

Tiziana Laureti

L'efficienza rispetto alla frontiera delle possibilità produttive

Modelli teorici ed analisi empiriche

Monografie
Scienze Sociali
14

MONOGRAFIE
SCIENZE SOCIALI

1. *Giovani Jeunes Jovenes*, a cura di Gianfranco Bettin Lattes, 2001
2. Francesco Ciampi, *Il governo delle risorse nell'ateneo: un modello per la valutazione ex ante dell'impatto economico e patrimoniale dei corsi di laurea*, 2001
3. Luciana Lazzeretti, Tommaso Cinti, *La valorizzazione economica del patrimonio artistico delle città d'arte: il restauro artistico a Firenze*, 2001
4. *Per leggere la società*, a cura di Gianfranco Bettin Lattes, 2003
5. Luciana Lazzeretti, *Nascita ed evoluzione del distretto orafa di Arezzo, 1947-2001: primo studio in una prospettiva ecology based*, 2003
6. *Art Cities, Cultural Districts and Museums*, edited by Luciana Lazzeretti, 2004
7. Saverio Migliori, *Lo studio e la pena. L'Università di Firenze nel carcere di Prato: rapporto triennale 2000-2003*, 2004
8. Franca Alacevich, *Promuovere il dialogo sociale. Le conseguenze dell'Europa sulla regolazione del lavoro*, 2004
9. Salvatore Curreri, *Democrazia e rappresentanza politica: dal divieto di mandato al mandato di partito*, 2004
10. Salvatore Curreri, *Partiti e gruppi parlamentari nell'ordinamento spagnolo*, 2005
11. *Produrre uomini. Procreazione assistita: un'indagine multidisciplinare*, a cura di Andrea Bucelli, 2006
12. *I sistemi museali in Toscana. Primi risultati di una ricerca sul campo*, a cura di Luciana Lazzeretti, 2006
13. *Constitutionalism and Democratic Transitions: Lessons from South Africa*, edited by Veronica Federico, Franco Fusaro, 2006

Tiziana Laureti

L'efficienza rispetto alla frontiera delle possibilità produttive

Modelli teorici ed analisi empiriche

Firenze University Press
2006

L'efficienza rispetto alla frontiera delle possibilità produttive : modelli teorici ed analisi empiriche / Tiziana Laureti. – Firenze : Firenze university press, 2006.

(Monografie. Scienze Sociali; 14)

<http://digital.casalini.it/8884534267>

Stampa a richiesta disponibile su <http://epress.unifi.it>

ISBN-10: 88-8453-426-7 (online)

ISBN-13: 978-88-8453-426-2 (online)

ISBN-10: 88-8453-427-5 (print)

ISBN-13: 978-88-8453-427-9 (print)

330.015195 (ed. 20)

Statistica economica

© 2006 Firenze University Press

Università degli Studi di Firenze

Firenze University Press

Borgo Albizi, 28 - 50122 Firenze, Italy

<http://epress.unifi.it/>

Printed in Italy

Ai miei genitori

INDICE

| | |
|---|----|
| PRESENTAZIONE | 11 |
| INTRODUZIONE | 15 |
| CAPITOLO PRIMO – L'EVOLUZIONE DEL CONCETTO E DELLA MISURA DELL'EFFICIENZA PRODUTTIVA | |
| 1.1 I concetti di produttività ed efficienza | 19 |
| 1.2 Il concetto di efficienza nella teoria microeconomica | 24 |
| 1.3 La funzione di produzione, la frontiera delle possibilità produttive e la misura di efficienza | 30 |
| CAPITOLO SECONDO – LA TECNOLOGIA DI PRODUZIONE | |
| 2.1 Premessa | 39 |
| 2.2 La rappresentazione della tecnologia di produzione: l'insieme di produzione e l'insieme di fabbisogno di input | 41 |
| 2.3 Le caratteristiche strutturali della tecnologia di produzione | 45 |
| 2.3.1 La descrizione della tecnologia di produzione | 45 |
| 2.3.2 Tecnologie regolari | 46 |
| 2.3.3 Tecnologie monotone | 47 |
| 2.3.4 Tecnologie convesse | 48 |
| 2.4 Il modello di analisi delle attività | 50 |
| 2.5 La rappresentazione della tecnologia mediante la funzione di produzione | 53 |
| 2.6 La rappresentazione della tecnologia a mezzo della funzione di trasformazione e della funzione di distanza | 62 |
| 2.7 I rendimenti di scala | 66 |
| 2.8 La frontiera dell'insieme di produzione e la misura di efficienza | 70 |
| CAPITOLO TERZO – L'EFFICIENZA RISPETTO ALLA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE: L'APPROCCIO PARAMETRICO | |
| 3.1 L'evoluzione dell'approccio econometrico: dalla funzione media alla frontiera di produzione | 77 |

| | | |
|---|---|-----|
| 3.2 | Alcuni sviluppi delle frontiere deterministiche: approccio probabilistico e approccio statistico | 85 |
| 3.3 | Le frontiere di produzione stocastiche, ovvero come superare i limiti delle frontiere deterministiche | 93 |
| 3.4 | La misura dell'efficienza tecnica nei modelli stocastici | 102 |
| 3.5 | La specificazione di forme distributive per il termine di errore unilaterale | 106 |
| 3.6 | Recenti sviluppi delle frontiere stocastiche | 110 |
| 3.6.1 | Qualche cenno all'analisi dei dati panel | 110 |
| 3.6.2 | L'introduzione dell'effetto di variabili esogene | 112 |
| 3.6.3 | La meta-frontiera di produzione | 121 |
| CAPITOLO QUARTO – L'EFFICIENZA RISPETTO ALLA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE: L'APPROCCIO NON PARAMETRICO | | |
| 4.1 | Premessa | 127 |
| 4.2 | L'analisi dell'efficienza rispetto a tecnologie convesse: la <i>Data Envelopment Analysis</i> | 130 |
| 4.2.1 | Il modello di base con rendimenti di scala costanti | 130 |
| 4.2.2 | Il modello a rendimenti di scala variabili | 140 |
| 4.2.3 | Misure di efficienza non radiali: il modello DEA additivo e il modello per la determinazione della misura di Russell | 148 |
| 4.3 | L'analisi dell'efficienza rispetto a tecnologie non monotone | 151 |
| 4.3.1 | Il concetto di congestione degli input e degli output | 151 |
| 4.3.2 | La determinazione delle misure di efficienza debole e della congestione | 159 |
| 4.4 | L'analisi dell'efficienza rispetto a tecnologie non convesse: il <i>Free Disposal Hull</i> | 165 |
| 4.5 | Recenti sviluppi della metodologia non parametrica | 168 |
| 4.5.1 | Considerazioni preliminari | 168 |
| 4.5.2 | Approcci stocastici alla DEA | 169 |
| 4.5.3 | L'introduzione di variabili esterne nell'analisi dell'efficienza | 173 |
| 4.5.4 | La presenza di strutture gerarchiche nei dati: la <i>Program Evaluation</i> | 178 |
| CAPITOLO QUINTO – LA MISURA DELL'EFFICIENZA NELL'ISTRUZIONE UNIVERSITARIA | | |
| 5.1 | Premessa | 185 |
| 5.2 | La rappresentazione formale del processo di produzione | 188 |
| 5.2.1 | La specificazione degli input e degli output | 188 |
| 5.2.2 | L'identificazione dell'unità produttiva | 193 |
| 5.2.3 | Metodologie per l'analisi della produttività e dell'efficienza dell'istruzione universitaria | 195 |
| 5.3 | L'università come unità produttiva | 197 |
| 5.3.1 | L'applicazione delle metodologie non parametriche | 197 |
| 5.3.2 | L'applicazione delle metodologie parametriche | 209 |
| 5.4 | Il dipartimento come unità produttiva | 210 |
| 5.5 | Il singolo laureato come unità produttiva | 226 |
| CAPITOLO SESTO – L'IDENTIFICAZIONE DELLA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE: DAI DATI ALLA STIMA | | |
| 6.1 | Premessa | 235 |
| 6.2 | La costruzione delle frontiere di produzione non parametriche | 236 |
| 6.2.1 | I programmi disponibili | 236 |

| | |
|---|-----|
| 6.2.2 L'utilizzazione del linguaggio AMPL | 238 |
| 6.3 La stima delle frontiere di produzione stocastiche | 252 |
| 6.3.1 I programmi disponibili | 252 |
| 6.3.2 L'utilizzazione del programma Intercooled Stata 9.0 | 253 |
| BIBLIOGRAFIA | 263 |

PRESENTAZIONE

Questa monografia sull'efficienza nella produzione rispetto alle frontiera delle possibilità produttive si colloca nel contesto della teoria duale della produzione, i cui sviluppi, a datare dai secondi anni '60 (con i contributi fondamentali di Ferguson del 1969 e di Gorman del 1976), hanno radicalmente trasformato l'analisi applicata della produzione.

La questione, in modo semplificato ma efficace e senza alcuna perdita di generalità, si può tradurre nei seguenti termini: se la tecnologia vincola il comportamento ottimizzante dei produttori, allora deve essere possibile utilizzare una rappresentazione accurata del comportamento ottimizzante per studiare la tecnologia.

Ci si può domandare subito: perché l'approccio duale? La risposta, che emerge da tutto l'elaborato stesso, è che si tratta di quello più moderno e che meglio si presta alla descrizione della produzione di capitale umano nell'università, il tema che interessa l'autrice, che lo sviluppa nella parte applicata del suo lavoro.

Il quale, come viene opportunamente annunciato dal sottotitolo, consta di una parte teorica e di una parte applicata.

Nella prima, l'autrice pone le basi concettuali e metodologiche di tutta la trattazione. Dopo aver illustrato i concetti di produttività ed efficienza ed averne discusso l'apparente analogia e la sostanziale difformità, viene diffusamente illustrata l'idea di efficienza nella teoria microeconomia, con la molto ben articolata definizione della funzione di produzione e le documentate illustrazioni dei concetti di efficienza tecnica ed allocativa.

Si tratta di una sorta di presentazione sommaria di temi, tutta giocata su un riuscito equilibrio di anticipazione dosata di idee e teorie che verranno riprese diffusamente poi, ad introduzione, appunto, dei pezzi forti che saranno costituiti, come vedremo tra un momento, da tecnologia di produzione, approccio parametrico e approccio non parametrico.

E parliamo subito del capitolo sulla tecnologia di produzione, allora. Vale a dire, di una parte del lavoro dove vengono trattati in modo formalizzato ed elegante, nell'approccio insiemistico, temi e concetti di base assai complessi, che, tuttavia, l'autrice riesce a rendere meno ostici, conciliandone le asperità sostanziali e formali con i contenuti sostantivi, facendo anche leva su illuminanti rappresentazioni grafiche.

Cosicché, capisaldi concettuali quali l'insieme di produzione, l'insieme di fabbisogno di input, l'isoquanto, il sottoinsieme efficiente, le diverse tecnologie, la funzione di distanza, i rendimenti di scala, la misura di efficienza di Farrell, risaltando in importanza e funzione, risultano agevolmente comprensibili, mi sembra di poter dire, anche ai non addetti ai lavori.

