziani potremmo dire che tutti gli istituti sono uguali tra loro, mentre se osservassimo il problema con grande attenzione al dettaglio arriveremmo ad identificare molte differenze tra le diverse case per anziani. Si apre, dunque, uno spazio di dibattito sulla corretta definizione di un modello, che possa rappresentare l'intero settore delle case per anziani.

Il nostro interesse si rivolge alla formulazione di un modello di costo relativamente semplificato che sappia cogliere gli aspetti fondamentali dell'attività di produzione degli istituti per anziani senza tuttavia trascurare le differenze osservabili in alcune caratteristiche importanti. Il processo di definizione del modello di costo è complicato dal fatto che non ci si ferma alla semplice enunciazione teorica dei fattori che spiegano i costi delle case per anziani, ma si compie un passo in più, determinando empiricamente il valore dei parametri in gioco. Si verifica, così, una continua interazione tra la rappresentazione teorica desiderata e la disponibilità dei dati per effettuare le stime, ed è per questo motivo che illustreremo il procedimento empirico che ha generato i risultati di questa ricerca, avendo cura di sottolineare tanto l'aspetto

teorico della formulazione del modello, quanto la rilevanza della banca dati che ci ha permesso di stimarlo.

In questo senso è importante sottolineare come, nel processo di scelta del modello di costo, si debbano fronteggiare due ordini di problemi: il primo è la concreta disponibilità di informazioni (dati), il secondo è il grado di dettaglio che si vuole attribuire al modello. Mentre la disponibilità di dati o la presenza di alcuni vincoli sulle informazioni statistiche è chiaramente un problema esogeno che sfugge alla possibile influenza del ricercatore, il grado di dettaglio espresso da un modello di costo è una variabile discrezionale che riflette gli obiettivi dell'analisi e la visione del ricercatore.

Il presente lavoro si prefigge di analizzare e di descrivere l'attività produttiva di una casa per anziani rappresentativa per tutta la Svizzera. Questo significa affrontare il problema della definizione del modello di costo da un'ottica generale. La scelta di essere generali, porta con sé alcune controindicazioni che riguardano, in primis, le necessarie semplificazioni da introdurre nel modello al fine di renderlo "universale" rispetto al campione di dati da osservare. Lo svantaggio di scegliere un modello generale, che rispecchi le caratteristiche di tutte le unità produttive considerate, riguarda la difficoltà nel cogliere tutte le differenze che certamente esistono nelle diverse case per anziani.

L'eterogeneità presente nel settore delle case per anziani svizzere si manifesta attraverso le numerose dimensioni, rispetto alle quali possiamo descrivere gli istituti. Per essere più concreti possiamo fare un esempio confrontando due case per anziani sulla base del solo criterio del numero di posti letto offerti. In questo modo potremmo concludere che non vi sono differenze tra i due istituti se riscontriamo la medesima capacità produttiva. Sappiamo bene che in questo semplice confronto dei letti sono state trascurate numerose altre dimensioni valutative che potrebbero essere importanti come: il grado di occupazione dei letti, l'offerta di servizi ai residenti, il personale (quantità e qualità), la collocazione geografica dell'istituto, la forma istituzionale dell'istituto, le modalità di finanziamento, il profilo sanitario dei residenti, etc.

Per gli scopi di questo studio, occorre confrontare le case per anziani sul piano dei costi e delle diverse caratteristiche di ogni istituto attraverso un modello, sufficientemente dettagliato da considerare il problema dell'eterogeneità del campione, ma relativamente semplificato per poter essere generalizzato a tutta la Svizzera. Questa

doverosa considerazione consente di valutare con il giusto spirito critico le scelte operate nella costruzione del modello di costo che sarà poi utilizzato per compiere tutte le analisi econometriche. L'attività produttiva di una casa per anziani si riflette nei costi (totali) che genera e che possono essere descritti attraverso una funzione di costo.²⁰³

Nello spirito di una funzione di costo neoclassica in senso ampio, per esprimere un sufficiente realismo, il modello di costo di una generica impresa rappresentativa del settore dovrebbe includere diverse dimensioni.²⁰⁴ Le variabili da considerare nella specificazione dell'attività produttiva si possono raggruppare nelle seguenti 4 tipologie:

- attività di produzione (misura degli output e delle sue caratteristiche qualitative);
- prezzi dei fattori produttivi impiegati;
- fattori caratteristici individuali;
- fattori "ambientali".

Dalla formulazione generale del modello di costo di una casa per anziani occorre passare, declinando con una serie di variabili specifiche le 4 categorie menzionate, ad una specificazione vera e propria della funzione di costo. La descrizione del processo produttivo risulta quindi conforme all'approccio neoclassico in senso ampio, poiché ai parametri strutturali (output e prezzi degli input) si aggiungono delle variabili esplicative per caratterizzare più fedelmente il comportamento degli istituti.

5.2.1 La misura dell'attività produttiva

Il livello produttivo, l'output, è la prima variabile sulla quale occorre riflettere, poiché risulta fondamentale definire con precisione cosa si sta producendo e misurarne la quantità. In Svizzera, il settore delle case per anziani, come già chiarito nel capitolo 2, si presenta come un insieme di istituzioni sufficientemente eterogenee (ad esempio in termini di dimensioni, profilo degli anziani, forma istituzionale, dotazioni mediche e strutturali, etc.) da legittimare una riflessione sul modo di misurarne l'attività. La descrizione qualitativa, sia essa generale o minuziosa, del settore delle case per anziani non si presta alla definizione di un modello econometrico, dato che queste tecniche

²⁰³ Per la discussione sulle proprietà della funzione di costo neoclassica si rimanda al capitolo 4.

²⁰⁴ Si veda Jacobs et al. (2006) per una presentazione dei problemi teorici e pratici di specificare il modello produttivo di un'impresa che opera nel settore sanitario.

devono esprimere relazioni di tipo quantitativo tra le variabili in gioco. Si tratta, quindi, di individuare una o più variabili che si possano misurare numericamente e che rappresentino in maniera adeguata l'attività produttiva dei singoli istituti.

Alcuni validi candidati, per cogliere il fenomeno nella sua accezione più generale possono essere: il numero di giornate di residenza registrate nel corso di un anno, il numero di persone ospitate, il numero di letti disponibili.²⁰⁵ Le giornate offerte da un istituto costituiscono senza dubbio un'indicazione sintetica e rappresentativa della produzione di servizi che una casa per anziani garantisce ai suoi ospiti. La presenza di residenti è infatti strettamente connessa ad altre attività che si possono rilevare indirettamente, attraverso la misurazione del numero di giornate. Tutti i servizi alberghieri di una casa per anziani, le attività infermieristiche, l'amministrazione ed i servizi tecnici sono intrinsecamente legati alle giornate di residenza che, pertanto, si utilizzano come indicatore della quantità di output prodotto. Il numero di residenti ospitati rappresenta una seconda possibile misura dell'output dell'istituto poiché, anche in questo caso, un maggior numero di persone presenti è, quantitativamente parlando, un chiaro indice di maggior attività.

Una differenza tra le due misure, importante dal punto di vista teorico, consiste nella maggior precisione offerta dal numero di giornate rispetto al numero di residenti che indicano semplicemente quanti ospiti hanno soggiornato nell'istituto senza cogliere la durata del loro ricovero. In realtà, nel settore delle case per anziani, dove i soggiorni degli ospiti superano quasi sempre la durata di un anno e si registrano raramente soggiorni brevi, enumerare gli ospiti o contare le giornate di residenza risulta sostanzialmente indifferente.

Un altro indicatore che potrebbe misurare l'output di una casa per anziani è rappresentato dal numero di letti disponibili. A ben vedere, questa misura coglie un aspetto ancora diverso e precisamente la capacità produttiva di un istituto. In linea di principio, sapere che la disponibilità di letti è grande non fornisce molte indicazioni sull'attività produttiva dal momento che non si conosce il grado di occupazione dei letti. In questo senso il numero di letti è un ottimo indicatore della dimensione dell'istituto e quindi della sua scala di produzione, mentre non sembra in grado di descrivere in modo

²⁰⁵ Per una discussione approfondita sui problemi legati alla scelta dell'output si veda Breyer (1987).

preciso l'effettivo volume di attività svolto. In Svizzera, il tasso di occupazione dei letti è molto vicino al 100% e quindi, anche in questo caso, la misura dell'output ottenuta mediante il numero di letti fornisce indicazioni molto simili a quelle generate dalle giornate di residenza e dal numero di ospiti. Tra le misure dell'output sopra elencate, che presentano caratteristiche diverse, quella maggiormente adeguata per il nostro tipo di analisi empirica sembra essere il numero di giornate di residenza.

Come detto, questa variabile non è in grado di cogliere tutte le sfumature dell'attività produttiva, ma rappresenta un indicatore preciso e comparabile dell'output delle case per anziani svizzere. Limitare la descrizione dell'attività di un istituto per anziani ad un unico indice di produzione potrebbe costituire una semplificazione eccessiva della realtà e sembra legittimo considerare altre possibili dimensioni nella misurazione dell'output prodotto da una casa per anziani. Molti istituti offrono, oltre al soggiorno, cure mediche, terapie riabilitative, attività sociali e ricreative, attività di sostegno religioso e psicologico, servizi accessori (farmacia, bar, parrucchiere, manicure, chiosco, ...). Ognuna di queste prestazioni si potrebbe misurare ed includere nel modello di costo, ma prima di farlo è bene considerare se questi servizi erogati costituiscano veramente un output, economicamente parlando, o se invece rappresentino semplicemente delle caratteristiche qualitative dell'unico prodotto offerto dalla casa per anziani: la giornata di residenza.

La differenza che esiste tra un'impresa che produce un unico prodotto (definita in gergo tecnico impresa single-output) e più prodotti congiunti (impresa multiple-output) è sostanziale da un punto di vista della teoria economica. Nel settore sanitario, ad esempio, è facile riscontrare questa caratteristica negli ospedali, che producono cure in ambiti clinici molto diversi tra loro (neurologia, ortopedia, maternità, pronto soccorso, ...) organizzandosi secondo un modello aziendale multiple-output. Si può parlare di produzione congiunta solo quando una singola impresa offre sul mercato beni e servizi che potrebbero essere prodotti da diverse imprese single-output.

Nel caso degli ospedali potremmo tranquillamente osservare la presenza di tante cliniche (imprese) indipendenti quante sono le specialità che di norma troviamo riunite

sotto un'unica organizzazione.²⁰⁶ Nel settore delle case per anziani, oltre al prodotto “giornata di residenza”, sembra difficile individuare un servizio che, secondo il ragionamento appena svolto, si possa affrancare dal contesto organizzativo dell'istituto. Non tanto perché sia difficile immaginare un'impresa che produce in modo esclusivo i servizi accessori che troviamo in una casa per anziani (riabilitazione, caffetteria, parrucchiere, farmacia, etc.), quanto per la natura strettamente funzionale al soggiorno in istituto, con la quale detti servizi vengono concepiti e organizzati.²⁰⁷

Le numerose attività collaterali svolte da una casa per anziani non possono e non devono essere ignorate nella formulazione del modello dal momento che impiegano risorse che si riflettono sui costi totali; tuttavia, sembra preferibile modellare l'attività produttiva dell'istituto secondo una logica single-output. L'eterogeneità presente nel settore verrà, per quanto consentito dai dati, considerata includendo nel modello una serie di variabile che specificano meglio la dimensione qualitativa dell'output o altre caratteristiche rilevanti.

5.2.2 La misura della qualità e delle caratteristiche ambientali

Un modello di costo neoclassico “in senso ampio” include, oltre alle variabili economiche strutturali (gli output e i prezzi dei fattori produttivi), un vettore di variabili che permettono di cogliere alcune caratteristiche importanti e/o situazioni ambientali particolari della singola osservazione. Nel settore delle case per anziani molti elementi giocano un ruolo essenziale per descrivere correttamente i servizi erogati dall'istituto e per contribuire a spiegarne i costi. Certamente il livello di qualità è un fattore determinante che, come visto nella rassegna dei lavori empirici, desta grande attenzione durante la fase di specificazione del modello di costo.

Risulta utile chiarire, per sommi capi, cosa si intenda per qualità nel settore dell'assistenza agli anziani, in modo da verificare anche l'esistenza di indicatori adatti alla misurazione empirica di questa variabile. Donabedian (1980; 1990), discutendo

²⁰⁶ In economia i vantaggi di costo che derivano dalla produzione congiunta di due o più beni si definiscono economie di scopo e sono alla base di molte scelte produttive multiple-output, specialmente nel settore sanitario, delle network utilities (elettricità, gas, acqua) e della formazione (università e politecnici).

²⁰⁷ Da un punto di vista aziendale è un problema di “make” or “buy”, secondo il quale la decisione di dotarsi internamente di un servizio (*make*) è preferibile all'acquisto da terzi dello stesso servizio (*buy*) a seconda delle caratteristiche dell'istituto (dimensione, profilo sociosanitario degli ospiti, etc.).

della qualità e delle tecniche per la sua valutazione nel settore sanitario, individua i tre pilastri costitutivi della qualità delle cure: (1) struttura, (2) processo e (3) risultato.

La qualità strutturale, nelle case per anziani si traduce sostanzialmente nel *comfort* alberghiero offerto ai residenti (accessibilità alla struttura, dimensione e stato di usura delle camere, qualità della mensa, disponibilità di spazi attrezzati, presenza di servizi accessori, etc.). Questo elemento, molto importante per definire il livello della qualità percepita dai residenti, è sicuramente un elemento che incide sui costi di produzione e che può essere misurato in termini quantitativi con opportuni indicatori.²⁰⁸ Molto spesso, ed è il caso della Svizzera, nel settore dell'assistenza agli anziani le normative sugli aspetti strutturali sono molto rigidi, obbligando al rispetto di alcuni standard minimi di qualità.

Per qualità nel processo si intende sia l'aspetto tecnico (appropriatezza delle procedure seguite) sia l'aspetto relazionale (soprattutto con l'utenza). Nei servizi alla persona, ed in particolare nelle case per anziani, la competenza (formazione) del personale ed il clima lavorativo sono elementi decisivi, uniti ad una gestione attiva della qualità (es. sistemi *Total Quality Management* o TQM).²⁰⁹ Questi aspetti sono più difficoltosi da misurare in modo accurato, ma non significa che non si possano trovare degli indicatori quantitativi da inserire in un modello di regressione. La regolamentazione del settore delle case per anziani in Svizzera, in molti cantoni, disciplina con dovizia di particolari le dotazioni di personale e le rispettive qualifiche, con l'obiettivo di fornire una qualità di processo minima nelle strutture pubbliche (e private al beneficio di un sussidio).

Il terzo elemento della qualità, il risultato, è alla fine ciò che interessa di più. In ambito medico, ovviamente la qualità di una prestazione si giudica dalla capacità di ripristinare lo stato di salute antecedente all'insorgere del problema trattato. Nell'assistenza agli anziani si può adattare questo concetto misurando la capacità del sistema casa per anziani di surrogare lo stile di vita degli ospiti pre-ricovero. Appare

²⁰⁸ Per un approfondimento sul tema della valutazione della qualità si vedano Gori e Vittadini (1999) e Franci e Corsi (1996; 2004).

²⁰⁹ Per una breve discussione sui sistemi di qualità in ambito sanitario si veda Pagano e Rossi (1999).

chiaro che il concetto di risultato deve essere inteso in senso relativo, giudicando in ultima analisi la qualità di vita offerta dall'istituto ai propri residenti.²¹⁰

La specificazione della qualità in un modello di costo è quindi soggetta alle normali difficoltà di reperire dei buoni indicatori (e i relativi dati), con l'aggravante di essere un attributo composito che incide sui costi di produzione senza tuttavia lasciarsi misurare in modo preciso. A questo riguardo nei lavori empirici la misura della qualità, nelle sue diverse accezioni, è generalmente approssimata con una o più variabili quantitative.

Laine et al. (2005), ad esempio, nel tentativo di spiegare l'efficienza tecnica con le differenze nel profilo qualitativo dei servizi di assistenza agli anziani erogati nei reparti di geriatria e negli istituti per anziani finlandesi, utilizzano tre indicatori clinici: piaghe da decubito, uso di calmanti e sonniferi, depressioni non trattate. Avendo accesso a dati molto disaggregati sulla realtà delle case per anziani giapponesi, Shimizutani and Suzuki (2007) propongono un indice sintetico di qualità basato su 12 items (valutati su 4 aspetti) che coprono le diverse componenti della qualità (strutturale, processo e risultato).²¹¹ Anderson et al. (2005), con riferimento agli Stati Uniti, descrivono gli indicatori qualitativi utilizzati nelle case per anziani di 6 Stato, nell'ambito di un programma pilota (*Nursing Home Public Reporting Quality Initiative*) condotto nel 2002. Si tratta di 11 indicatori (frequenze di eventi) per soggiorni prolungati e 3 indicatori per soggiorni brevi.²¹²

Da questi pochi esempi si capisce come la qualità sia un elemento difficile da ignorare, ma impossibile (almeno con i normali dati a disposizione) da misurare in modo preciso. Per questo motivo è forse l'elemento più appariscente dell'ampia categoria di variabili che si possono affiancare alle variabili strutturali (output e prezzi degli input) per rendere il modello di costo più realistico.

²¹⁰ Sul concetto di qualità della vita e dei problemi di misurazione si veda Zwinderman (1990).

²¹¹ I 12 aspetti considerati sono: stato di manutenzione, gestione del personale, formazione interna offerta, qualifiche ed esperienza del personale, localizzazione e accesso al servizio, politiche orientate al cliente, sistema di informazione e gestione dei reclami, protezione privacy, sistema di gestione delle emergenze, profilassi e norme igieniche, trasparenza gestionale, chiarezza amministrativa.

²¹² In questo caso gli indicatori si riferiscono esclusivamente alla qualità di processo, considerando ad esempio: piaghe da decubito, cateterizzazioni, infezioni all'apparato urinario, episodi di delirio, etc.

Di seguito discutiamo brevemente altri due fattori rilevanti per migliorare l'accuratezza della specificazione di un modello produttivo per il settore delle case per anziani: (1) il profilo degli ospiti e (2) i fattori "ambientali".

Il profilo degli ospiti (età media, livello di autonomia psicofisica, stato di salute, etc.) incide chiaramente sulla quantità di risorse assorbite da una giornata di residenza. Sembra quindi indispensabile una correzione del "case-mix" produttivo, quando si devono confrontare i costi di due istituti diversi. La disponibilità e la precisione di questa misura dipendono molto dal tipo di dati a disposizione, anche se è piuttosto raro non riuscire ad individuare almeno una variabile in grado di approssimare il fenomeno.

L'ultimo gruppo di variabili che è opportuno considerare riguarda la presenza di fattori "ambientali" che potrebbero incidere sulle condizioni di produzione, facendo crescere (o diminuire) i costi dei singoli istituti. Per fattori ambientali si intende qualunque tipo di vincolo o condizione particolare che non possono essere influenzati e/o decisi dalla casa per anziani ma che, in compenso, condizionano il processo produttivo. Queste variabili esogene, che spesso si presentano come variabili dicotomiche o categoriche (discrete), si riferiscono ad esempio alla posizione geografica (città, campagna, valle, etc.), all'assetto proprietario (istituto pubblico, privato non-profit, privato for profit), al regime di regolamentazione (appartenenza a questo o quel cantone), insomma a tutti gli aspetti che potrebbero, se ignorati, falsare il confronto tra le prestazioni degli istituti per anziani svizzeri.

Includere nel modello di costo un vettore di caratteristiche non solo contribuisce a migliorare la descrizione del processo produttivo e ad affinare la precisione delle stime dei parametri, ma riduce la quota di eterogeneità che finisce nei residui della regressione, dai quali, come spiegato nella sezione metodologica, si ricavano i punteggi di inefficienza degli istituti. Dopo aver indirettamente rafforzato l'importanza di una specificazione neoclassica "in senso ampio", ritorniamo alla parte strutturale (neoclassica) del modello di costo per discutere brevemente il problema di misurare i prezzi dei fattori produttivi.

5.2.3 La misura dei prezzi dei fattori produttivi

Rappresentare una casa per anziani come un'impresa che produce "giornate di residenza", significa identificare tutti gli input necessari per garantire la produzione desiderata. I fattori produttivi ed i beni intermedi che occorrono ad un istituto per

svolgere la propria attività sono numerosi e variegati, ma possiamo tentare un raggruppamento delle principali voci nelle seguenti quattro categorie:

- capitale: immobili, mobilia, attrezzature, apparecchiature mediche, etc.;
- lavoro: personale medico, infermieristico, tecnico, amministrativo, etc.;
- energia: elettricità, nafta, gas, etc.;
- materiale: alberghiero, medico-sanitario, economato, etc.

La formulazione del modello di costo richiede la presenza dei prezzi dei fattori produttivi che hanno contribuito a generare un dato output. Questo significa che, dal punto di vista teorico, la funzione di costo da stimare potrebbe contenere un vettore dei prezzi molto numeroso. Se tutti i mercati dei fattori fossero perfettamente concorrenziali, ogni istituto potrebbe acquistare al medesimo prezzo: terra, mano d'opera, energia, materiali, etc., ma poiché questa situazione si verifica solamente nei manuali di economia, le differenze nei prezzi degli input esistono. Per poter confrontare tra loro in modo corretto tutti gli istituti risulta, dunque, necessario disporre di informazioni sui prezzi dei fattori. La scelta dei fattori da impiegare nella funzione di costo deve garantire che i prezzi degli input siano sufficientemente confrontabili ed è per questo motivo si tende a prediligere l'uso poche classi generali di fattori produttivi.

In Svizzera, il prezzo di alcuni input utilizzati dalle case per anziani per la produzione è molto eterogeneo poiché riflette le particolari caratteristiche di domanda e offerta del relativo mercato di riferimento. Un esempio molto chiaro è rappresentato dal prezzo degli immobili (valore del terreno) che si determina sul mercato locale, secondo dinamiche molto diverse. Il prezzo di un terreno situato in una zona densamente popolata (centro città, cantoni urbani), sarà certamente molto più elevato rispetto al prezzo di un terreno corrispondente (per proprietà urbanistiche) ubicato in piena campagna. Per contro il prezzo di alcuni input è relativamente omogeneo a livello nazionale, poiché il mercato di riferimento presenta condizioni simili o, al limite, si tratta dello stesso mercato. Le apparecchiature mediche e tecniche, i medicinali, l'olio combustibile sono alcuni esempi di fattori produttivi per i quali è lecito attendersi una sostanziale omogeneità di prezzo.

Questa precisazione in merito alla possibile eterogeneità nel prezzo degli input è inutile per quanto riguarda i problemi di specificazione del modello teorico di costo, nel quale andrebbero inseriti tutti i prezzi dei fattori di produzione importanti. La

discussione assume, invece, un ruolo centrale se, come spesso accade nella ricerca empirica, si devono operare alcune scelte di specificazione del modello senza disporre di tutti i dati necessari. In questo frangente è molto utile presagire le conseguenze di un'omissione di una variabile sulla capacità del modello di spiegare le differenze nei costi.

Immaginiamo, per ipotesi, di voler includere nel modello il prezzo dell'energia elettrica ma di non avere alcun dato in merito. Se il prezzo dell'elettricità fosse costante in tutta la Svizzera (ad esempio fissato per legge come il prezzo delle sigarette), si potrebbe escludere il prezzo dell'elettricità dal modello empirico, dal momento che non darebbe alcun contributo nella spiegazione delle differenze di costo.²¹³ Appare chiara l'importanza di scegliere perlomeno quei fattori che possono contribuire a spiegare i costi di una casa per anziani tanto dal punto di vista teorico (a priori) quanto da quello empirico (a posteriori, osservando i risultati delle stime).

Da notare che, nella maggior parte dei casi, i prezzi dei fattori produttivi non sono osservabili e vengono costruiti dal ricercatore, derivandoli dalle informazioni contenute nella banca dati. Questa procedura è concettualmente molto semplice e consiste nel dividere la spesa monetaria riferita ad un certo fattore produttivo (ad esempio il monte salari) per le quantità fisiche impiegate (es. numero di lavoratori a tempo pieno). Da un punto di vista empirico, la corretta misurazione del prezzo del capitale costituisce un tema molto discusso in letteratura, che desideriamo approfondire.²¹⁴

Come per tutti gli altri input, il prezzo del capitale si può calcolare dividendo il valore monetario delle spese attribuite al capitale per una misura dello stock fisico di capitale, ottenendo così il prezzo di un'unità fisica di capitale. Tuttavia, in letteratura si riconoscono serie difficoltà metodologiche e pratiche nel formulare il prezzo del capitale per scopi empirici.²¹⁵ Occorre allora precisare: (1) quali spese si debbano attribuire al capitale e (2) come definire un indicatore per la misura dello stock di capitale. Da un punto di vista economico-contabile si possono considerare come spese

²¹³ Una variabile esplicativa che assume sempre lo stesso valore per tutte le osservazioni è, di fatto, una costante e non può essere inclusa nella stima di un modello di costo. In questo caso l'effetto della variabile omessa sarà catturato dalla costante del modello.

²¹⁴ Si vedano, ad esempio, Friedlander and Wang Chang (1983), Filippini e Maggi (1993), Filippini (1997) e Coelli et al. (2003).

²¹⁵ Per una discussione generale sul tema si veda ad esempio Jacobs et al. (2006).

di capitale gli ammortamenti del patrimonio immobiliare e mobiliare (deprezzamento dei cespiti aziendali), le spese di affitto e di leasing, le spese di manutenzione, le spese per macchinari ed attrezzature, il servizio del debito per finanziare l'acquisto di beni capitali, il costo opportunità del capitale proprio, etc. Solitamente queste informazioni, che si possono facilmente reperire dai conti economici delle imprese, sono disponibili in forma piuttosto aggregata.

Il vero problema è però costituito dalla misura dello stock di capitale, per la quale esistono tre vie empiriche praticabili: il valore monetario a bilancio, un indicatore fisico dello stock di capitale, il metodo dell'inventario permanente. Quest'ultimo approccio, suggerito da Jorgeson (1967, 1974), consiste nella ricostruzione, attraverso le informazioni desunte dai documenti contabili, del flusso di capitale dell'impresa e presenta l'indubbio vantaggio della precisione. Per contro, richiede dati molto particolareggiati sul patrimonio e sugli investimenti compiuti che, spesso, sono difficili da reperire. In alternativa, l'utilizzo di un indicatore (fisico o monetario) dello stock di capitale, in rappresentanza di tutte le "unità" di capitale fisico impiegate dall'impresa nel processo produttivo, rappresenta una via empiricamente percorribile, nonostante presenti qualche controindicazione.

La mancanza nella banca dati delle informazioni sui valori a bilancio rende inutilizzabile l'approccio della misura monetaria dello stock di capitale, suggerendoci di discutere in questa sede solamente l'utilizzo degli indicatori fisici. A questo proposito occorre sottolineare come il processo produttivo di una casa per anziani si focalizzi sul soggiorno e (se necessario) sulla cura degli ospiti. Per questo motivo sembra adeguato, e sufficientemente preciso, riferire tutte le spese di capitale (attrezzature, macchinari, apparecchi sanitari, immobili, etc.) al numero complessivo di letti. Da un punto di vista concettuale la scelta dei letti quali indicatore dello stock fisico di capitale è soddisfacente. Tuttavia, sotto il profilo empirico, si pone il problema, della correlazione tra l'indicatore della capacità produttiva (letti) e la misurazione dell'output (giornate di residenza). Questo problema, che è molto frequente nelle analisi empiriche basate sulla stima di funzioni di produzione e/o di costo, non si può mai risolvere completamente, e va quindi accettato.

Scegliendo il numero di letti come indicatore dello stock di capitale, il prezzo del capitale esprime il costo che ogni istituto ha sopportato per attrezzare, equipaggiare e predisporre all'erogazione del servizio di residenza, ogni singolo letto offerto agli ospiti

della casa per anziani. In letteratura questo concetto è noto con il termine di “*staffed beds*” e sta ad indicare che la misura dello stock di capitale scelta (il numero di letti) funge da denominatore comune a tutti gli impieghi di capitale necessari nella produzione del servizio residenziale.²¹⁶

Da questi esempi, che saranno ripresi e completati durante la discussione sulle singole variabili utilizzate nel modello (paragrafo 5.3), traspare come la progettazione teorica del modello di costo sia molto importante. Il ricercatore deve infatti ragionare sul processo produttivo tipico di una casa per anziani per distinguere gli elementi comuni ed imprescindibili da quelli particolari ed accessori. Inoltre, occorre valutare in termini pratici la fattibilità delle proprie decisioni, individuando le variabili concretamente candidate a descrivere l’attività di un istituto per anziani rappresentativo per tutta la Svizzera.

Il modello di costo teorico, prima di “trasformarsi” nel modello effettivamente utilizzato per le stime, deve superare la prova delle informazioni statistiche, poiché (caso estremo) senza dati non è possibile effettuare alcuna stima! La mancanza di dati può essere una ragione molto convincente per escludere una variabile dal modello, ma non prima di aver soppesato attentamente le implicazioni di una tale omissione e ricordando l’importanza di questa scelta, in sede di interpretazione dei risultati. Questo tipo di problema, che riguarda soprattutto la dimensione qualitativa per la quale è sempre difficile trovare degli indicatori precisi ed affidabili, non è risolvibile in modo “esatto” e lascia al ricercatore il compito di trovare la miglior approssimazione, compatibile con le informazioni ed i dati in suo possesso.

Le informazioni contenute nei dati svolgono quindi un ruolo chiave, ponendo vincoli e condizionando le scelte di specificazione del modello di costo. Pertanto, di seguito, discuteremo la banca dati a nostra disposizione che, nonostante presentasse alcune problematiche (valori mancanti in alcune informazioni chiave e aggregazione piuttosto ampia delle informazioni finanziarie), ha permesso di formulare un modello di costo in grado di descrivere efficacemente il processo produttivo di una casa per anziani rappresentativa.

²¹⁶ Si veda ad esempio Wagstaff (1989).

5.2.4 La banca dati disponibile ed il suo contenuto informativo

Il cardine di un'analisi empirica è costituito dai dati che si hanno a disposizione per verificare gli obiettivi della propria ricerca. In alcuni casi le osservazioni vengono appositamente raccolte attraverso indagini statistiche mirate per gli scopi della ricerca, mentre in altre circostanze i dati sono preesistenti, in quanto raccolti nell'ambito di progetti diversi ed indipendenti da quello in esame.

I dati statistici utilizzati in questa ricerca rientrano in quest'ultima fattispecie, poiché sono derivati da una banca dati costruita a partire dal 1997 dall'Ufficio Federale di Statistica (UFS), il quale ha dato vita alla pubblicazione annuale denominata "Statistiche degli stabilimenti sanitari non ospedalieri (svizzeri)".²¹⁷

La breve descrizione che si trova sul sito dell'UFS recita: *“la statistique des établissements de santé non hospitaliers est une statistique administrative, cherchant à décrire avant tout l'infrastructure et les activités des institutions accueillant des personnes âgées et handicapées. A cadence annuelle, les institutions soumises à l'obligation de renseigner procèdent au recensement de leurs prestations, de la population hébergée, du personnel assurant son accompagnement ainsi que de leurs comptes d'exploitation.”*

L'obiettivo di questa indagine è dichiaratamente di carattere amministrativo e descrittivo e si prefigge di fornire, sia a livello svizzero che cantonale, un quadro delle principali variabili che caratterizzano le strutture e l'attività degli istituti per persone anziane e disabili. Un aspetto fondamentale di questa indagine è rappresentato dall'obbligatorietà che assume nei confronti degli istituti, i quali ogni anno devono trasmettere le informazioni richieste.

Questo vincolo legale garantisce una copertura campionaria del fenomeno osservato molto prossima (99,6% nel 2002) all'intera popolazione statistica degli stabilimenti sanitari non ospedalieri svizzeri, a partire dai quali abbiamo costruito la banca dati che utilizzeremo per stimare la frontiera stocastica di costo per le case per anziani. La base statistica dalla quale siamo partiti è quindi costituita dalle informazioni,

²¹⁷ Tutte le informazioni che riguardano gli scopi, la metodologia e le definizioni utilizzate per questa pubblicazione sugli stabilimenti sanitari non ospedalieri si possono consultare sul sito internet dell'UFS:
http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen_quellen/blank/blank/ssmi/01.html

rilevate a livello di singolo istituto, che l'UFS ci ha messo a disposizione per gli anni 1998-2002. Si tratta di un (*unbalanced*) panel anonimizzato, composto da 11637 osservazioni, relative a 2565 istituti non ospedalieri su un periodo di 5 anni.²¹⁸ Per ogni istituto sono disponibili informazioni che riguardano:

- gli aspetti generali dell'istituto;
- il ventaglio di prestazioni offerte;
- il personale impiegato;
- le diverse tipologie di presa a carico degli ospiti;
- le caratteristiche individuali dell'utenza;
- gli aspetti economico finanziari dell'istituto.

Tra i dati amministrativi rilevati nel questionario dell'UFS, menzioniamo il cantone di appartenenza dell'istituto e la forma istituzionale, che sono due informazioni preziose per analizzare rispettivamente le eventuali differenze regionali e per classificare le case per anziani secondo la tassonomia introdotta nel capitolo 2.

Le prestazioni erogate dagli istituti sono suddivise in 5 categorie (cure medico-infermieristiche, servizi terapeutici, cure personali del corpo, servizi di sostegno socioccupazionale, servizi infrastrutturali), all'interno delle quali è possibile scendere in ulteriori dettagli.

Il personale è suddiviso rispetto al criterio di impiego nelle 5 aree produttive appena menzionate incrociato con il livello di formazione raggiunto dai dipendenti (classificato secondo 4 gradi: universitario, superiore, professionale, base). Inoltre, la registrazione del personale iscritto a libro paga rileva tanto il numero di dipendenti (*head count*) quanto l'equivalente numero di persone impiegate a tempo pieno (in inglese FTE acronimo per *Full Time Equivalent*).

Nella specificazione del tipo di presa a carico vengono descritti: (a) i posti letto disponibili, (b) le giornate di residenza erogate e (c) il numero di ospiti al 31 dicembre, riferiti a 4 grandi domini: (1) anziani e/o malati cronici, (2) psichiatria, (3) portatori di

²¹⁸ Nei dati panel il numero di osservazioni (11637) è dato dalla somma di tutte le osservazioni degli N individui (2565 istituti), ripetute nei T_n periodi (1-5 anni). Se tutti gli istituti fossero rappresentati in tutti i 5 anni, il numero di osservazioni sarebbe dato dal prodotto $N \times T$ pari a 12825.

handicap e (4) altri tipi di ricovero come, ad esempio, il recupero delle tossicodipendenze o dell'alcolismo. La presa a carico di persone anziane è la prestazione che ci interessa maggiormente e contiene un'ulteriore informazione circa la durata del soggiorno (breve se inferiore al mese e lungo altrimenti) e l'eventuale sistemazione degli ospiti in appartamenti protetti.