L'autrice passa poi ad illustrare l'efficienza rispetto alla frontiera nell'approccio parametrico. Qui, dopo aver tracciato un quadro storico sistematico dell'evoluzione dell'approccio stesso, vengono approfonditamente illustrate e discusse le frontiere deterministiche e quelle stocastiche, fino ai più recenti sviluppi di queste ultime, con largo uso di strumenti formali che sono, tuttavia, supportati da adeguate citazioni di risultati ottenuti dai più significativi lavori presenti in letteratura. È, questo, il punto di raccordo, che non verrà più abbandonato, tra la teoria e la pratica annunciate nel titolo ed è una cesura quanto mai opportuna e, direi, gradevole, che aiuta il lettore nella comprensione delle questioni che divengono sempre più complesse.

Il terzo dei capitoli, per così dire, teorici – ma ho del resto appena sostenuto che già nel precedente la trattazione ha in parte cambiato carattere, assumendo connotati misti – esamina l'approccio non parametrico; in buona sostanza, ma non solo, quello che viene indicato con l'acronimo DEA.

Anche qui, la trattazione è svolta ad un livello molto elevato, con l'uso di tecniche assai sofisticate, che vengono illustrate e discusse in modo sempre appropriato, con grande padronanza degli strumenti, sostenute, anch'esse da esemplificazioni tratte da lavori in letteratura e da illustrazioni grafiche. Anche questo capitolo è una testimonianza dell'alto grado di coinvolgimento nel dibattito teorico nazionale ed internazionale raggiunto dall'autrice. In questo ambito, vengono inoltre illustrati e discussi concetti nuovi nel dibattito sull'efficienza, quali l'efficienza debole e la congestione.

Molto interessante ed utile è altresì la discussione dei più recenti sviluppi della metodologia non parametrica, quali gli approcci stocastici alla DEA, l'introduzione di variabili esterne e la presenza di strutture gerarchiche nei dati.

Esaurita questa densa e completa parte sulle teorie, i metodi e le applicazioni che se ne sono date in letteratura, con i relativi risultati ottenuti, l'autrice passa alle applicazioni che se ne possono dare nel campo della produzione di capitale umano a livello universitario.

È, questa, una materia assai spinosa e in gran parte ancora da districare, che richiede una formalizzazione originale, dato che la produzione di capitale umano avviene in un contesto peculiare, con sì un produttore che è chiaramente identificabile, ma con una specificità dettata dai dati statistici disponibili che condizionano la preparazione teorica dell'analisi e impongono una modellizzazione tutta particolare.

L'autrice risolve brillantemente, così a me sembra, il problema, mettendo in primo piano la figura dello studente e modellando la tecnologia produttiva come una nella quale "lo studente produce sé stesso come laureato".

È un modo assai originale e, al tempo stesso, denso di implicazioni stimolanti, di porre la questione, che non ha altri precedenti se non quello di Catalano-Silvestri, peraltro affrontato in un contesto particolare e con finalità del tutto diverse.

Qui, l'idea viene sviluppata e portata alle estreme conseguenze sia concettuali che applicative, sostenuta com'è da una serie equilibrata e convincente di considerazioni ed ipotesi nel quadro della teoria neoclassica della produzione.

Viene, così, aperta una via interessante e, al tempo stesso, promettente, che può rappresentare una base solida per ulteriori sviluppi nello studio dell'efficienza nella produzione di capitale umano a livello universitario.

I risultati ottenuti dall'autrice in numerosi lavori di ricerca vengono illustrati e discussi e si palesa pienamente lo spazio di utilizzazione efficace delle metodologie illustrate in precedenza, con risultati interessanti, che fanno intravedere possibilità di aperture stimolanti anche sul piano dei raffinamenti metodologici.

La monografia si completa con un capitolo finale dove viene presentato e spiegato l'uso di opportuni *software* nell'analisi dell'efficienza.

Per riassumere e concludere, si tratta una trattazione di uniforme, elevato livello qualitativo, con molti e apprezzabili contributi originali.

L'analisi delle tecnologie produttive è svolta con rigore e profondità, non disgiunti, nei passaggi dove le intricatezze tecnico-scientifiche si fanno meno impervie, anche da una elegante presentazione.

Nessun aspetto rilevante dello schema teorico retrostante l'analisi applicata viene trascurato o anche soltanto sottovalutato. L'analisi dell'efficienza rispetto alla frontiera viene affrontata con rigore e lucidità, seguendo il disegno ormai consolidato e largamente accettato nella letteratura internazionale, ma anche sviluppando un proprio cammino esplicativo ed espositivo, con il dualismo degli approcci parametrico-non parametrico tutto giocato su analogie e differenziazioni che rendono la lettura, al tempo stesso, godibile e densa di contenuti avanzati e spesso alla frontiera della letteratura attuale sul tema, comprensibile anche negli aspetti più tecnici e di sostanza scientifica sempre rigorosa.

Gli aspetti applicativi, primo fra tutti, l'uso che viene fatto delle metodologie nel campo della produzione di capitale umano a livello universitario, sono esaminati con la dovuta cura e con grande proprietà, sia nell'uso delle diverse tecniche, sia nell'impiego dei dati disponibili, che vengono sempre adeguatamente analizzati criticamente, con risultati di assoluto rilievo e con diffusi connotati di originalità.

Il grado di aggiornamento della letteratura sugli argomenti trattati è assoluto. Nessun contributo di rilievo portato negli ultimi quindici-venti anni viene trascurato e ognuno viene visto come un tassello che ha la sua posizione ben definita nel quadro generale.

Guido Ferrari
Dipartimento di Statistica "G. Parenti"
Università di Firenze

INTRODUZIONE

Produttività ed efficienza sono termini comunemente utilizzati per caratterizzare l'abilità di una impresa nell'utilizzazione delle risorse, sia in ambito privato che pubblico.

Entrambi i concetti sono legati ad una teoria della produzione, che diventa essenziale per la determinazione dei criteri in base ai quali confrontare i risultati dell'attività produttiva e i fattori impiegati per ottenerli.

La misurazione dell'efficienza implica però il confronto con uno standard di ottimalità, differenziandosi in tal modo dalla valutazione della produttività.

Nel contesto della teoria neoclassica della produzione, il raggiungimento dell'efficienza produttiva è garantito dal concetto di massimizzazione del profitto, sulla base del quale, dati i prezzi esistenti sul mercato, il produttore è in grado di determinare quale delle tecniche possibili può essere adottata e questa scelta determina, simultaneamente, la quantità di output da produrre e le quantità dei fattori da impiegare.

In tale ottica, la capacità di scegliere programmi di produzione tecnicamente efficienti è normalmente presupposta come corollario della massimizzazione dei profitti, senza indagare sulle questioni organizzative dell'impresa.

L'obiettivo di questo lavoro è quello di offrire una panoramica, sicuramente incompleta, dell'evoluzione del concetto di efficienza, ed in particolare dell'efficienza tecnica, e delle metodologie attraverso le quali si affrontano i problemi della sua misurazione, alla luce della sempre maggiore attenzione rivolta a questi strumenti di valutazione.

Nel primo capitolo, si vuole offrire al lettore una visione complessiva delle problematiche concernenti la misurazione dell'efficienza rispetto alla frontiera delle possibilità produttive, secondo un percorso che intende fornire una rilettura delle interpretazioni teoriche del concetto di efficienza che si sono sviluppate nel tempo.

L'aspetto centrale della teoria della produzione, considerato nel secondo capitolo è, rappresentato dalla tecnologia nelle sue diverse proprietà e rappresentazioni, attraverso le quali si descrivono tutti i "modi" possibili di produrre beni da altri beni.

Privilegiando il concetto di efficienza tecnica, si illustrano le caratteristiche del "sistema" di riferimento, essenziale per l'individuazione della frontiera delle possibilità di produzione, rispetto alla quale definire una misura per ciascuna unità produttiva.

Lo sviluppo di specifiche metodologie per la costruzione della frontiera efficiente è strettamente legato alla scelta della particolare rappresentazione della tecnologia, che può privilegiare la funzione di produzione o l'insieme costituito da tutti i processi possibili.

Seguendo un'impostazione che cerca di ripercorrere l'evoluzione storica dei diversi contributi, coniugando l'avanzamento metodologico con l'applicazione empirica, nel terzo capitolo si concentrerà l'attenzione sugli aspetti tecnici delle metodologie parametriche di misurazione dell'efficienza, partendo dalle "tradizionali" funzioni medie fino a giungere alle "moderne" frontiere di produzione stocastiche.

Il quarto capitolo affronta il problema della valutazione dell'efficienza tecnica secondo un approccio non parametrico, caratterizzato dalla presenza di una moltitudine di tecniche attraverso le quali senza richiedere stringenti ipotesi sulla struttura della frontiera si perviene all'identificazione dell'insieme di unità assunte a riferimento.

L'ultima parte del lavoro è dedicata all'analisi delle questioni concernenti la "traduzione" dello strumento metodologico teorico in fasi operative, attraverso le quali si estrinseca l'analisi dell'efficienza di un insieme di unità produttive che operano in uno specifico settore di interesse.

Nel quinto capitolo si passano in rassegna alcuni lavori empirici riguardanti l'efficienza della formazione universitaria. Lo scopo è quello di illustrare criticamente le diverse scelte che possono essere effettuate al fine di tradurre i concetti propri di teoria della produzione, quali input, output, processo e unità produttiva, in misure quantitative ed indicatori operativi al fine di impostare un'analisi dell'efficienza di una delle principali attività attraverso le quali si investe nella formazione di capitale umano.

Nell'ottica di fornire uno strumento di supporto per l'implementazione delle tecniche per la stima delle frontiere delle possibilità produttive, nell'ultimo capitolo si offre qualche utile suggerimento per l'utilizzazione di due programmi, attraverso specifici esempi in relazione sia all'approccio non parametrico che a quello parametrico stocastico.

Desidero esprimere la mia gratitudine a Guido Ferrari che con la sua continua e preziosa opera di stimolo e di suggerimento ha reso possibile la realizzazione di questo lavoro.

Ringrazio, inoltre, James E. Storbeck per il contributo intellettuale di cui ho beneficiato durante i miei soggiorni di studio all'università di Warwick.

Mi preme infine riconoscere il debito che sento di avere verso il dottorato di ricerca in Statistica Applicata dell'Università di Firenze dove ho intrapreso i primi studi in questo settore traendo utili indicazioni che mi sono derivate dalle discussioni sull'argomento in occasione dei numerosi seminari ed incontri.

Capitolo Primo

L'EVOLUZIONE DEL CONCETTO E DELLA MISURA DELL'EFFICIENZA PRODUTTIVA

1.1 I CONCETTI DI PRODUTTIVITÀ ED EFFICIENZA

Nella letteratura economica e statistica, i termini “produttività” ed “efficienza”, che riflettono i due concetti comunemente utilizzati per caratterizzare l'abilità di una impresa nell'utilizzazione delle risorse, sono spesso impiegati come sinonimi. In realtà, essi denotano concetti in parte diversi, la cui misurazione può condurre ad indicatori cui sono associate diverse interpretazioni.

Un *indicatore di produttività* può essere definito mediante il rapporto tra il risultato dell'attività produttiva e i fattori impiegati per ottenerlo (Nisticò e Prosperetti, 1991). Facendo riferimento alla terminologia anglosassone, ormai invalsa nell'uso e di cui si farà particolare ricorso anche in seguito, il risultato dell'attività produttiva o prodotto viene definito *output*, mentre i fattori utilizzati nel processo per il suo ottenimento sono denominati *input*.

Come rileva Kuznets (1990), il concetto di produttività si compone di tre elementi: gli output, gli input e, infine, il processo tecnologico attraverso cui i primi due elementi sono connessi tra loro. La traduzione di questo concetto in strumenti di misurazione analitici implica che gli aggregati coinvolti, gli input e gli output, debbano essere noti e misurabili e, quindi, per la loro determinazione è necessario conoscere il processo di conversione dei primi nei secondi.

Il richiamo ad una teoria della produzione diventa, perciò, essenziale per la determinazione dei criteri in base ai quali confrontare i singoli output e i singoli input. La valutazione della produttività offre, perciò, la possibilità di impostare diverse analisi della struttura e del funzionamento di un'organizzazione economica, per valutare gli obiettivi raggiunti in rapporto ai mezzi utilizzati (Guarini e Tassinari, 1990).

In termini molto generali, la *misura dell'efficienza* di un'unità produttiva può essere definita per confronto tra il processo di produzione effettivamente realizzato e un altro processo, opportunamente scelto, corrispondente a uno standard di ottimalità, che può avere valenza nel tempo e nello spazio (Petretto, 1986).

È chiaro, quindi, che anche il concetto di efficienza è legato a una teoria della produzione. Come sottolineato dal suddetto Autore, la misurazione dell'efficienza deve avvenire tramite l'esame dell'evoluzione di indicatori specifici in una prospettiva *time series*, oppure attraverso il confronto degli indicatori con valori standard degli stessi, secondo una prospettiva *cross-section*.

Al fine di evidenziare gli elementi di diversità tra il concetto di efficienza e quello di produttività, si possono considerare come standard di riferimento i processi efficienti espressi dalla funzione di produzione, che indica, appunto, il massimo prodotto ottenibile da un dato livello di fattori produttivi, considerando la tecnologia esistente.

In una prospettiva *cross section*, ossia in riferimento a dati relativi a diverse unità produttive rilevate nello stesso momento, riferendosi al caso semplificato di due processi produttivi, A e B , concernenti due unità produttive, in cui viene utilizzato un unico input, x , che assume, rispettivamente, le specificazioni x_A e x_B , per produrre un solo output, y , la cui quantità è espressa, rispettivamente, da y_A e y_B , si può, innanzitutto, misurare la produttività di A e di B mediante il rapporto $P_A = y_A/x_A$ e $P_B = y_B/x_B$.

Se $P_A > P_B$ si conclude che il processo A è *più produttivo del processo B*, senza dover ricorrere a informazioni sulla tecnologia di produzione. Un indice di produttività di A relativo a B può essere facilmente costruito mediante il rapporto:

$$\frac{P_A}{P_B} = \frac{y_A/x_A}{y_B/x_B}.$$

Ipotizzando, invece, che la tecnologia sia descritta da una funzione di produzione, $y^* = f(x)$, si può individuare il massimo output producibile a partire dall'input x_A , ossia $y_A^* = f(x_A)$ e quello ottenibile utilizzando il livello x_B , ossia $y_B^* = f(x_B)$. In tal modo è possibile definire l'*Efficienza Tecnica (ET)* di ciascuna unità produttiva, confrontando l'output effettivamente prodotto con la quantità massima di output producibile a partire dalla quantità osservata dell'input, ottenendo misure di efficienza orientate nel senso degli output, definite, utilizzando la terminologia anglosassone, *output-oriented*.

Si può affermare che mentre la produttività è una misura assoluta, l'efficienza tecnica è una misura relativa. Nell'esempio considerato, si ottiene, per l'unità A , la misura di efficienza tecnica:

$$ET_A = \frac{y_A}{y_A^*} \leq 1$$

mentre per l'unità B l'efficienza tecnica è espressa dal rapporto:

$$ET_B = \frac{y_B}{y_B^*} \leq 1$$

Va osservato che, se un'unità produttiva, ad esempio A , producesse il massimo output ottenibile a partire dall'input utilizzato, x_A , la sua produttività sarebbe espressa da $P_A^* = y_A^*/x_A$. La misura di efficienza tecnica dell'unità A , può, quindi, essere espressa anche come indice di produttività relativo ad una unità ipotetica che produca il massimo output ottenibile data la tecnologia descritta dalla funzione di produzione $y^* = f(x)$, utilizzando il medesimo livello di input x_A , ossia mediante:

$$ET_A = \frac{y_A}{y_A^*} = \frac{P_A}{P_A^*} = \frac{y_A/x_A}{y_A^*/x_A}$$

In modo del tutto analogo si può definire la misura di efficienza di B attraverso l'espressione:

$$ET_B = \frac{y_B}{y_B^*} = \frac{P_B}{P_B^*} = \frac{y_B/x_B}{y_B^*/x_B}$$

La Figura 1.1 illustra graficamente i concetti precedenti: la produttività di A è uguale alla pendenza della retta OA , mentre quella di B è espressa dalla pendenza della retta OB . La determinazione dell'efficienza delle due unità considerate si basa sulla conoscenza dei punti A^* e B^* , che esprimono il massimo output producibile, descritto dalla funzione di produzione $y^* = f(x)$, a partire, rispettivamente, dalle quantità di input, x_A e x_B , effettivamente impiegate dalle due unità produttive.

Va osservato che, nel confrontare più processi di produzione attraverso un'analisi *cross-section*, un differenziale di produttività non comporta necessariamente un differenziale di efficienza e, d'altra parte, un differenziale di efficienza può non associarsi a un differenziale di produttività, salvo che la funzione di produzione sia caratterizzata da rendimenti di scala costanti, caratteristici di una tecnologia nella quale ad un incremento t dell'input corrisponde un pari aumento t dell'output¹. In tal caso i due concetti sono equivalenti (Nisticò e Prosperetti, 1991).

Continuando l'esempio precedente, nella Figura 1.2a, i due processi rappresentati sono entrambi pienamente efficienti poiché giacciono ambedue sulla funzione di produzione, che presenta rendimenti di scala decrescenti non uniformemente, per cui ad un incremento dell'input non corrisponde un incremento proporzio-

¹ Un'analisi formale dei rendimenti di scala è effettuata nel paragrafo 2.7.

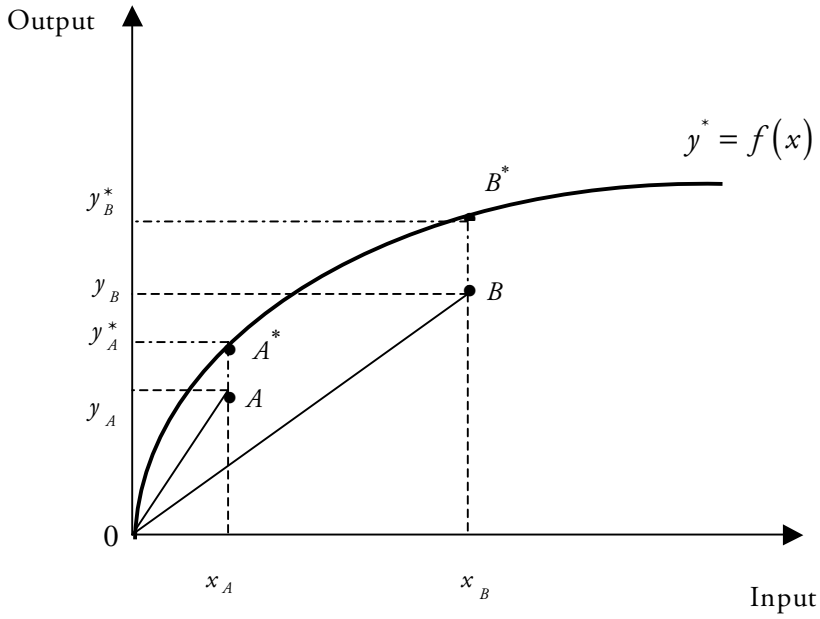


Figura 1.1 – Produttività ed efficienza.

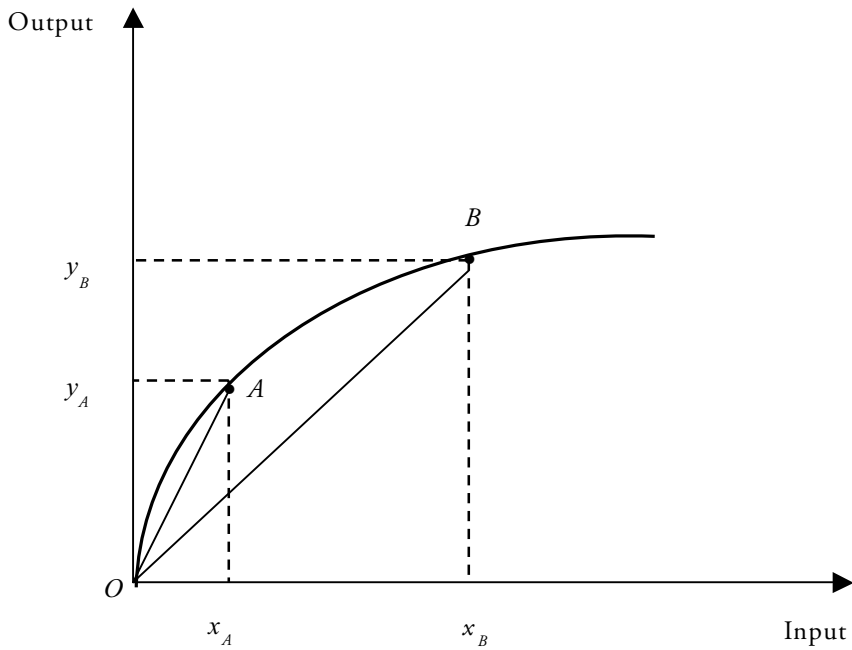


Figura 1.2a – Produttività ed efficienza: *identica efficienza diversa produttività*.