Le caratteristiche individuali degli ospiti riguardano il sesso, l'età e lo stato di salute. Con particolare riferimento a quest'ultimo aspetto, gli ospiti sono raggruppati in 4 classi di fabbisogno di cure: meno di un'ora, da 1 a 3 ore, da 3 a 5 ore e più di 5 ore di cure ricevute in media al giorno.²¹⁹ Questa informazione, che, come detto in precedenza, (capitolo 2) permette all'UFS di classificare le case per anziani secondo 3 diverse tipologie, ci consentirà di costruire una variabile che misuri il bisogno di cure dell'utenza di ciascun istituto. A completamento del profilo dei residenti ci sono informazioni sulle classi di invalidità (riprese dalle prestazioni del programma di assicurazione sociale contro l'invalidità AI) e sul tipo di eventuale infermità che contraddistingue ciascun ospite.

Agli aspetti economici dell'istituto è dedicata un'intera sezione del questionario che propone un conto economico piuttosto semplificato nel quale tuttavia sono presenti le principali voci di costo e di ricavo. In particolare, vengono misurate, per quanto concerne i costi, le spese complessive per il personale, per gli ammortamenti, per gli oneri finanziari e per le altre spese di esercizio, mentre dal lato dei ricavi troviamo le entrate per rette pagate dagli ospiti, i rimborsi provenienti dall'assicurazione malattia di base (LAMal), le prestazioni sociali a vario titolo (complementari AVS, sussidi per invalidità, etc.) ed infine i sussidi di provenienza cantonale e comunale.²²⁰

Da questa descrizione emerge un quadro abbastanza confortante per la costruzione vera e propria delle variabili da inserire nel modello, dal momento che vi sono tutte le informazioni essenziali per stimare una frontiera di costo, arricchite da alcuni elementi esplicativi utili per catturare, almeno in parte, l'estrema complessità del processo

²¹⁹ Da notare che per il calcolo del fabbisogno di cure sono considerate tutte le attività di sostegno rivolte all'ospite comprendendo: le cure mediche, le cure infermieristiche l'aiuto nelle attività quotidiane (vestirsi, lavarsi, etc), e nella deambulazione.

²²⁰ Il questionario completo, si può consultare nel sito dell'UFS:
http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen_quellen/blank/blank/ssmi/02.ContenPar.0001.DownloadFile.tmp/detailkonzept_sm_2001-11_f.pdf

produttivo che si desidera analizzare. Dopo aver verificato che la banca dati permetta, in linea teorica di costruire le variabili di un modello di costo, è opportuno spingersi oltre e discutere brevemente la qualità dei dati che si hanno a disposizione.

La prima considerazione da fare riguarda il grado di dettaglio delle informazioni, che in molti casi è molto particolareggiato. Avere a disposizione tante variabili per ogni istituto non è sempre un vantaggio, soprattutto se l'obiettivo è stimare una funzione di costo rappresentativa per tutta la Svizzera. A volte, per cogliere l'impatto sui costi di alcune caratteristiche è preferibile considerare un'aggregazione delle singole variabili. Per misurare, ad esempio, la complessità dei servizi offerti da una casa per anziani si potrebbe rilevare la presenza o l'assenza dei singoli servizi ottenendo così un vettore di variabili dicotomiche per ogni istituto. Procedendo in questo modo il numero di variabili esplicative potrebbe diventare molto alto, creando dei problemi econometrici legati ai gradi di libertà e alla correlazione delle variabili. In questi casi appare utile definire un indice che sintetizzi il numero dei servizi offerti in modo tale da cogliere ugualmente le differenze strutturali da un punto di vista quantitativo, senza scendere nel dettaglio.

Un altro aspetto fondamentale per valutare la qualità di una banca dati è il grado di completezza delle informazioni rilevate tramite i questionari dell'UFS. Sapere che il tasso di risposta degli istituti è elevatissimo non garantisce ancora che i questionari siano stati compilati in ogni loro parte. Il fenomeno delle osservazioni mancanti (*missing values*) non è da sottovalutare quando si vuole fare una stima econometrica. In particolare, non è possibile includere nell'analisi quegli istituti che difettano di informazioni chiave rispetto al modello di costo che si intende stimare.

Questo problema è di semplice soluzione da un punto di vista pratico, poiché basta eliminare tutte quelle osservazioni che sono incomplete in una qualsiasi variabile esplicativa. In realtà questo vincolo induce a fare molta attenzione nella specificazione del modello, poiché l'introduzione anche di una sola variabile afflitta dal problema dei *missing values* comporta una drastica riduzione del numero di osservazioni valide per la stima. Da notare che la nostra analisi econometrica non si riferisce all'intera banca dati, ma all'insieme delle case per anziani, così come definite dall'Ufficio Federale di Statistica. Si tratta cioè di quegli istituti nei quali almeno il 50% dei letti autorizzati è destinato ai soggiorni di lunga durata.

In termini numerici questa operazione riduce in modo significativo sia il numero di istituti (che passa dai complessivi 2'565 a 1'535) sia il numero delle osservazioni (che dal totale di 11'637 si abbassa a 6'608). Il numero di istituti da considerare nell'analisi si è ulteriormente ridotto a poco più di 1'300 istituti, dopo aver eliminato le osservazioni con valori mancanti nelle variabili strutturali imprescindibili (output, prezzo del lavoro e prezzo del capitale).²²¹

La disponibilità di un numero sufficientemente alto di osservazioni è un fattore positivo per assicurare una maggior precisione nelle stime e, in questo senso, la base di dati empiricamente utilizzabili è sufficientemente ampia visto che si lavora con circa 1'300 istituti per anziani per un totale complessivo di circa 5'000 osservazioni. Da notare che prima di procedere con le stime si è reso necessario procedere ad un'ulteriore scrematura dei dati sulla base del modello di costo specificato per le analisi.

Una proprietà importante di una banca dati è senza dubbio la qualità delle informazioni contenute nelle singole osservazioni. Questo aspetto è molto difficile da accertare dal momento che riguarda tutto il processo di rilevamento e raccolta dati che non è direttamente osservabile. Considerando le modalità con le quali l'UFS raccoglie annualmente i dati degli istituti sanitari non ospedalieri, si può desumere se le informazioni contenute nella banca dati siano sufficientemente precise per gli scopi dell'analisi. Lo strumento che permette di rilevare i dati è un questionario elaborato dall'UFS unitamente ad un manuale di spiegazione, nel quale vengono precisati gli obiettivi dell'inchiesta, le definizioni adottate, le variabili scelte e chiariti tutti gli aspetti che possono creare dubbi o difficoltà nella compilazione del questionario.

Il grado di dettaglio con il quale sono precisate le variabili, unito alle competenze specifiche dei responsabili degli istituti che devono compilare il questionario ed al carattere obbligatorio della compilazione, lasciano presupporre che la qualità dei dati sia elevata. Questo non significa che omissioni o errori non si possano verificare e non esclude l'errore materiale in fase di compilazione. La preoccupazione sulla qualità dei

²²¹ Il fenomeno dei *missing values* è più moderato nel sottoinsieme delle case per anziani. La definizione utilizzata per le case per anziani richiede, come visto in precedenza, la conoscenza di informazioni quantitativi basate sul numero di letti destinati a persone anziane. Le osservazioni con dati mancanti spesso non riportano l'informazione che permette di classificare un istituto come casa per anziani riducendo la numerosità di questo insieme di osservazioni. Per contro le informazioni relative agli istituti considerati come case per anziani tendono ad essere complete, limitando quindi il problema dei missing values all'interno di un campione già selezionato a monte.

dati, tuttavia, non riguarda quest'ultimo aspetto (che certamente va considerato) quanto piuttosto la ragionevole convinzione che le misure quantitative offerte dai dati contengano le informazioni che il ricercatore si aspetta.

Sulla base delle considerazioni esposte, possiamo dire che la qualità delle informazioni statistiche è buona e ci incoraggia a procedere nell'analisi definendo con precisione il modello di costo da utilizzare per le stime.

5.3 Specificazione del modello di costo utilizzato per le stime

Dopo aver discusso, su un piano ipotetico, le caratteristiche desiderabili di un modello di costo generico per il settore delle case per anziani ed aver analizzato la banca dati a disposizione per questa ricerca, è giunto il momento di presentare nei dettagli la funzione di costo che abbiamo scelto per descrivere l'attività produttiva di una casa per anziani svizzera rappresentativa.

Seguendo la letteratura prevalente in ambito sanitario, abbiamo specificato una funzione di costo neoclassica in senso ampio. Il motivo, come già ampiamente discusso, è legato alla necessità di cogliere almeno parte della complessità del processo produttivo di una casa per anziani. Dai parametri delle stime (e dai residui della regressione) discendono le valutazioni sull'efficienza produttiva, riferita però ad una rappresentazione del settore, la cui fedeltà dipende dalla precisione del modello. Nel settore dell'assistenza agli anziani sembra difficile poter escludere completamente dal modello di costo alcune dimensioni come la qualità, il profilo sociosanitario degli ospiti, le caratteristiche organizzative o altri aspetti particolari. Per questo motivo tra i regressori, in aggiunta all'output (giornate di residenza) ed ai prezzi dei fattori produttivi (lavoro e capitale), abbiamo incluso anche altre variabili che caratterizzano meglio il processo produttivo considerando alcuni attributi qualitativi, il profilo sanitario dei residenti e la posizione geografica dell'istituto.

Considerata l'eterogeneità a livello svizzero, ampiamente discussa nel capitolo 2, in merito all'assetto istituzionale ed ai vincoli regolatori (federalismo), è preferibile, seguendo Bös (1984), specificare una funzione di costo di tipo *behavioural*. Questo tipo di approccio, adottato in precedenza anche da Farsi e Filippini (2004), permette di temperare la rigidità strutturale di una funzione di costo neoclassica, ammettendo la presenza di fattori esterni (ad esempio le regolamentazioni cantonali) e di forme di

inefficienza. La volontà di ottimizzare (nel lungo periodo) gli input, in risposta a variazioni del loro prezzo di mercato, potrebbe scontrarsi con le norme di legge che regolamentano (direttamente o indirettamente) il settore delle case per anziani. Inoltre, l'inefficienza di costo, che è una palese deviazione dal postulato neoclassico di minimizzazione dei costi, si può certamente considerare come un elemento presente nella realtà. In altre parole, con l'approccio *behavioural* si cerca di unire il rigore della teoria economica (neoclassica) della produzione con la complessità del settore delle case per anziani, lasciando che alcuni elementi delle teorie economiche (menzionate nel capitolo introduttivo) possano spiegare il comportamento produttivo, descritto dalla frontiera di costo.

5.3.1 La scelta della specificazione di lungo periodo

Inquadro teoricamente l'approccio seguito nella specificazione del modello, il primo elemento concreto da discutere riguarda la specificazione di una funzione di costo di lungo periodo, utilizzando i costi totali come variabile dipendente, oppure di breve periodo, impiegando nella regressione i costi variabili. Dal punto di vista economico significa attribuire alle imprese analizzate un comportamento volto alla minimizzazione dei costi in assenza (lungo periodo) o in presenza (breve periodo) di vincoli. In letteratura entrambi gli approcci sono ben rappresentati e la decisione di utilizzare i costi totali o variabili può essere dettata dagli scopi dell'indagine, dalla disponibilità dei dati e dalla discrezionalità del ricercatore.²²² In questo lavoro utilizzeremo la funzione di costo totale di lungo periodo, giustificando la scelta con alcune considerazioni di tipo economico ed econometrico.²²³

In primo luogo, l'interesse della ricerca è quello di determinare empiricamente il valore delle (eventuali) economie di dimensione, che sono un concetto intrinsecamente di lungo periodo. Sembra preferibile quindi stimare una funzione di costo totale che descrive il problema produttivo dell'impresa considerando la possibilità di modificare

²²² In ambito sanitario alcuni lavori che hanno specificato la funzione di costo totale sono: Conrad and Strauss (1983), Grannemann et al. (1986), McKay (1988), Gertler and Waldman (1992), Vitaliano and Toren (1997), Filippini (2001), Crivelli et al. (2002), Farsi et al. (2005). Per contro una funzione di costo variabile è stata impiegata da: Cowing and Holtman (1983), Vita (1990), Gaynor and Anderson (1995), Aletras (1999), Blank and Eggink (2001), Christensen (2004).

²²³ In una fase preliminare della ricerca abbiamo considerato anche la specificazione con il costo variabile, optando però per l'impiego della funzione di costo totale.

tutti i suoi input. A ben vedere, utilizzando la [3.14], è possibile ricavare il valore delle economie di scala dai parametri della funzione di costo variabile. Questo approccio, è semplice solo in teoria, poiché non risolve il problema della misura dello stock di capitale (input fisso) e richiede la conoscenza del valore ottimale dell'input fisso per stimare le economie di dimensione.²²⁴

Una seconda ragione per specificare una funzione di costo totale è legata alla natura panel dei dati a nostra disposizione. Potendo osservare le stesse imprese su un periodo di 5 anni sembra preferibile non escludere a priori la possibilità che le imprese si trovino in equilibrio (statico) di lungo periodo. Inoltre, la natura *behavioural* del modello ammette la possibilità di deviare dal costo minimo (inefficienza), pur attribuendo alle imprese l'obiettivo della minimizzazione dei costi. Nel settore delle case per anziani questo significa individuare il tasso di occupazione dei posti letto che permette di minimizzare il costo di offrire un certo numero di giornate di residenza, esogenamente determinato. Quando il tasso di occupazione è vicino al 100%, la risposta produttiva ottimale agli aumenti dell'output è quella di aumentare anche la dimensione degli istituti, operando nuovi investimenti secondo una logica di lungo periodo.²²⁵ Viste le caratteristiche produttive del settore, incentrato sull'assistenza e sulla cura di ospiti che risiedono nella casa occupando un posto letto, è plausibile che il tasso di occupazione che minimizza i costi totali non si discosti molto dal pieno sfruttamento della capacità produttiva che si osserva nei dati.²²⁶

Nonostante il settore delle case per anziani sia caratterizzato dalla presenza di forme istituzionali molto diverse tra loro (pubblica, non profit, for profit), vi sono sufficienti argomentazioni per difendere l'ipotesi che le imprese si sforzino di raggiungere un obiettivo di produzione con il minimo consumo di risorse (minimizzazione dei costi). Gli istituti pubblici (comunali) sentono la pressione delle politiche di rigore finanziario che gravano soprattutto a livello locale e che mirano ad

²²⁴ Per una discussione su questo punto si vedano Vita (1990), Aletras (1999) e Smet (2002).

²²⁵ Il tasso di occupazione dei letti medio, riferito ai dati del campione usato per le stime, è pari a 97% per le case con meno di 25 posti letto, 96% per gli istituti con una capacità compresa tra 25 e 100 posti letto e 97% per le case per anziani con oltre 100 posti letto.

²²⁶ La capacità di riserva è molto importante per quelle attività soggette a forti oscillazioni della domanda unite all'impossibilità di razionamento sulla quantità (ad esempio l'elettricità). Il problema non si pone per le case per anziani in Svizzera, confrontate con un cronico eccesso di domanda, una domanda stabile ed un turnover lento.

eliminare le spese inutili. Lo spazio per eventuali gestioni orientate ad obiettivi diversi dalla minimizzazione dei costi risulta quindi minimo. Il settore privato che opera senza scopo di lucro è, molto spesso, finanziato e regolamentato dal cantone. La leva del controllo pubblico (che riguarda anche aspetti relativi alla qualità) unita alla motivazione intrinseca di erogare con continuità un servizio indispensabile per la società civile, fungono da stimolo per operare secondo la logica della minimizzazione dei costi. Le (poche) imprese private for profit presenti in questo mercato, sono per natura interessate a massimizzare i profitti che, dal punto di vista produttivo è la miglior garanzia di un atteggiamento volto all'ottimizzazione dei fattori produttivi nel breve e nel lungo periodo.

Alle considerazioni appena presentate si aggiunge la difficoltà di separare correttamente le spese di capitale fisse da quelle variabili. Nell'ottica di breve periodo, questa operazione è necessaria e, tuttavia, vi sono alcune voci di spesa riferite al capitale (attrezzature mediche, assicurazioni, servizi acquistati da terzi, etc.) che pur insensibili all'output nel breve periodo vengono incluse erroneamente nei costi variabili. Inoltre, nel settore delle case per anziani, è ragionevole ipotizzare che una buona parte delle voci di spesa di capitale non sia soggetta ad un vincolo temporale così stringente sull'arco dei 5 anni da noi osservati e quindi sembra più corretto specificare una funzione di lungo periodo.

Una critica ricorrente che viene mossa nei confronti dell'approccio di lungo periodo (stima di una funzione di costo totale) si basa sul seguente ragionamento: (a) non vi è certezza che le imprese si trovino in equilibrio statico (di lungo periodo), (b) ne segue che i costi (totali) potrebbero non essere minimizzati, (c) la teoria del duale diventa inutilizzabile per stabilire (oltre che i costi) anche le caratteristiche tecnologiche del settore in esame, (d) meglio usare la teoria del duale con una funzione di breve periodo. L'argomentazione (d) è neoclassicamente ineccepibile, anche se la scelta di utilizzare una funzione di breve periodo si basa sulle (presunte) debolezze di un'altra specificazione e non è, a sua volta, immune da altri tipi di critiche.

Ad ogni modo il nostro scenario di riferimento è leggermente diverso e chiama in causa alcune considerazioni di Bös (1986), riferite alla natura "*behvioural*" del modello

di costo.²²⁷ In quelle verifiche empiriche condotte in settori dove non si può assicurare la minimizzazione dei costi, il ricercatore deve comunque accontentarsi di un'approssimazione del concetto di funzione di costo neoclassica. Da un punto di vista teorico non è possibile scegliere quale specificazione dei costi (variabili o totali) sia la migliore, poiché entrambe permettono di analizzare la struttura dei costi, fornendo comunque indicazioni utili sul vero processo produttivo.

In aggiunta a questo, Farsi e Filippini (2004) offrono, in linea con la visione di Bös, un altro argomento per non abbandonare l'impostazione neoclassica nella stima di una funzione di costo totale, anche quando le condizioni di regolarità non dovessero essere completamente verificate. Si sostiene, infatti, che in presenza di vincoli regolatori (soprattutto sulle retribuzioni del personale), il mancato rispetto della concavità nei prezzi di tutti gli input può essere riconciliata con la specificazione neoclassica.²²⁸

L'ultimo punto che discuteremo per sostenere la scelta della specificazione di costo totale riguarda alcuni problemi econometrici che si incontrano, quando si stima una funzione di costo variabile. Il capitale di un'impresa, anche se non si aggiusta rapidamente a variazioni di breve periodo dell'output, non può essere considerato completamente esogeno.

Come suggerito da Grannemann et al. (1986), l'evidenza del settore sanitario è inconfutabile: (1) le strutture che devono soddisfare una forte domanda, impiegano anche grandi quantità di capitale (istituti con numero elevato di posti letto) e (2) in risposta alla variazione della domanda gli istituti reagiscono variando anche le dotazioni di capitale. Questo si traduce in una forte correlazione tra l'output e lo stock di capitale (comunque lo si misuri) che rende problematica la stima della funzione di costo variabile.

²²⁷ Dopo aver argomentato in senso neoclassico il comportamento delle imprese pubbliche, Bös propone una verifica empirica delle sue ipotesi attraverso la stima di una funzione di costo per un campione di imprese di trasporto (metropolitana) londinesi.

²²⁸ In verità, se si ammette che l'obiettivo delle imprese non è quello della minimizzazione dei costi, le condizioni di regolarità non dovrebbero essere rispettate neppure in teoria.

In questi frangenti un problema che si può verificare (lasciando poche possibilità di soluzione), è l'errore nel segno (positivo invece che negativo) del coefficiente della variabile che approssima lo stock di capitale.²²⁹

Anche quando tale evenienza non si riscontra nelle stime, resta il problema di trasformare i parametri della funzione di costo variabile in una misura delle economie di scala, ma a tale scopo è necessario impiegare il livello ottimale dello stock di capitale fisso (sconosciuto). Una possibilità, proposta da Caves et al. (1980), è quella di utilizzare il livello attuale dello stock di fattore fisso. In alternativa si può determinare il valore ottimale partendo dalle stime della funzione di costo variabile e, con una procedura analitica o numerica, individuare il valore dello stock di capitale che minimizza i costi.²³⁰ Appare evidente come le difficoltà, concettuali e pratiche, che si incontrano nella specificazione di un modello di lungo periodo, si ritrovano, sotto altre forme, anche nella stima di una funzione di costo variabile.

L'ultima motivazione che ci ha spinto a specificare un modello di costo totale riguarda l'interesse a verificare il livello di inefficienza di costo delle singole imprese. Se infatti ragioniamo in termini pragmatici (o seguendo la teoria dell'inefficienza X), la nostra aspettativa è quella di ritrovare nei dati la presenza di comportamenti inefficienti, misurati dall'eccesso di costo rispetto alla frontiera stimata. Di conseguenza, specificando una funzione di costo variabile, escluderemmo dalla stima alcune voci di spesa legate al capitale, perdendo traccia dell'inefficienza legata a questi fattori produttivi di lungo periodo.

Visto l'interesse esplicito a valutare le economie di scala e le inefficienze di costo delle case per anziani svizzere, e preso atto della limitata disponibilità di informazioni finanziarie, abbiamo optato per una specificazione di costo totale, le cui variabili esplicative saranno discusse a seguire.

²²⁹ In letteratura si trovano due possibili giustificazioni a questo problema: Guyomard and Vermersch (1989) interpretano il risultato come il segnale di un eccessivo stock di capitale, mentre Cowing and Holtmann (1983) attribuiscono il segno errato alla forte correlazione che si verifica tra l'output e l'indicatore dello stock di capitale.

²³⁰ Per una spiegazione dei dettagli di questa procedura si veda Smet (2002).

5.3.2 Il modello di costo utilizzato nelle stime

Alla luce delle discussioni precedenti abbiamo optato per un modello semplice e che illustri le principali caratteristiche del processo produttivo, con lo scopo di spiegare i costi totali generati dall'attività di una casa per anziani rappresentativa per la Svizzera. Sulla base delle caratteristiche tipiche del processo produttivo di un istituto per anziani e dei dati statistici in nostro possesso, i costi totali di una casa per anziani sono, pertanto, descritti dalla seguente equazione:

$$CT = f(Y, P_L, P_K, H, D_{RM}, Q_M, D_{TP}, Q_S, D_{FO}, D_{REGIO}, T) \quad [5.1]$$

La variabile dipendente, ossia i costi totali (CT), rappresentano il totale delle spese di una casa per anziani, comunicato ogni anno dall'istituto all'UFS, comprendente tutte le voci di costo imputate all'esercizio concluso. Di queste spese fanno ovviamente parte i salari corrisposti al personale, gli ammortamenti delle attrezzature e dei macchinari, gli affitti degli immobili, il servizio del debito (interessi passivi ed oneri finanziari), l'energia, il materiale di consumo, le prestazioni fornite da terzi, e tutte le altre spese generate dall'attività produttiva.

Il costo totale è una variabile che non presenta grandi difficoltà concettuali dal momento che esprime l'insieme delle spese sostenute per produrre ed è misurata con ragionevole precisione, provenendo i dati dalla contabilità degli istituti. L'unica precisazione che occorre fare riguarda il modo di valutare i costi totali rispetto alla dimensione temporale che, di solito, porta con sé una modifica dei prezzi. I costi, che sono una grandezza monetaria (espressa in franchi svizzeri) riferita all'anno d'esercizio, si possono misurare ai prezzi correnti oppure, per eliminare gli effetti dell'inflazione, a prezzi costanti. La nostra scelta è stata di esprimere i costi, e tutte le altre variabili monetarie considerate nel modello, a prezzi costanti, utilizzando per la deflazione l'indice dei prezzi al consumo pubblicato dall'UFS.²³¹

Per spiegare i costi di una casa per anziani è necessario poter misurare correttamente il livello di produzione dell'istituto considerato e in questa ricerca abbiamo utilizzato il numero totale di giornate di residenza offerte dall'istituto in un

²³¹ La base utilizzata per deflazionare le variabili monetarie è il mese di maggio 2000 nel quale l'indice dei prezzi al consumo è posto pari a 100. Nei cinque anni considerati dall'analisi i valori dell'indice risultano i seguenti: 98 (1998), 98,8 (1999), 100,3 (2000), 101,3 (2001), 102 (2002).

anno (Y). Come discusso in precedenza, esiste più di una possibilità per misurare l'output di una casa per anziani e in quest'analisi utilizzeremo il numero di giornate di residenza offerte dall'istituto nel corso di un anno. Rispetto ad altri possibili indicatori di attività come il numero di letti ed il numero di ospiti, la quantificazione delle giornate effettive di residenza sembra in grado di cogliere con maggiore precisione l'utilizzo delle risorse consumate dal processo produttivo. Il numero dei letti è, infatti, un indicatore della capacità produttiva piuttosto che della reale produzione (i letti potrebbero anche essere vuoti!), mentre il numero di ospiti non considera la durata del soggiorno ed è quindi una misura imprecisa dell'attività svolta da una casa per anziani. Inoltre, le giornate di residenza sono un indicatore affidabile dato che si rilevano nel momento stesso in cui viene generata la prestazione e sono ricavate dalle informazioni amministrative sull'ammissione/dimissione degli ospiti.

Il vero problema nell'utilizzare le giornate di residenza come regressore dei costi totali è di carattere econometrico dal momento che si richiede l'esogenità delle variabili indipendenti da quella dipendente. In altre parole, sul numero di giornate offerto da un istituto grava il sospetto che questa variabile sia strategicamente sotto il controllo dell'impresa che sceglie un determinato livello di output anche (o soprattutto) in considerazione dei costi di produzione che genera. In parte questa preoccupazione è certamente condivisibile, ma, nel caso specifico delle case per anziani, visti il tasso di occupazione molto vicino al 100% ed il forte eccesso di domanda, l'output si può considerare esogeno. Inoltre, in questo settore sono presenti norme di legge molto stringenti per quanto riguarda l'autorizzazione all'esercizio dell'attività. In diversi cantoni l'intervento regolatorio si spinge ad una vera e propria pianificazione dell'offerta, limitando di fatto l'autonomia decisionale in merito alla scala di produzione e garantendo (dal nostro punto di vista) l'esogenità dell'output.

Stabilito che le giornate di residenza offrano una buona misura dell'output, occorre precisare che dalle informazioni dell'UFS è possibile suddividere il totale delle giornate in quattro categorie: soggiorni lunghi, soggiorni brevi, soggiorni in appartamenti protetti (*foyers*), altri tipi di soggiorno (non precisati). Queste

informazioni sulla composizione dell'output saranno utilizzate per definire meglio il campione da utilizzare per le stime e per specificare meglio il modello di costo.²³²

Le due variabili riferite al prezzo dei fattori produttivi (P_L) e (P_K) necessitano diverse spiegazioni dal momento che si tratta di misure ricavate rielaborando altre informazioni della banca dati. Se consideriamo il fattore lavoro possiamo affermare con relativa certezza che la spesa per salari è determinata dal numero di unità lavorative impiegate (personale equivalente al tempo pieno) moltiplicato per il rispettivo salario annuale corrisposto. Il prezzo del lavoro è quindi un parametro che misura, per ogni singolo dipendente, la remunerazione monetaria delle risorse da lui conferite nel processo produttivo. Il nostro punto di osservazione è tuttavia l'impresa nel suo complesso e quindi è necessario passare da una definizione individuale di prezzo del lavoro ad una più generale.

Secondo questa visione potremmo individuare alcune grandi categorie di impiego all'interno dell'istituto (personale amministrativo, medico, infermieristico, tecnico, etc.) e definire il prezzo medio all'interno di ciascuna categoria, dividendo la quota del monte salari per il numero di unità lavorative. Questo modo di procedere sarebbe in grado di cogliere le normali differenze che esistono nella remunerazione di incarichi diversi, prevedendo per ogni categoria di impiego un prezzo specifico. Pur disponendo dei dati circa l'impiego di unità a tempo pieno in diverse categorie funzionali, il questionario dell'UFS non suddivide le spese per il personale, rendendo vano ogni tentativo di costruire un prezzo specifico per ogni tipologia di lavoratore di un istituto per anziani. Il prezzo del lavoro che utilizzeremo è, pertanto, il salario medio generale ed è costruito dividendo l'intero monte salari per il totale di unità equivalenti ad un impiego a tempo pieno.²³³

La scelta della variabile deputata alla misurazione del prezzo del capitale comporta, come già accennato in precedenza, qualche difficoltà in più, legata al tipo di informazioni disponibili. Innanzitutto, vista l'impossibilità di misurare lo stock di capitale con il valore a bilancio o con il metodo dell'inventario permanente, si deve

²³² Nel campione di case per anziani la distribuzione delle giornate sulle quattro categorie menzionate è fortemente sbilanciata in favore dei soggiorni di lunga durata (superiori ad un mese). In pratica le altre tre categorie sono spesso assenti o hanno un peso molto contenuto.

²³³ La misura del prezzo del lavoro risulta quindi la media ponderata dei diversi salari corrisposti per funzione, dove il peso è costituito dalla numerosità dei lavoratori in ciascuna categoria salariale.

costruire il prezzo del capitale, normalizzando le spese attribuite al capitale per un indicatore fisico dello stock di capitale, nel nostro caso il numero di posti letto. Sono allora possibili due approcci: il metodo diretto ed il metodo residuale. Nel primo caso si analizzano le voci di costo del conto economico e si raggruppano solamente quelle riferite al capitale, ad esempio: ammortamenti, interessi passivi, spese per la manutenzione, pigioni e canoni leasing, etc. Il metodo residuale, invece, parte dall'idea che le spese di un'impresa si possano suddividere in due grandi categorie: lavoro e capitale. Secondo questa logica, è possibile sottrarre ai costi totali le spese per i salari, ottenendo così una misura delle spese relative al capitale (in senso ampio).

Nella banca dati a nostra disposizione le categorie di spesa sono solamente 4: salari, altre spese di esercizio, ammortamenti, oneri finanziari (affitti ed interessi passivi). Attribuendo al capitale le ultime due voci (ammortamenti e oneri finanziari), la differenza tra i due metodi si riduce all'inclusione o meno delle spese di esercizio. In questa categoria, che di fatto è un insieme indistinto di spese, possono nascondersi i costi legati ad input come le materie o l'energia, ma anche voci attribuibili al capitale. Di conseguenza entrambi gli approcci forniscono una misurazione approssimata (in direzioni opposte) del prezzo del capitale. Dovendo scegliere tra due misure imperfette del prezzo del capitale occorre verificare quale delle due opzioni permetta di costruire un prezzo del capitale maggiormente affidabile. In questa ricerca abbiamo deciso di adottare una misurazione delle spese del capitale seguendo l'approccio residuale.

Il motivo principale di questa scelta è legato al peso esiguo delle spese dirette (ammortamenti ed interessi-affitti) per il capitale, che ammontano in media all'8% delle spese totali, mentre le "altre spese di esercizio" di una casa per anziani coprono circa il 20% dei costi totali. Seguire l'approccio diretto significa escludere dalla regressione il prezzo di un "input composito" (energia, materiale di consumo, servizi acquistati da terzi, spese di manutenzione, etc.) che genera 1/5 di tutti i costi.

Nella funzione di costo totale devono figurare tutti gli input e quindi, scegliendo di utilizzare la misura diretta del capitale, occorre definire anche il prezzo di questo "input composito", necessario per garantire ad ogni ospite: luce, cibo, calore, pulizia,

medicamenti, etc.²³⁴ Secondo questa interpretazione, dividendo le spese di esercizio di una casa per anziani per il numero di ospiti, si ottiene il prezzo dell'input composito. Nel settore delle case per anziani, dove il tasso di occupazione è molto vicino al 100%, questa separazione tra spese di capitale e spese di esercizio è trascurabile dato che il numero di ospiti coincide con il numero di letti. Considerando il nostro scenario empirico, il metodo residuale corrisponde quindi al metodo diretto, specificato con una restrizione sui parametri dei due input.²³⁵ Vista questa considerazione, abbiamo svolto alcune analisi econometriche preliminari per verificare se i risultati delle stime ottenute con i due approcci fossero molto diversi.²³⁶ Considerata la sostanziale uguaglianza tra le diverse specificazioni, seguendo il concetto di “*staffed bed*”, abbiamo misurato il prezzo del capitale con il metodo residuale, esprimendolo in franchi (a prezzi costanti) per letto “attrezzato”.²³⁷

Da notare che questa misura del prezzo del capitale rappresenta, con ogni probabilità, un'approssimazione per difetto del vero prezzo del capitale. Nel settore delle case per anziani, infatti, può capitare che alcune voci di spesa chiaramente imputabili al capitale non vengano registrate correttamente (ad esempio le donazioni che spesso riguardano terreni o immobili o capitali finanziari). Nel caso degli istituti comunali il problema è, invece, strutturale nel senso che la registrazione delle spese, soprattutto le voci di investimento, è inserita nel bilancio del comune rendendo molto difficile (se non impossibile) una corretta misurazione delle spese per il capitale. Questi problemi si riflettono inevitabilmente nelle informazioni sulle voci di spesa del capitale

²³⁴ In alternativa si potrebbe imporre l'omogeneità lineare nei prezzi, immaginando di dividere i costi ed i prezzi di lavoro e capitale per il prezzo (sconosciuto) dell'input composito. In questo modo l'input composito funge da numerario. Questa procedura, utilizzata ad esempio da Nyman (1989), si basa sull'ipotesi che l'input non osservato abbia un prezzo uniforme cosa irrealistica se consideriamo un input composito come le spese di esercizio di una casa per anziani.

²³⁵ I due metodi sono equivalenti se la somma dei coefficienti del capitale (definito con il metodo diretto) e dell'input composito corrisponde al valore del coefficiente del capitale (definito) con l'approccio residuale. Per verificare questa condizione è sufficiente stimare il modello a tre input con una restrizione lineare sui parametri degli input.