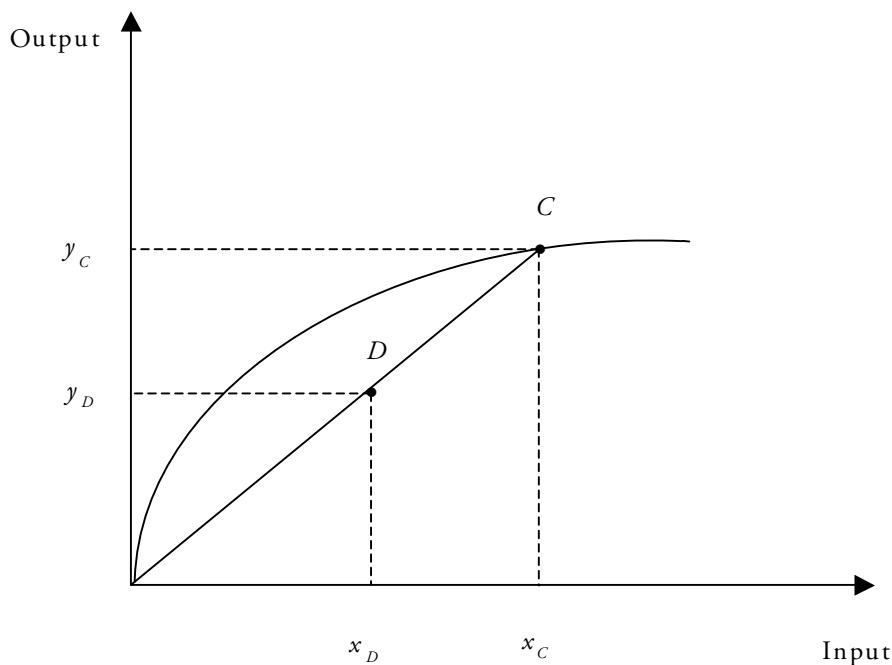


Figura 1.2b – Produttività ed efficienza: *identica produttività diversa efficienza*.

nale dell'output, ma la produttività dell'unità *A* risulta superiore a quella di *B*. Quindi, in questo caso semplificato, che coinvolge il confronto di due soli processi, ad un identico livello di efficienza, corrispondono due livelli di produttività.

La situazione opposta, caratterizzata da una situazione in cui le due unità produttive valutate presentano identici livelli di produttività ma diversa efficienza, è rappresentata nella figura 1.2b.

In quest'ultima situazione, la produttività delle due unità *C* e *D* è identica ed è misurata, rispettivamente, dai rapporti y_C/x_C e y_D/x_D , dove $y_C/x_C = y_D/x_D$. Ma l'unità *D*, posizionandosi al di sotto della funzione di produzione, presenta una minore efficienza rispetto a *C*, che invece giace sulla funzione di produzione.

L'impiego di un indicatore di produttività come *proxy* di una misura di efficienza può condurre, quindi, a risultati fuorvianti nei casi in cui i rendimenti di scala non possono considerarsi costanti. Nonostante ciò, misure di produttività media dei fattori, come, ad esempio, il lavoro, sono state ampiamente utilizzate nelle applicazioni empiriche in qualità di indici di efficienza.

Come osservano Gazzei, Lemmi e Viviani (1997), diversi Autori, soprattutto nell'ambito di studi di carattere microeconomico, esprimono il concetto di

produttività in forma strettamente dipendente da quello di efficienza, tanto da utilizzarlo, talvolta, nella definizione stessa.

Solo per fornire qualche esempio, si possono considerare le interpretazioni di Schmookler (1952), secondo cui la produttività è “un indice di efficienza”, e quella di Kendrick (1956), per il quale la produttività è un “cambiamento nell’efficienza produttiva”.

Zappa (1957) sottolinea che la “produttività è spesso nota come efficienza economica e che l’efficienza è non di rado conosciuta come produttività tecnica”, mentre Hatry (1982) sostiene che “l’efficienza riguarda il rapporto tra risorse utilizzate e quantità di beni o servizi prodotta. Le misure di efficienza si definiscono come la relazione tra la quantità di risorse impiegate ed il prodotto che ne deriva. Il rapporto tra prodotto e risorse impiegate viene chiamato produttività. Il suo inverso, il rapporto risorse/produzione, è definito efficienza o costo unitario. Le due forme sono equivalenti”.

L’analisi del concetto di efficienza e il conseguente sviluppo di strumenti metodologici adeguati alla misurazione, che costituiscono l’oggetto del presente lavoro, prendono avvio da Farrell (1957), che, come si vedrà più avanti, pose l’accento sul problema consistente nell’individuare la tecnologia di riferimento rispetto alla quale definire, poi, un’appropriata misura di efficienza².

La funzione di produzione appare così come una *frontiera* tra i processi tecnicamente possibili, che giacciono al di sotto di essa, e quelli impossibili (Thiry e Tulkens, 1987).

L’efficienza di un’impresa può essere, quindi, valutata considerando la posizione del processo produttivo posto in essere rispetto alla frontiera delle possibilità di produzione.

1.2 IL CONCETTO DI EFFICIENZA NELLA TEORIA MICROECONOMICA

L’efficienza, considerata in un’accezione generale che include anche il concetto di efficienza tecnica appena discusso, rappresenta una nozione estremamente importante nella teoria economica e in particolare nella teoria della produzione.

² I primi approcci alla misurazione della produttività si basano sulla costruzione di indici parziali, definiti come rapporto tra quantità prodotta ed uno solo dei fattori produttivi utilizzati. Solow (1957) pose le basi per lo sviluppo di indici di produttività globali, ottenuti rapportando il prodotto all’insieme dei fattori impiegati.

Le moderne teorie della produzione si rifanno agli approcci tradizionali, ossia a quello marginalista e a quello del sovrappiù. Dopo la pubblicazione del saggio di Sraffa (1960), è stato ripreso il concetto di produzione inteso come fenomeno circolare, ossia come *riproduzione*, che ha stimolato nuovo interesse per l'analisi della produzione tipica degli economisti classici e di Marx.

Nel modello fondi-flussi di Georgescu-Roegen (1973), si propone un'analisi della produzione che concentra l'attenzione sulla durata dei processi produttivi, mettendo in evidenza la possibilità che un dato procedimento di fabbricazione possa essere realizzato con differenti forme di attivazione, ognuna delle quali connessa con uno specifico problema di utilizzazione efficiente di elementi fondo della produzione, al fine di eliminare, o attenuare, eventuali sottoutilizzazioni che si riflettono nell'efficienza.

Nel contesto dell'impostazione neoclassica, nasce una teoria della produzione della scuola di Losanna, grazie ai contributi di vari Autori, quali Walras (1874) e Pareto (1906), i quali focalizzano la loro attenzione sull'intero sistema economico e sul suo equilibrio generale piuttosto che sulla singola unità di produzione, il cui processo decisionale e organizzativo viene perciò relegato in un ruolo sussidiario e considerato come una sorta di scatola nera il cui contenuto rimane inesplorato. Allo sviluppo della scuola paretiana contribuiscono diversi Autori, tra i quali Samuelson (1947) e Frisch (1965) ed in particolare, per l'uso della teoria della dualità, Shephard (1953) e McFadden (1966).

Il modello di *analisi delle attività*, o nella terminologia anglosassone *activity analysis*, rappresenta un modo alternativo di esaminare i processi di produzione e costituisce una formulazione particolare all'interno del più generale modello degli insiemi di produzione.

In questo contesto, il problema della scelta tra possibilità alternative è direttamente espresso come problema di utilizzazione ottimale di risorse date; inoltre, si fa esplicito riferimento allo strumento della programmazione lineare, introdotto precedentemente da von Neumann (1938, 1945).

L'associazione di ciascuna situazione efficiente con un vettore di prezzi trova una formulazione particolarmente interessante, essendo definita una procedura di determinazione di tali valori. Lo sviluppo del modello di analisi delle attività si deve soprattutto ai contributi di Koopmans (1951) e Debreu (1959), i quali rappresentano due esponenti della cosiddetta teoria della produzione *neo-walrasiana*, sviluppatasi dopo la seconda guerra mondiale.

Nella formulazione neoclassica della teoria della produzione, si utilizza prevalentemente una rappresentazione analitica del processo produttivo, definito come un processo di trasformazione regolato da uomini (Frisch, 1965) e descritto mediante una n -pla di numeri reali, \mathbf{z} , gli elementi negativi della quale misu-

rano gli input del processo, x , mentre gli elementi positivi ne misurano gli output, y , oppure mediante una coppia di vettori input-output (x, y) .

La costruzione di un modello di scelta del processo “ottimo” è basata sul criterio di massimo risultato netto, o in via subordinata, di minimo costo (Tani, 1986). Pertanto, ad ogni processo deve essere associata una valutazione del risultato netto, oppure del costo del vettore x degli input.

Nel caso solitamente considerato, ossia nel cosiddetto modello di produzione classico, si suppone dato un vettore di prezzi dei fattori, w , e dei prodotti, p , ai quali è possibile acquistare gli input e vendere gli output, in qualunque quantità.

Il costo di produzione relativo al processo (x, y) è definito, quindi, dal prodotto wx , mentre il risultato netto del processo è ottenuto mediante l'espressione: $py - wx$.

Il modello utilizzato per la rappresentazione delle possibilità di scelta che la tecnica offre è costituito dalla funzione di produzione, che identifica per ogni possibile vettore di input la quantità massima di output producibile.

L'obiettivo del massimo profitto consente, dati i prezzi esistenti sul mercato, data la funzione di produzione, di determinare quale delle tecniche possibili possa essere adottata dall'imprenditore e questa scelta determina, simultaneamente, la quantità di output da produrre e le quantità dei fattori da impiegare.

In tal modo, risulta individuato il processo produttivo ottimo per ciascuna impresa, che consente ad ogni unità di operare in una situazione di *efficienza produttiva*. La medesima situazione si verifica per il sistema della produzione nel suo complesso, purché si consideri un'economia walrasiana perfettamente concorrenziale, in cui tutte le unità produttive risultano *price-takers*.

Un elemento estremamente importante da sottolineare è che la teoria della produzione neoclassica, in condizioni di perfetta concorrenza, studia il comportamento razionale delle unità produttive *prima* del reale svolgimento del processo produttivo.

La funzione di produzione è, infatti, parte di una teoria che richiede la determinazione contemporanea di tutte le incognite, e quindi dei prezzi dei fattori e delle quantità prodotte da ogni impresa, prima che ogni atto produttivo abbia luogo. Anche nel caso del lungo periodo è necessario supporre che tutte le modificazioni, che intervengono in relazione alla tecnologia e alle dotazioni presenti e future dei fattori, siano note al momento in cui le imprese decidono le modalità con cui allocare le loro risorse, all'inizio del processo produttivo (Vaggi, 1987).

Il grande vantaggio di questa impostazione, che ne giustifica anche il largo impiego nelle applicazioni a livello econometrico, risiede proprio nelle buone qualità analitiche che possono essere attribuite alle funzioni di produzione, le

quali consentono l'impiego di strumenti matematici potenti e di comode specificazioni econometriche.

Il limite maggiore di un tale impianto teorico consiste nel fatto che la tradizionale funzione di produzione non descrive tutte le possibili relazioni tecnologiche tra gli input e l'output, ma solo quelle tecnicamente output efficienti (Walters, 1963).

Va osservato comunque che, sebbene la funzione di produzione individui le tecniche efficienti, ciò non implica che vi sia un unico metodo efficiente di produzione che lega l'output ai fattori produttivi.

Le imprese possono, infatti, scegliere con quali proporzioni dei fattori produrre ogni data quantità di output; la funzione di produzione indica, per ogni y , i metodi produttivi che comportano una sostituzione fra i fattori, per cui, in generale, se si riduce un input, il livello del prodotto può restare inalterato solo se aumenta la quantità impiegata di almeno un altro fattore.

La scelta della particolare combinazione di fattori impiegata, che massimizza il risultato netto, dipende dai prezzi relativi dei fattori e dalle relazioni di sostituibilità tecnica, descritte dalla funzione di produzione.

La concezione neoclassica della produzione si fonda sull'idea di allocazione efficiente delle risorse a prezzi dati. L'ipotesi di continuità e differenziabilità della funzione di produzione consentono di determinare *quanto* produrre uguagliando la produttività marginale del fattore j -esimo al prezzo del fattore stesso in termini del prodotto, ossia:

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \frac{w_j}{p}$$

e *come* produrre scegliendo una particolare combinazione di input j e s , ossia:

$$\frac{x_j}{x_s} .$$

L'allocazione efficiente delle risorse implica anche che ogni fattore riceva parte del prodotto, in relazione al proprio contributo produttivo.

Spostando l'attenzione dall'analisi della singola impresa al sistema complessivo, l'allocazione delle risorse esistenti in una situazione di concorrenza perfetta viene effettuata attraverso le scelte degli imprenditori che determinano, non solo la massimizzazione dei profitti e la minimizzazione dei costi, ma anche il livello dei prezzi relativi dei fattori.

In una tale situazione si realizza, quindi, la situazione di *efficienza paretiana di tipo allocativo*, in relazione alla quale non può esistere alcun modo di aumentare la soddisfazione di un individuo senza ridurre la soddisfazione di qualcun altro.

Contro l'impostazione della teoria neoclassica, che identifica il concetto di efficienza con quello di efficienza allocativa derivante dalle condizioni concorrenziali che caratterizzano il mercato, si muove il lavoro di Leibenstein (1966), che introduce un nuovo concetto di efficienza connesso alle decisioni interne all'impresa in relazione alla scelta dei processi produttivi.

Il tentativo è diretto al superamento delle limitazioni insite nell'utilizzazione della funzione di produzione che incorpora, da quanto emerso in precedenza, non solo le caratteristiche tecniche dei processi di produzione, ma anche un'ipotesi di comportamento dell'agente, le cui decisioni producono effetti nella gestione dell'attività produttiva dell'unità considerata.

Leibenstein denomina questo concetto di efficienza "*efficienza x*" (*x-efficiency*) per indicare la capacità non dei mercati, ma dell'organizzazione aziendale di allocare le risorse in modo efficiente, ossia adottando un comportamento tale da rendere il saggio marginale di sostituzione tecnica uguale al rapporto tra i prezzi dei fattori, nonché di scegliere programmi di produzione tecnicamente efficienti.

Come già sottolineato in precedenza, nella teoria economica neoclassica una simile capacità viene normalmente presupposta come corollario della massimizzazione dei profitti, senza indagare sulle questioni organizzative dell'impresa.

Tuttavia, alcuni economisti erano intervenuti precisando che questo presupposto non è sempre valido nella realtà delle imprese non sottoposte alla pressione della concorrenza. L'allontanamento da una situazione di concorrenza perfetta e, quindi, l'esistenza sul mercato, ad esempio, di situazioni monopolistiche, può causare un'allocazione delle risorse non efficiente e, conseguentemente, la presenza di inefficienza allocativa.

Ad esempio Hicks (1935) con la sua "*quiet life*" del monopolista aveva affrontato la problematica, asserendo che l'assenza di pressione concorrenziale può indurre i produttori a non seguire pienamente un comportamento di massimizzazione del profitto ma, piuttosto, a posizionarsi in una situazione vicina a quella di massimizzazione.

La teoria della *efficienza x* di Leibenstein fu criticata da Stigler (1976), il quale ricondusse il concetto di *x-efficiency* alla nozione classica di efficienza allocativa.

Secondo il suddetto Autore, l'esistenza di eventuali differenze nell'output prodotto da diverse imprese, che impiegano una medesima quantità di input, può essere dovuta al fatto che gli imprenditori considerati sono dotati di una diversa conoscenza tecnologica e ciò si riflette nell'utilizzazione di specifici input. L'effetto di tali variazioni viene attribuito interamente ai fattori produttivi, per cui il problema si traduce in una situazione di allocazione delle risorse non efficiente.

Al di là delle critiche, il merito della teoria della *x-efficiency* è stato quello di aver messo in luce l'esistenza di tipologie di comportamento del produttore,

definibili come razionali ma compatibili con situazioni anche tecnicamente inefficienti, conducendo gli studiosi a prendere sempre più in considerazione, anche nell'analisi empirica ed econometrica, la possibilità di osservare situazioni produttive che non si posizionano sulla funzione di produzione, e stimolando in tal modo l'analisi di situazioni non efficienti, la cui presenza era sostanzialmente negata nel contesto della teoria neoclassica.

Un ruolo fondamentale nello sviluppo dell'analisi dell'efficienza viene esercitato dall'*analisi delle attività*, attraverso la quale vengono introdotti strumenti alternativi alla funzione di produzione per affrontare i problemi di efficiente allocazione di risorse date fra produzioni alternative.

Infatti, la funzione di produzione consente di trattare solo i casi in cui, per un dato insieme di prezzi, la scelta dell'unità produttiva è unica, mentre non possono essere considerate situazioni per cui la ricerca del massimo profitto porta a più di una possibile scelta produttiva, vale a dire a un insieme di scelte produttive.

Koopmans (1951) introduce una definizione formale di *efficienza tecnica* sviluppando e adattando il concetto di efficienza paretiana, che si traduce nell'affermazione per cui in caso di efficienza perfetta in riferimento ai beni finali, nessun bene può essere incrementato senza con ciò provocare una diminuzione o uno spreco di altri output.

Segue la definizione di processo produttivo efficiente come di un processo rispetto al quale qualsiasi incremento in qualunque output richiede una riduzione in almeno un altro output o un incremento in almeno un input, e nel quale una riduzione in qualsiasi input richiede un incremento in almeno un altro input o una riduzione in almeno un output.

Pur costituendo uno strumento per distinguere i processi efficienti da quelli non efficienti, la definizione di Koopmans non consente di sviluppare una soluzione operativa per la determinazione del grado di efficienza di un processo non efficiente, o per l'identificazione di un processo efficiente o di una combinazione di processi efficienti, rispetto alla quale confrontare gli altri processi.

Questa problematica è affrontata da Debreu (1951), il quale si pone il problema di trovare una misura della perdita connessa ad una situazione non ottima, che esprima "la distanza" dalla situazione effettiva del sistema produttivo a quella di ottimo paretiano, introducendo la prima misura di efficienza produttiva, ossia il noto coefficiente di utilizzazione delle risorse (*coefficient of resource utilization*).

Farrell (1957) estende il lavoro iniziato da Koopmans e Debreu evidenziando che l'efficienza produttiva si compone di un altro elemento importante che riflette l'abilità dei produttori di selezionare la giusta combinazione di input per la produzione di una determinata quantità di output alla luce dei prezzi prevalenti sul mercato.

In un tale contesto, Farrell nel suo “*The Measurement of Productive Efficiency*” definisce l’efficienza produttiva complessiva come il prodotto dell’efficienza tecnica e dell’efficienza allocativa, denominata *price efficiency*.

La determinazione di queste misure di efficienza viene effettuata confrontando la *performance* osservata di una certa unità produttiva con uno standard di perfetta efficienza, definito secondo specifici criteri. Va osservato, comunque, che nella definizione di Farrell di efficienza allocativa si ritrova implicitamente il concetto di massimizzazione del profitto o di minimizzazione del costo da parte del produttore in mercati competitivi.

1.3 LA FUNZIONE DI PRODUZIONE, LA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE E LA MISURA DI EFFICIENZA

Il modello della funzione di produzione, oltre a rappresentare il primo modello in senso storico ad essere utilizzato dalla teoria economica, è anche quello che consente con maggiore semplicità formale di fornire, almeno in prima istanza, risposte a questioni economiche inerenti il processo produttivo, come ad esempio al fatto se esistano diverse tecniche o modi di produrre gli stessi prodotti e di quali relazioni sussistano fra le proporzioni di impiego dei fattori produttivi utilizzati (Zamagni, 1994).

I primi studi empirici sulla funzione di produzione si basano su serie storiche di dati relativi agli input impiegati e agli output prodotti da ciascuna unità produttiva.

La funzione di produzione Cobb-Douglas (1928) è applicata per la prima volta su dati rilevati in diversi periodi di tempo (serie temporali o *time series*), per verificare empiricamente la teoria della produttività marginale per la distribuzione del prodotto totale.

Poiché con il trascorre del tempo si assiste in genere ad una crescita della popolazione, cui si accompagna un certo sviluppo tecnologico, la critica principale mossa dagli studiosi verso una tale procedura consiste nell’affermare che le eventuali relazioni stimate tra il prodotto e i fattori produttivi, capitale e lavoro, potrebbero essere esclusivamente il risultato dell’agire del tempo.

Bronfenbrenner e Douglas (1939) utilizzano, per la prima volta, dati *cross section* per la stima della funzione di produzione mediante il metodo dei minimi quadrati ordinari (*Ordinary Least Squares, OLS*), ipotizzando che tutte le deviazioni dalla funzione stimata siano dovute ad errori casuali di misurazione della variabile dipendente, o ad operazioni aleatorie di vario genere non incluse nel modello.

Secondo i suddetti Autori, i risultati ottenuti confermano i precedenti studi basati su dati temporali.

Nei lavori empirici concernenti la stima dei parametri della funzione di produzione da dati *cross-section*, la forma deterministica del modello di produzione viene, quindi, “modificata” attraverso l’introduzione di disturbi stocastici, al fine di considerare eventuali errori casuali non sistematici, dovuti, in parte, all’agire dei produttori ed effettuati nel tentativo di adattare gli input per soddisfare le condizioni necessarie per la massimizzazione del profitto.

D’altra parte, l’interpretazione di questi disturbi stocastici, che si assume abbiamo media nulla e varianza costante, non risulta chiaramente delineata in letteratura sino allo studio di Marschak e Andrews (1944), che affrontano il trattamento esplicito degli aspetti probabilistici del modello della funzione di produzione.

I suddetti Autori definiscono il termine stocastico come una componente che riflette l’*efficienza tecnica* dell’unità produttiva, e che dipende dalle conoscenze tecnologiche, dagli sforzi, dai desideri e dalla fortuna di un dato imprenditore.

In una tale situazione, la misura dell’efficienza tecnica di una certa unità produttiva i può essere descritta da uno o più parametri, che entrano nell’espressione generale della funzione di produzione specificandola per l’unità i -esima.

In altre parole, ciascuna unità produttiva è caratterizzata da una propria funzione di produzione, che differisce da quella delle altre unità per il termine di efficienza, ma che è identica in tutti gli altri aspetti.

Marschak e Andrews spiegano, inoltre, la variabilità dei risultati tra le imprese, riconoscendo l’esistenza di differenze nell’*efficienza economica*, ossia nell’abilità del produttore di scegliere la combinazione degli input più redditizia, in risposta ad eventuali modificazioni dei prezzi sul mercato dei fattori e dei prodotti. Un’ulteriore fonte di variabilità nell’output delle imprese può derivare, in una situazione di non concorrenza perfetta, dalla presenza di differenze nei prezzi pagati o ricevuti dalle diverse unità produttive.