²³⁶ In tutte le verifiche econometriche effettuate, l'input composito è stato normalizzato sia per il numero di ospiti sia per il numero di letti. In questo secondo caso la verifica della restrizione sui parametri è stata sottoposta a test.

²³⁷ la scelta del numero di letti come proxy per lo stock di capitale è conforme all'approccio suggerito da Wagstaff (1989) e da Filippini (2001, 2004).

(ammortamenti, affitti ed interessi passivi) che, per alcuni istituti, sono assenti o rilevano importi pari a 0.²³⁸

Seguendo l'approccio neoclassico in senso ampio, l'output ed il vettore dei prezzi degli input, non sono sufficienti a descrivere con precisione i costi di un settore complesso e caratterizzato da una forte eterogeneità come quello delle case per anziani. Per ovviare a questo problema si rende necessario specificare all'interno del modello alcune variabili che siano in grado di ridurre l'eterogeneità presente nel campione, cogliendo determinati aspetti delle caratteristiche dell'output e/o del processo produttivo che lo genera. Secondo questa logica abbiamo inserito nel modello alcune variabili che mirano a catturare gli effetti sui costi provocati da: il profilo degli ospiti (case-mix), il livello di qualità del servizio erogato, la posizione geografica (per grandi aree), il progresso tecnologico.

La variabile H, che si propone di cogliere l'effetto sui costi dovuto alle possibili differenze nel fabbisogno di cure da parte degli ospiti residenti nei diversi istituti, corrisponde all'indicatore di fabbisogno medio di cure definito nel capitolo 2. Per gli scopi della regressione, la variabile fornisce una rappresentazione sintetica, attraverso un numero che varia (nel continuo) tra 0,5 e 6, del case-mix dei residenti ai quali gli istituti erogano i propri servizi. Le differenze nel profilo sanitario degli ospiti incidono sull'assorbimento di risorse e, di conseguenza, sui costi generati da una casa per anziani che produce un determinato numero di giornate. In altre parole, la variabile H si presenta come un attributo qualitativo dell'output e permette di specificare meglio la dimensione produttiva.

In merito alla variabile utilizzata per rilevare il "case-mix" degli ospiti occorre sottolineare che esistono modi più precisi per misurare lo stato di salute del residente medio di ogni istituto, dal momento che è prassi consolidata quella di definire una misura delle capacità psico-fisiche degli ospiti e valutare nel dettaglio lo stato di salute generale degli ospiti. Un indicatore spesso utilizzato è il grado di dipendenza dell'ospite che esprime, secondo una scala numerica, lo stato di salute di ogni residente. Purtroppo, queste misure che sintetizzano attraverso un indicatore quantitativo il profilo sanitario

²³⁸ Come discusso nel paragrafo 5.2.4, è necessario eliminare dal campione le osservazioni che presentano valori mancanti nelle variabili chiave. La riduzione imposta dall'assenza di informazioni sul capitale è di circa 1000 osservazioni su 4000 (circa 200 istituti su 800).

dei residenti non sono disponibili nella nostra banca dati ed è quindi necessario utilizzare una misura indiretta (la variabile H) dello stato di salute.²³⁹

Il profilo sanitario degli ospiti è un fattore importante, soprattutto nell'attuale realtà svizzera, in cui le case per anziani possono erogare prestazioni molto simili a quelle ospedaliere. Molte persone anziane ricoverate negli istituti necessitano cure costose, che sono coperte (in parte o integralmente) dall'assicurazione malattia di base. Tali spese sono sostenute dall'istituto e vengono rimborsate dalle casse malattia (assicurazioni), secondo gli accordi e le leggi vigenti in ogni cantone. Dividendo il rimborso annuo totale (ricevuto ai sensi della legge federale LAMal sull'assicurazione malattia) per il numero di residenti si ottiene una variabile continua (da zero a infinito), che esprime il profilo medico di ogni casa per anziani. Tra le spese mediche e lo stato di salute esiste, infatti, una correlazione positiva molto elevata e quindi le informazioni sul profilo sanitario degli ospiti si riescono a desumere indirettamente dalle spese mediche generate all'interno dell'istituto.

Un possibile problema causato dall'utilizzo della spesa sanitaria in rappresentanza delle prestazioni sanitarie è legato all'eventuale differenza tra cantone e cantone nelle tariffe ammesse al rimborso delle cure erogate. Inoltre, il valore del rimborso medio per ospite è molto correlato con il fabbisogno di cure dei residenti (variabile H) e con il numero di medici impiegati dall'istituto (variabile Q_M che verrà discussa tra breve). Queste ragioni di carattere econometrico, unite al problema delle differenze di prezzo tra i diversi cantoni ci hanno spinto a costruire la variabile *dummy* (D_{RM}) che assume valore pari a 1 quando il livello del rimborso per le spese mediche assicurate ai sensi LAMal è superiore al valore mediano del campione.²⁴⁰ Questa variabile binaria, insieme alla variabile continua H, ha il compito di catturare le differenze nel case-mix degli istituti.

L'utilizzo della variabile *dummy* D_{RM} ci obbliga ad aprire una breve parentesi metodologica. Nel modello vengono considerate complessivamente 6 *dummies*, 3 delle quali sono state "dicotomizzate" (la variabile può assumere solo il valore 0 o 1), anche

²³⁹ Filippini (2001) ha stimato la funzione di costo per le case per anziani ticinesi utilizzando come variabile di controllo per lo stato di salute il grado di dipendenza medio di ogni istituto.

²⁴⁰ Da alcune verifiche preliminari, inserendo la variabile "Rimborso medio per ospite" (R) in modo continuo, il valore dei parametri principali del modello non cambiano.

se la variabile originaria era discreta o continua. Nel nostro caso riteniamo questa procedura preferibile per tre ragioni, che si rifanno a problemi: econometrici, di dati ed economici. Il primo punto da considerare è il numero di parametri aggiuntivi che sono richiesti per stimare il modello con le variabili continue. Utilizzando la forma funzionale Translog, la sostituzione delle tre variabili *dummies* (D_{RM} , D_{TP} e D_{FO}), comporta l'aggiunta di 15 parametri che, oltre a ridurre i gradi di libertà (ma non è questo il nostro problema), accentuano la multicollinearità già presente nella translog.

Inoltre, l'uso di variabili continue con una specificazione translog non ammette la presenza di valori nulli per uno o più regressori.²⁴¹ Nel nostro caso l'uso delle variabili originali comporterebbe un'ulteriore riduzione (indesiderata) del campione a causa dell'incompletezza dei dati (*missing values*) riferiti tali variabili.²⁴²

L'ultima delle motivazioni a sostegno della versione binaria delle variabili riguarda il ruolo che implicitamente si attribuisce ad una variabile esplicativa quando la si trasforma in *dummy*. Nella forma dicotomica una variabile è in grado di misurare se la presenza di un certo attributo o di una data caratteristica incide al rialzo o al ribasso sui costi totali di una casa per anziani. In altre parole, una variabile *dummies*, pur contribuendo a descrivere in modo più accurato il processo produttivo (e di conseguenza a ridurre l'eterogeneità), mantiene un ruolo ausiliario nello spiegare i costi di produzione di un istituto per anziani.

Dopo questo chiarimento, proseguiamo la descrizione del modello di costo con la dimensione qualitativa che, alla luce dei discorsi fatti nel paragrafo 5.2.2, abbiamo considerato nella duplice veste di qualità di processo e di qualità delle dotazioni strutturali. Per ciascun aspetto abbiamo definito una variabile continua (indicatore principale) ed una binaria (indicatore ausiliario). La variabile Q_M è stata introdotta per cercare di cogliere, almeno parzialmente e in modo indiretto, il livello di qualità delle cure erogate da un istituto. Questa variabile si basa sul rapporto tra il personale medico-infermieristico (espresso in unità equivalenti a un tempo pieno) e gli ospiti residenti.

²⁴¹ Il problema si può superare utilizzando la translog generalizzata (specificazione Box-Cox) come suggerito da Caves et al. (1980).

²⁴² Questo problema non scompare completamente con le variabili binarie, poiché in assenza di dati il numero di osservazioni che non rispecchiano definizione scelta per assegnare alla variabile il valore 1 (che quindi assume il valore 0) aumenta. In questo modo è tuttavia possibile stimare i parametri delle caratteristiche considerate dalle variabili *dummies* senza ridurre la dimensione del campione.

Questo indicatore ha un campo di variazione (teorico) da zero ad infinito ed esprime la densità di personale di cura per ospite. Poiché il modello include sia il numero di giornate erogate (e quindi il numero di residenti), sia un indicatore del case-mix degli ospiti, la variabile Q_M esprime, *ceteris paribus*, un elemento oggettivo (oltre che percepito) di qualità delle prestazioni.²⁴³

Rivolgendo la nostra attenzione agli aspetti strutturali della qualità abbiamo costruito la variabile Q_S , che misura il numero complessivo di servizi attivati presso l'istituto. I dati sui servizi erogati contemplano una suddivisione molto dettagliata sulla tipologia dei servizi erogati rilevando: 5 voci relative ai servizi medici (servizio medico, servizio psichiatrico, cure ambulatoriali, cure a domicilio, altre cure), 8 principali servizi di cure ed assistenza alla persona (parrucchiere, pedicure, odontotecnico, apparecchi acustici, servizio di assistenza psicologica, servizio di animazione, servizi sociali) e 6 servizi tecnici (amministrazione, cucina, lavanderia, pulizia, economato). A questi servizi di base si aggiunge una lista di servizi accessori, che contempla: terapie (fisioterapia, ergoterapia, idroterapia, psicoterapia, etc.), servizi alla persona (assistenza religiosa, formazione, orientamento professionale, atelier di lavoro, etc.) e servizi infrastrutturali (farmacia, laboratori di analisi, trasporto, etc.). Queste attività si definiscono accessorie poiché, nonostante alcune di esse siano importanti e diffuse, non sono strettamente necessarie per offrire un servizio di residenza e cura agli anziani. Grazie alla variabile Q_S è possibile verificare se gli istituti con un livello di qualità alberghiera-infrastrutturale maggiore producono a costi più elevati.²⁴⁴

Per cogliere un'altra sfumatura qualitativa, diversa dalle precedenti, abbiamo considerato un aspetto organizzativo che potrebbe incidere indirettamente nella qualità dei processi gestionali e, di riflesso, sulle motivazioni del personale a contatto con gli ospiti. Si tratta di una variabile binaria costruita sulla base dell'importanza del lavoro full-time nell'organico dell'istituto. Il rapporto tra il numero di personale equivalente a tempo pieno e l'effettivo delle persone stipendiate rappresenta la percentuale di impiego media dell'istituto e fornisce una chiara indicazione di quanto l'istituto ricorra a

²⁴³ In letteratura la scelta del rapporto tra personale di cura e residenti è un indicatore di qualità dell'output abbastanza consolidato. Nella stima della funzione di costo è utilizzato da numerosi autori, ad esempio, McKay (1988), Crivelli et al. (2002), Vitaliano (2003).

²⁴⁴ Il numero di servizi offerti come proxy del livello di qualità alberghiera-infrastrutturale è stato impiegato in letteratura da Vitaliano (2003).

contratti a tempo parziale.²⁴⁵ Una maggior presenza di personale con contratti a tempo parziale si può spiegare con una moltitudine di fattori (es. il ricorso all'outsourcing, l'organizzazione dei turni di lavoro, una diversa piramide demografica dello staff, etc.) riconducibili (almeno in parte) a differenze nel processo produttivo. La variabile D_{TP} , che assume il valore 1 quando la percentuale di impiego medio di un'unità lavorativa è superiore alla mediana, ci permetterà di verificare se e come la tendenza ad impiegare il personale a tempo pieno incida sui costi.²⁴⁶

Nel modello abbiamo incluso anche la variabile *dummy* D_{FO} , che si prefigge di verificare se gli istituti che dispongono di appartamenti protetti, producono (*ceteris paribus*) a costi diversi. Questa variabile assume valore 1 se nell'istituto viene erogata almeno una giornata di residenza in un appartamento protetto (indicando l'esistenza della struttura distaccata), valore 0 negli altri casi. Il senso di distinguere le strutture che operano in autonomia (ma sotto la gestione della casa per anziani) da quelle centralizzate è quello di offrire un complemento alle tre dimensioni (case-mix, qualità nei processi e qualità nelle infrastrutture) già considerate.

La pratica di mantenere in un appartamento protetto (di solito) viene adottata con quegli anziani che mostrano un sufficiente grado di autonomia fisica e relazionale e che hanno presumibilmente un sanitario "leggero". In questo senso la variabile D_{FO} è un indicatore del (buono) stato di salute degli ospiti. Vista in un altro modo, la presenza di appartamenti distaccati esprime una scelta organizzativa dell'istituto (qualità di processo ed alberghiera) volta ad erogare il servizio di residenza nel miglior modo possibile. Per queste ragioni il coefficiente della *dummy* potrebbe assumere un segno negativo o positivo, anche se riteniamo la seconda possibilità poco probabile.

Il modello di costo fin qui descritto, oltre alla parte strutturale (output e prezzi degli input), contempla gli aspetti legati al profilo sanitario degli ospiti (case-mix) e la qualità (di processo ed alberghiera) dei servizi erogati. I dati a nostra disposizione avrebbero permesso di considerare questi aspetti in modo diverso o più dettagliato. Consapevoli che le misure proposte sono semplicemente un'approssimazione, nelle fasi

²⁴⁵ Ad esempio, se in una casa per anziani lavorano 15 persone per un equivalente a tempo pieno di 9 unità significa che in media (9/15) ogni persona lavora al 60%.

²⁴⁶ Il valore mediano del grado di occupazione medio dei dipendenti è pari a 0,672.

preliminari della definizione del modello abbiamo valutato molte alternative.²⁴⁷ In assenza di un indicatore molto preciso del case-mix e della qualità, ogni misura surrogata (le variabili da noi costruite) fornisce un contributo alla spiegazione dei costi. Tuttavia, aggiungendo al modello altre variabili “*proxy*”, le singole misure tendono a confondersi e disturbarsi a vicenda, offrendo un contributo modesto alla spiegazione dei costi e quindi alla specificazione del modello. Per questa ragione abbiamo preferito l’impiego del minimo numero indispensabile di variabili esplicative, optando per i 6 regressori (3 continui e 3 dicotomici) descritti in precedenza.

Parte dell’eterogeneità presente nel settore delle case per anziani è dovuta a caratteristiche specifiche dei singoli istituti (fattori individuali) e alla presenza di vincoli esterni di varia natura (fattori ambientali). Con riferimento ai fattori ambientali, in Svizzera (stato federale) la dimensione geografica può avere un peso rilevante poiché cattura in parte gli aspetti legati alla regolamentazione ed alle caratteristiche locali del mercato. Dall’analisi svolta nel capitolo 2, erano emersi con chiarezza alcuni tratti distintivi a carattere regionale che, tuttavia, non coincidevano con i confini amministrativi dei cantoni.

Di conseguenza abbiamo idealmente suddiviso il campione di dati in 4 regioni, aggregando i cantoni in unità territoriali più vaste e secondo un certo criterio geografico/linguistico. Questa decisione è motivata in parte dal basso numero di osservazioni di determinati cantoni, in parte dal desiderio di cogliere alcuni aspetti legati alla “cultura regionale” in senso più ampio.

Per queste ragioni, la variabile D_{RT} raggruppa tutti i cantoni romandi di lingua francese (Ginevra, Vaud, Neuchâtel, Jura) oltre al Ticino in quella che potremmo definire (ignorando una piccola parte del canton Grigioni) regione “latina”;²⁴⁸ la variabile D_{LB} unisce i cantoni limitrofi all’area bernese (Berna, Friburgo Vallese, Soletta, Basile Città e Campagna); la variabile D_{CSE} è composta dai cantoni della Svizzera centrale (Lucerna, Nidwaldo, Obwaldo, Svitto, Uri, Zugo) e dai cantoni della

²⁴⁷ Alcuni esempi delle variabili considerate sono: l’età media, la composizione uomo-donna, il livello di formazione dello staff, la percentuale di residenti con gravi handicap, la quota del volontariato, indicatori di ricorso all’outsourcing, etc.

²⁴⁸ I cantoni Friburgo e Vallese sono stati esclusi da questa regione poiché bi-lingue. Abbiamo comunque verificato la stabilità delle stime rispetto all’aggregazione dei due cantoni nella regione romanda.

zona sud-orientale (San Gallo, Appenzello Interno ed Esterno, Glarona, Grigioni);²⁴⁹ la variabile D_{NZ} considera i cantoni nell'area a Nord di Zurigo (Zurigo, Argovia, Turgovia, Sciaffusa);²⁵⁰ Il vettore di *dummies* regionali $D_{REGIONO}$ (D_{RT} , D_{LB} , D_{CSE} , D_{NZ}) permette di considerare almeno in parte l'impatto sui costi dei differenti sistemi di regolamentazioni cantonali e consente di filtrare l'influsso di qualche fattore inosservato a carattere regionale.²⁵¹

Il modello di costo è completato dalla presenza di un trend lineare (variabile T) che ha il compito di catturare l'impatto del progresso tecnologico sul costo di produzione di una casa per anziani. Nel modello non sono presenti variabili riferite in modo esplicito alle caratteristiche individuali degli istituti. Questa mancanza è in parte voluta per analizzare il comportamento dei punteggi di inefficienza di costo, raggruppati secondo alcune tipologie specifiche, come la forma istituzionale o la dimensione degli istituti. L'obiettivo è quello di esplorare in modo qualitativo le possibili determinanti dell'inefficienza, senza effettuare una regressione a due stadi che, come discusso nella sezione metodologica, comporta seri problemi econometrici.

Una seconda ragione è quella di non appesantire ulteriormente un modello abbastanza ricco, inserendo variabili esplicative che consentirebbero di catturare in minima parte l'eterogeneità presente nei dati. Utilizzando il modello TREM, proposto da Greene (2005), saremo comunque in grado di filtrare una parte degli effetti dovuti alle caratteristiche non osservate. Con la specificazione dettagliata del modello di costo si conclude la parte teorica della ricerca per lasciare il posto alle verifiche empiriche. Le scelte fin qui compiute ci permettono di proseguire, rivolgendo la nostra attenzione all'analisi delle economie di dimensione e alla misurazione dell'inefficienza di costo. Nel prossimo capitolo descriveremo le principali statistiche del campione di case per anziani selezionate per le stime e discuteremo i risultati empirici.

²⁴⁹ In principio avevamo separato la zona centrale da quella sud-orientale. Purtroppo, in quest'ultima area la mancanza di dati provenienti dai cantoni Appenzello Interno e San Gallo (numericamente il più importante) hanno reso necessaria questa aggregazione.

²⁵⁰ Anche se i dati del canton Ticino (125 osservazioni) rappresentano circa il 7% del campione e sono sufficienti a costruire una *dummy* autonoma, abbiamo preferito l'aggregazione con la Romandia. A titolo di informazione elenchiamo (in parentesi) la numerosità delle 4 regioni: D_{LB} (655) D_{RT} (615), D_{NZ} (360), e D_{CSE} (140).

²⁵¹ Per poter stimare il modello con le *dummies* regionali è necessario escluderne una dalla lista dei regressori. I risultati sono insensibili alla scelta della *dummy* esclusa, che funge da termine di paragone per i coefficienti delle altre.

*“Give me the fruitful error any time, full of seeds,
bursting with its own corrections.
You can keep your sterile truth for yourself.”*

(Vilfredo Pareto, Economista)

6 Economie di scala ed inefficienza di costo: analisi dei risultati empirici

Questo capitolo è dedicato alla presentazione dei risultati empirici e alla loro discussione. Dopo avere esposto il problema economico (capitolo 2) e la teoria della produzione (capitolo 3), spiegato la metodologia utilizzata (capitolo 4), presentato una rassegna della letteratura e descritto le scelte di “design” della ricerca (capitolo 5), è arrivato il momento di analizzare i risultati delle stime.

Per facilitare la chiarezza espositiva e la comprensione del percorso di indagine svolto, abbiamo strutturato il capitolo in quattro paragrafi. Il primo ha il compito di illustrare la costruzione e le caratteristiche del campione di dati concretamente impiegato nelle stime. La descrizione dei risultati delle stime e la discussione di alcune varianti completano il paragrafo. Nella parte centrale del capitolo (secondo e terzo paragrafo) presenteremo l’analisi dell’efficienza di costo e delle economie di scala delle case per anziani in Svizzera. Gli elementi di riflessione che emergono dall’analisi economica sulla funzione di costo permetteranno di tracciare alcune considerazioni, anche di carattere normativo, riguardo ai problemi di gestione della politica sanitaria nel settore dell’assistenza agli anziani. È quindi tempo di esporre i risultati della stima della frontiera di costo, ripercorrendo tutte le fasi dell’analisi empirica: preparazione dei modelli e dei dati, esecuzione delle stime, analisi dei risultati e considerazioni critiche.

6.1 L'analisi empirica

Le tecniche econometriche a disposizione per stimare una funzione frontiera di costo sono, come discusso in precedenza, molteplici e presentano vantaggi e svantaggi che devono essere pesati e messi in relazione agli obiettivi dell'indagine e ai dati disponibili. La scelta del campione è un argomento molto delicato in una ricerca empirica ed è quindi importante dedicargli la giusta attenzione. Inoltre, la descrizione delle problematiche incontrate nella selezione del campione utilizzato nelle stime è un'occasione per ribadire ed avvalorare alcune delle scelte di "design" compiute. I risultati delle stime sono il cardine di una ricerca empirica e, nonostante la loro esposizione si esaurisca nel commento di poche tabelle, presuppongono tutta una serie di tentativi e prove che merita una discussione. Per questo motivo abbiamo scelto di mettere in risalto i risultati econometrici dedicandogli uno spazio autonomo nel quale discutere soprattutto i valori dei parametri, la loro coerenza economica, la significatività statistica e la stabilità dei risultati ottenuti con modelli econometrici diversi. Iniziamo allora ricordando quali sono i 5 modelli econometrici per dati panel che abbiamo stimato:

- FEM: *Fixed Effect Model* – Schmidt and Sickles (1984);
- REM: *Random Effect Model* – Schmidt and Sickles (1984);
- ML-REM: *Maximum Likelihood Random Effect Model* – Pitt and Lee (1981);
- TREM: *"True" Random Effect Model* – Greene (2005);
- POOL: *Maximum Likelihood Cross-Section Model* – Aigner et al. (1977).

La discussione delle proprietà teoriche dei cinque modelli considerati per le stime è già stata presentata, ma deve, tuttavia, essere riferita alla realtà delle case per anziani, cosa che faremo in sede di commento ai risultati empirici. Prima, però, è necessario affrontare il delicato problema della costruzione della banca dati da utilizzare per le stime.

6.1.1 La scelta del campione utilizzato per le stime

La stima di una funzione di costo richiede che le osservazioni relative alle variabili del modello siano disponibili ed esprimano valori economicamente plausibili. Per poter procedere con le stime econometriche è, pertanto, necessario svolgere un'analisi statistica preliminare ed eventualmente procedere ad una verifica e "pulizia"

dei dati. Questo processo è sempre delicato dal momento che mira ad escludere alcune osservazioni dall'analisi ed è opportuno dichiarare con grande trasparenza, tutte le operazioni di "pulitura" che hanno permesso di costruire il campione utilizzato. In questo processo di esclusione si possono seguire almeno tre criteri diversi: la mancanza di dati, la plausibilità economica dei dati, la pertinenza del dato rispetto all'obiettivo dell'analisi.

Dei tre criteri menzionati solamente il primo non implica scelte discrezionali dal momento che la mancanza di un dato è un problema oggettivo del quale occorre semplicemente prendere atto scartando le osservazioni che ne sono colpite. In merito alla plausibilità economica dei dati, si apre uno spazio di dibattito senza regole certe, eccezion fatta per il buon senso e per qualche indicatore statistico. La decisione di scartare alcune osservazioni è, in ultima analisi, arbitraria e merita una breve giustificazione ed un commento sulle possibili conseguenze sui risultati delle analisi. L'intervento sulla banca dati si concluderebbe a questo stadio se la raccolta delle informazioni statistiche fosse avvenuta nell'ambito della ricerca. Molto spesso, tuttavia, la ricerca empirica deve ricorrere a dati che provengono da contesti molto diversi rispetto a quelli in esame, rendendo necessaria un'ulteriore operazione di filtraggio dei dati, giustificato dalla scarsa pertinenza delle informazioni con gli obiettivi della ricerca.

La banca dati originale contava 11'637 osservazioni, riferite ad istituti non ospedalieri attivi in tutto il territorio svizzero nell'arco di cinque anni, dal 1998 al 2002. Come discusso nel capitolo 2, l'UFS utilizza una tassonomia degli istituti non ospedalieri seguendo un criterio prevalentemente statistico. La prima scrematura rispecchia quindi la necessità di non considerare istituti che, pur operando nel settore sociosanitario, non svolgano a titolo principale l'attività di assistenza agli anziani. Per questo motivo tutte le strutture che non assegnavano almeno la metà dei propri posti letto ai malati cronici e/o agli anziani sono state scartate. Questa operazione ha ridimensionato il campione ad una cifra pari a circa 6'500 osservazioni.

Il passo successivo prevede l'esclusione delle osservazioni che presentano dati incompleti nelle variabili principali utilizzate per descrivere la funzione di costo. Sono state scartate le osservazioni con valori mancanti relativamente a: letti, giornate di residenza, ospiti, personale impiegato, informazioni finanziarie ed informazioni sul profilo sanitario dei residenti. Per concentrare la nostra attenzione sulle case per anziani medicalizzate, cercando di escludere istituti di vario genere presenti nel campione,

abbiamo deciso di scartare le osservazioni i cui ricavi provenienti dal rimborso delle prestazioni sanitarie di base fossero nulli. Il tipico profilo sanitario di una casa per anziani, nel senso comune del termine, tende ad escludere l'evenienza che nel corso di un anno non vi siano prestazioni mediche da poter rimborsare. Se questo accade sono possibili due spiegazioni: non si tratta di una casa per anziani (se non dal punto di vista della definizione statistica) oppure si tratta di una casa di riposo per anziani (*altersheime*). In entrambi i casi è preferibile escludere queste osservazioni probabilmente generate da un processo produttivo diverso da quello che si vuole descrivere. Dopo questa selezione, eseguita per allineare i dati con l'oggetto di studio, le osservazioni disponibili scendono a circa 3'500 unità.

Tuttavia, il numero di osservazioni non è bilanciato rispetto al tempo e diversi istituti sono rappresentati solamente in uno o due anni sui cinque disponibili. La presenza di queste osservazioni potrebbe costituire un ostacolo all'utilizzo delle tecniche panel che necessitano una certa variabilità temporale. Dopo aver analizzato i risultati di alcune stime preliminari, svolte con un panel sbilanciato, abbiamo deciso di bilanciare i dati, escludendo dal campione gli istituti con meno di 5 anni di dati.²⁵² Questa operazione di natura "tecnica" ha ridotto il campione a 1'900 unità, costituite da 380 istituti osservati su 5 anni.

Analizzando le caratteristiche del campione si rileva la presenza di osservazioni che, da un punto di vista economico, esprimono valori poco plausibili. Per questa ragione abbiamo scartato gli istituti che presentavano alternativamente: dotazioni inferiori a 10 posti letto, un tasso di occupazione dei letti inferiore al 10% o organizzate secondo una forma istituzionale non chiaramente identificabile.²⁵³ Il primo criterio riflette situazioni in cui un "istituto" è rappresentato con ogni probabilità da reparti o distaccamenti di strutture più complesse, diverse dalle case per anziani tradizionali. Nel secondo caso, vista la cronica carenza di posti letto nell'assistenza agli anziani, è

²⁵² Il confronto dei risultati ottenuti stimando un campione sbilanciato (nel quale sono presenti le osservazioni con almeno 4 anni) sarà presentato in seguito nella sezione (paragrafo 6.1.3) dedicata alla discussione di alcune stime svolte in alternativa a quelle finali.

²⁵³ Il problema delle forme istituzionali difficili da identificare riguarda poche unità (circa una decina). Inoltre, da alcune stime preliminari, l'esclusione di queste osservazioni non modifica in alcun modo i risultati delle stime.

difficile includere nell'analisi, considerandole alla stregua di case per anziani, quelle strutture con un tasso di occupazione dei letti così basso.

In aggiunta a queste restrizioni di tipo concettuale è necessaria un'operazione di "pulitura statistica" finale, dato che alcune osservazioni esprimono valori economicamente poco sensati o logicamente inaccettabili (ad esempio errori di compilazione dei dati originali che generano situazioni irrealistiche). Per ridurre al minimo indispensabile questa fastidiosa e criticabile operazione di "potatura" dei dati (*data-trimming*) ci siamo basati sulla distribuzione statistica delle variabili scartando solamente i valori estremi e contrari al buon senso economico. Quest'ultimo blocco di operazioni, nonostante abbia ridotto il campione di poche unità, ha contribuito ad eliminare una eterogeneità statistica "patologica" (outliers) che non trova riscontro e giustificazione nella realtà che si desidera analizzare.²⁵⁴

Il campione finale da utilizzare per le stime è quindi pari a 1'770 osservazioni riferite a 354 case per anziani nel corso dei 5 anni coperti dallo studio. Si può notare come le ultime due fasi di preparazione dei dati abbiano ridotto il campione di 230 osservazioni (da 1'900 a 1'770), eliminando solamente 46 istituti. Prima di rendere definitiva l'esclusione di questi dati abbiamo verificato i valori delle stime, che risultano sostanzialmente uguali.²⁵⁵

In Svizzera la dimensione del federalismo è sempre una chiave di lettura importante ed è opportuno verificare, rispetto a questa caratteristica, come si presenta la banca dati in esame. I 26 livelli regionali di governo, che dal punto di vista della regolamentazione sanitaria godono di grande autonomia, non sono adeguatamente rappresentati nel campione. Il motivo principale è da iscriversi alla mancanza di dati completi, vincolo non superabile, che produce delle distorsioni nella rappresentatività del campione: alcuni cantoni sono eccessivamente rappresentati, mentre altri sono addirittura assenti. Nonostante le probabili/possibili differenze che si riscontrano a livello cantonale (parzialmente catturate dalle *dummies* regionali), l'obiettivo principale di questa tesi è la stima di una funzione di costo generale e valida per tutta la Svizzera.

²⁵⁴ In quest'ultima fase sono state scartate solamente 10 case per anziani.

²⁵⁵ Un confronto dei coefficienti stimati con i due campioni è stato fatto in sende di analisi preliminare e non ha evidenziato differenze importanti.

Per questo motivo l'appartenenza ad un determinato cantone resta comunque un'informazione che può essere utilizzata in modo descrittivo permettendo alcune considerazioni, sia qualitative sia quantitative in sede di valutazione dell'inefficienza delle singole case per anziani. Dal punto di vista econometrico si potrebbe anche pensare di suddividere il campione rispetto ai singoli cantoni e procedere alle stime dei diversi gruppi di dati. Questo approccio è tuttavia fortemente limitativo dal momento che compromette l'attendibilità delle stime cantonali vista la bassa numerosità delle osservazioni dei singoli gruppi che, per alcuni cantoni, impedisce la possibilità di giungere ad una stima.²⁵⁶ La precisazione sulla struttura cantonale dei dati è importante per sottolineare ancora una volta il ruolo dell'eterogeneità nella corretta determinazione dei punteggi di inefficienza, cosa della quale ci occuperemo nel prosieguo del capitolo.

Con riferimento al modello di costo [5.1], specificato nel capitolo precedente, nella Tabella 6.1 presentiamo le principali statistiche descrittive relative alle variabili utilizzate nelle stime, oltre a qualche indicatore descrittivo generale come la dimensione, il costo medio per istituto e l'età media dei residenti.

Gli indicatori presentati in aggiunta alle variabili esplicative del modello esprimono un tratto caratteristico della tipica (mediana) casa per anziani svizzera: un istituto di piccole dimensioni, con un profilo anagrafico marcatamente elevato e con un costo medio difficilmente finanziabile dall'utenza (il reddito mensile richiesto sarebbe circa pari a 5'800 CHF, che è di molto superiore al valore mediano dei redditi delle persone anziane). La rappresentazione descritta da questi parametri è in linea con la percezione che il settore suggerisce all'osservatore esterno.

I valori medi di queste cifre, tuttavia, celano una fortissima variabilità che si manifesta in diverse dimensioni: case per anziani molto piccole (10-15 posti letto) oppure molto grandi (oltre 300 posti letto), pazienti che generano costi sanitari limitati (meno di 2'000 CHF all'anno rimborsati dall'assicurazione obbligatoria) o molto elevati (più di 70'000 CHF all'anno rimborsati dall'assicurazione obbligatoria), pochi o molti servizi all'utenza, basso o alto fabbisogno di cure degli ospiti, etc. Con queste statistiche descrittive in mente, diventa interessante verificare il significato ed il contributo che

²⁵⁶ Ci riferiamo a 12 cantoni (Appenzello Interno, Appenzello Esterno, Basilea Città, Glarona, Grigioni, Jura, Lucerna, Nidwaldo, Obwaldo, San Gallo, Svitto e Zugo) il cui numero di osservazioni è inferiore a 30, quando il numero di parametri da stimare è superiore a 20!

ciascuno dei parametri, selezionati per descrivere il processo produttivo di una casa per anziani, fornisce alla spiegazione dei costi totali.

Tabella 6.1 - Statistiche descrittive delle principali variabili utilizzate nelle stime.

Variabile	Descrizione	Media (Dev. Std.)	Min.	1° Q.le	Mediana	3° Q.le	Max
CT	Costi Totali	4'560'320 (3'652'410)	292'510	2'081'600	3'595'000	6'076'000	33'775'000
Y	Giornate	22'618.2 (16'054.1)	2'419	11'959	18'763	27'968	126'210
PL	Prezzo del Capitale	21'292.2 (10'440.1)	5'161.9	62'227	18'373	76'867	70'311
PK	Prezzo del Lavoro	69'669.2 (13'101.5)	16'067	13'967	69'415	25'485	138'108
H	Ore cura per ospite	2.78 (1.11)	0.5	1.95	2.635	3.612	6
Q _M	Staff cure per ospite	0.42 (0.17)	0.021	0.302	0.424	0.552	1.132
Q _S	Numero di servizi	16.76 (5.67)	2	12	16	21	36
Rimborso LAMal (CHF/ospite)		20'209.8 (11'235.8)	1'400.61	21'219	17'843	25'821	78'431
Percentuale media di impiego dello staff		0.676 (0.133)	0.124	0.588	0.672	0.759	1
Numero di letti		64.66 (46.46)	10	35	53	81	385
Costo medio per giornata		201.27 (62.73)	66.45	154.45	191.95	237.10	450.75
Età media ospiti per istituto		84.20 (4.39)	79.98	82.99	84.93	86.48	92.5
Valore medio delle dummies (numerosità % del gruppo rispetto al campione)²⁵⁷							
D _{RM}	D _{TP}	D _{FO}	D _{RT}	D _{LB}	D _{NZ}	D _{CSE}	
0.5	0.5	0.09	0.35	0.37	0.20	0.08	

Per determinare i parametri del modello di costo abbiamo deciso di impiegare una specificazione Translog che, come abbiamo avuto modo di precisare nel capitolo 4, presenta il vantaggio di poter misurare in modo flessibile le economie di scala.²⁵⁸

²⁵⁷ Le dummies D_{RM} e D_{TP} sono ripartite equamente tra il valore 0 ed il valore 1 poiché sono costruite a partire dal valore mediano delle rispettive variabili continue “Rimborso LAMal” e “Percentuale media di impiego dello staff”, i cui valori sono riportati nella Tabella 6.1.

²⁵⁸ Per legittimare la nostra scelta, in una fase preliminare dell’analisi abbiamo stimato il modello di costo impiegando le seguenti forme funzionali: Cobb-Douglas, Translog e Translog “ridotta” (modello adottato in questa analisi). Il test sul rapporto di (log)verosimiglianza conferma la legittimità del modello impiegato.

Quando si utilizza una forma funzionale flessibile (translog) che approssima localmente la “vera” relazione, un passaggio importante che precede le operazioni di stima è quello di definire il punto di approssimazione. L’idea stessa di rappresentare una generica casa per anziani suggerisce di utilizzare il valore centrale (mediana) delle distribuzioni dei regressori. In questo modo i parametri stimati dal modello esprimono le relazioni tecnologiche di un’ipotetica casa per anziani con le caratteristiche dell’istituto mediano, offrendo un termine di paragone generale (ma rappresentativo) per le considerazioni economiche che scaturiscono dall’analisi empirica. Per contro, quando ci si allontana molto dal punto di approssimazione, occorre cautela nell’interpretazione dei risultati.

6.1.2 I risultati delle stime

Sulla base del modello di costo specificato e descritto in precedenza e in virtù dell’omogeneità lineare nei prezzi imposta alla funzione di costo, il modello da stimare risulta quindi il seguente:

$$\begin{aligned}
 \ln\left(\frac{CT_{it}}{P_{K_{it}}}\right) &= \alpha_0 + \alpha_Y \ln(Y_{it}) + \alpha_L \ln\left(\frac{P_{L_{it}}}{P_{K_{it}}}\right) + \alpha_H \ln(H_{it}) + \alpha_{Qm} \ln(Q_{M_{it}}) + \alpha_{Qs} \ln(Q_{S_{it}}) \\
 &+ \alpha_{YY} \frac{1}{2} \ln^2(Y_{it}) + \alpha_{LL} \frac{1}{2} \ln^2\left(\frac{P_{L_{it}}}{P_{K_{it}}}\right) + \alpha_{HH} \frac{1}{2} \ln^2(H_{it}) + \alpha_{Qm} \frac{1}{2} \ln^2(Q_{M_{it}}) + \alpha_{Qs} \frac{1}{2} \ln^2(Q_{S_{it}}) \\
 &+ \alpha_{YL} \ln(Y_{it}) \ln\left(\frac{P_{L_{it}}}{P_{K_{it}}}\right) + \alpha_{YH} \ln(Y_{it}) \ln(H_{it}) + \alpha_{LH} \ln\left(\frac{P_{L_{it}}}{P_{K_{it}}}\right) \ln(H_{it}) \\
 &+ \alpha_{DRM}(D_{RM}) + \alpha_{DTP}(D_{TP}) + \alpha_{DFO}(D_{FO}) \\
 &+ \alpha_{DR1}(D_{RT}) + \alpha_{DR2}(D_{NZ}) + \alpha_{DR3}(D_{CSE}) + \alpha_T(T) + [\alpha_i] + \varepsilon_{it}
 \end{aligned}
 \tag{6.1}$$

A beneficio del lettore richiamiamo la denominazione delle variabili impiegate per spiegare il livello dei costi totali (CT) delle case per anziani svizzere:

- Y = numero di giornate di residenza registrate nel corso dell’anno;
- P_L = prezzo medio del lavoro (salario annuo di un’unità a tempo pieno);
- P_K = prezzo medio del capitale (importo annuo per letto attrezzato);
- H = indicatore del fabbisogno medio di cure di un residente;
- Q_M = indicatore di qualità delle cure (staff medico/infermieristico per residente);
- Q_S = indicatore di qualità strutturale (numero di servizi attivati);
- D_{RM} = indicatore binario (*dummy*) del livello dei rimborsi per spese mediche;

- D_{TP} = indicatore binario (*dummy*) del livello di contratti a tempo pieno;
 D_{FO} = indicatore binario (*dummy*) della presenza di *foyers*;
 D_{RT} = indicatore binario (*dummy*) regione “Romanda” e cantone Ticino;
 D_{NZ} = indicatore binario (*dummy*) regione “Nord e Zurigo”;
 D_{CSE} = indicatore binario (*dummy*) regione “Centro e Sud-orientale”;
 T = trend lineare.

Nell’equazione [6.1], i pedici i e t indicano rispettivamente la casa per anziani (i) e l’anno di riferimento (t). La struttura del termine di errore (ε_{it}), così come l’eventuale presenza dell’effetto specifico di ogni istituto (α_i), saranno discussi in modo più approfondito commentando le stime generate dai diversi modelli econometrici utilizzati.

Nel modello [6.1] i prodotti incrociati che coinvolgono i due indicatori di qualità (variabili Q_M e Q_S) sono stati tralasciati e quindi la specificazione è equiparabile ad un modello translog completo nel quale viene imposta la seguente restrizione:

$$\alpha_{YQ_m} = \alpha_{YQ_s} = \alpha_{LQ_m} = \alpha_{LQ_s} = \alpha_{HQ_m} = \alpha_{HQ_s} = \alpha_{Q_mQ_s} = 0 \quad [6.2]$$

In letteratura, la pratica di ridurre la Translog, eliminando alcuni prodotti incrociati o restringendo i parametri di determinati coefficienti, è abbastanza diffusa e trova a sostegno argomentazioni di carattere economico o econometrico.²⁵⁹ Nel nostro caso, in assenza della restrizione [6.2], emergono dei problemi di convergenza durante la stima del modello TREM, che si basa su una procedura iterativa (massima verosimiglianza simulata).²⁶⁰ Per escludere i coefficienti che incrociano gli indicatori di qualità con le altre variabili, si può addurre come ulteriore giustificazione (secondaria) il desiderio di considerare nel modello di costo solo i coefficienti diretti (1° e 2° ordine) di queste due *proxy* della qualità.²⁶¹

²⁵⁹ Alcuni esempi relativi al settore delle case per anziani si trovano in Gertler e Waldman (1992) e Rebba e Rizzi (2002).

²⁶⁰ Per motivi simili Farsi e Filippini (2008), nella stima di un modello di frontiera di costo TREM per il settore ospedaliero svizzero, scelgono la forma funzionale Cobb-Douglas al posto della Translog.

²⁶¹ Come discusso in precedenza, infatti, non disponiamo di indicatori precisi della qualità delle cure e può essere preferibile omettere le interazioni di queste *proxy* con le altre variabili. Per scrupolo, in una fase preliminare abbiamo stimato (con la tecnica GLS) anche il modello Translog completo e, dei 7 coefficienti relativi ai prodotti incrociati delle variabili Q_M e Q_S , solamente due sono risultati significativamente diversi da 0.

In una specificazione flessibile come la Translog, la scelta di escludere alcuni coefficienti dei prodotti incrociati è quindi il frutto di un compromesso di diverse esigenze: realismo nella specificazione, contenimento del numero di parametri da stimare e attenzione a eventuali problemi econometrici (in questo caso, la convergenza della procedura di stima iterativa prevista dal TREM). Visto l'interesse ad impiegare il modello di frontiera stocastica TREM, abbiamo quindi optato per una specificazione Translog "ridotta".

Utilizzando i 5 modelli scelti per le analisi empiriche, abbiamo stimato la funzione di costo [6.1] i cui parametri, come era auspicato, hanno prodotto numerose analogie, ma anche qualche differenza importante. Nella Tabella 6.2 sono riportati i valori dei parametri stimati che, trattandosi di una specificazione log-log, si possano interpretare con il concetto di elasticità della variabile dipendente rispetto al regressore in esame. In particolare, tutti i coefficienti del primo ordine esprimono l'elasticità dei costi totali rispetto ai diversi fattori che spiegano e determinano i costi stessi. Tale misura, si riferisce ad un contesto (*ceteris paribus*) nel quale si immagina di variare solamente una grandezza alla volta, tenendo le altre variabili costanti.

Dall'analisi delle stime notiamo che in generale per tutti i modelli considerati il livello di significatività statistica dei parametri è molto buono. I valori dei coefficienti del primo ordine rispettano le aspettative sul segno e risultano positivi in tutti i modelli, indicando come un aumento dell'output, del prezzo dei fattori produttivi, dell'intensità di cure richiesta dagli ospiti e della qualità, si riflettano in un aumento dei costi totali.

In generale i risultati delle stime sono molto incoraggianti sia per il livello di significatività dei principali parametri sia, soprattutto, per la stabilità delle stime nei diversi modelli. Va sottolineata un'importante differenza nel coefficiente dell'output (α_Y) del modello FEM che registra un valore più contenuto (circa la metà) rispetto agli altri modelli.

Questo dato, pur non destando preoccupazioni particolari, suggerisce richiamo ad alcuni aspetti econometrici (discussi nel capitolo 4) utili a sottolineare il significato ed il peso della dissonanza che si riscontra tra il modello FEM e gli altri quattro modelli.

Cameron e Trivedi (2005) evidenziano la scarsa precisione del modello ad effetti fissi (FEM), quando la dimensione temporale del panel è limitata e la variabilità nel tempo (*within variation*) dei regressori è debole.

Tabella 6.2 - Stima dei parametri della funzione di costo con 5 diversi modelli.

		FEM ²⁶²	REM	ML-REM	TREM	POOL
Costante	α_0	---	14.98*** (.012)	14.78*** (.015)	14.91*** (.005)	14.92*** (.011)
Coeff. 1°ordine	α_Y	0.472*** (.029)	0.912*** (.009)	0.942*** (.007)	0.923*** (.003)	0.978*** (.005)
	α_L	0.598*** (.010)	0.630*** (.009)	0.631*** (.010)	0.606*** (.004)	0.669*** (.009)
	α_H	0.014 (.009)	0.038*** (.009)	0.036*** (.011)	0.030*** (.005)	0.091*** (.010)
	α_{Qm}	0.154*** (.011)	0.192*** (.010)	0.175*** (.008)	0.188*** (.004)	0.241*** (.011)
	α_{Qs}	0.060*** (.012)	0.049*** (.011)	0.054*** (.014)	0.055*** (.006)	0.029** (.011)
Coeff. 2°ordine	α_{YY}	0.275*** (.030)	0.299*** (.015)	0.201*** (.011)	0.246*** (.005)	0.054*** (.010)
	α_{LL}	0.065*** (.021)	0.054** (.018)	0.070*** (.017)	0.051*** (.010)	0.017 (.023)
	α_{HH}	-0.026 (.021)	-0.011 (.020)	-0.008 (.024)	-0.040*** (.011)	0.028 (.024)
	α_{QQm}	0.074*** (.010)	0.086*** (.009)	0.071*** (.008)	0.091*** (.005)	0.121*** (.011)
	α_{QQs}	0.061* (.028)	0.056* (.026)	0.030 (.035)	0.031 (.019)	-0.074* (.034)
Coeff. Prodotti	α_{YL}	0.014 (.013)	0.003 (.011)	0.015** (.010)	0.016*** (.004)	0.021 (.011)
	α_{YH}	-0.025* (.012)	-0.036** (.011)	-0.047*** (.008)	-0.046*** (.005)	0.004 (.011)
	α_{LH}	0.032* (.014)	0.046*** (.013)	0.045** (.016)	0.047*** (.009)	0.087*** (.017)
Coeff. Dummies Ausiliarie	α_{DRM}	-0.011 (.007)	0.001 (.007)	-0.004 (.008)	0.015*** (.004)	0.047*** (.008)
	α_{DTP}	0.023*** (.006)	0.033*** (.005)	0.034*** (.007)	0.034*** (.003)	0.013* (.006)
	α_{DFO}	-0.037** (.014)	-0.065*** (.013)	-0.074*** (.013)	-0.056*** (.006)	-0.057*** (.012)
Coeff. Dummies Regionali	α_{DR1}	---	0.028 (.017)	0.045*** (.013)	0.030*** (.005)	0.060*** (.010)
	α_{DR2}	---	-0.061*** (.018)	-0.046*** (.017)	-0.042*** (.005)	-0.036*** (.009)
	α_{DR3}	---	-0.032 (.025)	-0.051** (.021)	-0.028*** (.007)	-0.042*** (.013)
Trend	α_T	0.016*** (.001)	0.010*** (.001)	0.010*** (.001)	0.011*** (.001)	0.007*** (.002)
Scomposizione dei residui			λ	3.359*** (.087)	2.725*** (.477)	1.767*** (.123)
SOLO modelli di Massima Verosimiglianza (ML-REM, TREM e POOL)			α_i	nd	0.140*** (.002)	nd
			σ_v	0.077	0.038	0.074
			σ_u	0.261	0.103	0.166

NOTE: in parentesi sono riportati gli errori standard;
 *, **, ***, indicano la di significatività del parametro rispettivamente del 95%, 99% e 99,9%;
 il parametro λ (con $\lambda = \sigma_u / \sigma_v$) misura l'asimmetria nelle componenti della varianza dei residui.

²⁶² Il modello FEM non consente l'uso di variabili binarie costanti nel tempo. Le *dummies* ausiliarie utilizzate nel modello sono però costruite a partire da variabili continue. Alcune case per anziani durante il periodo osservato hanno modificato il valore della variabile di riferimento, determinando una variabilità temporale sufficiente per la stima del parametro.

Nel nostro caso il panel si estende solo per 5 anni, ma ciò che più conta è la ridotta variabilità temporale dei regressori principali, con particolare riferimento all'output (variabile Y).²⁶³

Constatata la presenza di una bassa variazione temporale delle variabili esplicative, abbiamo deciso di non enfatizzare troppo il commento dei coefficienti ottenuti con il modello FEM, attribuendo maggiore importanza agli altri quattro modelli, meno affetti da questo problema.²⁶⁴

Per contro, a causa dell'eterogeneità non osservata (*heterogeneity bias*), è possibile che le stime dei modelli REM, ML-REM, TREM e POOL risultino distorte. Questo problema riguarda in minima parte il modello ad effetti fissi, le cui stime possono servire da termine di paragone per stabilire la direzione delle (eventuali) distorsioni presenti negli altri modelli.²⁶⁵ Con riferimento a queste ultime specificazioni, il problema del *heterogeneity bias* si accentua in modo particolare nel modello POOL che, al contrario dei tre modelli ad effetti casuali (REM, ML-REM e TREM), non sfrutta le informazioni provenienti dalla struttura panel dei dati.

Seguendo le argomentazioni di Cameron e Trivedi (2005) riteniamo che le stime del modello ad effetti casuali REM, nonostante l'indicazione contraria del test di Hausman-Taylor, siano preferibili. Inoltre, visto il dichiarato interesse a stimare una frontiera di costo, desideriamo contenere l'impatto dell'eterogeneità non osservata sulle stime di inefficienza. Come abbiamo già avuto modo di sottolineare, il modello TREM è in grado, rispetto alle altre specificazioni, di filtrare (parzialmente) la presenza nei residui dell'eterogeneità non osservata, misurando l'inefficienza di costo con maggiore precisione. Questa importante caratteristica econometrica non va, però, a scapito dell'affidabilità nel rilevare le economie di scala, dal momento che i coefficienti stimati

²⁶³ Il peso percentuale della variazione intra gruppo (*within variation*) è in generale basso ed oscilla tra il 10% ed il 15% per tutte le variabili, ad eccezione dell'output, che presenta una variabilità temporale estremamente bassa (1,3%).

²⁶⁴ In aggiunta a ciò, come evidenziato da Hsiao (2003), il modello ad effetti fissi soffre dell'*incidental parameter problem*, che, non si manifesta solo nella stima delle costanti (*dummies* individuali), ma affligge anche gli altri parametri.

²⁶⁵ Nel nostro caso il test di Hausman-Taylor (1980) favorisce il modello FEM rispetto al modello REM. Va ricordato, però, che il test non giudica l'adeguatezza complessiva dei due modelli, ma solamente una specifica proprietà econometrica. Inoltre, la bassa variazione temporale dei regressori potrebbe rendere meno attendibile il test.

con tale modello risultano molto simili a quelli ottenuti con le altre specificazioni ad effetti casuali.

Per queste ragioni, ma senza abbandonare la presentazione in parallelo dei risultati di più modelli, nel prosieguo dell'analisi empirica ci baseremo principalmente sui coefficienti del modello TREM per elaborare le analisi sulle economie di scala e sull'efficienza di costo. Per rendere più scorrevole l'esposizione, faremo riferimento ai tre modelli ad effetti casuali, evidenziando le differenze riscontrate con i modelli FEM e POOL solo se necessario.

Il parametro α_Y , che esprime l'elasticità dei costi totali rispetto all'output, assume un valore compreso tra 0,9 (REM) e 0,94 (ML-REM), indicando una risposta dei costi meno che proporzionale alla variazione delle giornate di residenza. Alle luce delle considerazioni econometriche esposte in precedenza, riteniamo che il valore stimato dai tre modelli ad effetti casuali sia, in verità, sovrastimato. Si può infatti notare che il coefficiente dell'output assume i valori estremi in corrispondenza dei due modelli che risentono in modo diametralmente opposto del problema dell'eterogeneità non osservata. Tale coefficiente è molto basso (0,47) nel modello che non soffre del problema (FEM), mentre raggiunge il massimo valore (0,98) nel modello POOL. La grande distanza tra i due valori suggerisce che la distorsione provocata dal *heterogeneity bias* su tutti i modelli (ad eccezione del FEM preso a riferimento) sia verso l'alto con una tendenza a sovrastimare il parametro dell'output. Ragionando in modo conservativo possiamo quindi affermare che, alla mediana dei dati, un aumento delle giornate di residenza erogate determina una crescita dei costi meno che proporzionale.²⁶⁶ In Svizzera, dove il tasso di occupazione è vicino al 100%, aumentare l'output significa anche aumentare la capacità produttiva di un istituto, evidenziando così la presenza di economie di dimensione nel settore delle case per anziani.

Il coefficiente α_L è positivo ed indica che la funzione di costo è monotonicamente crescente nel vettore dei prezzi, come previsto dalle condizioni di regolarità neoclassiche. Vista l'imposizione dell'omogeneità lineare e la forma funzionale log-log, il parametro si interpreta come quota percentuale del costo del lavoro sul totale dei costi

²⁶⁶ Il test di Student indica che il coefficiente dell'output del modello POOL (pari a 0,979 con un errore standard di 0,0047) è comunque diverso da 1, ad un livello di significatività superiore a 99,9%.

di una casa anziani rappresentativa. In base alle stime rileviamo come poco più del 60% dei costi sia generato dal fattore lavoro, mentre il restante 40% circa sia riferito alle spese per il capitale. Con ogni probabilità l'importanza del fattore lavoro è leggermente sottostimata a causa dell'approccio residuale adottato per la definizione del prezzo del capitale. Considerando come valore di riferimento la quota media campionaria del fattore lavoro (pari a 69% circa), la stima del prezzo del lavoro, che nei modelli REM e ML-REM raggiunge il 63%, si rivela comunque un parametro sufficientemente attendibile.

I parametri delle variabili H e Q_M e Q_S mostrano un segno positivo (come era atteso), ad indicare che i costi totali crescono con l'aumentare dei bisogni di cura da parte degli ospiti (catturato dalla variabile H : ore di cura per ospite) e con l'erogazione di un livello qualitativamente più elevato (variabili Q_M e Q_S , rispettivamente: personale di cura per ospite e numero medio di servizi attivati). La sensibilità dei costi a queste variabili è moderata, ma non trascurabile visto che l'aumento del 10% di H , di Q_S o di Q_M determinano una crescita dei costi nell'ordine di: 3% (H) 5% (Q_S) e 18% (Q_M). Questa informazione può essere molto importante, ad esempio, per valutare se il differenziale di costi tra istituti simili in molte caratteristiche, ma con profili di ospiti e livelli qualitativi differenti, sia ragionevolmente accettabile. Per i modelli FEM e POOL, i coefficienti di queste variabili (H e Q_M e Q_S), leggermente diversi nella cifra e nella significatività, forniscono indicazioni analoghe a quelle discusse per gli altri modelli.

Ogni variabile compare nel modello anche nella sua forma quadratica (coefficienti del 2° ordine), le cui stime risultano molto simili nei modelli panel (compreso il FEM), ma non sempre significative. I coefficienti del 2° ordine non hanno un'interpretazione economica così immediata come nel caso dell'elasticità di costo espressa dai parametri del primo ordine, tuttavia alcuni di essi sono molto importanti per determinare il valore delle economie di scala (α_{YY}) e per verificare empiricamente il rispetto ex-post delle condizioni di regolarità (α_{LL}) della funzione di costo.²⁶⁷ Il livello di significatività di

²⁶⁷ I coefficienti del secondo ordine (compresi i prodotti incrociati) si interpretano come variazioni percentuali del coefficiente del primo ordine considerato, in risposta a variazioni percentuali di un determinato regressore. Ad esempio, il coefficiente α_{YY} si interpreta come variazione dell'elasticità di costo rispetto all'output (Y) al variare dell'output stesso.

questi due parametri è molto elevato (superiore al 99,9%) nelle 4 specificazioni panel che, inoltre, esprimono valori molto simili.²⁶⁸

Nel complesso, dal confronto tra i coefficienti del primo e del secondo ordine, stimati con i diversi modelli, emerge una modesta variabilità dei parametri che conferisce all'interpretazione economica della funzione di costo una maggior consistenza e credibilità.

Anche i coefficienti dei prodotti incrociati presentano poche differenze, anche se in alcune specificazioni esprimono un livello di significatività basso o nullo. Questo tipo di situazione, che si manifesta secondo uno schema “a macchia di leopardo”, è piuttosto comune quando si stimano funzioni flessibili con molti parametri, e non è un problema legato alla specificazione del modello o alla tecnica econometrica.²⁶⁹

Per gli scopi di questa tesi, sono di particolare rilevanza economica i due parametri che incrociano l'output con le altre variabili (α_{YL} e α_{YQS}), poiché concorrono a determinare il valore delle economie di scala, lontano dal punto di approssimazione. I parametri in questione, tuttavia, esprimono un livello di significatività non sempre accettabile, assumono segni discordi e valori numerici molto contenuti.

Di sicuro interesse economico sono invece i parametri delle variabili *dummies*, ausiliarie (D_{RM} , D_{TP} , D_{FO}) inserite nel modello per migliorarne la capacità esplicativa. Abbiamo definito “ausiliarie” le tre *dummies*, per il fatto di aiutare le rispettive variabili continue (H , Q_M e Q_S) a cogliere in modo complementare le caratteristiche del processo produttivo legate all'utenza, all'organizzazione o alla struttura degli istituti per anziani. Con un buon livello di precisione i parametri si possono interpretare direttamente come variazione percentuale dei costi totali, quando la variabile assume valore 1 rispetto a quando assume valore 0.²⁷⁰

²⁶⁸ Nel modello POOL il coefficiente α_{LL} non è significativo, mentre il coefficiente α_{YY} è risultato molto più basso (0,05) rispetto al parametro dei modelli panel (da 0,2 a 0,3).

²⁶⁹ I 5 prodotti incrociati esclusi dal modello (per le ragioni esposte all'inizio del paragrafo) sono quelli maggiormente colpiti da questo problema, avvalorando la nostra decisione di non considerarli.

²⁷⁰ Per ottenere il costo totale in presenza o in assenza della variabile *dummy* è necessario invertire il logaritmo, esprimendo in modo moltiplicativo l'impatto della *dummy* sui costi. Ad esempio, la variabile D_{FO} , che ha un valore (media dei 5 modelli) di -0,06 circa, si interpreta come un minor costo del 6% in presenza di foyers ($D_{FO} = 1$). Il valore corretto dell'impatto sui costi è invece 1.0618 (pari a $e^{0,06}$), ed è sostanzialmente identico ai fini pratici dell'interpretazione.

La variabile D_{RM} è stata inserita per considerare alcuni aspetti legati al profilo degli ospiti, sfuggiti al controllo della variabile continua (H) che misura il case-mix dei residenti. A dipendenza del livello medio (alto o basso) del rimborso delle spese mediche, questa *dummy* suddivide il campione in due gruppi, associando il valore 1 a quelle osservazioni che presentano un rimborso (delle spese mediche “LAMal”) per ospite superiore al valore mediano. Il parametro risulta positivo (e significativamente diverso da zero) nel modello di riferimento (TREM) e nel modello POOL, risultando rispettivamente pari a 0,015 e 0,047 e contribuendo a spiegare la differenza nei costi di produzione tra i due gruppi. Il valore del parametro è tuttavia molto contenuto ed il confronto con gli altri 3 modelli mostra alcuni contrasti, parte dei quali era attesa come discusso nel capitolo 5.²⁷¹

La *dummy* D_{TP} è costruita per suddividere le osservazioni del campione in due gruppi di eguale numerosità, ma con due livelli diversi (alto e basso) di impiego di personale con contratti a tempo pieno. Questa variabile, in grado di distinguere alcune scelte organizzative (es. la struttura dei turni di lavoro) degli istituti, suggerisce una relazione diretta tra un elevato ricorso a contratti a tempo pieno e l’aumento dei costi. Il parametro risulta significativo e molto stabile (circa 0,03) nei 4 modelli panel, indicando che le case per anziani con una percentuale di contratti *full-time* superiore al valore mediano producono a costi più elevati (3%) rispetto agli altri istituti.

Questo suggerisce, in modo contro intuitivo, che la gestione di un minor numero di dipendenti faccia aumentare i costi. Per interpretare questa variabile si deve riconoscere che i contratti a tempo parziale sono spesso associati a remunerazioni più basse e che la funzione di costo non riesce a cogliere tutti i differenziali nel prezzo del lavoro. In modo particolare pensiamo ad alcuni oneri contributivi legati alle prestazioni sociali che potrebbero spiegare il maggior costo delle case con alti livelli di contratti a tempo pieno. Una spiegazione alternativa si può trovare nella diversa organizzazione dei processi produttivi degli istituti che utilizzano un maggior numero di dipendenti a percentuali di impiego parziali. Il costo addizionale, che le case per anziani devono

²⁷¹ Specificando il modello con la variabile continua (R_M) “Rimborso Lamal medio per ospite” insorgono problemi di multicollinearità con gli indicatori di case-mix (H) e di qualità “medica” (Q_M). Come discusso in precedenza (cfr. nota 242) abbiamo preferito la specificazione binaria, poiché l’inclusione della variabile R_M , il cui parametro risulta sempre statisticamente non significativo, disturba gli altri regressori senza fornire alcun contributo alla spiegazione dei costi.

sostenere per amministrare un numero di dipendenti relativamente più grande, si potrebbe tradurre in una maggiore produttività lavorativa, legata all'ottimizzazione dei turni di lavoro, oltre ad un minor onere per le ore di straordinario. A prescindere dalla spiegazione della *dummy* D_{TP} , confinata nel campo di ipotesi la cui verifica non è praticabile, una prevalenza di assunzioni a tempo parziale superiore alla media incide sui costi totali, abbassandoli.

L'ultima *dummy* ausiliaria considerata è la variabile D_{FO} , che indica la presenza di *foyers* esterni alla struttura principale, nei quali trovano alloggio alcuni ospiti della casa anziani. In tutti i modelli stimati questa variabile risulta molto significativa, con i valori del parametro che oscillano da $-0,037$ (modello FEM) a $-0,074$ (modello ML-REM), indicando (*ceteris paribus*) una riduzione dei costi tra il 4% e il 7% in presenza di questo attributo. Una possibile spiegazione del minor livello dei costi associato a queste strutture é da ricercare nel profilo sanitario degli ospiti che risiedono nelle strutture distaccate. Generalmente, infatti, in virtù di un maggior grado di autosufficienza degli ospiti, l'erogazione del servizio di assistenza nei *foyers* assorbe meno risorse e genera quindi costi inferiori.²⁷²

Il contributo che le tre *dummies* ausiliarie apportano al modello di costo, seppure piccolo numericamente e non sempre statisticamente significativo, è importante per due ragioni.²⁷³ Dal punto di vista teorico, avvalora la necessità di un modello ricco in grado di cogliere l'eterogeneità presente nel settore. Dal punto di vista della misurazione dell'inefficienza di costo, si riesce a filtrare una parte della variabilità dei costi che finirebbe altrimenti nei residui della regressione, alterando le prestazioni individuali delle singole case per anziani. Ogni *dummy*, se presa isolatamente, contribuisce in modo non decisivo nella spiegazione dei costi e potrebbe essere anche esclusa (soprattutto se poco significativa dal punto di vista statistico).²⁷⁴ Tuttavia, il vettore complessivo di

²⁷² Le variabili che catturano il case-mix o la qualità delle cure sono *proxy* imperfette e lasciano inspiegata una certa variabilità di costo di provenienza sanitario-organizzativa. Se la parte di questi aspetti non considerata dal modello è correlata con la presenza di *foyers*, la *dummy* D_{FO} è in grado di catturare l'effetto addizionale sui costi.

²⁷³ La scelta di mantenere le variabili, nonostante una bassa significatività, è supportata dai test statistici sulle restrizioni dei parametri in discussione. Il modello che impone l'uguaglianza a 0 di questi parametri viene infatti scartato nel confronto con il modello senza restrizioni (con le variabili).

²⁷⁴ Per arrivare alla versione definitiva del modello sono state fatte numerosi tentativi che consideravano anche l'esclusione di una o più *dummies*. I parametri principali della funzione di costo si sono rivelati poco sensibili alle diverse specificazioni, mitigando le conseguenze di un'eventuale scelta imperfetta.

variabili binarie arricchisce il modello dal punto di vista economico, meritando di figurare nella frontiera di costo utilizzata per le stime.

Le 3 *dummies* regionali incluse nel modello forniscono alcuni spunti di riflessione interessanti. Osservando i modelli ad effetti casuali (REM, ML-REM, e TREM), si nota una diminuzione dei costi spostandosi da Ovest a Est. Il valore della dummy D_{RT} è infatti positivo, indicando un maggior costo rispetto alla regione circostante il canton Berna, presa a riferimento. Quest'ultima, però esprime costi più elevati se confrontata con le altre aree di lingua tedesca: le regioni della svizzera centrale e sudorientale (D_{CSE}) e la regione a Nord di Zurigo (D_{NZ}). Difficile dare una spiegazione dei diversi livelli di costo che caratterizzano le 4 aree senza disporre di informazioni approfondite sulle diverse realtà regionali. A prescindere dai diversi sistemi di regolamentazione cantonale, impossibili da considerare nel nostro modello, le *dummies* regionali sembrano confermare quella sottile, ma innegabile, differenza di approccio che esiste tra la Svizzera "latina" e la Svizzera "tedesca", nell'erogare il servizio di assistenza agli anziani. In termini numerici i coefficienti delle *dummies* regionali risultano abbastanza contenuti (circa 5%), ma non per questo trascurabili, in quanto permettono di filtrare dai residui della regressione una parte dell'eterogeneità.

Per completare il modello resta da commentare il parametro della variabile trend (T), che è responsabile di catturare la variazione dei costi provocata dal "progresso tecnologico". Il valore stimato indica (per i modelli panel) un incremento annuo dei costi che oscilla tra l'1,5% e il 2%, spiegato quindi da fattori esogeni e non meglio precisati. Inutile dire che il fattore maggiormente indiziato di causare tale crescita dei costi nel tempo è la diffusione (giustificata in parte da un profilo sanitario degli ospiti via via più pesante) di tecnologie mediche all'avanguardia che permettono di offrire più servizi ed il costante aumento generalizzato del costo del lavoro. Nonostante il valore del coefficiente sia contenuto, l'elevata significatività statistica del coefficiente (oltre a 99,9%) conferisce al parametro un ruolo importante nella spiegazione dell'aumento con il passare del tempo dei costi di produzione. Questo "eccesso di costo" è provocato da ragioni che sfuggono al controllo delle imprese e sarebbe rilevato nei punteggi di inefficienza delle imprese, se non fosse considerato all'interno del modello.

Per concludere la parte dedicata all'analisi ed al commento dei parametri stimati con i diversi modelli econometrici, resta da completare la verifica delle condizioni di regolarità, con riferimento alla concavità della funzione di costo rispetto ai prezzi dei

fattori produttivi. Per effettuare questa verifica si deve analizzare la matrice Hessiana che, utilizzando le equazioni [4.6] e l'imposizione dell'omogeneità lineare nei prezzi, nel punto di approssimazione si riduce a:

$$A_{L,K} = \begin{vmatrix} \alpha_{LL} - \alpha_L(1 - \alpha_L) & \alpha_L(1 - \alpha_L) \\ \alpha_L(1 - \alpha_L) & \alpha_{LL} - \alpha_L(1 - \alpha_L) \end{vmatrix} \quad [6.2]$$

Questo significa che per avere una matrice negativa semidefinita è sufficiente che α_{LL} sia minore di $\alpha_L(1 - \alpha_L)$. In tutti i modelli la condizione è soddisfatta, verificando con successo la concavità nei prezzi della funzione di costo. Per appurare se la condizione di regolarità è valida anche per tutte le osservazioni è necessario richiamare la matrice [4.7] che, declinata in base al modello di costo scelto, assume la forma seguente:

$$A(y, w) = \begin{vmatrix} \alpha_{LL} - S_L(1 - S_L) & -\alpha_{LL} + S_L(1 - S_L) \\ -\alpha_{LL} + S_L(1 - S_L) & \alpha_{LL} - S_L(1 - S_L) \end{vmatrix} \quad [6.3]$$

In virtù dell'imposizione dell'omogeneità lineare nei prezzi, gli elementi sulla diagonale principale sono identici e, affinché la matrice risulti negativa semidefinita, è sufficiente che l'elemento a_{11} della matrice [6.3] sia negativo. Dall'equazione [4.6] ricaviamo la quota di costo del lavoro che risulta:

$$S_L = \alpha_L + \alpha_{LL} \ln\left(\frac{P_L / \bar{P}_L}{P_K / \bar{P}_K}\right) + \alpha_{YL} \ln(y / \bar{y}), \quad \text{dove } \bar{P}_L, \bar{P}_K, \bar{y} \text{ rappresentano i valori mediani.}$$

La sostituzione dei valori campionari di (Y), (P_L) e (P_K) ha permesso di verificare come le condizioni di regolarità della funzione di costo stimata siano rispettate per tutte le 1'770 osservazioni e per tutti i 5 modelli. Questo risultato è importante nella misura in cui le considerazioni fatte per il punto di approssimazione vengano estese e generalizzate a realtà più lontane dal valore mediano. Occorre però riconoscere che il buon esito di questa verifica dipende in parte dalla semplicità del modello (due soli input) e dall'imposizione dell'omogeneità lineare (simmetria della matrice).²⁷⁵

²⁷⁵ In pratica basta verificare che l'elemento a_{11} della matrice $A(y, w)$ sia negativo. il segno di tale elemento dipende in modo cruciale dal valore di α_{LL} che, è risultato inferiore a 0.07 in tutti i modelli stimati, rendendo molto difficile la violazione della concavità. Sarebbe infatti necessario trovare empiricamente una ripartizione delle quote di costo molto sbilanciata (almeno 92,5% e 7,5%) in favore di uno dei due input, evenienza improbabile in questo settore ed assente nel nostro campione.