È necessario, a questo punto, evidenziare il problema fondamentale connesso alla stima della funzione di produzione da dati *cross-section* effettuata utilizzando tecniche “classiche” come gli OLS.

La funzione ottenuta da una tale procedura, non rappresenta, in effetti, una “vera e propria” funzione di produzione dal punto di vista della teoria economica, secondo la quale la funzione identifica una frontiera, nel senso che esprime il massimo output ottenibile da una data combinazione di input, massimo che non può essere superato da nessuna unità produttiva.

A tale proposito si instaura un dibattito teorico, a cui partecipano numerosi economisti, concernente la differenza concettuale tra la funzione di produzione, utilizzata nella teoria microeconomica neoclassica, e la funzione sti-

mata da dati *cross-section*, che può definirsi, come si vedrà più avanti, una funzione “media”.

Reder (1943) mette in risalto il fatto che solamente un punto sulla funzione di produzione si riferisce ad una situazione reale, ossia alla combinazione di fattori utilizzati e al livello di output prodotto, corrispondente alla situazione in cui l'unità produttiva massimizza il profitto. Tutti gli altri punti sono puramente teorici o ipotetici, ossia esprimono l'output che l'impresa avrebbe ottenuto se avesse trovato redditizio utilizzare altre quantità di fattori. Da qui la diversità concettuale con la funzione di produzione Cobb-Douglas stimata a partire dai dati osservati, che rappresenta combinazioni di fattori e livelli di output scelti da diverse unità produttive, come conseguenza del comportamento massimizzante.

Il suddetto Autore distingue la funzione di produzione teorica per ciascuna unità produttiva, che definisce “*intrafirm*”, dalla funzione stimata mediante l'utilizzazione di dati provenienti da processi posti in essere da diverse unità di produzione, denominata “*interfirm*”.

Bronfenbrenner (1944) sottolinea che la determinazione della funzione di produzione “*interfirm*” è logicamente e temporalmente posteriore alla determinazione dell'equilibrio, essendo il luogo dei punti che indicano per ogni impresa le quantità dei fattori utilizzati e il livello di output raggiunto attraverso la massimizzazione del profitto, dati i prezzi dei fattori produttivi sul mercato e l'offerta dei fattori stessi.

La denominazione “*interfirm*” deriva dal fatto che tale funzione costituisce l'unione dei punti di equilibrio per diverse unità produttive e rappresenta uno strumento teorico legittimo, che può essere utilizzato per la verifica empirica della teoria della distribuzione.

È chiaro, quindi, che nei primi studi che riguardano la stima della funzione di produzione si assume che l'output ottenuto dall'impresa possa essere maggiore, o minore, di quello indicato dalla funzione di produzione dell'intera industria, ossia per l'insieme delle unità considerate.

Si ipotizza, in altre parole, che la funzione da stimare sia una sorta di *funzione media* per l'industria, per cui è ragionevole ammettere che alcune unità siano in grado di produrre più della media, altre meno. Sul significato di questa “media” si sono alternate diverse interpretazioni.

Alcuni economisti si riferiscono alla funzione media come alla funzione di una unità produttiva di “dimensioni medie”, mentre altri considerano la funzione media come un legame matematico riflettente una sorta di “tecnologia media”. Un'ulteriore interpretazione vede la funzione di produzione media come la funzione esprime l'output sostenibile, ottenuta mediante l'eliminazione di fluttuazioni casuali dovute all'agire di coincidenze più o meno fortunate.

Al di là delle critiche mosse alle diverse interpretazioni del concetto di “media”, è necessario sottolineare che la funzione di produzione stimata da dati *cross-section*³ mediante gli OLS può essere validamente utilizzata solo per determinati scopi. In particolare, la funzione media potrebbe essere impiegata adeguatamente come approssimazione della funzione aggregata, oppure se l'interesse fosse sulla stima di quanto output, in media, può essere ottenuto per una unità nell'industria, con un dato insieme di input.

Per riprodurre i concetti teorici della funzione di produzione, e quindi, per ottenere un'adeguata determinazione della misura di efficienza, il termine di errore del modello econometrico dovrebbe essere caratterizzato da una forma distributiva unilaterale, o quanto meno asimmetrica. Questo tipo di specificazione identifica le cosiddette *funzioni frontiera di produzione*.

A tale riguardo, si può affermare che il lavoro di Farrell (1957) getta le basi per lo sviluppo di un nuovo approccio nello studio dell'efficienza a livello di singola impresa, in quanto focalizza l'attenzione su due elementi fondamentali: come definire il concetto di efficienza e come calcolare la tecnologia di riferimento rispetto alla quale definire misure di efficienza.

Il contributo determinante di Farrell è rappresentato dal riferimento non ad una funzione media, ma ad una *funzione empirica costruita a partire dai risultati migliori osservati nella pratica*, che si possa identificare con il concetto di frontiera di produzione. Attraverso l'ipotesi di rendimenti costanti di scala e di convessità⁴, secondo la quale anche combinazione lineari di due vettori di input possono produrre la medesima quantità di output, Farrell identifica la frontiera di produzione come la “*most pessimistic or conservative estimate of the isoquant*”.

Per quanto riguarda la stima della frontiera da un punto di vista statistico, il suddetto Autore osserva, dapprima, che esistono alcune funzioni efficienti, rispetto alle quali tutti i punti osservati deviano casualmente ma in un'unica direzione, e si riferisce poi alle analogie esistenti con la stima dei parametri di distribuzioni estreme.

L'isoquanto unitario utilizzato da Farrell può, quindi, considerarsi il precursore delle frontiere di produzione, sia nel contesto parametrico, dove riproduce il concetto economico di funzione di produzione, sia in quello non parametrico, dove descrive la superficie dell'insieme delle possibilità produttive.

³ Hoch (1955), riferendosi alla “*average firm*”, di cui modella il comportamento economico, stima la funzione di produzione Cobb-Douglas combinando dati *time-series* e *cross section* utilizzando l'analisi della covarianza.

⁴ Le proprietà della tecnologia di produzione saranno oggetto di un'analisi formale nel paragrafo 2.3.

Un primo approccio, sviluppato in letteratura per la costruzione delle frontiere di tipo parametrico, rappresenta un'evoluzione del metodo econometrico tradizionale per la stima delle funzioni di produzione ed ha origine da un suggerimento dello stesso Farrell, quando afferma: “*There exists some efficient function, from which all the observed points deviate randomly but in the same direction.*”

Il suggerimento contribuisce dapprima allo sviluppo del cosiddetto **approccio parametrico deterministico** (Aigner e Chu, 1968), che successivamente matura originando il concetto di frontiere stocastiche, attraverso le quali viene introdotto nel modello oltre al termine di errore unilaterale anche un termine bilaterale. Farrell riconosce, infatti, che: “*Errors of observation will introduce an optimistic bias, which can only be eliminated if the distributions of both errors and efficiencies are known. This is an interesting problem for any theoretical statistician.*”

In particolare, la presenza di differenze nell'*efficienza tecnica* di diverse imprese appartenenti ad una medesima industria, intesa come settore produttivo, viene ricondotta da Aigner e Chu alla diversità esistente nelle dimensioni delle imprese e alla conseguente diversa disponibilità e al diverso uso dei fattori produttivi.

Nonostante sia possibile riconoscere l'esistenza di una funzione di produzione dell'industria (*industry production function*), la cui forma è identica per un insieme di unità produttive omogenee, i parametri tecnici della funzione stessa possono variare tra le unità, in conseguenza della presenza di differenze nell'*efficienza tecnica*. Agli effetti esercitati da fattori puramente casuali, quali condizioni ambientali sfavorevoli e shock esterni, si aggiungono le diversità esistenti tra le imprese nei livelli di *efficienza economica* raggiunti, che determinano il livello di output ottenuto e, quindi, specificano una particolare funzione di produzione per ciascuna unità produttiva.

L'efficienza economica di un'impresa viene ricollegata all'abilità dell'imprenditore nell'effettuare aggiustamenti delle quantità dell'output e degli input, in conseguenza di mutamenti nei prezzi di mercato, al fine di raggiungere la massimizzazione del profitto.

L'agire dei fattori precedenti può determinare il raggiungimento da parte dell'impresa di un livello inferiore di output rispetto a quello espresso dalla funzione di produzione, ma mai di un livello superiore. In quest'ultimo caso, infatti, si sarebbe in presenza di un cambiamento della tecnologia, di cui si dovrebbe tener conto nella valutazione dell'efficienza.

Aigner e Chu, mantenendo una specificazione parametrica della frontiera di produzione, descritta, in particolare, dalla funzione Cobb-Douglas, propongono di utilizzare metodi di programmazione matematica (lineare e quadratica) per la stima della frontiera, in modo tale che i punti osservati giacciono al di sotto della frontiera stessa.

Il metodo introdotto dai suddetti Autori intende caratterizzare le differenze che possono manifestarsi nell'output tra unità produttive che impiegano identici vettori di input, e spiegare perché l'output di una data impresa possa essere inferiore a quello espresso dalla frontiera, attraverso l'introduzione esplicita di un termine di errore che esprime l'efficienza tecnica delle unità produttive.

La specificazione di una particolare forma distributiva del termine di errore, che descrive l'efficienza tecnica, si deve ad Afriat (1972), che ipotizza, in particolare, una distribuzione beta.

L'introduzione di particolari metodi di stima statistici (Førsund, Lovell e Schmidt, 1980) denominati dei minimi quadrati corretti (*Corrected Ordinary Least Squares, COLS*) e dei minimi quadrati modificati (*Modified Ordinary Least Squares, MOLS*) si deve, rispettivamente, a Richmond (1974) e a Greene (1980).

Per superare il limite principale dell'approccio parametrico deterministico, ossia l'estrema sensibilità ai valori anomali (*outliers*), Timmer (1971) propone di stimare la frontiera utilizzando le medesime tecniche di programmazione lineare di Aigner e Chu, ma consentendo ad una proporzione specificata delle osservazioni di posizionarsi al di sopra della frontiera. Anche se l'obiettivo di introdurre nel modello una certa variabilità casuale nei dati si può considerare in parte raggiunto, va osservato che, essendo la specificazione della proporzione essenzialmente arbitraria, il metodo risulta privo di esplicite giustificazioni economiche o statistiche.

Il tentativo di superare l'ulteriore limitazione del metodo della programmazione lineare, ossia l'incapacità di ottenere stime dotate delle usuali proprietà statistiche, si deve a Schmidt (1976) il quale afferma che gli stimatori ottenuti rappresentano, in realtà, stimatori di massima verosimiglianza, nel caso in cui le variabili che descrivono l'efficienza delle unità produttive seguano una distribuzione esponenziale o una distribuzione metà normale. Questa osservazione non si dimostra però sufficiente a migliorare le proprietà statistiche della frontiera di produzione, stimata con il metodo della programmazione lineare, in quanto è facilmente verificabile che le usuali condizioni di regolarità per l'applicazione della massima verosimiglianza non vengono soddisfatte.