Con riferimento al modello di costo specificato, le stime econometriche dei parametri hanno prodotto complessivamente dei buoni risultati, che incoraggiano a procedere oltre con la descrizione delle implicazioni economiche suggerite dall'analisi empirica. Prima, però, desideriamo presentare al lettore i risultati di alcune tra le molte varianti stimate nel corso della ricerca che hanno contribuito all'attuale specificazione econometrica del modello di costo. L'obiettivo principale è quello di verificare in modo trasparente due alternative che, pur essendo praticabili, non sono state seguite: (1) utilizzare la forma funzionale Cobb-Douglas e (2) utilizzare un maggior numero di variabili (continue) nel modello. Questo confronto, che mostra un sostanziale allineamento dei risultati ottenuti con le diverse specificazioni presentate, ci permette di legittimare le scelte compiute nella definizione del modello di costo.

6.1.3 Discussione critica su alcune scelte di specificazione del modello di costo

Le stime empiriche del modello [6.1] sono state ottenute con diversi modelli econometrici e sono state presentate in modo comparativo per evidenziare vantaggi e svantaggi dei vari modelli. Alla fine, tenuto conto del problema dell'eterogeneità non osservata e delle caratteristiche del dataset, abbiamo scelto il modello TREM (Greene, 2005) come riferimento principale per le analisi empiriche e per le considerazioni che ne scaturiscono.

Tuttavia, prima di arrivare al confronto ed alla scelta della tecnica econometrica abbiamo effettuato alcune verifiche preliminari sulla correttezza nella scelta della forma funzionale Translog "ridotta" e nell'utilizzo di alcune variabili dicotomiche (al posto delle rispettive variabili continue). Questi confronti tra il modello impiegato nelle stime e specificazioni alternative, forniscono un riscontro sulla stabilità dei parametri stimati, rispetto a leggere modifiche della forma funzionale o del modello di costo.

Discutere il comportamento di tutti i 5 modelli econometrici sotto le diverse specificazioni eccede gli scopi di questa sezione. Pertanto, presenteremo i confronti dei risultati empirici, utilizzando solo le stime del modello REM (stimato con la tecnica GLS) che, rispetto agli altri modelli di frontiera (ML-REM e TREM), non richiede particolari ipotesi sulla distribuzione del termine di errore.²⁷⁶

²⁷⁶ I confronti tra numerose specificazioni alternative, oltre a quelle descritte in questo paragrafo, sono stati verificati singolarmente per i 5 modelli econometrici con risultati sostanzialmente analoghi.

Il primo confronto, illustrato dalla Tabella 6.3, considera come base di partenza il modello [6.1], che risulta una translog “ridotta” le cui stime dei parametri sono riportate nella prima colonna. La seconda colonna identifica, invece, la specificazione Cobb-Douglas del modello [6.1].

Tabella 6.3 - Stabilità delle stime rispetto alla forma funzionale.

Parametro		Mod. [6.1] Translog “ridotta”	Cobb-Douglas
Costante	α_0	14.98***	15.05***
Coeff. 1°ordine	α_Y	0.912***	0.832***
	α_L	0.630***	0.622***
	α_H	0.038***	0.059***
	α_{Qm}	0.192***	0.129***
	α_{Qs}	0.049***	0.069***
Coeff. 2°ordine	α_{YY}	0.299***	[= 0]
	α_{LL}	0.054**	[= 0]
	α_{HH}	ns	[= 0]
	α_{QQm}	0.086***	[= 0]
	α_{QQs}	ns	[= 0]
Prodotti incrociati Mod. [6.1]	α_{YL}	ns	[= 0]
	α_{YH}	-0.036**	[= 0]
	α_{LH}	0.046***	[= 0]
Prodotti incrociati rimanenti	α_{YQm}	[= 0]	[= 0]
	α_{YQs}	[= 0]	[= 0]
	α_{LQm}	[= 0]	[= 0]
	α_{LQs}	[= 0]	[= 0]
	α_{HQm}	[= 0]	[= 0]
	α_{HQs}	[= 0]	[= 0]
	α_{QmQs}	[= 0]	[= 0]
Dummies		Valori non presentati	
Trend	α_T	0.010***	0.011***
Wald test (χ^2)		67.76***	548.02
N° restrizioni		7	8

NOTE: *, **, ***, indicano la di significatività del parametro rispettivamente del 95%, 99% e 99,9%; l'ipotesi nulla del Wald test è che i parametri analizzati siano pari a zero. Quando il valore del χ^2 è significativo ad un livello superiore al 99,9% tale ipotesi viene respinta.

L'allineamento delle stime nelle diverse specificazioni è piuttosto buono, anche se il coefficiente dell'output α_Y , di particolare interesse per il calcolo delle economie di scala, è più basso nella specificazione Cobb-Douglas.

Una possibile chiave interpretativa per spiegare tale differenza risiede nella rigidità strutturale della Cobb-Douglas che impone un solo valore all'elasticità di costo rispetto all'intero spettro dei possibili output. In altre parole, mentre la translog è in grado di misurare le economie di scala in corrispondenza di un qualsiasi livello produttivo, la Cobb-Douglas impone lo stesso valore delle economie di scala per qualsiasi output. Di conseguenza tale misura tiene indirettamente conto delle possibili differenze nelle economie di scala in corrispondenza di capacità produttive molto distanti tra loro, fornendo una sorta di indicazione media.

Un semplice Wald test, condotto sul modello translog “ridotto”, ci permette di stabilire se la restrizione che annulla i coefficienti del 2° ordine, trasformando il modello [6.1] in una specificazione Cobb-Douglas, è statisticamente accettabile. Dal confronto emerge chiaramente come il valore del χ^2 (548.02) sia molto elevato, rigettando, con ogni probabilità, l'ipotesi che la forma funzionale più adatta sia la Cobb-Douglas.

Preso atto dell'esito positivo del test, che ci rassicura sull'adeguatezza della forma funzionale scelta per stimare il modello di costo, ci rivolgiamo ad indagare un secondo aspetto, che riguarda il possibile effetto esercitato dalla specificazione in forma binaria (*dummies*) di alcune variabili costruite partendo da valori continui.

Come discusso nel capitolo 5, il modello di costo impiega alcune variabili trasformate in forma binaria, nonostante l'informazione originale fosse continua. Siamo dunque interessati a verificare il possibile effetto sulle stime esercitato dalla specificazione in forma binaria (*dummies*) di alcune variabili “ausiliarie” del modello, deputate cioè a catturare gli effetti sui costi del case-mix e della qualità.

Nella Tabella 6.4 presentiamo, quindi, il modello [6.1] messo a confronto con 4 varianti che considerano, al posto delle due *dummies* D_{RM} e D_{TP} presenti nel modello [6.1], le sottostanti variabili continue R_M e T_P .²⁷⁷

Nella seconda colonna, le due variabili continue (R_M e T_P) sono aggiunte al modello di base, mentre nella terza e quarta colonna troviamo le due combinazioni con

²⁷⁷ Le due variabili R_M e T_P esprimono rispettivamente il valore medio per ospite del rimborso delle spese mediche (variabile di *case-mix*) ed il grado medio di impiego rispetto al tempo pieno della forza lavoro (*proxy* per la qualità di processo).

una variabile continua ed una variabile *dummy*. La quinta colonna raffigura il modello che utilizza le quattro variabili binarie (D_{RM} , D_{TP} , D_{QM} e D_{QS}), in sostituzione degli indicatori originali (R_M , T_P , Q_M e Q_S), definiti nel continuo.

Tabella 6.4 - Stabilità delle stime rispetto alla specificazione delle variabili ausiliarie.

Parametro		Mod. [6.1]	$R_M + T_P$	$R_M + D_{TP}$	$T_P + D_{RM}$	4 Dummies
Costante	α_0	14.98	15.01	14.98	15.01	14.91
Coeff. 1°ordine	α_Y	0.912	0.908	0.921	0.898	0.915
	α_L	0.630	0.643	0.634	0.639	0.608
	α_H	0.038	0.037	0.037	0.038	0.052
	α_{Qm}	0.192	0.172	0.188	0.175	-
	α_{Qs}	0.049	0.048	0.049	0.049	-
Coeff. 2°ordine	α_{YY}	0.299	0.258	0.243	0.314	0.312
	α_{LL}	0.054	0.046	0.052	0.047	ns
	α_{HH}	ns	ns	ns	ns	ns
	α_{QQm}	0.086	0.076	0.087	0.075	-
	α_{QQs}	ns	0.071	ns	ns	-
Prodotti incrociati (Y, L, H) + D_{FO} , $D_{REGIONO}$ + Trend: inclusi ma non presentati						
Dummies ausiliarie: D_{RM} D_{TP} D_{QM} D_{QS}	α_{DRM}	ns	-	-	ns	ns
	α_{DTP}	0.033	-	0.031	-	0.046
	α_{DQM}	-	-	-	-	0.063
	α_{DQS}	-	-	-	-	ns
Coeff. 1° 2°:	α_{RM}	-	0.019	ns	-	-
	α_{RRM}	-	ns	ns	-	-
Rimborso (R_M) e Tempo-Pieno (T_P)	α_{TP}	-	0.167	-	0.168	-
	α_{TTP}	-	0.134	-	ns	-

NOTE: ns = non significativo (livello di confidenza pari a 1%)

La prima considerazione da fare osservando, i coefficienti delle due variabili continue aggiunte al modello [6.1], è che il rimborso medio per ospite (R_M) ha un basso impatto ed è quasi sempre non significativo, mentre la variabile T_P mostra un alto livello di significatività ed un peso notevole. La prima variabile (R_M) non contribuisce molto alla spiegazione dei costi di una casa per anziani e, dal punto di vista pratico, può essere anche omessa. Per la variabile T_P , al contrario, si pone il problema di conferire (utilizzando il regressore nella forma continua) troppa importanza ad un indicatore, la cui efficacia nel misurare differenze di organizzazione nel processo produttivo non è certa. Per motivi opposti abbiamo quindi preferito trasformare le due variabili nella forma dicotomica, conferendo a questi regressori un significato maggiormente qualitativo.

In merito alla stabilità dei parametri, appare chiaro come in generale vi sia un buon allineamento dei coefficienti, anche se il parametro che misura il coefficiente dell'output (α_Y) esprime un campo di variazione non trascurabile. Preso come riferimento il coefficiente dell'output stimato con il modello [6.1], il cui valore è pari a 0.912, le differenze tra le varie specificazioni sono nell'ordine del 1-2% confermando una sostanziale stabilità delle stime.

Questa solidità empirica, che emerge in modo sistematico anche dalle molteplici varianti considerate nel corso della ricerca, è un ulteriore elemento di tranquillità per decidere su quale modello econometrico basare le analisi successive. Le indicazioni forniteci dalle stime della frontiera di costo ci permettono di sviluppare nel prosieguo del capitolo l'analisi dell'inefficienza di costo e delle economie di scala nel settore delle case per anziani in Svizzera.

6.2 Misurazione e analisi dell'inefficienza di costo

Nel capitolo 3 abbiamo definito [3.16] l'inefficienza di costo come il rapporto tra il costo minimo possibile (frontiera) ed il costo osservato per produrre un certo output. In altre parole, si misura la contrazione che il livello di costo osservato dovrebbe subire per raggiungere il costo minimo (la frontiera di costo).

L'inefficienza di costo è quindi definita nello spazio [0,1] in corrispondenza delle due situazioni estreme di piena efficienza (efficienza di costo = 1) e di completa inefficienza (efficienza di costo = 0). In alternativa si potrebbe considerare l'approccio inverso, misurando in quale proporzione il costo effettivo (osservato) di una casa per anziani è superiore a quello efficiente (previsto dalla frontiera). Tale misura è più intuitiva da interpretare e varia da 0 (perfetta efficienza) ad infinito.

In pratica, utilizzando la forma funzionale translog, è sufficiente invertire la [4.11] per costruire il punteggio di (in)efficienza di costo di ogni singola osservazione (PEC_{it}) nel seguente modo:

$$PEC_{it} = 1 - \frac{C_{it}}{C_{it}^*} = \frac{\exp[\beta_0^* + c(y_{it}, w_{it}; \hat{\beta}) + u_{it}^*]}{\exp[\beta_0^* + c(y_{it}, w_{it}; \hat{\beta})]} = 1 - \exp[u_{it}^*] \quad [6.4]$$

L'inefficienza di costo di ogni singola impresa (in un determinato anno) viene quindi misurata partendo dal rapporto tra la componente deterministica dei costi osservati (C_{it}) ed il valore indicato dalla frontiera (C_{it}^*).²⁷⁸ La formula [6.4] indica, per costruzione, direttamente l'eccesso di costo in termini percentuali rispetto al valore efficiente misurato sulla frontiera. I confronti tra i punteggi di (in)efficienza, ottenuti a partire dai parametri dei diversi modelli sono quindi possibili e di particolare rilevanza per valutare in modo critico la portata dei risultati sul livello (medio) di (in)efficienza di costo nel settore delle case per anziani.

6.2.1 Confronto dell'(in)efficienza di costo nei diversi modelli

I residui delle regressioni permettono, con tecniche diverse, di ottenere la stima (u_{it}) della componente individuale e sistematica dell'errore, stima che determina, utilizzando la [6.4], i punteggi di (in)efficienza di costo.

Prima di presentare i dettagli dell'analisi comparativa, occorre ricordare come i diversi modelli impiegati si basino su ipotesi differenti in merito alla struttura dell'eterogeneità non osservata e dell'inefficienza di costo. L'interpretazione dei risultati deve quindi pesare il valore assoluto dei punteggi ottenuti, alla luce delle proprietà econometriche del modello che li ha generati. I due aspetti sono inscindibile e quindi verranno discussi entrambi in sede di commento ai risultati dei punteggi di (in)efficienza di costo.

Un quadro sintetico del confronto tra i modelli è offerto dalla Tabella 6.5, nella quale presentiamo le principali statistiche descrittive dei punteggi di (in)efficienza di costo delle case per anziani svizzere.

La prima constatazione da fare è la grande differenza tra il livello di inefficienza medio misurato dai diversi modelli: il modello TREM esprime un eccesso medio di costo pari a 8,1% (valore più basso), mentre il valore più alto, pari ad un eccesso di costo medio pari al 265,4%, è rilevato dal modello FEM. L'enorme distanza tra i due valori non deve allarmare poiché, anche se si tratta della stessa misura (l'inefficienza di

²⁷⁸ Per componente deterministica dei costi si intende il costo effettivamente osservato nei dati, dedotta la parte stocastica del residuo ε_{it} che, come spiegato nel capitolo 4, viene scomposto in due componenti: la parte sistematica (che misura l'inefficienza) e la parte casuale (*white noise*).

costo) ottenuta dallo stesso modello di costo, la taratura dello strumento (ma non lo strumento) è molto diversa ed il risultato era atteso.

Tabella 6.5 - Statistiche descrittive dei punteggi di (in)efficienza di costo.

Modello	Media	Dev. Std.	Min.	25° Perc.le	Mediana	75° Perc.le	Max
FEM	2.654	1.241	0	1.758	2.617	3.572	5.745
REM	0.733	0.256	0	0.569	0.748	0.917	1.595
ML-REM	0.266	0.153	0.007	0.148	0.251	0.374	0.891
TREM	0.081	0.064	0.011	0.048	0.067	0.096	1.044
POOL	0.129	0.077	0.018	0.078	0.111	0.154	0.656

La stima con il modello FEM, infatti, non distingue tra eterogeneità ed inefficienza, convogliando nel vettore delle caratteristiche individuali le differenze di costo non spiegate dai regressori. Questo significa che i confronti tra i costi delle diverse case per anziani avvengono sulla base dell'eterogeneità nelle caratteristiche più che sull'efficienza di costo. In altri termini, il modello FEM rileva come inefficienza anche tutte le differenze di costo dovute invece alle caratteristiche individuali delle osservazioni. Per contro, il modello TREM è in grado di misurare e scorporare (almeno parzialmente) il contributo dell'eterogeneità non osservata da quello dell'inefficienza, rendendo la misura di quest'ultima più affidabile. I punteggi di inefficienza calcolati con questo metodo potrebbero, però, risultare sottostimati, dal momento che la presenza di inefficienza costante nel tempo viene etichettata come eterogeneità non osservabile. L'allineamento tra risultati empirici e aspettative teoriche, riscontrato nella misura dell'inefficienza con il metodo rispettivamente meno e più favorevole, è un chiaro indicatore che l'eterogeneità non osservabile gioca un ruolo importante, meritevole di un successivo approfondimento.

Tracciati, così, i confini empirici (ma teoricamente fondati) dell'inefficienza media, notiamo che negli altri modelli ad effetti casuali si raggiunge un livello medio di inefficienza non trascurabile e pari a 73,3% (REM) e 26,6% (ML-REM). Il modello POOL, che ignora la struttura temporale dei dati, indica un livello medio di inefficienza contenuto (12,9%) ed è il più vicino al valore misurato dal modello TREM.

I risultati presentati nella Tabella 6.5 sottolineano come la misura assoluta dell'inefficienza sia sempre associata ad un certo margine di incertezza. Per questa

ragione sembra necessario discutere, da una parte i motivi che potrebbero determinare la forte oscillazione osservata nelle diverse misure di inefficienza e, dall'altra, il comportamento dell'inefficienza relativa. Questo secondo aspetto, che sarà discusso in seguito, riguarda la stabilità dei ranghi delle case per anziani nella graduatoria di inefficienza stilata in base ai diversi modelli.²⁷⁹ Prima, infatti, desideriamo occuparci delle differenze di inefficienza (assoluta) rilevate dai diversi modelli, espressione del problema di misurazione e separazione di due componenti non osservabili: inefficienza ed eterogeneità.

Le differenze appena evidenziate sono, infatti, il riflesso dell'interazione tra la realtà che si cela dietro i dati e la predisposizione dei modelli utilizzati ad enfatizzare un aspetto specifico della stessa. Con ogni probabilità, non tutte le case per anziani svizzere operano sulla frontiera di costo, così come i dati a nostra disposizione per misurarla non rivelano tutta l'eterogeneità presente. Il problema è quello di capire quale sia la struttura temporale delle due variabili latenti, che meglio si adatta al settore dell'assistenza agli anziani.

Nell'analisi dell'inefficienza di costo, il problema dell'eterogeneità non osservata viene accentuato in modo estremo dal modello ad effetti fissi (FEM), che d'ora in avanti sarà considerato nei commenti solo per segnalare aspetti econometrici di particolare interesse. Pertanto, i rimanenti 4 modelli stimati si possono suddividere in due categorie nette: i modelli (REM e ML-REM) che ipotizzano un'inefficienza costante nel tempo, ed i modelli (TREM e POOL) che assumono l'inefficienza variabile nel tempo.²⁸⁰

Il confronto tra i due approcci mostra una chiara differenza nei risultati dei punteggi medi di efficienza, indicando un livello maggiore per i modelli che ipotizzano l'inefficienza costante nel tempo. Questo risultato è coerente nella misura in cui: (1) ci sia una presenza di eterogeneità non osservata importante e (2) il modello TREM sia in grado di rilevarla meglio degli altri modelli. La seconda condizione è sicuramente vera

²⁷⁹ In questa sede facciamo solo notare come la casa per anziani che esprime il minimo punteggio di inefficienza coincida per i modelli REM, ML-REM e TREM (nel POOL è 5^a e nel FEM 15^a), mentre l'istituto con il più alto valore (354°) di inefficienza coincide per i modelli REM, ML-REM e POOL (nel FEM è 349° e nel TREM è 310°).

²⁸⁰ Ad essere precisi, il modello POOL, ignorando la struttura temporale dei dati, non attribuisce all'inefficienza nessun tipo di comportamento teorico. Tuttavia, trattando tutte le osservazioni in modo distinto, i valori annuali riferiti allo stesso istituto possono risultare diversi tra loro, assimilando il modello alla tipologia con inefficienza variabile.

per il modello TREM sviluppato da Greene (2005) appositamente per questo scopo, mentre resta da capire (situazione non nuova in letteratura) come mai il modello POOL, che si basa semplicemente sull'asimmetria dei residui, esprima un livello di inefficienza media nettamente inferiore ai modelli REM e ML-REM. In molti lavori empirici sull'analisi dell'inefficienza di costo con le frontiere stocastiche, le stime del livello medio di inefficienza del modello POOL sono leggermente superiori a quelle del modello TREM, ma chiaramente inferiori agli altri modelli panel.²⁸¹ Una possibile spiegazione è legata al fatto che il modello POOL ignora la struttura panel presente nei dati, mancando di considerare nei residui l'effetto delle caratteristiche specifiche degli individui e costanti nel tempo.²⁸² Questa omissione, se da un lato esaspera il problema del “*Heterogeneity Bias*” dei coefficienti della funzione di costo, dall'altro tende a produrre livelli di inefficienza più contenuti rispetto ai modelli panel tradizionali.

A prescindere dalle considerazioni di tipo tecnico-econometriche, il cuore del problema, la scelta del metodo più accurato per misurare le performance di costo delle imprese, è ben illustrato dalla seguente domanda:

“qual è la vera struttura temporale dell'inefficienza e dell'eterogeneità non osservata nel settore delle case per anziani svizzere?”

La risposta va cercata nei residui della regressione, dove inefficienza ed eterogeneità (non osservata) si mischiano, seguendo ciascuna la propria dinamica temporale. L'econometria propone alcuni modelli che aiutano a scomporre il termine di errore (ε_{it}) e determinare la componente di inefficienza. Tuttavia, per poter giungere ad una stima dell'inefficienza è necessario attribuirle, a priori, una struttura temporale (costante o variabile) ben precisa.

Il modello TREM, ad esempio, ipotizza che l'inefficienza sia un fenomeno che si manifesta di anno in anno con valori diversi, causata da fattori imponderabili (errori di gestione da parte degli amministratori, introduzione di nuovi processi e/o tecnologie,

²⁸¹ A titolo di esempio citiamo alcuni lavori, focalizzati sul confronto dell'inefficienza stimata con diversi metodi econometrici e riferiti a settori molto diversi (case per anziani, distribuzione di gas e di elettricità, ospedali, etc.): Greene (2005), Farsi et al. (2005), Farsi et al. (2007) Farsi et al. (2008).

²⁸² Nel modello POOL il termine di inefficienza u_{it} è considerato *iid* rispetto al tempo e agli individui. Per questa ragione non è in grado di catturare le differenze (eterogeneità) che esistono tra gli individui, ma sono costanti nel tempo.

turnover del personale in ruoli strategici, etc.) e sempre diversi in ogni periodo. In altre parole, gli istituti non perpetuano eventuali errori commessi e sono in grado di reagire, in tempi brevi, al mutare delle condizioni del settore (es. regolamentazione). A questa componente idiosincratICA dell'inefficienza, si sovrappone un elemento strutturale (cioè costante nel tempo) che cattura una parte dell'eterogeneità non osservata. Questo significa immaginare che la maggior parte delle caratteristiche produttive non osservate (qualità delle cure, infrastrutture, livello di formazione del personale, norme di legge, etc.) mutino lentamente. Il modello TREM ipotizza quindi l'eterogeneità (non osservata e specifica per ogni individuo) costante nel tempo, mentre attribuisce gli eventuali eccessi di costi a forme di inefficienza sempre nuove e diverse ad ogni periodo.

L'alternativa a questa "visione del mondo" è costituita dai modelli REM e REM-ML, che reputano l'inefficienza come permanente e imm modificabile nel tempo. Appare ragionevole considerare l'eterogeneità non osservata (o perlomeno i fattori più importanti come la qualità, il contesto economico di riferimento, etc.) come una caratteristica specifica (e quindi relativamente costante nel tempo) di ogni istituto. Viceversa, sulla natura dell'inefficienza riteniamo che entrambe le componenti (costante e variabile nel tempo) siano giustificabili e presenti in un panel di appena 5 anni. A ben vedere, l'intreccio di questi modi di considerare l'inefficienza costituisce l'anello di congiunzione tra elementi contrastanti come l'approccio al lungo periodo scelto (che non può prevedere inefficienza sistematica e persistente) con la breve durata del panel (che si concilia con la presenza di inefficienze di breve periodo).

In conclusione, consapevoli dei margini di errore contenuti in questa scelta, riteniamo il modello TREM maggiormente adatto a misurare il livello medio di inefficienza di costo nel settore delle case per anziani in Svizzera. Occorre dire che il modello TREM, nonostante riesca a filtrare l'eterogeneità non osservata, tende a sottostimare il valore effettivo dell'inefficienza di costo. Se, infatti, una parte di inefficienza fosse costante nel periodo considerato (ma non necessariamente su tempi più lunghi), verrebbe etichettata come eterogeneità e quindi non conteggiata come eccesso di costo ingiustificato.

Il confronto tra i valori medi dei punteggi di inefficienza misurati con i diversi modelli ha messo in luce importanti differenze assolute, che abbiamo legittimato in massima parte con le caratteristiche econometriche dei modelli utilizzati. A meno di un fattore di scala sembrerebbe quindi esistere una frontiera per valutare, almeno in termini

relativi, le prestazioni di costo di tutte le singole case per anziani, stilando la classifica dal più efficiente al meno efficiente. La coerenza delle indicazioni fornite dai diversi modelli è un requisito imprescindibile, qualora si desideri trasformare l'analisi econometrica delle frontiere stocastiche in uno strumento operativo, ad esempio nell'ambito della regolamentazione.²⁸³

Il problema è interessante anche da un punto di vista metodologico per verificare se la distorsione provocata dall'eterogeneità presente nei dati, altera il comportamento dei modelli rispetto al problema di ordinare i punteggi di inefficienza in una graduatoria. A tale scopo nella Tabella 6.6 presentiamo la correlazione tra i punteggi di inefficienza (e tra i relativi ranghi) calcolati con i diversi modelli. Le correlazioni tra i valori di inefficienza sono disposte nella semi-matrice inferiore, mentre le celle che si trovano sopra la diagonale principale indicano la correlazione tra i ranghi (generati dai punteggi di inefficienza).

Tabella 6.6 - Correlazione dei punteggi di inefficienza (valori e ranghi).

Modello	FEM	REM	ML-REM	TREM	POOL
FEM	1	0.623	0.547	0.030	0.289
REM	0.641	1	0.973	0.244	0.725
ML-REM	0.555	0.961	1	0.280	0.806
TREM	-0.023	0.047	0.069	1	0.473
POOL	0.174	0.525	0.621	0.570	1

Per confrontare i modelli ad inefficienza costante con quelli ad inefficienza variabile nel tempo abbiamo considerato la media temporale dei punteggi in modo da avere un unico valore di inefficienza per ogni casa per anziani, come nei modelli ad inefficienza costante. La correlazione tra i modelli è sistematicamente più elevata nei ranghi (risultato atteso), ma risulta piuttosto bassa per alcune combinazioni. Il modello TREM è quello che evidenzia i valori più bassi, mentre i modelli REM e REM-ML

²⁸³ Secondo la filosofia della *yardstick competition*, il regolatore potrebbe utilizzare la stima di una frontiera di costo per negoziare i budget globali corrisposti alle imprese private che forniscono servizio pubblico nel settore dell'assistenza agli anziani. Nel cantone Ticino a partire dall'anno 2002 sono entrati in vigore i contratti di prestazione e, nella fase preliminare di scelta del metodo di definizione dei budget, è stato considerato anche l'approccio econometrico. Il Dipartimento della Sanità e Socialità ha infatti commissionato uno studio (Farsi et al. 2003) di frontiera stocastica di costo per gli istituti ticinesi.

sono altamente correlati tra loro (0,96) e discretamente correlati sia con il modello ad effetti fissi (0,64 e 0,55) sia con il modello POOL (0,52 e 0,62).

Dal punto di vista metodologico è interessante notare come il comportamento dell'unico modello ad inefficienza variabile nel tempo (TREM) sia molto diverso dagli altri modelli. Ancora una volta emerge il problema dell'eterogeneità non misurata che incide pesantemente sui livelli medi di inefficienza e anche nell'ordinamento delle graduatorie. Su questo tema, in letteratura, il dibattito è aperto e l'evidenza empirica contrastante. Le basse correlazioni tra modelli ad inefficienza costante e variabile sono in linea con lo studio di Farsi et al. (2005), mentre contrastano con i risultati di Greene (2005). Da notare che in questo studio (basato su un panel di 5 anni relativo ad un campione di banche americane) la distorsione causata dall'eterogeneità non osservata è risultata, a giudizio dell'autore, molto contenuta. L'analogia di risultati con il paper di Farsi et al. (2005), che analizza panel di case per anziani ticinesi, è confortante, dal momento che i due lavori condividono le stesse problematiche relative al settore dell'assistenza agli anziani.²⁸⁴

Alla luce di queste considerazioni metodologiche, che evidenziano una certa difficoltà nell'ordinare in modo univoco le prestazioni economiche dei singoli individui, sembra rischioso tradurre in modo meccanico i risultati empirici di una frontiera di costo in termini di provvedimenti operativi di dettaglio. Questo non significa però che i risultati dell'analisi dell'efficienza di costo non forniscano indicazioni attendibili, al contrario.

Nonostante i limiti discussi, le analisi dell'efficienza di costo indicano in modo chiaro la rilevanza dell'eccesso di costo nel settore, per il quale il modello TREM costituisce un riferimento certamente conservativo (sottostima dell'inefficienza), e individuano con buona precisione il posizionamento delle singole performance in un segmento specifico della graduatoria dei punteggi. La sensibilità del rango ai modelli decresce in modo marcato fin quasi a scomparire se consideriamo, ad esempio, una

²⁸⁴ Le correlazioni tra i modelli analizzati da Farsi et al. (2005) sono decisamente più elevate rispetto a quelle evidenziate in questo studio. Una possibile motivazione è legata al minor impatto dell'eterogeneità non osservata sui risultati econometrici. La banca dati a disposizione presenta infatti differenze importanti rispetto a tre dimensioni: (1) è prodotta dal Dipartimento di Sanità e Socialità del cantone Ticino, (2) consiste in un panel di 9 anni e (3) contiene informazioni più dettagliate sul profilo medico degli ospiti.

suddivisione in decili della distribuzione dei punteggi di inefficienza calcolati con i diversi modelli. In altre parole, l'esatta posizione di ogni singola casa per anziani nella "classifica degli inefficienti" è molto instabile, ma difficilmente i diversi modelli divergono nell'attribuire agli istituti un'informazione meno specifica (ma più indicativa), sotto forma di scala valutativa.²⁸⁵

6.2.2 (In)efficienza di costo ed eterogeneità

Nonostante abbiamo individuato alcuni motivi che sconsigliano la formulazione di raccomandazioni di dettaglio in merito alle politiche cantonali, riteniamo che la stima di una frontiera di costo unica per tutta la Svizzera sia una base sufficientemente solida per analizzare l'inefficienza di costo delle case per anziani, sotto altre prospettive. Come discusso nella sezione metodologica, l'inefficienza di costo di un singolo istituto esprime l'eccesso di costo, rispetto alla frontiera di riferimento stimata, per produrre un dato livello di output. Per questa ragione l'analisi dell'efficienza di costo ha un'ottica prevalentemente individualistica che permette di elaborare, alla luce dei punteggi di efficienza di tutto il campione, anche considerazioni di tipo comparativo tra diverse dimensioni, ad esempio: tra cantoni, tra capacità produttive diverse e tra tipologie di forme istituzionali.

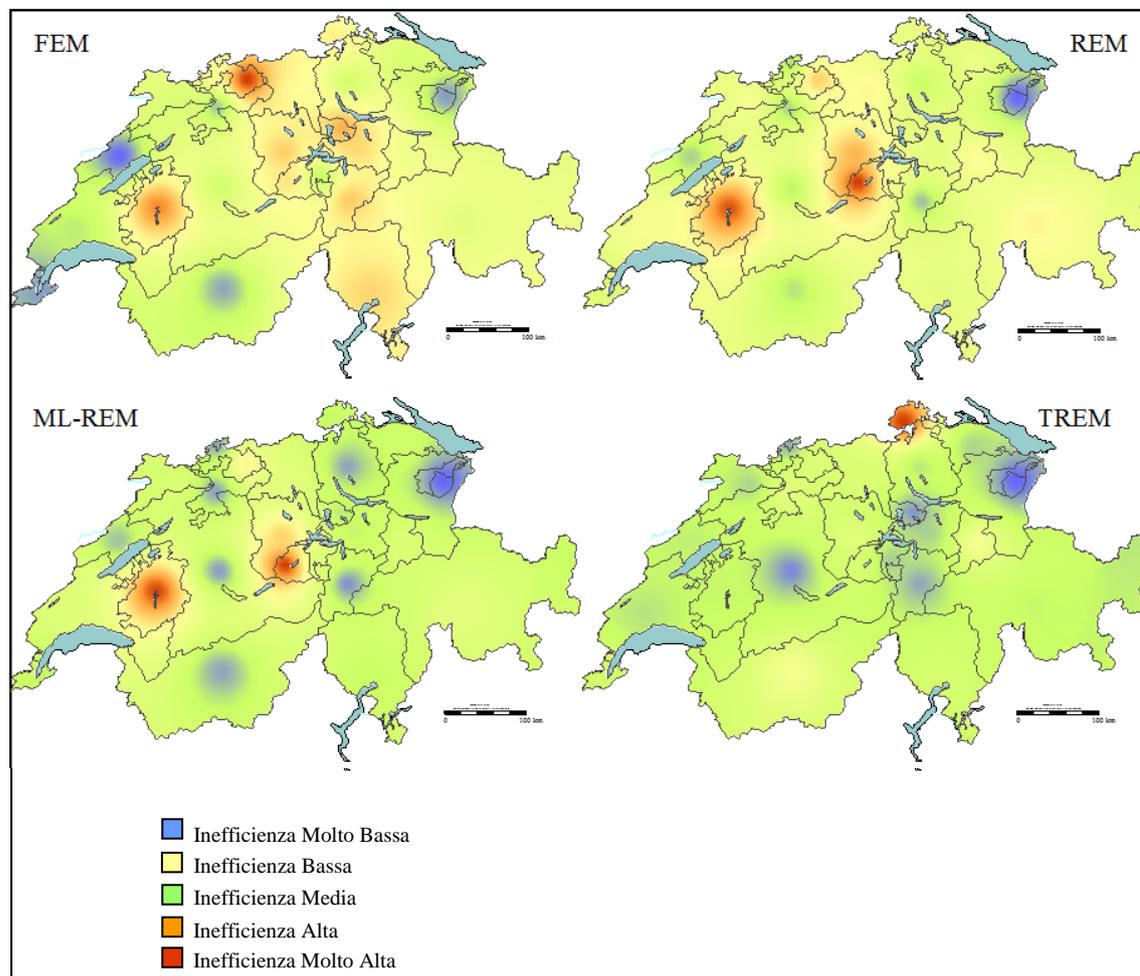
Per completare l'analisi dell'inefficienza di costo delle case per anziani, desideriamo, quindi, approfondire da un'altra prospettiva il difficile rapporto tra eterogeneità ed efficienza, raggruppando le osservazioni in gruppi di omogenei e confrontando in modo qualitativo i punteggi di inefficienza dei diversi gruppi.

Viene spontaneo considerare come primo criterio distintivo la dimensione geografica, separando i punteggi di efficienza registrati a livello cantonale. L'obiettivo principale è quello di stabilire se, a prescindere dal metodo utilizzato, i punteggi di inefficienza delle singole case per anziani seguano un criterio territoriale, determinando differenze sistematiche tra i cantoni. La Figura 6.1 propone un confronto tra i 4 metodi panel utilizzati, nel quale i punteggi di inefficienza di ogni cantone sono messi in

²⁸⁵ Ad esempio, basandosi sulla posizione dei quintili, si potrebbero convertire i singoli punteggio in una scala con 5 attributi per la valutazione dell'inefficienza: molto bassa, bassa, media, alta, molto alta.

relazione con i quintili della distribuzione a livello nazionale e trasformati in una mappa geografica a gradienti.²⁸⁶

Figura 6.1 - Eterogeneità territoriale dei punteggi di inefficienza.



In pratica l'intervallo di variazione (differenza tra il valore massimo ed il valore minimo del campione) viene diviso in 5 classi di inefficienza alle quali si assegna un colore specifico (cfr. legenda Figura 6.1).

All'interno dei confini cantonali tutti i punteggi concorrono a posizionare il cantone rispetto alla scala nazionale, "colorandolo" appropriatamente. Se la struttura dell'inefficienza di costo differisse molto da cantone a cantone, avremmo un'immagine

²⁸⁶ La mappa a gradienti associa al centroide di ogni singola area (confini cantonali) uno specifico colore, tra quelli ammessi dallo spettro predefinito. Il colore è determinato da un processo di interpolazione di tutti i dati riferiti ad ogni singola area, utilizzando un algoritmo specifico (in questo caso l'inverso della distanza elevato a potenza). Questo algoritmo, che sostanzialmente riflette la densità delle osservazioni intorno alla media associata al centroide, tende a generare un effetto grafico a sfere quando vi sono importanti differenze di inefficienza (colore) tra due cantoni limitrofi. L'elaborazione grafica è stata realizzata con il programma MapViewer 7. Per un approfondimento su questa tecnica di interpolazione si vedano Davis (1986) e Franke (1982).

a macchie colorate, mentre una distribuzione cantonale che rispecchia alla perfezione quella nazionale, darebbe origine ad una figura monocromatica (di colore verde).

Filtrando progressivamente l'eterogeneità non osservata dai punteggi di inefficienza, costruendo cioè la stessa mappa con i punteggi generati dai diversi modelli econometrici, si nota come l'immagine della cartina si vada uniformando verso il valore mediano nazionale, producendo un'immagine molto omogenea nel caso del modello TREM.²⁸⁷ Questo dato suggerisce che la forte eterogeneità presente nei 26 cantoni svizzeri, pur giocando un ruolo importante nel definire i costi di produzione di una casa per anziani, non sembra determinare differenze sistematiche nei livelli di inefficienza di costo a livello cantonale.

Per confermare le suggestioni visive, suscitate dal modello TREM nella Figura 6.1, abbiamo verificato se i punteggi cantonali di inefficienza fossero generati da distribuzioni diverse e se le medie cantonali fossero diverse da quella nazionale. Il confronto tra le varie distribuzioni cantonali dei punteggi di inefficienza si è svolto per mezzo del test non parametrico Kruskal-Wallis che, in modo molto netto, ha rifiutato l'ipotesi di situazioni cantonali diverse da quella nazionale. La verifica delle distanze dei punteggi di inefficienza medi cantonali rispetto alla media generale ha sostanzialmente ribadito il concetto, dal momento che solo 3 su 24 cantoni presenti nel campione, si scostano dalla media nazionale (verso il basso) in modo statisticamente significativo (ad un livello di confidenza del 95%).²⁸⁸

Con le dovute cautele, legate sostanzialmente ad una valutazione statistica prevalentemente descrittiva, questi risultati evidenziano come le differenze tra i punteggi di inefficienza non appaiano legate in modo stretto ai confini geografici cantonali, quanto piuttosto ad altri elementi strutturali. L'eterogeneità, introdotta da non meglio specificate caratteristiche, si manifesta con intensità diverse nei cantoni (cfr. capitolo 2.3) e viene trasmessa ai differenziali di (in)efficienza tra i cantoni, in

²⁸⁷ Da notare che la rappresentazione grafica non tiene conto del peso che i diversi cantoni hanno in termini di numero di osservazioni. Con l'eccezione del cantone Berna, il numero di osservazioni nei cantoni associati ad un livello di inefficienza molto bassa (colore blu) o molto alta (colore rosso) è molto contenuto. Nel dettaglio troviamo: Sciaffusa – inefficienza molto alta – (7 osservazioni), Zugo, Appenzello Esterno e Uri – inefficienza molto bassa – (rispettivamente 6, 3 ed 1 osservazioni).

²⁸⁸ I tre cantoni che mostrano un'inefficienza media inferiore a quella nazionale (7,4%) sono: Berna (6,9%), Neuchâtel (6,8%) e Argovia (6,7%).

proporzione inversa a quanto il modello riesca a filtrare l'eterogeneità dalla misura di inefficienza cantonale.

L'analisi intercantonale dell'inefficienza non ha evidenziato differenza nel livello delle performance economiche degli istituti per anziani, ma questo non significa che non vi siano chiavi di lettura alternative in grado di rilevarne. Tra le molte dimensioni che si potrebbero considerare, siamo interessati a due aspetti specifici: (1) la forma istituzionale, e (2) la dimensione (capacità produttiva) degli istituti. Questi criteri di raggruppamento travalicano i confini cantonali e analizzano le possibili differenze di inefficienza (media) espresse da gruppi di osservazioni omogenei rispetto alla caratteristica scelta.

Il criterio istituzionale mira a verificare se, come proposto dalla letteratura economica sul principale-agente e sulla teoria dei diritti di proprietà, l'assetto organizzativo determina delle differenze significative di inefficienza così come ipotizzato dalla teoria economica.²⁸⁹

L'ipotesi da testare è quella che, *ceteris paribus*, le imprese pubbliche siano mediamente meno efficienti di quelle private e, all'interno di quest'ultima categoria, le imprese for profit più efficienti rispetto a quelle non profit. Per quanto riguarda la dimensione è interessante verificare il comportamento dell'inefficienza a seconda della capacità produttiva che è stata suddivisa in tre classi a dipendenza del numero di posti letto disponibili.²⁹⁰

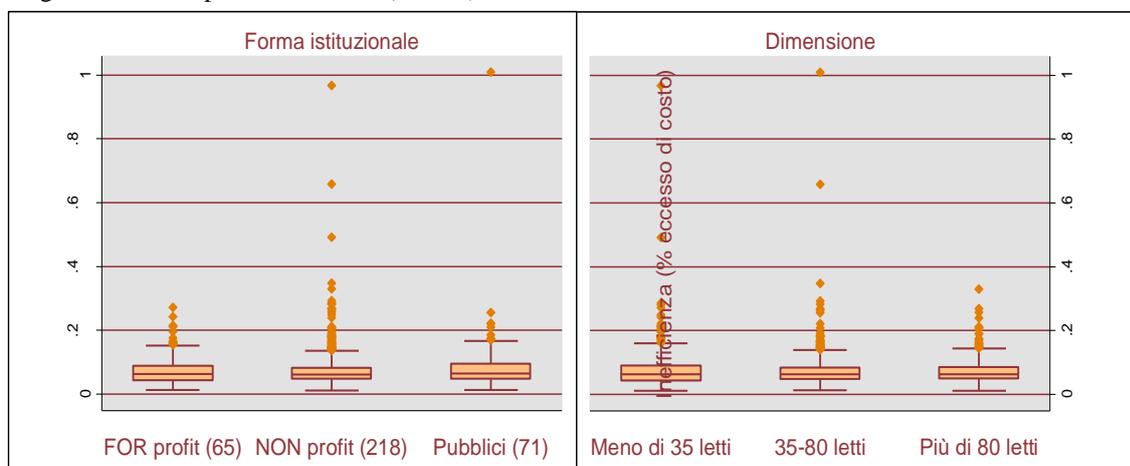
Nella Figura 6.2 presentiamo il box-plot dei punteggi di inefficienza calcolati con il modello TREM e suddivisi secondo la classificazione appena descritta. Dal confronto emerge un dato (forse) inatteso: una sostanziale somiglianza nella forma e nella posizione delle distribuzioni dei 6 gruppi considerati. In altre parole, non sembrano emergere, a livello generale, differenze importanti nelle performance di case per anziani

²⁸⁹ La letteratura in questo ambito è molto ampia ed articolata e citiamo a titolo di esempio alcuni riferimenti teorici quali Demsetz (1967) Shleifer (1985) Laffont and Tirole (1993), Kaserman and Mayo (1995), Weisbrod (1988) ed altri più applicati come Frech (1985), Sloan (2000), Vitaliano (2003).

²⁹⁰ Le tre classi dimensionali (piccola, media e grande) sono costruite utilizzando il 1° ed il 3° quartile in modo tale che il 50% delle osservazioni appartenga alla classe media e l'altro 50% sia simmetricamente distribuito nelle due classi estreme.

con caratteristiche differenti per quanto attiene alla dimensione e alla struttura organizzativa.

Figura 6.2 - Box-plot inefficienza (TREM), secondo la forma istituzionale e la dimensione.



Anche nel confronto tra questi gruppi di osservazioni abbiamo eseguito il test di Kruskal-Wallis, che ha dato per entrambi gli aspetti considerati (forma istituzionale e dimensione) un esito negativo.

Questo risultato va però riletto alla luce delle considerazioni fatte in precedenza sulla difficoltà di separare l'eterogeneità non osservata dalla misura dell'inefficienza. La modifica della dimensione o della forma istituzionale, pur certamente possibile e presente in minima misura nel campione, è un evento abbastanza infrequente, rendendo queste due caratteristiche stabili nel tempo.

Se accettiamo che alcuni problemi insiti nella forma organizzativa scelta e/o nella scala di produzione possano generare una certa inefficienza (giustificata ad esempio dagli elevati costi da sostenere per la sua totale eliminazione), dovremmo riconoscere che i punteggi di inefficienza misurati con il modello TREM non sono in grado di misurare tali differenze sistematiche.

Nonostante la possibilità che il modello TREM sottostimi il valore (assoluto) dell'inefficienza di costo, l'analisi comparativa ha messo in luce la maggior affidabilità di questo modello rispetto agli altri. Per questa ragione, nel prossimo paragrafo, presenteremo l'analisi delle economie di scala del settore, basandoci principalmente sulle stime ottenute con il modello TREM.

6.3 Analisi delle economie di scala

I concetti di economie di scala, di dimensione e di efficienza di scala, che sono stati introdotti e presentati nel capitolo 3 nella loro forma generale, trovano ora spazio per un'analisi empirica basata sui risultati delle stime della funzione di costo. Il concetto di economie di scala, già presentato nell'espressione [3.12], può essere calcolato come il reciproco dell'elasticità di costo rispetto all'output:

$$E_{Scala} = \frac{1}{\varepsilon_y(y, w)} = \frac{1}{\frac{\delta \ln C}{\delta \ln y}} = \frac{1}{\alpha_y + \alpha_{yy} \ln(y/\bar{y}) + \alpha_{yL} \ln\left(\frac{P_L}{P_K} / \frac{\bar{P}_K}{\bar{P}_L}\right) + \alpha_{yH} \ln(H/\bar{H})} \quad [6.5]$$

Nel punto di approssimazione dei dati l'espressione [6.5] si semplifica al reciproco dell'elasticità di costo rispetto all'output ($1/\alpha_y$), mentre per calcolarne i valori al di fuori di questo punto occorre considerare tutti i parametri coinvolti nella formula [6.5]. Appare quindi chiaro che i diversi modelli possono dare risposte diverse ed è questo il motivo che ci spinge a presentare in modo comparativo i risultati empirici ottenuti. Si hanno economie di scala quando i valori di E_{Scala} risultano superiori ad 1, mentre parleremo di diseconomie di scala nel caso contrario. In altre parole, la presenza delle economie di scala implica che la funzione del costo medio unitario sia decrescente rispetto alla dimensione (misurata dall'output), rendendo conveniente un'espansione della scala produttiva.

Per descrivere l'andamento dei costi unitari quando ci si allontana dal punto mediano di approssimazione dei dati, abbiamo calcolato il valore delle economie di scala in altri due punti di approssimazione: il 1° ed il 3° quartile. Non esiste un motivo specifico nella scelta di questi due punti alternativi di approssimazione se non quello della simmetria rispetto al valore mediano ed il fatto che siano state presentate le statistiche descrittive anche per il 1° ed il 3° quartile della distribuzione delle variabili, offrendo al lettore un riferimento concreto nell'interpretazione delle economie di scala.

Nell'analisi delle economie di scala abbiamo deciso di presentare i risultati ottenuti con i due modelli seguenti: REM e TREM. Come discusso in precedenza, il modello preso come riferimento principale è il modello TREM al quale affianchiamo come termine di confronto alternativo il modello da esso più distante. Il REM, infatti, non nasce come modello di frontiera e inoltre viene stimato con la procedura GLS

(invece che con la massima verosimiglianza) senza imporre al termine di errore alcuna distribuzione.

Per questi due modelli (TREM e REM) abbiamo calcolato le economie di scala, la dimensione ottimale ed il costo unitario minimo corrispondente. Per il calcolo delle economie di scala nei due punti specifici di approssimazione diversi dalla mediana si è preferito ragionare in termini di *ceteris paribus* (variando cioè solamente il livello dell'output), riferendo tutte le altre variabili coinvolte nel calcolo delle economie di dimensione al loro valore mediano.²⁹¹

Prima di commentare in dettaglio i risultati dei diversi modelli occorre precisare che il termine dimensione, misurato in numero di letti, è in realtà riferito al numero di giornate di cura erogate in un anno. Vista l'ottica di lungo periodo considerata e l'eccesso cronico di domanda presente nel settore, ipotizzando quindi un tasso di occupazione dei letti molto vicino al 100%, la capacità produttiva in termini di posti letto si ricava dividendo le giornate di residenza (Y) per i 365 giorni dell'anno.

Nella Tabella 6.7 presentiamo i valori delle economie di scala, stimate a partire da 2 dei 5 modelli econometrici considerati (REM e TREM) e calcolate con riferimento a tre diversi livelli dell'output.

Tabella 6.7 - Economie di dimensione calcolate per diverse capacità produttive.

Modello	Livello scelto dell'output (Y)			Capacità ottimale in posti letto ^a	Costo medio minimo ^b	Costo medio per 30 posti letto
	Mediana (52 letti) (Int. confidenza-99%)	1° quartile (33 letti)	3° quartile (77 letti)			
REM	1.096 (1.126 – 1.069)	1.286	0.969	69	163.80	183.- (+8.9%)
TREM	1.083 (1.093 – 1.074)	1.231	0.979	71	157.50	172.- (+9.5%)

NOTE:

^a La dimensione ottimale corrisponde al numero di giornate (output) che minimizza i costi unitari ed è espressa in termini di posti letto dividendo per 365 giorni l'output ottimale.

^b Il valore è calcolato utilizzando i parametri della stima della funzione di costo totale ed imponendo il valore mediano a tutte le variabili eccetto l'output (giornate di cura) che corrisponde al valore ottimale.

²⁹¹ Se modifichiamo solamente il punto di approssimazione dell'output (Y), l'espressione [6.4] si riduce a $1/\alpha_y + \alpha_{yy} \ln(y/\bar{y})$. Modificare anche il punto di approssimazione delle altre variabili significa immaginare che vi sia una correlazione positiva tra tutti i regressori, ossia che una casa per anziani piccola debba anche fronteggiare prezzi inferiori, case-mix più leggeri, etc. Per completezza abbiamo effettuato il calcolo delle economie di scala con tutte le variabili approssimate al 1° o 3° quartile, ottenendo risultati molto simili a quelli presentati.

Il primo commento che emerge è che in entrambi i modelli constatiamo la presenza di economie di scala per livelli di produzione superiore alla mediana. Questo fatto suggerisce come più della metà delle case per anziani presenti nel campione non sfrutti completamente le economie di scala presenti nel settore.

Se consideriamo il comportamento della funzione di costo in corrispondenza di livelli produttivi lontani dalla mediana, osserviamo come nei pressi del 1° quartile della distribuzione delle giornate di residenza (Y) il valore di E_{Scala} sia chiaramente superiore a 1 ($E_{Scala} = 1,23$ nel modello TREM e $E_{Scala} = 1,28$ nel modello REM). Il dato si interpreta come la forte presenza di economie di scala espresse dagli istituti con una ridotta capacità produttiva. Pertanto, secondo l'indicazione fornita da entrambi i modelli, quando il livello produttivo non raggiunge i 35 letti (circa il primo quartile) è possibile beneficiare di una sostanziale riduzione del costo medio di produzione, aumentando la scala produttiva. Approssimando le giornate di cura al 3° quartile si può fare un discorso analogo, ma con conclusioni diverse: per gli istituti di grandi dimensioni (a partire da 80 letti circa) le economie di scala sono completamente sfruttate, visto che il valore di E_{Scala} è inferiore a 1, indicando costi di produzione unitari maggiori, rispetto a processi produttivi su scala più contenuta.

Dall'analisi delle economie di scala, calcolate in corrispondenza di alcuni livelli produttivi di riferimento (1°, 2° e 3° quartile), emerge con sufficiente chiarezza l'andamento della curva dei costi unitari. In corrispondenza di una limitata capacità i costi unitari decrescono rapidamente, continuano a decrescere in modo meno marcato raggiungendo il minimo per una capacità produttiva di molto superiore a quella mediana, per poi iniziare a crescere in prossimità della soglia indicata dal 3° quartile della capacità produttiva.

Da questa constatazione discende anche la possibilità di utilizzare i coefficienti dei modelli stimati per individuare quel preciso livello di produzione (misurato dalla variabile Y in termini di giornate di residenza) che minimizza i costi, definendo quindi la dimensione ottimale per l'impresa. La capacità ottimale, la cui misura di dettaglio per i singoli modelli è indicata sempre nella Tabella 6.7, si aggira intorno ai 70 posti letto e quindi piuttosto lontano dal centro della distribuzione (50 letti), indicando come la casa per anziani rappresentativa (quella mediana) operi ad una scala di produzione troppo ridotta.

Un'analisi più approfondita, basata sul confronto dei costi medi nel punto di approssimazione e sulla cautela richiesta nel valutare situazioni lontane da tale punto, permette di tracciare una linea interpretativa comune. Sulla base delle stime dei modelli selezionati emerge in modo univoco che:

1. in corrispondenza di una capacità produttiva mediana (pari a circa 50 posti letto), nel settore delle case per anziani esistono economie di dimensione non completamente sfruttate;
2. la scala produttiva ottimale in grado di minimizzare i costi è individuata in corrispondenza di una capacità di posti letto maggiore di quella mediana (e verosimilmente intorno a 70-75 posti letto);
3. la funzione dei costi medi presenta una forte asimmetria rispetto alla dimensione ottimale tale per cui, scale di produzione inferiori al 1° quartile (ad esempio 30 posti letto) generano importanti eccessi di costo (da 3% a 10% a dipendenza del modello considerato);
4. il costo medio della casa per anziani rappresentativa (considerando tutte le variabili esplicative al loro valore mediano), pur non raggiungendo il minimo possibile, non si discosta molto (circa 1% in più) da tale livello;
5. il riferimento assoluto per il costo medio di un istituto rappresentativo è difficile da stabilire, poiché varia a seconda del modello, ma è compreso tra 140 e 165 franchi per giornata di residenza. Un riferimento può essere il valore medio dei 4 modelli ad effetti casuali, pari a 156.30 franchi per giornata, che corrisponde in pratica al livello indicato dal modello TREM (157.5), da noi indicato quale modello più attendibile.²⁹²

Il quadro che emerge è abbastanza semplice da interpretare in chiave di politica sanitaria, dal momento che tutte le indicazioni empiriche fornite dalle stime, sia pure con diversa intensità, sollevano il problema della scala di produzione. Bisogna però ricordare che la dimensione delle case per anziani spesso non può essere modificata e quindi occorre circostanziare in modo preciso la portata dei risultati di questo studio.

²⁹² I calcoli per i modelli ML-REM e POOL non sono riportati nella tabella 6.2, ma sono stati comunque eseguiti, mentre il modello FEM non è stato considerato in questa sede per le ragioni già discusse in precedenza.

In linea con l'approccio di lungo periodo scelto per la definizione della funzione di costo, la prima raccomandazione che scaturisce dai risultati empirici sulle economie di scala è quella di progettare le nuove case per anziani (posto che si decida di costruirle) considerando l'eccesso di costo provocato da una dimensione ridotta. Da un punto di vista operativo significa ragionare, per quei casi dove la densità della domanda di riferimento è bassa, in termini economici piuttosto che geografico-politici, promuovendo progetti consortili che permettano di raggiungere una capacità produttiva adeguata. In Svizzera l'assistenza agli anziani è storicamente legata al territorio o all'istituzione comunale e ancora oggi vi sono resistenze a discutere progetti che esulino dalla dimensione comunale, nonostante ci si renda conto dei costi da sopportare.

Se la corretta dimensione di una nuova casa per anziani non rappresenta (almeno dal punto di vista teorico) particolari problemi, ampliare un istituto esistente per sfruttare al meglio le economie di dimensione sembra una via percorribile, ma non sempre conveniente.²⁹³ Rimanendo su un piano generale si può comunque osservare come le economie di scala si possano sfruttare creando reti di collaborazione tra istituti sotto un'unica direzione aziendale. In una rete di case per anziani la condivisione di risorse, di competenze di attrezzature e di servizi potrebbero permettere una razionalizzazione ed una riduzione di costi, nonostante gli stabili (responsabili di una parte delle economie di scala) restino separati. È chiaro che il nostro modello non fornisce risultati precisi sull'importanza di questo aspetto, poiché non disponiamo di informazioni sull'eventuale organizzazione degli istituti sotto forma di reti di case per anziani.²⁹⁴

Una seconda riflessione suggerita dai risultati empirici è l'importanza attribuita all'accesso ai servizi di assistenza agli anziani rispetto al (potenziale) risparmio di costo associato alle economie di scala. In tutti i modelli si osserva che in un intorno abbastanza ampio della dimensione ottimale (soprattutto in direzione di capacità maggiori) l'aumento dei costi è piuttosto contenuto.

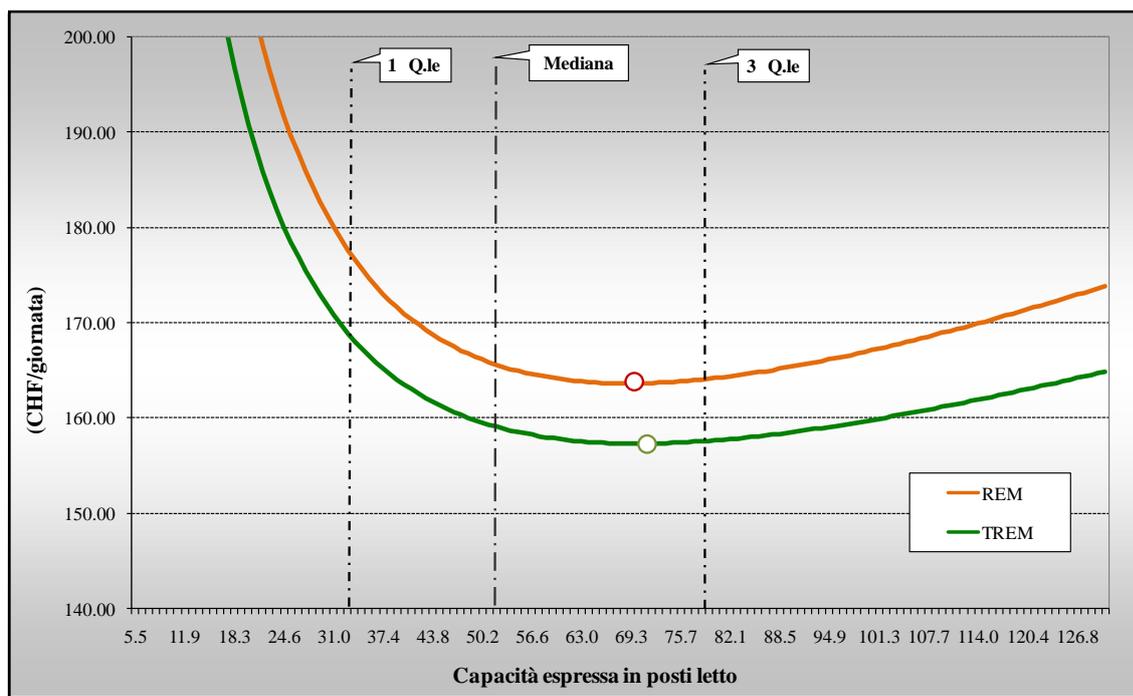
²⁹³ A questo proposito ogni singolo caso va discusso separatamente, valutando l'investimento secondo il criterio del valore attuale netto del progetto di ampliamento. I costi iniziali per espandere la scala produttiva devono essere ripagati dai risultati di esercizio futuri adeguatamente scontati.

²⁹⁴ In letteratura l'impatto delle reti di case per anziani sia sui costi, ma soprattutto sulla qualità delle cure è un aspetto che viene considerato con attenzione. Si veda ad esempio Mc Kay (1991), Chen and Shea (2004), Knox (2006) e Ozcan (2008).

La contrazione della capacità ottimale, situata nei pressi del 3° quartile (77 posti letto), fino alla dimensione mediana (51 posti letto) determina un aumento dei costi nell'ordine dell'1-2%, indicando una sostanziale proporzionalità nei costi, e quindi libertà di scegliere la scala all'interno di questo intervallo produttivo. A coronamento di quanto esposto, sembra sensato affermare che l'argomento della dimensione minima efficiente debba avere priorità sulle altre dimensioni valutative, solamente al di sotto di una determinata soglia, individuabile verosimilmente nei pressi del 1° quartile della distribuzione dell'output.

Il riepilogo grafico di quanto è emerso dall'analisi empirica sul tema delle economie di scala è rappresentato nella Figura 6.2, che mostra l'andamento della funzione dei costi medi calcolata per i modelli REM e TREM, approssimando tutte le variabili (tranne l'output) al loro valore mediano.

Figura 6.2 - Rappresentazione grafica della funzione di costo medio.



Dal punto di vista delle raccomandazioni di “policy” si visualizza bene la necessità di favorire il raggiungimento di una scala minima (sul cui livello esistono ampi margini per una discussione), a partire dalla quale le economie di scala, ancorché non completamente sfruttate, potrebbero passare in secondo piano in nome di obiettivi più importanti, come l'accesso territoriale o la qualità di vita all'interno dell'istituto.

In un intorno della mediana, i modelli stimati non individuano un costo medio per giornata identico, ragione per la quale possiamo delimitare i confini di tale valore

considerando attendibile l'intervallo compreso tra 155 e 165 franchi. Da notare come la differenza numerica nel costo medio, stimato dai diversi modelli, non influisca sulle considerazioni di policy, dal momento che la capacità ottimale e la scala minima efficiente sono ben allineate.

Un aspetto che merita una verifica ulteriore riguarda l'applicabilità delle riflessioni fatte in merito alle economie di dimensione alle diverse realtà cantonali. Il "microfederalismo" svizzero, che ha certamente contribuito alla diffusione su tutto il territorio nazionale di molti istituti di piccole dimensioni, non è necessariamente in contrasto con l'efficienza di scala dei singoli istituti. Tuttavia, le raccomandazioni di politica sanitaria sono spesso indirizzate al livello di governo superiore (i cantoni), responsabile di definire le linee guida che operativamente i comuni dovranno attuare. Il decisore pubblico cantonale, nell'utilizzo dei risultati per l'intero campione, dovrebbe interpretarli sulla scorta delle possibili specificità che caratterizzano il singolo cantone, specificità che non sempre vengono colte dall'aggregazione su scala nazionale.

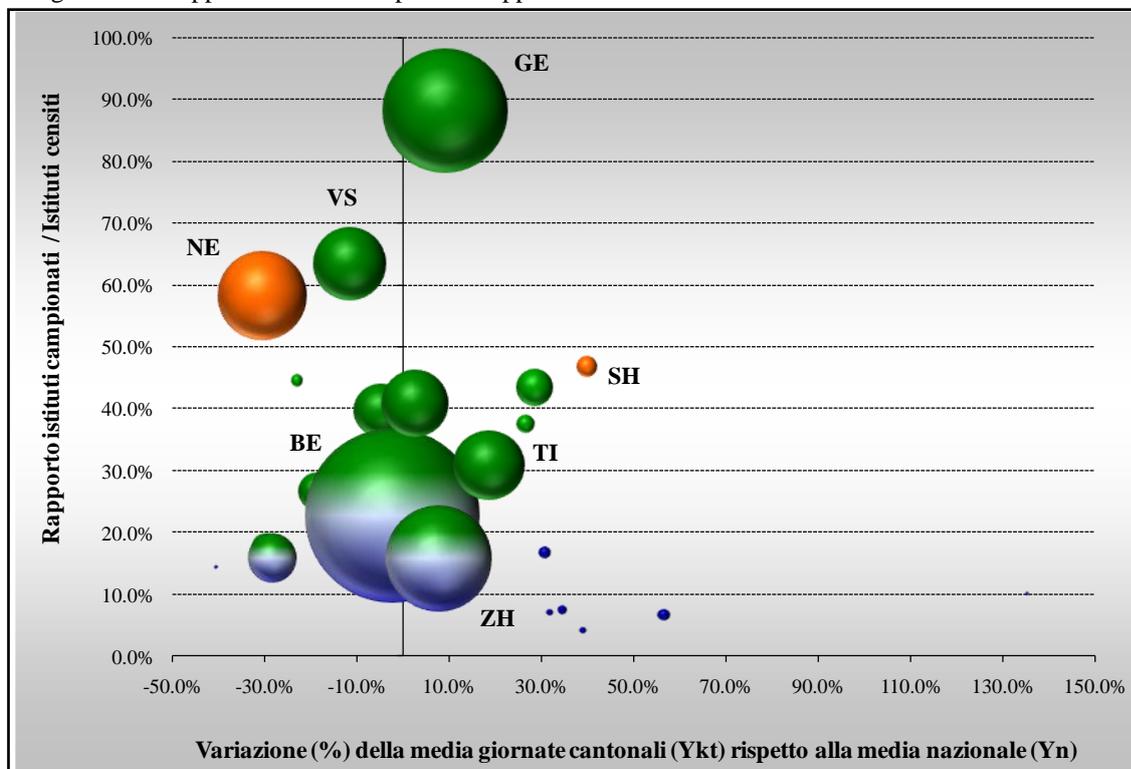
La decisione metodologica di aggregare tutti i dati cantonali per stimare una funzione unica, come discusso in precedenza, obbliga la scelta di un unico punto di approssimazione e può essere utile discutere la capacità di tale punto di rappresentare le economie di scala nei diversi cantoni svizzeri. La mediana (o la media) della capacità produttiva riferita ad un singolo cantone potrebbe differire sensibilmente dalla mediana (o dalla media) dell'output calcolato per tutta la Svizzera, rendendo le interpretazioni di politica sanitaria più incerte.

Per illustrare la situazione di partenza dei dati cantonali abbiamo costruito la Figura 6.3, che mostra contemporaneamente tre dimensioni: (1) la dispersione dell'output medio cantonale intorno alla media nazionale, (2) la dimensione del sottocampione cantonale rispetto alla popolazione cantonale, (3) il peso di ogni cantone misurato con le quote relative all'interno del campione. Osservando la dimensione delle sfere, si nota come solo pochi cantoni possano contare su un numero di osservazioni tale da consentire una verifica empirica (stima della funzione di costo) specifica.²⁹⁵ Peraltro,

²⁹⁵ Da notare che alcuni cantoni non sono rappresentati nella banca dati (AI e SG), mentre altri cantoni (AR, BS, GL, GR, JU, LU, NW, OW, SH, SO, SZ, UR, ZG) presentano un numero di istituti inferiore a 10 unità. Anche volendo, per questi cantoni sarebbe impossibile una stima separata, vista la mancanza di gradi di libertà. Di fatto, la stima con i sottocampioni cantonali è risultata possibile solamente per 6 cantoni (BE, GE, ZH, NE, VS, TI).

quasi la totalità dei cantoni è addensata a cavallo dell'asse verticale, che indica il valore dell'output mediano nazionale, suggerendo una buona rappresentatività del modello nazionale.