Si può, quindi, ragionevolmente affermare che l'approccio econometrico riesce a fornire una stima della funzione di produzione che rappresenti il concetto economico di funzione frontiera e che, al contempo, consenta di ottenere stime dei parametri con le usuali proprietà statistiche, solo in conseguenza dei contributi di Aigner, Lovell e Schmidt (1977) e di Meeusen e van den Broeck (1977).

L'innovazione di tale approccio, che sarà denominato **parametrico stocastico**, consiste nell'ipotizzare che il processo di produzione sia sottoposto a due distinti disturbi casuali, economicamente distinguibili e con differenti caratteristiche.

La prima componente rappresentata il termine di inefficienza u , con distribuzione unilaterale, mentre la seconda componente v , viene descritta da una distribuzione normale e rappresenta una componente puramente aleatoria.

Per comprendere la portata e l'intensità dell'innovazione generata dal nuovo approccio nel contesto delle frontiere di produzione, è necessario ricordare che nella letteratura economica vi erano stati, in precedenza, diversi tentativi diretti a trasformare il modello di produzione in un modello stocastico, come il già citato lavoro di Marschak e Andrews (1941).

Lo studio che più si "avvicina" al concetto di frontiera di produzione stocastica, è rappresentato dal lavoro di Zellner, Kmenta e Drezé (1966), nel quale alla funzione di produzione si aggiunge un termine stocastico descrivente l'agire di fattori legati ad eventi puramente aleatori, rappresentati, ad esempio, da variazioni imprevedibili nelle *performance* dei macchinari utilizzati o da eventi meteorologici.

Le più recenti versioni della teoria neoclassica non ricorrono alla funzione di produzione, e quindi allo strumento del calcolo differenziale, ma alla teoria degli insiemi. Ciò non modifica, comunque, la concezione della teoria della produzione come allocazione delle risorse a mezzo di decisioni decentrate, dirette alla massimizzazione dell'utile netto, a un dato sistema di prezzi (Koopmans, 1951).

L'esistenza di diverse alternative di produzione, ossia di processi diversi che permettono di ottenere lo stesso prodotto, ovvero di utilizzazioni diverse degli stessi input, viene rappresentata mediante l'*insieme di produzione* \tilde{Z} , (o Z) costituito da tutti i vettori z (o da tutte le coppie di vettori (x, y)), che rappresentano, secondo la convenzione prescelta, processi di produzione possibili in una data situazione.

Le proprietà formali che caratterizzano l'insieme Z rispetto ai legami esistenti tra i diversi processi possibili, sono dirette alla semplificazione dell'analisi delle scelte di processi ottimali e alla rappresentazione analitica di caratteristiche che possono trovarsi nelle alternative tecnologiche reali (Tani, 1986).

L'insieme delle produzioni possibili, che descrive le caratteristiche della tecnologia, ossia le relazioni fra input e output, prende il posto della funzione di produzione.

Nei casi di produzione singola questa funzione può essere considerata come la rappresentazione della frontiera efficiente dell'insieme Z (Koopmans, 1951).

Le imprese devono scegliere una delle attività comprese in Z , determinando le quantità prodotte e la domanda di fattori, e fissando, in tal modo, i rapporti relativi tra le quantità impiegate dei vari input e il metodo di produzione.

Nella scelta del processo di produzione da attivare, le unità produttive sono guidate sempre dal criterio della massimizzazione del profitto, il cui raggiungimento è garantito dalla condizione di convessità di Z , attraverso la quale si assicura che, per ogni vettore di prezzi, si trovi la giusta combinazione di fattori e di prodotti.

In un tale contesto, un processo di produzione appartenente all'insieme Z è detto tecnicamente efficiente se è massimale in Z rispetto alla relazione di ordine \leq , definita tra i vettori, per cui non è possibile ridurre la quantità di alcun input senza aumentare la quantità di un altro input e/o ridurre la quantità di uno o più output; né è possibile aumentare la quantità di qualche output senza ridurre la quantità di qualche altro output o aumentare la quantità di qualche input (Tani, 1986).

Tale definizione si inserisce chiaramente nell'ottica paretiana, ripresa successivamente da Koopmans. Anche in questa impostazione, al concetto di efficienza tecnica, collegato al buon uso degli input disponibili a livello di singola unità produttiva, si accompagna il concetto di efficienza economica, connesso ad una opportuna allocazione delle risorse disponibili tra le diverse produzioni, il cui raggiungimento viene garantito dall'esistenza di un sistema di prezzi che agisce a livello aggregato.

Per lo sviluppo di metodologie che definiranno l'**approccio non parametrico** all'analisi dell'efficienza è di nuovo fondamentale il lavoro di Farrell (1957), che estende il concetto utilizzato da Pareto e Koopmans in riferimento all'intero sistema economico, agli input così come agli output di qualsiasi organizzazione produttiva ed, esplicitamente, evita qualunque uso dei prezzi e dei relativi meccanismi di scambio.

Ancora più sostanziale appare la decisione di utilizzare le *performance* delle altre unità produttive per valutare il comportamento di ogni unità in relazione all'output prodotto e agli input utilizzati.

In tal modo è possibile procedere empiricamente nel determinare l'efficienza relativa delle unità di produzione sviluppando il concetto di frontiera sia da un punto di vista econometrico, come già visto, che da un punto di vista non parametrico.

Ispirato da Debreu (1951), Farrell introduce, a livello "micro", misure di efficienza basate sulla contrazione radiale che collega i punti non efficienti osservati per le unità di produzione con i punti di riferimento (non osservati) sulla frontiera di produzione, stimata attraverso l'impiego di sistemi di equazioni lineari.

Farrell e Fieldhouse (1962), nel tentativo di generalizzare il metodo precedente al caso di rendimenti crescenti di scala, suggeriscono l'utilizzazione di problemi di programmazione lineare. Boles (1967) fornisce, successivamente, un'indicazione per la formulazione dei problemi di programmazione lineare al caso di output multipli.

Ma è il lavoro di Charnes, Cooper e Rhodes (1978) che, ispirandosi agli studi di Farrell, contribuisce in modo fondamentale allo sviluppo di strumenti

non parametrici per la determinazione della frontiera di produzione, attraverso l'introduzione di una metodologia che sarà successivamente denominata *Data Envelopment Analysis* (DEA).

I suddetti Autori introducono, per ciascun'unità produttiva, una misura di efficienza definendola come il massimo del rapporto tra una media ponderata degli output e una media ponderata degli input, che caratterizzano il processo dell'unità analizzata, con il vincolo che gli altri rapporti definibili per le restanti unità produttive siano minori o uguali ad uno. La determinazione dell'efficienza relativa di ciascuna unità viene anche formulata mediante un ordinario problema di programmazione lineare.

Nel fornire un'interpretazione della funzione di produzione che deriva dall'applicazione del metodo proposto, Charnes, Cooper e Rhodes affermano di aver introdotto un nuovo tipo di funzione di produzione che vagamente si ricollega al concetto di impresa rappresentativa, utilizzato tempo prima da Marshall.

Quest'ultimo Autore proponeva di caratterizzare le possibilità produttive attraverso la media degli output ottenuti da dati input e di definire l'utilizzatore di questa relazione media come l'unità rappresentativa. Eventuali variazioni tra i costi sostenuti dalle imprese per l'ottenimento di una data quantità di output erano riconducibili all'età dell'impresa e ad eventuali differenze esistenti nelle capacità imprenditoriali.

Nell'approccio suggerito da Charnes, Cooper e Rhodes, l'ottimo rispetto al quale determinare l'efficienza di ciascuna unità produttiva è costituito, piuttosto che da imprese rappresentative "medie", da imprese rappresentative efficienti.

Questo insieme di metodologie non parametriche presenta un carattere puramente deterministico, perciò, in genere, i disturbi casuali sono inclusi nella misura di efficienza. D'altra parte, nell'approccio econometrico, dove si impone una struttura parametrica sia alla tecnologia, descritta attraverso una funzione di produzione, sia alla distribuzione del termine di efficienza, si può verificare una commistione tra errori di specificazione ed efficienza (Ferrier e Lovell, 1990).

Se da una parte, quindi, l'approccio non parametrico, assicura una maggiore flessibilità nella descrizione della tecnologia e nella misurazione dell'efficienza tecnica, dall'altra, il non tener conto di una certa variabilità stocastica, determina un'eccessiva dipendenza dei risultati dai dati osservati, che si traduce in una forte sensibilità ai valori anomali, come nel caso dell'approccio deterministico, sebbene recenti sviluppi suggeriscano l'introduzione di un approccio stocastico alla DEA (Simar e Wilson, 2000, Cazals, Florens e Simar, 2002).

Capitolo Secondo

LA TECNOLOGIA DI PRODUZIONE

2.1 PREMESSA

L'efficienza produttiva rappresenta un'importante caratteristica della *performance* del produttore che, come sottolineato nel capitolo precedente, è stata ingiustamente trascurata nella letteratura economica.

Il più volte citato lavoro pionieristico di Farrell (1957) ha determinato un progressivo sviluppo delle tematiche connesse al concetto di efficienza produttiva e alle conseguenze che il suo riconoscimento avrebbe potuto determinare nella formulazione delle metodologie di rappresentazione del processo produttivo. Gli studi hanno interessato sia gli aspetti teorici sia quelli applicativi.

Come già anticipato, definendo l'efficienza produttiva come l'abilità di una organizzazione di produrre un ben specificato output al minimo costo, emerge la presenza di una componente tecnica, riflettente l'efficienza nella trasformazione fisica degli input in output, e di una componente allocativa, che esprime l'efficienza di tipo economico nell'allocazione ottima dei fattori.

Non escludendo la possibilità di interazioni tra le due misure di efficienza produttiva, derivanti dalle complesse azioni decisionali poste in essere dagli agenti economici in relazione allo svolgimento della produzione, Farrell introduce una misura dell'efficienza allocativa indipendente dall'efficienza tecnica.

Le misure di efficienza, essendo basate su uno standard osservato, ossia determinato dalle organizzazioni produttive che condividono una tecnologia comune, rappresentano misure relative, nel senso che ciascuna unità produttiva è valutata in riferimento alle migliori *performance* manifestate dall'insieme di unità produttive che presentano caratteristiche simili a quella in esame.

Se le *performance* del gruppo di riferimento cambiano, conseguentemente si modificano anche le misure di efficienza. Come afferma Kopp (1981), l'adozio-

ne da parte di Farrell di un isoquanto unitario, rispetto al quale determinare la distanza di ciascun processo osservato, implica che la tecnologia presenti rendimenti di scala costanti.

Dallo studio attento del lavoro di Farrell, si evidenziano, quindi, gli elementi e i concetti fondamentali che intervengono nell'analisi dell'efficienza rispetto alla frontiera delle possibilità produttive.

L'elemento principale è rappresentato dalla tecnologia di produzione, condivisa dalle unità produttive costituenti il gruppo di riferimento, e dalla sua conseguente rappresentazione, che può avvenire attraverso l'impiego di diversi strumenti, quali, ad esempio, la funzione di produzione e l'insieme delle possibilità produttive.