Figura 6.3 - Rappresentatività del punto di approssimazione dei dati a livello cantonale.



In altri termini, i valori rappresentativi dell'output mediano dei singoli cantoni non si discosta molto dal riferimento scelto per stimare la funzione di costo nazionale, rendendo le indicazioni di quest'ultima sufficientemente attendibili anche per i singoli cantoni. Gli unici cantoni che si discostano in misura rilevante (oltre il 30%) dalla media nazionale sono solamente due: Neuchâtel per difetto (-30,4%) e Sciaffusa per eccesso (+ 40,0%).

L'analisi di dettaglio delle singole realtà cantonali non rientra (come più volte sottolineato) negli obiettivi di questo lavoro. Inoltre, vista l'esigua consistenza del numero di osservazioni, per molti cantoni si è rivelato impossibile stimare separatamente le funzioni di costo e poco efficace l'introduzione nel modello di un vettore di *dummies* per i singoli cantoni.²⁹⁶

²⁹⁶ Inserendo nella funzione di costo un vettore di *dummies* cantonali non si ottengono differenze importanti nei parametri, mentre la maggioranza delle *dummies* cantonali non risulta significativa.

Alla luce delle considerazioni appena presentate la scelta di stimare una funzione di costo nazionale appare corroborata anche da queste semplici analisi descrittive che, tuttavia, hanno evidenziato nuovamente il problema dell'eterogeneità presente nel campione di dati.

La descrizione ed il commento dei risultati scaturiti dall'analisi dei punteggi di inefficienza e delle economie di scala nel settore delle case per anziani in Svizzera, che ci ha comunque permesso di inquadrare i termini generali del problema, deve cedere il passo ad alcune riflessioni più ampie e sfaccettate sugli insegnamenti maturati dall'evidenza empirica. Il prossimo ed ultimo capitolo è dedicato quindi ad un riepilogo del lavoro, mettendo in evidenza sia gli aspetti contenutistici sia quelli metodologici, cercando inoltre di inquadrare in chiave prospettica il valore aggiunto di questa tesi di dottorato.

“There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.”

(John Ronald Reuel Tolkien, Scrittore)

7 Conclusioni

Per chiudere in modo coerente ed organico il lavoro svolto, proponiamo un ultimo capitolo, il cui scopo è quello di valorizzare gli sforzi compiuti durante la ricerca, presentando sinteticamente i risultati (ma anche i limiti) dell'evidenza empirica ed offrendo un pizzico di prospettiva oltre gli stretti confini di questa tesi. Pertanto, dopo una sintesi degli argomenti trattati e dei principali risultati ottenuti, proporremo alcune riflessioni personali per situare lo studio rispetto alla letteratura del settore e discuteremo alcuni aspetti di politica sanitaria. La tesi si chiude con una riflessione, tesa a traguardare i possibili sviluppi futuri in questo ambito di ricerca.

7.1 Riepilogo

In Svizzera la spesa sanitaria ogni anno cresce in modo più o meno marcato ed ha raggiunto nel 2006 un livello pari a 52,7 miliardi di franchi. Questa cifra assume particolare rilievo se confrontata con un indicatore della produzione del paese (PIL), che nel medesimo anno ammontava a (circa) 487 miliardi di franchi. Il costo complessivo del sistema sanitario equivale quindi al 10,8% del PIL svizzero e, nel confronto internazionale, questo indicatore posiziona la Svizzera al secondo posto (a pari merito con la Francia), dietro agli Stati Uniti (15,0%) e davanti alla Germania (10,2%). L'importanza del settore sociosanitario, all'interno del quale la maggior parte del peso

grava sulle case per anziani rappresentava nel 2006 poco più di un quinto della spesa sanitaria complessiva. Inoltre, la complessa struttura organizzativa del settore, frutto in parte della presenza del federalismo, e la forte presenza dello Stato rendono il settore delle case per anziani un campo di studio molto interessante.

L'oggetto di studio principale di questo dottorato è costituito dall'analisi dell'efficienza produttiva delle case per anziani svizzere, realizzata attraverso la stima econometrica di una funzione di costo. In pratica l'obiettivo del presente lavoro consiste nel rappresentare il processo produttivo tipico del settore delle case per anziani, attraverso un modello di costo econometrico che sia in grado di fornire risposte, empiricamente fondate, ad una serie di domande di carattere economico. Gli aspetti econometrici applicati in questa tesi si inquadrano nel filone di letteratura denominato "Analisi delle Frontiere Stocastiche" (SFA dall'inglese *Stochastic Frontier Analysis*), che offre numerosi modelli econometrici contraddistinti da caratteristiche econometriche molto diverse tra loro. Applicando le diverse procedure di stima a disposizione si ottengono indicazioni non sempre allineate che impongono una verifica empirica della stabilità e plausibilità dei risultati.

Gli obiettivi dell'analisi empirica erano mirati a verificare principalmente i seguenti aspetti legati all'(in)efficienza produttiva del settore delle case per anziani:

- analizzare il fenomeno dell'inefficienza di costo, fornendo un'indicazione sul livello medio misurato in Svizzera e discutendo criticamente i risultati ottenuti con diversi modelli;
- verificare, in modo descrittivo, le differenze nei livelli medi di inefficienza tra gruppi di istituti con caratteristiche simili (forma istituzionale, appartenenza ad un determinato cantone, dimensione, etc.);
- misurare il livello delle economie di scala in questo settore, determinando il valore della dimensione ottimale per un istituto;
- proporre alcune raccomandazioni di politica sanitaria che emergono dalla discussione dei risultati empirici.

In questa tesi, gli aspetti di valutazione empirica dell'efficienza produttiva si intrecciano a quelli metodologici, legati alla scelta del modello econometrico adatto per effettuare una corretta misurazione sia dei parametri della funzione di costo, sia dei

punteggi di (in)efficienza di costo. Con particolare interesse all'aspetto econometrico, possiamo riassumere i principali obiettivi metodologici:

- verificare se l'impiego delle tecniche econometriche proposte da Greene (2005), che considerano maggiormente l'eterogeneità non osservata, permette di misurare l'inefficienza di costo in modo più preciso;
- confrontare i risultati delle stime dei parametri della funzione di costo ottenuti con i diversi metodi econometrici per valutarne la stabilità e quindi la capacità di fornire indicazioni attendibili sulla tecnologia produttiva del settore;

In estrema sintesi l'analisi empirica ha evidenziato che in Svizzera il settore delle case per anziani è caratterizzato da un certo livello di inefficienza. Il modello *True Random Effect Model* (TREM), preso come riferimento principale per la sua capacità di depurare (almeno parzialmente) l'effetto dell'eterogeneità non osservata dalla misura dell'inefficienza, indica che la metà degli istituti presenti nel campione eccede il livello di costo indicato dalla frontiera in una misura uguale o inferiore al 6,7% (mediana). Gli altri modelli, che rilevano un livello di inefficienza sostanzialmente più elevato, obbligano a non sottovalutare il problema dell'eccesso di costo in questo settore.

Per quanto riguarda la descrizione dell'inefficienza per diversi gruppi di case per anziani, nell'analisi abbiamo considerato tre possibili criteri: l'appartenenza ad un determinato cantone, la diversa forma istituzionale (pubblica, privata non profit, privata for profit) e la dimensione degli istituti (con meno di 35 posti letto, da 35 a 80 posti letto e con più di 80 posti letto). Per ciascuna dimensione valutata, il livello di inefficienza media espresso dai diversi sottogruppi è risultato piuttosto simile. In qualche (raro) caso è emersa una differenza, significativa dal punto di vista statistico, ma di poco conto ai fini pratici vista la ridotta distanza dei punteggi di inefficienza medi in termini assoluti. In conclusione, i tre criteri considerati non contribuiscono a spiegare i diversi livelli di inefficienza che caratterizzano le case per anziani.

L'aspetto della dimensione produttiva e dei costi ad essa associati, ha fornito risposte molto precise. L'analisi empirica ha evidenziato come, partendo dalla dimensione più piccola considerata nel campione (10 posti letto), l'aumento della scala di produzione genera un abbassamento dei costi medi per giornata di residenza. Le economie di scala presenti nel settore sono molto importanti in corrispondenza a bassi livelli di capacità produttiva e persistono, seppure con intensità minore, fino al

raggiungimento della dimensione ottimale. Questo valore, da un confronto tra i risultati dei 5 modelli stimati suggerisce che la dimensione ottimale di un istituto per anziani dovrebbe situarsi in prossimità di 70-75 posti letto.

Gli obiettivi metodologici della tesi si concentravano soprattutto sul problema di considerare l'impatto che l'eterogeneità non osservata può avere sulle stime dei parametri e sui punteggi di inefficienza. L'analisi empirica ha suggerito un possibile problema di "taratura" del modello nella misurazione del livello (medio) di inefficienza, come evidenziato dalle differenze tra i valori indicati dai diversi modelli. Parte del problema è legato alla presenza nei dati di caratteristiche specifiche delle imprese che non sono rilevate dal modello (problema dell'eterogeneità non osservata). Per questa ragione, il modello TREM, riuscendo a filtrare parzialmente le caratteristiche individuali degli istituti (eterogeneità non osservata), risulta maggiormente adatto a misurare l'inefficienza di costo nel settore delle case per anziani in Svizzera. Questo modello rappresenta quindi un valido riferimento anche se i valori dell'inefficienza potrebbero essere sottostimati, visto che il modello classifica ogni eventuale forma di inefficienza costante nel tempo come eterogeneità.

Rispetto al problema della stabilità delle stime dei coefficienti del modello di costo, l'analisi empirica ha mostrato una generale convergenza dei risultati, nonostante qualche differenza significativa su alcuni parametri. Considerati i vantaggi e gli svantaggi econometrici dei diversi modelli e, in ragione della brevità temporale del panel (5 anni), risulta quindi preferibile utilizzare un ventaglio di modelli, per poter confrontare in seguito le eventuali differenze nei risultati. Nonostante la presenza di alcune differenze, parzialmente giustificabili dal punto di vista metodologico, i risultati delle stime sono comunque stabili ed indicano, con relativa precisione, quali siano i parametri della tecnologia produttiva nel settore dell'assistenza agli anziani in Svizzera.

Una visione complessiva dei risultati ottenuti indica che la tesi, rispetto agli obiettivi metodologici ed empirici, ha saputo fornire validi elementi di comprensione del processo produttivo tipico del settore. Per ottenere questi risultati abbiamo basato l'analisi empirica su alcune precise scelte metodologiche, che abbiamo ampiamente descritto, mettendone in luce i principali pregi. A questo punto sembra opportuno proporre alcune riflessioni più generali sui risultati della tesi e menzionare anche alcuni limiti e problemi connessi con l'impiego di queste tecniche, così come sono stati evidenziati da una certa parte della letteratura.

7.2 Riflessioni conclusive sui risultati dell'analisi empirica

A conclusione di tutte le analisi empiriche svolte desideriamo rielaborare i risultati ottenuti, partendo da un punto di osservazione esterno alla ricerca per trarre indicazioni più generali e con accenti diversi rispetto a quelli già discussi. In primo luogo, è nostra intenzione riportare i risultati ottenuti con la letteratura esistente, per contestualizzare meglio le indicazioni di *policy* scaturite dall'evidenza empirica, in funzione delle problematiche attuali (e future) che caratterizzano il settore. Seguono alcune considerazioni di carattere metodologico che anticipano la discussione sui limiti della tesi, presentata nel paragrafo successivo.

7.2.1 Confronto con la letteratura e aspetti di *policy*

Nella rassegna della letteratura presentata nel capitolo 5, abbiamo discusso i risultati di numerosi lavori empirici sulla stima di funzioni di costo nel settore delle case per anziani, che desideriamo porre in relazione ai risultati di questa tesi.

La misurazione del livello medio di inefficienza, ottenuto con il modello di riferimento TREM, è risultato pari a 8,1% ed è certamente più elevato rispetto al minimo registrato nella rassegna (4%), rilevato da Farsi et al. (2005) utilizzando lo stesso modello econometrico. Occorre sottolineare, però, come in quest'ultimo studio si analizzassero dati di un cantone specifico (Ticino), cosa che contribuisce a ridurre l'eterogeneità presente nei dati rispetto ad un campione di case per anziani nazionale. Per contro il campo di variazione del punteggio medio di inefficienza misurato dagli altri studi è molto ampio, raggiungendo il massimo (40%) nello studio di Rebba e Rizzi (2002). Pur consapevoli delle difficoltà nel confrontare stime di inefficienza ottenute con tecniche econometriche e modelli diversi, riteniamo che i valori di inefficienza misurati in questo studio (in particolare quelli del modello di riferimento TREM) siano in linea con le indicazioni della letteratura. Inoltre, il nostro studio mostra come, nel punto di approssimazione dei dati (51,4 posti letto), le economie di scala non siano completamente sfruttate. Il dato è in sintonia con la maggior parte dei lavori empirici, che sostengono esplicitamente la presenza di economie di scala nel settore. Va comunque rilevato come in alcuni studi siano stati riscontrati valori delle economie di scala superiori a quelli ottenuti nel presente lavoro.

Le implicazioni di politica sanitaria che discendono dall'analisi empirica meritano una contestualizzazione settoriale *ex-ante* (prima cioè di conoscere i valori delle stime)

ed *ex-post* (discutendo i risultati e proponendo, se del caso, alcune soluzioni concrete per migliorare la situazione nel settore). Il quadro di riferimento iniziale è stato presentato a più riprese, sottolineando come:

- la presente analisi si concentri esclusivamente sulla dimensione economica;
- i fattori storico-culturali abbiano un grande ruolo nel determinare gli attuali assetti del settore;
- il settore delle case per anziani sia caratterizzato da una forte presenza dello Stato, legittimata in parte da ragioni di efficienza, ma soprattutto per questioni di equità;
- la stretta interconnessione tra i diversi livelli di regolamentazione abbia certamente un impatto sui costi, ma soprattutto sul modello organizzativo che prevale in un dato cantone;
- la grande autonomia (garantita dal federalismo) nelle scelte dei cantoni potrebbe aver contribuito ad esasperare l'eterogeneità, comunque presente nel settore dell'assistenza agli anziani.

Con queste premesse e alla luce dei risultati dell'analisi empirica, ampiamente commentati nei paragrafi precedenti, ci sentiamo di fare alcune considerazioni generali di carattere normativo.

Il settore delle case per anziani è confrontato con un eccesso di domanda cronico che, osservando l'attuale ritmo di crescita della popolazione anziana rispetto all'incremento di offerta di nuovi istituti, è destinato a peggiorare. In considerazione di questa problematica, è lecito attendersi che l'orientamento della società civile e delle autorità pubbliche cantonali e comunali si indirizzi maggiormente (e in parte lo sta già facendo) verso il potenziamento dell'offerta dei servizi agli anziani. Preso atto che le forme di intervento che la società mette a disposizione per soddisfare i bisogni della popolazione anziana non si esauriscono con le strutture residenziali, la costruzione di nuove case per anziani è sicuramente una misura concreta ed efficace per migliorare le condizioni di accesso agli anziani bisognosi, riducendo i tempi di attesa.

In questo studio abbiamo cercato di insistere non tanto sulla misura puntuale delle economie di scala o della dimensione ottimale dal punto di vista dei costi, quanto piuttosto, sull'importanza di non commettere errori (in sede di progettazione) i cui

effetti negativi si ripercuoteranno per molti anni a venire. Il messaggio di politica sanitaria (o, in un'ottica più allargata, di responsabilità sociale) che ci sentiamo di trasmettere è quello della soglia minima. Le economie di scala nel settore sono forti in corrispondenza delle piccole dimensioni, determinando grandi risparmi di costo se correttamente sfruttate.

In Svizzera, basandoci sui risultati di questo studio, ma anche in vista del crescente aumento della speranza di vita e quindi della domanda, le nuove case per anziani dovrebbero sfruttare appieno le economie di scala, dotandosi di 70-75 posti letto. Il motivo di questa raccomandazione è ovvio per un'economista, ma vale la pena richiamare il concetto di costo opportunità che sta dietro a questo ragionamento. Case per anziani troppo piccole, che quindi producono ad un costo medio troppo elevato, precludono la possibilità di finanziare un maggior numero di posti letto "sprecati".

Questo discorso è certamente vero nelle zone con una sufficiente densità abitativa che permette una razionalizzazione dell'offerta di servizi residenziali alla popolazione anziana, mentre occorre maggiore cautela per le zone più isolate e con elevata dispersione abitativa. In questi frangenti, infatti, il conflitto tra efficienza (di scala) ed equità (di accesso) si accentua fino a spostare (soprattutto per ragioni politiche) l'accento sull'opportunità di avere un beneficio locale (la casa per anziani piccola) in cambio di un costo collettivo (l'eccesso di costo provocato dalla bassa capacità produttiva). Questa riflessione sui possibili svantaggi connessi con una produzione tesa a considerare solamente gli aspetti dell'efficienza deve rendere attenti sul fatto che si devono valutare tutte le dimensioni rilevanti di un problema per risolverlo nel miglior modo possibile dal punto di vista della società.

L'aspetto probabilmente più delicato risiede nell'impossibilità di conciliare una dimensione efficiente, dal punto di vista della scala, con l'equità territoriale. In effetti l'utilità di un centro residenziale per anziani è legata al bacino di utenza che l'istituto è in grado di soddisfare e tende a diminuire quanto più i confini di tale bacino si ampliano. Inoltre, per gli ospiti, vivere in una casa per anziani troppo grande può essere fonte di disagio e anonimità, e questo ostacola l'importante obiettivo dell'integrazione sociale nella vita comune in istituto. Va comunque precisato che la dimensione ottima stimata, 70-75 letti, non dovrebbe comportare disagi ed anonimità molto diversi rispetto a quelli vissuti in case per anziani di piccole dimensioni.

Un aspetto che discende in parte dalla considerazione fatta in precedenza riguarda la possibilità che si verifichi un trade-off anche tra efficienza di scala e qualità nelle prestazioni offerte. Il motivo di questa affermazione risiede nella constatazione che una casa per anziani ha lo scopo di accogliere delle persone non più in grado di vivere da sole e bisognose di assistenza medica, ma soprattutto umana. Per questa semplice ragione, quanto più una casa per anziani si allontana da uno schema familiare, organizzandosi in modo asettico e impersonale, tanto più difficile diventa perseguire standard di qualità sia nei processi sia nei livelli di soddisfazione dell'ospite. Una caratteristica che influenza negativamente la soddisfazione degli utenti è certamente un'eccessiva dimensione dell'istituto, che si traduce molto spesso in ambienti impersonali con difficoltà di comunicazione tra gli ospiti.

Una possibile via per conciliare i trade-off enunciati, potrebbe essere la produzione di un servizio organizzato sotto forma di reti sanitarie, costituite da istituti di piccole dimensioni (in grado di raggiungere gli obiettivi di qualità e di equità territoriale). Parte delle economie di scala sono legate ad attività (mensa, lavanderia, amministrazione, dotazioni tecniche, etc.) che si possono esternalizzare o comunque razionalizzare a livello di rete (e non di singolo istituto), ottenendo importanti risparmi di costo.

La crescente difficoltà nel finanziare i modelli di welfare attualmente presenti in Svizzera ha determinato, a molti livelli, un ripensamento di alcune scelte di fondo ed una maggior attenzione all'efficienza nell'uso delle risorse. Questo atteggiamento lo si può cogliere, ad esempio, nella nuova impostazione della perequazione finanziaria tra la Confederazione ed i cantoni (entrata in vigore il 1° gennaio 2008, dopo 15 anni di preparazione), nel fenomeno delle aggregazioni comunali, nella diffusione in molte amministrazioni comunali di modelli organizzativi ispirati al "*New Public Management*". Non vi è ragione di credere che il settore dell'assistenza agli anziani, direttamente toccato dalle trasformazioni menzionate, resti impermeabile a questi sforzi tesi a promuovere l'efficienza produttiva.

L'eccesso di costo medio di una casa per anziani svizzera, misurato con il modello ritenuto più accurato (TREM), ma anche il più conservativo, non è trascurabile. Il giudizio sulle performance degli istituti elvetici non è comunque negativo, dal momento che oltre il 75% delle case per anziani presenti nel campione non eccede la soglia del 10% rispetto al costo minimo calcolato dalla frontiera. Ad ogni modo, secondo le

indicazioni del modello TREM, in termini monetari l'eccesso di costo medio si traduce, per un istituto che opera alla scala ottimale (71 posti letto), in una maggiore spesa di circa 250'000 franchi per anno (prezzi del 2002), pari ad un eccesso di costo dell'8%.

In questo senso, alcune linee di tendenza emergenti nel settore (adozione di sistemi *full costing*, utilizzo della contabilità analitica, allineamento dei sistemi di rilevamento del fabbisogno di cure degli ospiti, etc.), suggeriscono una maggior propensione da parte degli istituti ad accettare una misurazione comparativa delle prestazioni (*benchmarking*). In altre parole, sembra che il concetto di efficienza come cultura, in grado quindi di stimolare attivamente la gestione ottimale delle risorse, stia prendendo il posto dell'inefficienza come misura, volta in qualche modo a sanzionare comportamenti o decisioni errate.

In sintesi, possiamo affermare che i risultati dell'analisi empirica, condotta con la stima di una frontiera stocastica di costo per il settore delle case per anziani, forniscono almeno due indicazioni difficili da ignorare, anche dopo aver considerato tutti i limiti e le attenuanti del caso. La prima considerazione riguarda la necessità di attribuire il giusto peso alla dimensione di un istituto che, secondo la stima della dimensione ottima, dovrebbe attestarsi intorno a 70-75 posti letto. La seconda osservazione è quella di prestare una concreta attenzione al problema dello spreco di risorse provocato dall'inefficienza di costo in questo settore.

7.2.2. Impiego delle frontiere stocastiche nella regolamentazione

L'ultimo aspetto che desideriamo trattare riguarda le problematiche metodologiche da considerare, per rendere l'analisi dell'efficienza uno strumento decisionale concreto da utilizzare nella regolamentazione di un settore complesso come quello delle case per anziani. L'impiego delle frontiere stocastiche in ambito sanitario (ospedali e case per anziani) risale agli inizi degli anni '90 ed ha immediatamente animato il dibattito scientifico originando correnti di sostenitori e oppositori circa la reale capacità di questa tecnica di produrre risultati affidabili per la politica sanitaria e la regolamentazione. Tra gli oppositori spicca, in particolare, il pensiero autorevole di Newhouse (1994) che sottolinea la difficoltà di misurare correttamente l'inefficienza di costo in presenza di un processo produttivo complesso e multidimensionale come quello dei servizi sanitari.

Secondo questa visione ogni tentativo di valutare con precisione e di confrontare l'eccesso di costo delle singole strutture è vanificato dall'incompleta specificazione del modello di costo e dalla forte presenza di eterogeneità non osservata. Skinner (1994) constata, invece, l'arbitrarietà ed i rischi di specificazione connessi con l'imposizione di un termine di errore asimmetrico a due componenti (*Normal* e *Half-Normal*), mentre Dor (1994) evidenzia la superiorità dei metodi panel tradizionali (LSDV e GLS) rispetto alle stime di massima verosimiglianza, che richiedono ipotesi restrittive circa la distribuzione delle componenti del termine di errore.

La distanza tra oppositori e sostenitori, pur restando ben visibile, è andata progressivamente riducendosi e, all'inizio degli anni 2000, il dibattito sul tema specifico dell'uso a scopo di *policy* delle stime di inefficienza di costo calcolate con una frontiera stocastica, pur rimanendo presente, ha assunto toni più moderati. Nei lavori di Folland and Hofler (2001), Street (2003) e Smith and Street (2005) viene riconosciuta l'importanza e l'utilità di questi metodi per fornire indicazioni generali (livello medio di inefficienza) o per grandi gruppi (confronti tra livelli medi di efficienza per tipologia). Questi autori sottolineano, tuttavia l'estrema sensibilità dei punteggi di inefficienza delle singole imprese alle scelte di specificazione del modello, ai dati, alle tecniche econometriche impiegate, suggerendo grande cautela nel tradurre meccanicamente i risultati empirici in provvedimenti di regolamentazione.

Tenendo a mente le considerazioni appena presentate, possiamo comunque affermare che le stime econometriche hanno permesso di individuare la frontiera di costo, che rappresenta quindi il termine di paragone (*Best Practice*) per le prestazioni delle case per anziani. Come ampiamente discusso, si pone un problema empirico-metodologico, che consiste nello stabilire quale modello permetta di misurare correttamente l'inefficienza. Non si può dare una risposta definitiva, anche se i risultati delle analisi (e della letteratura in materia) forniscono indirettamente due criteri operativi piuttosto chiari. Il primo riguarda la valutazione dell'eterogeneità che si nasconde nei dati, mentre il secondo consiste nel riconoscere i limiti metodologici in sede di interpretazione dei risultati. L'analisi ha evidenziato che i modelli in grado di rilevare con maggior precisione l'eterogeneità non osservata indicano livello di inefficienza minori. Tuttavia, il problema non riguarda solo la sensibilità della scala di inefficienza al modello, quanto la possibile inconsistenza delle graduatorie di inefficienza.

A questo proposito bisogna riconoscere il contributo fondamentale che la qualità dei dati apporta nel processo di stima, risolvendo alla fonte il problema dell'affidabilità dei risultati. Nel caso specifico del settore dell'assistenza agli anziani è evidente come le informazioni a disposizione dei Dipartimenti di Sanità cantonali siano molto più dettagliate e meglio aderenti alla cornice istituzionale/regolatoria, rispetto ai dati raccolti dall'ufficio federale di statistica nell'ambito dell'indagine annuale su tutti gli istituti non ospedalieri svizzeri.

Riguardo al problema dell'eterogeneità questo non significa che il dato particolare valga più del dato generale, poiché nel caso di 26 banche dati cantonali si sposterebbe la discussione ad un livello superiore, senza trovare una soluzione. Le eventuali (e probabili) differenze nei dati cantonali palesano l'eterogeneità costringendo ad adeguare i modelli di costo, magari più precisi, ai dati disponibili. Una banca dati nazionale risolve il problema della confrontabilità tra i risultati cantonali, pagando un prezzo in termini di precisione.

Una volta deciso che la prospettiva della ricerca è quella nazionale ed accertata (ed accettata) la presenza di eterogeneità nei dati, la scelta di un unico metodo di stima diventa un esercizio da equilibrista che sconsigliamo vivamente. Riteniamo preferibile lavorare in parallelo, analizzando gli stessi dati con approcci econometrici diversi, enfatizzando la complementarità delle tecniche e cercando di cogliere un risultato di sintesi, tanto dalle analogie quanto dalle differenze riscontrate nelle stime dei diversi metodi. Per contro, se il contesto di ricerca è basato su dati più omogenei (ad esempio di un cantone specifico) e più dettagliati, le stime di una funzione frontiera di costo potrebbero trasformarsi in un valido strumento di supporto decisionale per il regolatore, così come accade nell'ambito della *yardstick regulation*.

I cantoni che si sono orientati verso forme di regolamentazione che prevedono una modalità di finanziamento prospettico (ad esempio i contratti di prestazione), potrebbero utilizzare i risultati delle analisi empiriche per ridurre le asimmetrie informative nei confronti degli istituti regolamentati. I costi stimati dalla frontiera per ogni singolo istituto si possono paragonare ai costi documentati dall'istituto, fornendo una preziosa indicazione per stabilire (e/o negoziare) il livello del budget da assegnare.

Questo tipo di esperienza è stata fatta dal canton Ticino che, a partire dal 2006, ha introdotto i contratti di prestazione per gli istituti del settore sociosanitario sottoposti a

regolamentazione. Durante la fase che ha preceduto l'attuazione dei contratti, il Dipartimento di Sanità e Socialità (DSS) ha voluto sperimentare lo strumento delle frontiere stocastiche, commissionando uno studio econometrico sulle case per anziani ticinesi.²⁹⁷ I budget previsti dalla frontiera stocastica per gli istituti analizzati dallo studio si sono rivelati molto simili a quelli calcolati dal DSS, utilizzando costi standard molto dettagliati.

Alla luce di queste considerazioni appare evidente come la scelta della metodologia non possa mai essere disgiunta dal contesto empirico nel quale si opera, limitandosi a valutare in astratto pregi e difetti dei diversi metodi disponibili. Questo non significa, però, che ogni scelta sia legittima o inattaccabile ed è per questo che nel prossimo paragrafo desideriamo presentare anche alcuni limiti di questo lavoro.

7.3 Possibili sviluppi futuri della ricerca

In questo paragrafo conclusivo desideriamo presentare una breve riflessione, il cui scopo è quello di posizionare i risultati di questa ricerca in una prospettiva più ampia rispetto ai confini editoriali della tesi.

In questo studio abbiamo puntato con decisione verso le metodologie parametriche e, in particolare, abbiamo considerato le frontiere stocastiche di costo. Oltre alla scelta di fondo tra i due approcci (parametrico vs. non parametrico), lungo il percorso di ricerca abbiamo intrapreso alcune strade e scelto alcune opzioni rispetto ad altre. Nello specifico, desideriamo presentare alcune possibili alternative di ricerca che non hanno trovato spazio in questa tesi, ma che potrebbero costituire il punto di partenza per ricerche future.

Nonostante siano molte le questioni da affrontare, riteniamo che le linee di sviluppo per approfondire l'analisi sul tema dell'(in)efficienza delle case per anziani siano ben definite: (1) seguire approcci metodologici diversi (2) raffinare ulteriormente il modello di costo, (3) impiegare nuovi e/o diversi modelli econometrici. Sul primo punto è bene sottolineare come la tesi abbia scelto in modo preciso di utilizzare i metodi parametrici, considerandoli più adatti per il tipo di analisi empirica da svolgere. Non

²⁹⁷ Lo studio (non pubblicato) di Farsi, Filippini e Lunati (2003) è disponibile presso l'Istituto Mecop dell'Università della Svizzera Italiana come rapporto interno di ricerca.

significa, tuttavia, che un'indagine sull'(in)efficienza produttiva condotta con le moderne tecniche non parametriche (in particolare alcuni modelli DEA che considerano anche gli aspetti stocastici) non potesse rivelare aspetti interessanti e sconosciuti sul settore delle case per anziani in Svizzera. La letteratura mostra, infatti, una decisa convergenza sui metodi ed una maggior valorizzazione delle complementarità che le diverse tecniche sono in grado di offrire, come testimonia il recente volume curato da Fried et al. (2008).

Rispetto al problema di specificazione del modello, è bene sottolineare come la qualità dei dati, benché la fonte sia assolutamente affidabile e le rilevazioni ragionevolmente precise, presenti alcuni limiti. La banca dati utilizzata nella tesi include le rilevazioni degli istituti non ospedalieri da quando (anno 2000) l'ufficio Federale di Statistica ha istituito questo tipo di pubblicazione. Da allora, e fino al 2004, i criteri di classificazione impiegati e le modalità di rilevamento sono evoluti per dare origine alla situazione di stabilità procedurale dei giorni nostri. Ciò significa che verosimilmente, i dati riferiti agli anni successivi al 2002 (ultimo anno considerato nello studio) presenteranno, oltre un minor numero di informazioni mancanti, un livello qualitativamente più elevato. Inoltre, il panel si è arricchito di altri quattro periodi (2003-2006), permettendo analisi di migliore qualità da un punto di vista econometrico.

Avere a disposizione molti dati secondari di ottima qualità (come lo sono quelli dell'UFS), non sempre può risolvere determinati problemi, in particolare quelli legati alla specificazione del modello di costo. Nel caso concreto, un punto debole della banca dati è certamente costituito dai pochi dettagli del conto economico e dalla mancanza di un rilevamento sistematico della qualità delle cure degli istituti. I problemi di disponibilità/qualità delle informazioni statistiche portano a discutere altri aspetti legati alla specificazione del modello, che si potrebbero considerare in futuro qualora i dati lo permetteranno empiricamente. Ci riferiamo alla possibilità di stimare una funzione di costo variabile per un confronto dei risultati con la specificazione di lungo periodo (costi totali) e/o alla scelta di definire il processo produttivo di una casa per anziani come multiprodotto. Un ultimo aspetto da enfatizzare riguarda il possibile impatto di un miglioramento della qualità (nel senso appena discusso) dei dati sull'affidabilità delle stime, in particolare per gli scopi di politica sanitaria e di regolamentazione.

A questo proposito vale la pena ricordare il lavoro di Farsi et al. (2003), svolto per il Dipartimento della Sanità e Socialità del Canton Ticino, durante il periodo di

sperimentazione che ha preceduto l'impiego dei contratti di prestazione per il settore sociosanitario ticinese. Lo studio è stato svolto con la medesima metodologia della presente tesi (stima di una frontiera di costo stocastica per le case per anziani), ma utilizzando una banca dati con minori problemi di eterogeneità. I risultati e le considerazioni di politica sanitaria (previsione di un budget globale) si sono rivelati molto affidabili. La banale conclusione che ci sentiamo di proporre su questo delicato argomento è che la rilevanza in termini di *policy* di uno studio non dovrebbe dipendere solo dall'interesse in tal senso del ricercatore, ma anche dalla disponibilità di dati e di metodologie adatte allo scopo. Nel caso della politica sanitaria e della regolamentazione del settore delle case per anziani ci sembra che il problema risieda soprattutto nella debolezza delle banche dati, piuttosto che nei limiti delle frontiere di costo stocastiche.

Come discusso nel capitolo 4, nell'ambito della metodologia delle frontiere stocastiche sono disponibili numerosi modelli di frontiera, alcuni dei quali scelti ed applicati in questa tesi. Appare evidente che in futuro si potrebbe estendere l'analisi ad altri modelli econometrici già in uso in letteratura (es. specificazione "Mundlak", *True Fixed Effect*) per un confronto più esteso sui risultati delle stime dei coefficienti e dei livelli inefficienza. Da un punto di vista strettamente metodologico, la tesi non propone nuove tecniche o modelli, ma permette di riflettere in modo critico sulle difficoltà nel declinare l'applicazione di questi modelli ad un contesto empirico complesso, come quello delle case per anziani svizzere. Pertanto, riteniamo che, in aggiunta ai contenuti dell'analisi, la ricerca contribuisca a sviluppare quel *corpus* di evidenze empiriche che spesso sono di ispirazione per nuovi lavori.

In conclusione, riteniamo che la presente tesi di dottorato abbia il merito di non nascondere e, al contrario, discutere i problemi esistenti (economici, metodologici e di dati), fornendo in compenso alcune risposte empiriche precise circa le caratteristiche produttive del settore. Dal punto di vista di chi scrive, vi è poi la consapevolezza di aver completato un percorso di crescita che abbraccia sia la dimensione professionale sia quella personale e che, inevitabilmente, non può trasparire dal presente documento. Resta la speranza che i contenuti di questa ricerca rappresentino, anche dal punto di vista di chi legge, un tangibile valore aggiunto.

BIBLIOGRAFIA

- Afriat, S.N., (1972), “*Efficiency Estimation of Production Function*”, *International Economic Review*, 8, 568–598.
- Ahn, S.C., Y.H. Lee and P. Schmidt, (1994), “GMM Estimation of a Panel Data Regression Model with Time-Varying Individual Effects”, Working Paper, Department of Economics, Michigan State University, East Lansing, MI.
- Aigner, D.J., C.A.K. Lovell and P. Schmidt, (1977), “*Formulation and Estimations of Stochastic Frontier Production Function Models*”, *Journal of Econometrics*, 6, 21–37.
- Alchian, A.A., (1965), “*Some economics of property rights*”, *Il Politico*, 30, 816–829.
- Alchian, A.A. and H. Demsetz, (1972), “*Production, Information Costs, and Economic Organization*”, *The American Economic Review*, Vol. 62, 777–795.
- Aletras, V.H., (1999), “*A comparison of hospital scale effects in short-run and long-run cost functions*”, 8, 521–530.
- Anderson, R.I., B.K. Hobbs, H.S. Weeks and J.R. Webb, (2005), “*Quality of Care and Nursing Home Cost-Efficiency Research*”, *Journal of Real Estate Literature*, 13, 325–335.
- Baltagi, B.H., (2008), *Econometric Analysis of Panel Data*, 4th ed., Chichester (UK): Wiley.
- Baltagi, B.H. and J.M. Griffin, (1984), “*Short and Long Run Effects in Pooled Models*”, *International Economic Review*, 25, 631–645.
- Banca d’Italia (a cura di), (2003), *L’efficienza nei servizi pubblici*.
- Battese, G.E. and T.J. Coelli (1992), “*Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India*”, *Journal of Productivity Analysis*, 3, 153–169.

- Battese, G.E. and T.J. Coelli (1995), “*A model for technical inefficiency effects in stochastic frontier production function for panel data*”, *Empirical Economics*, 20, 325–332.
- Battese, G.E. and T.J. Coelli and D.S. Prasada Rao, (1998), *An Introduction to Efficiency and Productivity Analysis*, Boston: Kluwer Academic Publishers.
- Baumol, W.J., J.C. Panzar, and R.D. Willig, (1982), *Contestable Markets and The Theory of Industry Structure*, New York: Harcourt Brace Jovanovich, Inc.
- Bauer, P.W., A.N. Berger, G.D. Ferrier and D.B. Humphrey, (1998), “*Consistency Conditions for Regulatory Analysis of Financial Institutions: A comparison of Frontier Efficiency Methods*”, *Journal of Economics and Business*, 50, 857–114.
- Becker, G.S., (1983), “*A Theory of Competition Among Pressure Groups for Political Influence*”, *The Quarterly Journal of Economics*, 98, 371–400.
- Berger, A.N., D. Hancock and D.B. Humphrey, (1993), “*Bank efficiency derived from the profit function*”, *Journal of Banking & Finance*, 17, 317–347.
- Bishop, C.E., (1980), “*Nursing Home Cost Studies and Reimbursement Issues*”, *Health Care Financing Review*, 1, 47–64.
- Bjørnerud, S., C. De la Maisonneuve and J. Oliveira Martins, (2006), “*Projecting OECD Health and Long-term Care Expenditures: what are the main drivers?*”, *Economics Department WP n° 477*, OECD.
- Blank, J.L. and E. Eggink, (2001), “*A Quality-Adjusted Cost Function in a Regulated Industry: The Case of Dutch Nursing Homes*”, *Health Care Management Science*, 4, 201–211.
- Borzaga, C., G. Fiorentini e A. Matacena (a cura di), (1998), *Non-profit e sistemi di welfare*, Roma: Carocci.
- Bös, D., (1986), *Public Enterprise Economics: Theory and Applications*, Amsterdam: North-Holland.
- Box, G.E. and D.R. Cox, (1964), “*An analysis of transformations*”, *Journal of the Royal Statistical Society*, 26, 211–246.
- Breyer, F., (1987), “*The Specification of a Hospital Cost Function*”, *The Review of Economics and Statistics* 79, 443–453.
- Bua, R.N., (1997) *The Inside Guide to America's Nursing Homes: Rankings & Ratings for Every Nursing Home in the U. S., 1998-1999*, New York: Warner Books.
- Buchanan, J.M. and G. Tullock, (1962), *The Calculus of Consent*, Ann Arbor: University of Michigan Press.

- Cameron, A.C. and P.K. Trivedi, (2005), *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.
- Cardani, A.M., (1988), *Lezioni di teoria della produzione e domanda*, Milano: Unicopli.
- Carrington, R., T.J. Coelli and D.S. Prasada Rao, (2005), “*The Performance of Australian Universities: Conceptual Issues and Preliminary Results*”, *Economic Papers*; 24, 145–163.
- Caudill, S.B., J.M. Ford and D.M. Gropper, (1995), “*Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroskedasticity*”, *Economic Letters*, 41, 17–20.
- Caves D.W., L.R. Christensen and J.A. Swanson, (1980), “*Productivity in U.S. Railroads, 1951-1974*”, *Bell Journal of Economics*, 11, 166–185.
- Caves, D.W., L.R. Christensen and J.A. Tretheway, (1984), “*Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ*”, *Rand Journal of Economics* 15, 471–489.
- Caves, D.W., L.R. Christensen, M.W. Tretheway and R.J. Windle, (1985), “*Network effects and the measurement of economies of scale and density for U.S. railroads*”, in: Doughtey, A.F. (ed.), *Analytical Studies in Transport Economics*, Cambridge: Cambridge University Press
- Cawley, J., D.C. Grabowski and R.A. Hirth, 2006, “*Factor substitution in nursing homes*”, *Journal of Health Economics*, 25, 234–247.
- Chamberlain, G., (1982), “*Multivariate Regression Models for Panel Data*”, *Journal of Econometrics*, 18, 5–46.
- Chambers, R.G., (1988), *Applied Production Analysis*, Cambridge: Cambridge University Press.
- Chen, L-W and D.G. Shea, (2004), “*The Economies of Scale for Nursing Home Care*”, *Medical Care Research and Review*, 61, 38–63.
- Chew, L.C., K. Hsein and Y. Jonsay, (2005), “*Airline code-share alliances and costs: imposing concavity on translog cost function estimation*”, *Review of Industrial Organization* 26, 461–487.
- Christensen, E.W, (2004), “*Scale and scope economies in nursing homes: A quantile regression approach*”, *Health Economics*, 13, 363–377.
- Christensen, L.R., W.D. Jorgenson and L.J. Lau, (1971), “*Conjugate Duality and the Transcendental Logarithmic Function*”, *Econometrica*, 39, 255–256.
- Christensen, L.R., W.D. Jorgenson and L.J. Lau, (1973), “*Transcendental Logarithmic Production Frontiers*”, *The Review of Economics and Statistics* 55, 28–45.

- Coase, R.H., (1937), “*The Nature of the Firm*”, *Economica*, 4, 386–405.
- Coase, R.H., (1960), “*The Problem of Social Cost*”, *Journal of Law and Economics*, 3, 1–44.
- Cobb, C.W. and P.C. Douglas, (1928), “*A Theory of Production*”, *American Economic Review* 18, 139–165.
- Coelli, T.J., (1995), “*Recent Development in Frontier Modelling and Efficiency Measurement*”, *Australian Journal of Agricultural Economics*, 39, 219–245.
- Coelli, T., A. Estache, S. Perelman and L. Trujillo (2003), *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, Washington DC: World Bank.
- Conrad, R. and R. Strauss, (1983), “*A multiple-output multiple-input model of the hospital industry in North Carolina*”, *Applied Economics*, 15, 341–352.
- Cooper, W.W., L.M. Seiford and K. Tone, (2007), *Data envelopment Analysis: a Comprehensive Text with Models, Applications, References and a DEA-Solver Software*, Boston: Kluwer.
- Cornes, R., (1992), *Duality and Modern Economics*, Melbourne: Cambridge University Press.
- Cornwell, C., P. Schmidt and R.C. Sickles, (1990), “*Production frontiers with cross-sectional and time-series variation in efficiency levels*”, *Journal of Econometrics*, 46, 185–200.
- Cowing, T. and A.G. Holtmann, (1983), “*Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data*”, *Southern Economic Journal* 49, 637–653.
- Crivelli, L., M. Filippini e D. Lunati, 2001, “*Dimensione ottima degli ospedali in uno Stato federale*”, *Economia Pubblica*, 5, 97–119.
- Crivelli, L., M. Filippini e D. Lunati, (2001), *Efficienza nel settore delle case per anziani in Svizzera*, Berna: SECO.
- Crivelli, L., M. Filippini and D. Lunati (2002), “*Regulation, ownership and efficiency in the Swiss nursing home industry*”, *International Journal of Health Care Finance and Economics*, 2, 79–97.
- Daraio, C. and L. Simar, (2008), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer.
- Davis, J., (1986), *Statistics and data analysis in geology (2nd ed.)*, New York: John Wiley and Sons.
- Debreu, G., (1951), “*The Coefficient of Resource Utilisation*”, *Econometrica*, 19, 225–234.

- Demsetz, H., (1967), "*Toward a Theory of Property Rights*", *The American Economic Review*, 57, 347–359.
- Demsetz, H., (1988), "*The Theory of the Firm Revisited*" *Journal of Law, Economics and Organization*, 4, 141–61.
- Deprins, D. and L. Simar (1989), "*Estimating Technical Inefficiencies with Correction for Environmental Conditions, with an application to railway companies*", *Annals of Public and Cooperative Economics*, 60, 81–102.
- Dervaux, B., H. Leleu, H. Nogues and V. Valdmanis, (2006), "*Assessing French nursing home efficiency: An indirect approach via budget-constrained DEA models*", *Socio-Economic Planning Sciences*, 40, 70–91.
- Diewert, W.E., (1971), "*An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function*", *Journal of Political Economy* 79, 481–507.
- Diewert, W.E., (1982), "*Duality approaches to microeconomic theory*", in *Handbook of Mathematical Economics*, Amsterdam: North-Holland Publishing.
- Diewert, W.E. and T.J. Wales, (1987), "*Flexible Functional Forms and Global Curvature Conditions*", *Econometrica*, 55, 43–68.
- Donabedian, A. (1980), *Explorations in quality assessment and monitoring - Volume I - The definition of quality and approaches to its assessment*, Ann Arbor: University of Michigan.
- Donabedian, A. (1990), *La qualità dell'assistenza sanitaria. Principi e metodologie di valutazione*, Roma: La Nuova Italia Scientifica.
- Dor, A., (1989), "*The Costs of Medicare Patients in Nursing Homes in the United States: A Multiple Output Analysis*", *Journal of Health Economics*, 8, 253–270.
- Dor, A., (1994), "*Non-minimum cost function and the stochastic frontier: On applications to health care providers*", *Journal of Health Economics*, 13, 329–334.
- Erlandsen, E., and F.R. Førsund, (1999), "*Efficiency in the Provision of Municipal Nursing-and Home-Care Services: the Norwegian Experience*", Department of Economics, Memorandum 32/99, University of Oslo.
- Evans, R.G., (1971), "*Behavioural Cost Function for Hospitals*", *The Canadian Journal of Economics*, 4, 198–215.
- Fabrizi, D., R. Fazioli e M. Filippini, (1996), *L'intervento pubblico e l'efficienza possibile*, Bologna: Il Mulino.
- Färe, R., S. Grosskopf, and C.A.K. Lovell, (1988), "*Scale Elasticity and Scale Efficiency*", *Journal of Institutional and Theoretical Economics*. 144, 721–729.

- Färe, R., S. Grosskopf and C.A.K. Lovell, (1994), *Production frontiers*, Cambridge: Cambridge University Press.
- Farrell, M.J., (1957), “*The Measures of Productive Efficiency*”, *Journal of the Royal Statistical Society*, 120, 253–281.
- Farsi, M. and M. Filippini (2004), “*An Empirical Analysis of Cost Efficiency in Non-profit and Public Nursing Homes*”, *Annals of Public and Cooperative Economics*, 75, 339–365.
- Farsi, M. and M. Filippini, (2008), “*Effects of Subsidization and Teaching Activities on Hospital Costs in Switzerland*”, *Health Economics*, 17, 335–350.
- Farsi, M., M. Filippini and W.H. Greene (2005), “*Efficiency Measurement in Network Industries: Application to the Swiss Railway Companies*,” *Journal of Regulatory Economics*, 28, 69–90.
- Farsi, M., M. Filippini and M. Kuenzle, (2005), “*Unobserved Heterogeneity in Stochastic Cost Frontier Models: An Application to Swiss Nursing Homes*”, *Applied Economics*, 37, 2127–2141.
- Farsi, M., M. Filippini and M. Kuenzle, (2006), “*Cost efficiency in regional bus companies: An application of new stochastic frontier models*”, *Journal of Transport Economics and Policy*, 40, 95–118.
- Farsi, M., M. Filippini and M. Kuenzle, (2007), “*Cost Efficiency in the Swiss Gas Distribution Sector*”, *Energy Economics*, 29, 64–78.
- Farsi, M., M. Filippini e D. Lunati, (2003), *Sperimentazione dell’impiego di un modello econometrico di funzione di costo, nell’ambito del progetto contratti di prestazione nel settore sociosanitario ticinese, Rapporto per il Dipartimento della sanità e socialità del Cantone Ticino, Lugano e Zurigo.*
- Feger, F., (2000), “*A Behavioral Model of the German Compound Feed Industry. Functional Form, Flexibility, and Regularity*”, PhD Dissertation, Essen.
- Fehr, E., U. Fischbacher and S.Gächter, (2002), “*Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms*”, *Human Nature*, 13, 1–25.
- Feldstein, M.S. and J. Shuttinga, (1977), “*Hospital costs in Massachusetts: A methodological study*”, *Inquiry*, 14, 22–31.
- Filippini M., (1991), “*La struttura dei costi delle ferrovie private svizzere secondo la teoria del duale: implicazioni per una politica delle fusioni*”, *Diss. der Rechts- und staatswissenschaftlichen Fakultät der Universität, Zürich.*
- Filippini, M., (1996), “*Dimensioni e struttura dei costi delle ferrovie private svizzere*”, *Economia Pubblica*, 2, 93–118.

- Filippini, M., (1997), *Elements of the Swiss Market for Electricity*, Berlin: Physica-Verlag.
- Filippini, M., (1998), “*Efficienza di costo nei servizi assistenziali residenziali per anziani*”, *Economia Pubblica*, 5, 5–25.
- Filippini, M., (2001), “*Economies of Scale in the Swiss Nursing Home Industry*”, *Applied Economics Letters*, 8, 43–46.
- Filippini M. and R. Maggi. (1993), “*Efficiency and Regulation in the Case of the Swiss Private Railways*”, *Journal of Regulatory Economics*, 5, 199–216.
- Folland, Sherman T. and R. A. Hofler (2001), “How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production”, *Health Economics*, 10: 683-698.
- Fox, K.J., (2002), *Efficiency in the public sector*, Boston: Kluwer Academic Publishers.
- Franci A. e M. Corsi, (1996), *La qualità soggettiva dell'assistenza residenziale agli anziani*, Padova: Summa.
- Franci A. e M. Corsi, (2004), *La soddisfazione del personale e le sue componenti nelle residenze per anziani*, Padova: Summa.
- Franke, R., (1982), “*Scattered Data Interpolation: Test of Some Methods*”, *Mathematics of Computations*, 33, 181–200.
- Frech, H.E., (1985), “*The Property Rights Theory of the Firm: Some Evidence from the U.S. Nursing Home Industry*”, *Zeitschrift für die gesamte Staatswissenschaft*, 141, 146–166.
- Fried, H.O., C.A.K. Lovell and S.S. Schmidt (eds.), (1993), *The Measurement of Productive Efficiency and Productivity Change*, New York: Oxford University Press.
- Fried, H.O., C.A.K. Lovell and S.S. Schmidt (eds.), (2008), *The Measurement of Productive Efficiency and Productivity Change*, New York: Oxford University Press.
- Friedlaender, A.F. and S.J. Wang Chiang. (1983), “*Productivity Growth in the Regulated Trucking Industry*”, *Research in Transportation and Economics*, 1, 149–184.
- Frisch, R., (1965), *Theory of Production*, Chicago: Rand-McNally.
- Gallant, A.R., (1981), “*On the Bias in Flexible Functional Forms and Essentially Unbiased Form*”, *Journal of Econometrics*, 15, 211–245.
- Gaynor, M. and G.F. Anderson, (1995), “*Uncertain demand, the structure of hospital costs and the cost of empty hospital bed*”, *Journal of Health Economics*, 14, 291–317.

- Gerdtham, U.G., M. Löthgren, M. Tambour and C. Rhenberg, (1999), “*Internal Markets and Health Care Efficiency: A Multiple-Output Stochastic Frontier Analysis*”, *Health Economics*, 8, 151–164.
- Gertler P. and D.M. Waldman, (1992), “*Quality Adjusted Cost Functions and Policy Evaluation in the Nursing Home Industry*” *Journal of Political Economy*, 100, 1232–1256.
- Gori, E. e G. Vittadini (a cura di), (1999), *Qualità e valutazione nei servizi di pubblica utilità*, Milano: Etas.
- Grannemann, W., S. Brown and M.V. Pauly, (1986), “*Estimating Hospital Cost: A Multiple Output Analysis*”, *Journal of Health Economics*, 5, 107–127.
- Greene, W.H., (1980a), “*Maximum Likelihood Estimation of Econometric Frontier Functions*”, *Journal of Econometrics* 13, 27–56.
- Greene, W.H., (1980b), “*On the Estimation of a Flexible Frontier Production Model*”, *Journal of Econometrics* 13, 101–115.
- Greene, W.H. (1997). *Frontier Production Functions*, In Pesaran, M.H. and Schmidt, P. (eds.). *Handbook of Applied Econometrics, Volume II: Microeconomics*, Boston: Blackwell Publishers Ltd, 81–166.
- Greene, W.H., (2002), *Econometric Analysis* (5rd ed.), Upper Saddle River: Prentice Hall.
- Greene, W.H., (2005), “*Reconsidering heterogeneity in panel data estimators of the stochastic frontier model*”, *Journal of Econometrics*, 126, 269–303.
- Greene, W.H., (2008), *The Econometric Approach to Efficiency Analysis*, in Fried, H.O., C.A.K. Lovell and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency and Productivity Change*, New York: Oxford University Press.
- Gui, B., (1991), “*Le organizzazioni mutualistiche e senza fine di lucro. Un approccio unificato al terzo settore*”, *Stato e mercato*, 31, 143–157.
- Gulley, D.O. and R.E. Santerre, (2003), “*The Effect of Public Policies on Nursing Home Care in the United States*”, *Eastern Economic Journal*, 29, 93–104.
- Guyomard H. and D. Vermersch, (1989), “*Derivation of Long-Factor Demands from Short-Run Responses*”, *Agricultural Economics* 3, 213–230.
- Hall, R. E., (1973), “*The Specification of Technology with Several Kinds of Output*”, *Journal of Political Economy*, 81, 878–92.
- Hausman, J.A. and W.E. Taylor (1981), “*Panel Data and Unobservable Individual Effect*,” *Econometrica*, 49, 1377–1398.

- Heckman, J. and T. MaCurdy, (1980), “*A Life Cycle Model of Female Labor Supply*”, *Review of Economic Studies*, 47, 47–74.
- Hicks, J.R., (1935), “*Annual Survey of Economic Theory: The theory of monopoly*”, *Econometrica*, 3, 1–20.
- Hsiao, C., (2003), *Analysis of Panel Data (2nd ed.)*, Cambridge: Cambridge University Press.
- Huang, C.J. and Liu, J.-T, (1994), “*Estimation of a Non-Neutral Stochastic Frontier Production Function*”, *Journal of Productivity Analysis* 5, 171–180.
- Jacobs, R., P.C. Smith and A. Street, (2006), *Measuring Efficiency in Health Care. Analytic Techniques and Health Policy*, Cambridge: Cambridge University Press.
- Jondrow, J., I.S. Materov, P. Schmidt and C.A.K. Lovell (1982), “*On Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model*”, *Journal of Econometrics*, 19, 233–238.
- Jorgenson, D.W., (1967), “*The Theory of Investment Behaviour*”, National Bureau of Economic Research, Columbia University Press.
- Jorgenson, D.W. (1974), “*The Economic Theory of Replacement and Depreciation*”, in: Sellekaerts, W. (ed.), *Econometric and Economic Theory, Essays in Honour of Jan Tinbergen*, London: MacMillan.
- Kahneman, D., J. L. Knetsch and R. H. Thaler, (1991), “*Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*”, *Journal of Economic Perspectives*, 5, 193–206.
- Kaserman, D.L. and J.W. Mayo, (1995), *Government and Business. The Economics of Antitrust and Regulation*, Fort Worth: The Dryden Press.
- Koenker, R.W. and G.J. Bassett, (1978), “*Regression Quantiles*”, *Econometrica*, 46, 33–50.
- Koopmans, T.C., (1951), “*An Analysis of Production as an Efficient Combination of Activities*”, In Koopmans, T.C. (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph 13. New York, Wiley.
- Kooreman, P., (1994), “*Nursing home care in The Netherlands: a nonparametric efficiency analysis*”, *Journal of Health Economics*, 13, 301–316.
- Knox, K.J., E.C. Blankmeyer and J.R. Stutzman, (2006), “*Comparative Performance and Quality Among Nonprofit Nursing Facilities in Texas*”, *Nonprofit and Voluntary Sector Quarterly*, 35, 631–667.
- Kruskal W.H. and W.A. Wallis, (1952), “*Use of ranks in one-criterion variance analysis*”, *Journal of the American Statistical Association*, 47, 583–621.

- Kumbhakar, S.C., (1990), "*Production frontiers, panel data, and time-varying technical inefficiency*", *Journal of Econometrics*, 46, 201–211.
- Kumbhakar, S.C., S. Ghost and J.T. McGuckin, (1991), "A *Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms*", *Journal of Business and Economic Statistics*, 9, 279–286.
- Kumbhakar, S.C. and C.A.K Lovell, (2000), *Stochastic Frontier Analysis*, New York: Cambridge University Press.
- Kuenzle, M., (2005), *Cost Efficiency in Network Industries: Application of Stochastic Frontier Analysis*, PhD Dissertation, Swiss Federal Institute of Technology, Zürich, Switzerland.
- Kurowosky, B.D., (1982), *The Measurement and Assurance of Quality. Long Term Care: Perspective from research and Demonstration*, Health Care Financing Administration, Washington DC: US Department of Health and Human Services.
- Laffont, J.-J. e J. Tirole, (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge, MS: The MIT Press.
- Laine, J., M. Linna, U. Häkkinen and A. Noro, (2005), "Measuring the productive efficiency and clinical quality of institutional long-term care for the elderly", *Health Economics*, 14, 245–256.
- Lau, L.J., (1986), "Functional Forms in Econometric Model Building", In Griliches, Z. and M.D. Intriligator (eds.), *Handbook of Econometrics, Volume III*, Amsterdam: North-Holland Publishing Company, 1515–1565.
- Lave, J.R., L.B. Lave and L.P. Silverman, (1972), "Hospital cost estimation controlling for case-mix", *Applied Economics*, 4, 165–180.
- Lee, Y.H. and P. Schmidt, (1993), "A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency", In Fried, H.O., C.A.K. Lovell and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, New York: Oxford University Press, 237–255.
- Leibenstein, H., (1966), "Allocative vs. X-inefficiency", *American Economic Review*, 56, 392–415.
- Leibenstein, H., (1978), "X-inefficiency Exists – Reply to an Xorcist", *American Economic Review*, 68, 203–211.
- Linn, M.W., L. Gruel and B.S. Linn, (1977), "Patient Outcome as a Measure of Quality of Nursing Home Care", *American Journal of Public Health*, 67, 337–344.
- McFadden, D., (1978), "Cost, Revenue and Profit Functions", In Fuss, M. and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, New York: North-Holland, 3–109.

- Marshall, A., (1890), *Principles of economics: an introductory text*.
- Mas-Colell, A., M.D. Whinston and J.R. Green, (1995), *Microeconomic Theory (2nd ed)*, Oxford: Oxford University Press.
- McKay, N.L., (1988), “*An Econometric Analysis of Cost and Scale Economies in the Nursing Homes Industry*”, *Journal of Human Resources*, 23, 57–75.
- McKay, N.L., (1991), “*The effects of ownership and ownership change on nursing home industry costs*”, *Health Services Research*, 31, 327–346.
- McKay, N.L., M.E. Deily, (2002), “*Ownership and changes in hospital inefficiency, 1986-1991*”, *Inquiry: a journal of medical care organization, provision and financing*, 39, 388–399.
- McKay, N.L. and M.E. Deily, (2008), “*Cost inefficiency and hospital health outcome*”, *Health Economics*, 17, 833–848.
- Meeusen, W. and J. Van der Broeck, (1957), “*Efficiency estimation from Cobb–Douglas production functions with composed error*”, *International Economic Review*, 8, 435–444.
- Mundlak, Y., (1978), “*On the Pooling of Time Series and Cross-Section Data*”, *Econometrica*, 46, 69–85.
- Musgrave, R.A., (1959), *The Theory of Public Finance*, New York: McGraw Hill.
- Newhouse, J.P., (1994), “*Frontier Estimation: How Useful a Tool for Health Economics*”, *Journal of Health Economics*, 13, 317–322.
- Niskanen, W.A., (1968), “*The Peculiar Economics of Bureaucracy*”, *American Economic Review*, 58, 293–305.
- Niskanen, W.A., (1971), *Burocracy and Representative Government*, Chicago: Aldine Press.
- Nyman, J.A., (1988), “*The marginal cost of nursing home care*”, *Journal of Health Economics*, 7, 393–412.
- OECD, (2006), “*Reviews of Health System – Switzerland*”.
- Ozcan, Y.A., (2007), *Health Care Benchmarking and Performance Evaluation. An Assessment using Data Envelopment Analysis (DEA)*, New York: Springer.
- Pagano A. e C. Rossi, (1999), “*La valutazione dei servizi sanitari*”, in Gori, E. e G. Vittadini (a cura di), *Qualità e valutazione nei servizi di pubblica utilità*, Milano: Etas.
- Panzar, J.C. and R.D. Willig, (1977), “*Economies of Scale in Multi-Output Production*”, *Quarterly Journal of Economics*, 91, 481–493.

- Pareto, V., (1916), *Trattato di sociologia generale*.
- Pitt, M.M. and L.-F. Lee (1981), “*The measurement and sources of technical inefficiency in Indonesian weaving industry*”, *Journal of Development Economics*, 9, 43–64.
- Polachek, S. and B. Yoon, (1996), “*Panel Estimates of a Two-Tiered Earnings Frontier*”, *Journal of Applied Econometrics*, 11, 169–178.
- Posner, R.A., (1974), “*Theories of Economic Regulation*”, *The Bell Journal of Economics and Management Science*, 5, 335–358.
- Powell, W.W. and R. Steinberg, (2006), *The Nonprofit Sector: a research handbook (2nd ed.)*, New Haven, CT: Yale University Press.
- Rebba, V., e D. Rizzi, (2002), “*Un’analisi dell’efficienza e dei costi delle strutture residenziali per anziani della regione Veneto*”, *Politiche Sanitarie*, 3, 62–83.
- Reifschneider, D. and R. Stevenson, (1991), “*Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency*”, *International Economic Review* 32, 715–723.
- Richmond, J., (1974), “*Estimating the Efficiency of Production*”, *International Economic Review*, 15, 515–521.
- Ryan, D.L. and T.J. Wales, (2000), “*Imposing local concavity in the translog and generalized Leontief cost function*”, *Economics Letters*, 67, 253–260.
- Sabbatini, D., (2008), I servizi pubblici locali tra mercato e regolazione, in *Questioni di Economia e Finanza*, Occasional papers 19, Banca di Italia.
- Salvatore, D., (1985), *Statistica ed econometria. Teoria e problemi*, Milano: ETAS.
- Samuelson, P.A., (1947), *Foundation of economic analysis*, Cambridge: Harvard University Press.
- Shephard, R.W., (1953), *Cost and Production Functions*, Princeton: Princeton University Press.
- Shleifer, A., (1985), “*A Theory of Yardstick Competition*”, *Rand Journal of Economics*, 16, 319–327.
- Schmidt, P. and R.E. Sickles, (1984), “*Production Frontiers and Panel Data*”, *Journal of Business and Economic Statistics*, 2, 367–374.
- Sickles, R.E., (2005), “*Panel Estimators and the Identification of Firm-Specific Efficiency Levels in Parametric, Semiparametric and Nonparametric Settings*”, *Journal of Econometrics*, 126, 305–34.

- Shimizutani, S. and W. Suzuki, (2007), “*Quality and efficiency of home help elderly care in Japan: Evidence from micro-level data*”, *Journal of the Japanese and International Economies*, 21, 287–301.
- Simar, L., (1992), “*Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-parametric and Semi-parametric Methods with Bootstrapping*”, *The Journal of Productivity Analysis*, 3, 171–203.
- Simar, L. and P.W. Wilson (2000), “*Statistical Inference in Nonparametric Frontier Models: The State of the Art*”, *Journal of Productivity Analysis*, 13, 49–78.
- Simar, L., C.A.K. Lovell and P. Vanden Eeckaut, (1994), “*Stochastic Frontiers Incorporating Exogenous Influences on Efficiency*”, Discussion Paper No.9403, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Simon, H.A., (1947), *Administrative behavior : a study of decision-making processes in administrative organizations*, 4th ed. (1997), New York: The free press.
- Simon, H.A., (1982), *Models of Bounded Rationality*. Cambridge, MS: The MIT Press.
- Skinner, J., (1994), “*What do stochastic frontier cost functions tell us about inefficiency?*”, *Journal of Health Economics*, 13, 323–328.
- Sloan, F.A., (2000), “*Not-for-profit Ownership and Hospital Behaviour*”, in: Culyer, A.J. and J.P. Newhouse, *Handbook of Health Economics; Vol. 1*, Amsterdam: North-Holland.
- Smet, M., (2002), “*Cost characteristics of hospitals*”, *Social Science and Medicine*, 55, 895–906.
- Smith, A., (1776), *The Wealth of Nations*, London: London editions.
- Spady, R.H. and A.F. Friedlaender, (1978), “*Hedonic Cost Functions for the Regulated Trucking Industry*”, *The Bell Journal of Economics*, 9, 159–179.
- Stevenson, R.E., (1980), “*Likelihood Functions for Generalized Stochastic Frontier Estimation*”, *Journal of Econometrics*, 13, 57–66.
- Stigler, G.J., (1971), “*The Theory of Economic Regulation*”, *Bell Journal of Economics and Management Science*, 3, 3–18.
- Stigler, G.J., (1976), “*The Xistence of X-Efficiency*”, *American Economic Review*, 66, 213–216.
- Stiglitz, J.E., (2003; 2004), *Economia del settore pubblico (2^a ed.)*, Milano: Hoepli.
- Tani, P., (1986), *Analisi microeconomica della produzione*, Roma: La Nuova Italia Scientifica.

- Thaler, R., (1980), “*Toward a positive theory of consumer choice*”, *Journal of Economic Behavior and Organization*, 1, 39–60.
- Travaglini, C., (1998), “Un modello di osservazione economico-aziendale di unità di produzione e di erogazione di servizi sociali”, in Borzaga, C., G. Fiorentini e A. Matacena (a cura di), *Non-profit e sistemi di welfare*, Roma: Carocci.
- Varian, H.R., (1992), *Microeconomic Analysis (3rd ed.)*, New York: W.W. Norton & Company.
- Varian, H.R., (2005), *Intermediate Microeconomic: A Modern Approach (7th ed.)*, New York: W.W. Norton & Company.
- Vita, M.G., (1990), “*Exploring Hospital Production Function Relationships with Flexible Functional Forms*”, *Journal of Health Economics*, 9, 1–21.
- Vitaliano D. and M. Toren, (1994), “*Cost and efficiency in nursing homes: a stochastic frontier approach*”, *Journal of Health Economics*, 13, 281–300.
- Vitaliano, D.F., (2003), “*Do not-for-profit firms maximize profit?*”, *The Quarterly Review of Economics and Finance*, 43, 75–87.
- Wagstaff, A., (1989), “*Estimating Efficiency in the Hospital Sector: A comparison of three Statistical Cost Frontier Models*”, *Applied Economics*, 21, 659–672.
- Walras, L., (1874), *Éléments d'économie politique pure, ou théorie de la richesse sociale*.
- Wang, H.-J., (2002), “*Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model*”, *Journal of Productivity Analysis*, 18, 241–253.
- Weisbrod, B.A., (1988), *The Nonprofit Economy*, Cambridge: Harvard University Press.
- Weisbrod, B.A. (ed.), (1998) “*To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*”, New York: Cambridge University Press.
- Weisbrod, B.A. (2004), “*Institutional Form and Organizational Behavior*”, in Steinberg, R. (ed.), *The economics of nonprofit enterprises*, Cheltenham: Elgar.
- Winsten, C.B., (1957), “*Discussion on Mr. Farrell's Paper*”, *Journal of the Royal Statistical Society*, 120, 282–284.
- Zamagni, S. (a cura di), (1998), *Non profit com e economia civile*, Bologna: Il Mulino.
- Zellner, A., (1989), “*An efficient method for estimating seemingly unrelated regressions and test for aggregation bias*”, *Journal of the American Statistical Association*, 57, 585–612.
- Zhu, J., (2003), *Quantitative Models for Performance evaluation and Benchmarking, Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Boston: Kluwer Academic Publishers.

- Zoric, J., (2006), Cost Efficiency and Regulation of Slovenian Water Distribution Utilities: an Application of Stochastic Frontier Methods, PhD Dissertation, University of Lugano, Switzerland.
- Zweifel, P. and F. Breyer, (1997), *Health Economics*, New York: Oxford University Press.
- Zwinderman, A.H., (1990), “*The measurement of change of quality of life in clinical trials*”, *Statistics in medicine*, 9, 931–942.