Alcune rappresentazioni, definite *rappresentazioni primali*, fanno un esplicito e diretto riferimento ai processi di produzione, mentre altre presuppongono la soluzione di un problema di scelta, e, più precisamente, di un problema di ottimizzazione vincolata.

Queste ultime vengono denominate *rappresentazioni duali*, in quanto originate dalla teoria della dualità (Shephard, 1970), e comprendono la funzione di costo (minimo valore degli input necessari a produrre dati output, a dati prezzi degli input medesimi), la funzione di profitto, ristretta e generale (massimo valore attribuibile a dati prezzi degli output con un dato ammontare di input), e la funzione indiretta di produzione (massimo output ottenibile a un dato costo, dati i prezzi degli input).

L'analisi della tecnologia di produzione, che sarà sviluppata nelle pagine seguenti, riguarda lo studio delle sole *relazioni tecniche* che formano il sistema dei vincoli al quale è sottoposto lo svolgimento dell'attività produttiva. Ovviamente, la realizzazione dell'efficienza tecnica a livello delle singole unità è condizione necessaria ma non sufficiente perché si realizzi un'utilizzazione efficiente delle risorse a livello di intero sistema produttivo. Ipotizzando che la scelta degli input avvenga a livello di singola unità secondo il criterio di massimizzazione del risultato netto, o di minimizzazione del costo di produzione della specificata quantità di prodotto, dato un certo vettore di prezzi, si può affermare che, in assenza di esternalità, è possibile raggiungere la piena efficienza produttiva dell'intero sistema. Tale risultato, come già anticipato in precedenza, rappresenta l'essenza della teoria neoclassica della produzione.

Dalla rappresentazione della tecnologia attraverso l'insieme di produzione, utilizzata nei modelli di equilibrio economico generale e sviluppata come generalizzazione dell'impostazione in termini di analisi delle attività, si possono derivare logicamente le altre, anche se storicamente nate prima.

Alcune rappresentazioni prendono in considerazione tutti i processi possibili, come la *famiglia degli insiemi di fabbisogno di input* o la *funzione di distanza*;

altre, invece, fanno riferimento ai soli processi efficienti, come la *frontiera dell'insieme di produzione*, la *funzione di produzione* e la *funzione di trasformazione*.

La rappresentazione della tecnologia coinvolge anche la definizione di specifiche proprietà, attraverso le quali si descrivono le caratteristiche della tecnologia stessa.

Se a partire dalla rappresentazione della tecnologia in termini di insieme di produzione, o di insieme di fabbisogno di input o di funzione di produzione, è possibile, impostando il problema della minimizzazione del costo, ottenere la funzione di costo, dalla dualità esistente tra funzione di distanza e funzione di costo è possibile, seguendo Farrell (1957), individuare non solo il grado di efficienza tecnica dell'unità produttiva che adotta il processo (x, y) ma anche il grado di efficienza della stessa rispetto ai prezzi degli input utilizzati (*efficienza allocativa*).

Sebbene la funzione di costo e quella di profitto rappresentino due concetti di estremo interesse, particolare attenzione sarà dedicata, nel corso del presente lavoro, alle problematiche connesse alla misurazione della sola efficienza tecnica, e quindi alla rappresentazione della tecnologia e alla selezione dei metodi parametrici e non parametrici per l'individuazione della frontiera delle possibilità produttive.

2.2 LA RAPPRESENTAZIONE DELLA TECNOLOGIA DI PRODUZIONE: L'INSIEME DI PRODUZIONE E L'INSIEME DI FABBISOGNO DI INPUT

La più generale delle rappresentazioni che fanno esplicito riferimento ai processi produttivi, posti in essere da unità produttive che impiegano un vettore di input $x = (x_1, x_2, \dots, x_s) \in R_+^s$ per ottenere un vettore di output $y = (y_1, y_2, \dots, y_m) \in R_+^m$, è costituita certamente dall'**insieme di produzione** o **insieme delle possibilità di produzione** (*production possibilities set*) dell'impresa (Varian, 1978) che, come già detto, rappresenta l'insieme di tutti i processi o piani di produzione (x, y) tecnicamente realizzabili.

Tale insieme, indicato, come già anticipato, utilizzando il simbolo Z , è un sottoinsieme di $R_+^s \times R_+^m$, e fornisce una descrizione completa della tecnologia di un'impresa. Nel caso in cui per definire l'insieme di produzione vengano utilizzati i vettori¹ z di ordine r , l'insieme di produzione, denominato \tilde{Z} , rappresenta un sottoinsieme R^r .

¹ Come già anticipato, gli elementi negativi del vettore z misurano gli input, mentre gli elementi positivi specificano gli output del processo (Tani, 1986).

L'impostazione in termini di insiemi di fabbisogno di input, nata all'interno dell'approccio duale per fornire una trattazione di carattere generale, ma maggiormente finalizzata alle applicazioni (Tani, 1989), considera un tipo particolare di insieme delle possibilità di produzione ristretto (Varian, 1978) definito **insieme di fabbisogno di input** (*input requirement set*) relativo al vettore \mathbf{y} di output, costituito dall'insieme dei vettori \mathbf{x} di input che consentono di produrre il vettore \mathbf{y} di output.

In termini formali si ha:

$$L(\mathbf{y}) = \{ \mathbf{x} \mid \mathbf{x} \in R_+^s, (\mathbf{x}, \mathbf{y}) \in Z \}$$

La Figura 2.1 rappresenta un insieme composto da tre diverse combinazioni \mathbf{x}^A , \mathbf{x}^B , e \mathbf{x}^C dei due input x_1 e x_2 che consentono di produrre il livello \mathbf{y} di output.

Si può considerare anche la rappresentazione della tecnologia basata sull'**insieme degli output** o **insieme degli output producibili** (*output set*), $P(\mathbf{x})$, che contiene tutti i vettori \mathbf{y} di output producibili a partire dal vettore di input \mathbf{x} , ossia, in termini formali:

$$P(\mathbf{x}) = \{ \mathbf{y} \mid \mathbf{y} \in R_+^m, (\mathbf{x}, \mathbf{y}) \in Z \}$$

I legami esistenti tra le tre precedenti rappresentazioni sono espressi, in primo luogo, dall'**input** e dall'**output corrispondenza**, definite rispettivamente da:

$$L: R_+^m \rightarrow L(\mathbf{y}) = \{ \mathbf{x} \mid \mathbf{x} \in R_+^s, (\mathbf{x}, \mathbf{y}) \in Z \}$$

$$P: R_+^s \rightarrow P(\mathbf{x}) = \{ \mathbf{y} \mid \mathbf{y} \in R_+^m, (\mathbf{x}, \mathbf{y}) \in Z \}$$

e dalle espressioni seguenti attraverso le quali si può ottenere l'insieme di produzione a partire da $L(\mathbf{y})$ o da $P(\mathbf{x})$, ossia:

$$Z = \{ (\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in L(\mathbf{y}), \mathbf{y} \in R_+^m \}$$

$$Z = \{ (\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in P(\mathbf{x}), \mathbf{x} \in R_+^s \}$$

L'equivalenza delle tre rappresentazioni della tecnologia di produzione può, inoltre, essere illustrata graficamente (Färe e Grosskopf, 1996).

Nella Figura 2.2, l'insieme di produzione è rappresentato dai vettori input-output che risultano delimitati dalla linea continua e dall'asse delle ascisse.

L'insieme $P(\mathbf{x})$ associato ad un dato input x^0 è uguale a $P(x^0) = [0, y^0]$ mentre l'insieme $L(\mathbf{y})$ associato a un dato output y^0 è espresso da $L(y^0) = [x^0, +\infty)$.

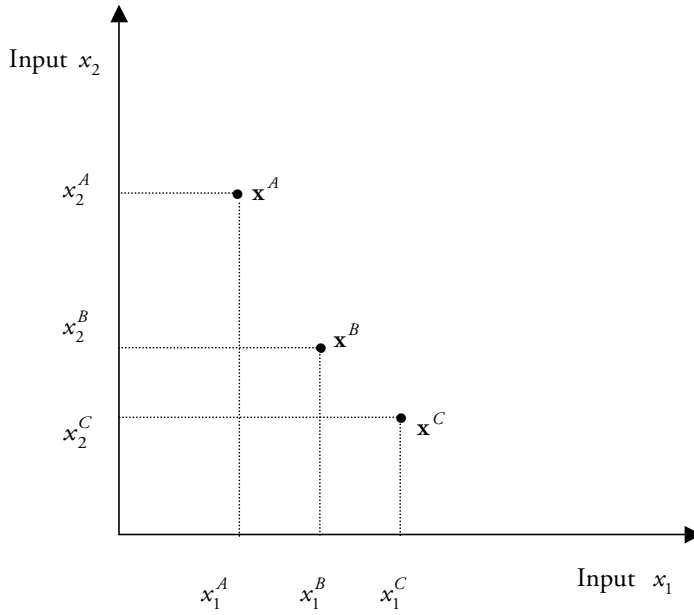


Figura 2.1 – Insieme di fabbisogno di input.

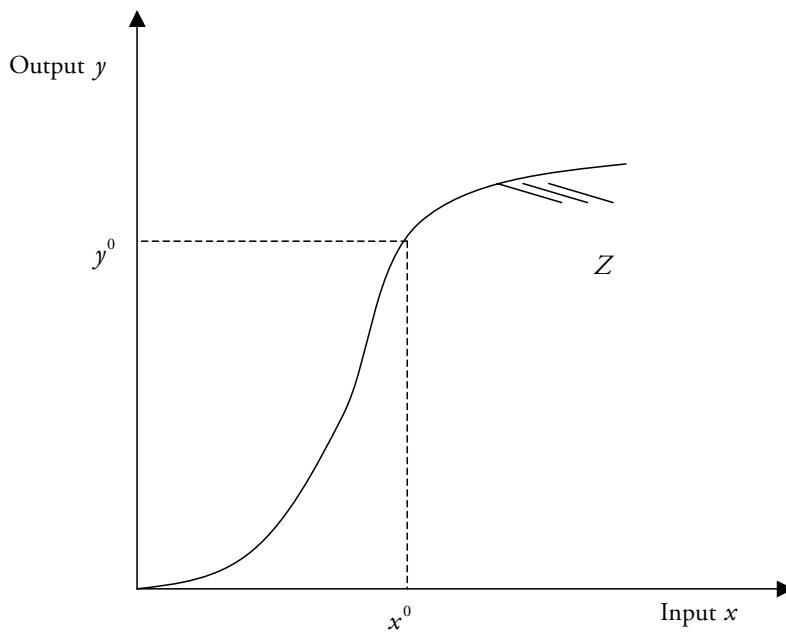


Figura 2.2 – Rappresentazioni della tecnologia di produzione.

Al concetto di insieme di fabbisogno di input è collegato quello di **isoquanto**, di più lunga e consolidata tradizione, che, in termini formali, è definito dalla seguente espressione:

$$IsoqL(y) = \left\{ x \mid x \in L(y), \lambda x \notin L(y), \forall \lambda \in [0,1) \right\}$$

In altre parole, appartiene all'isoquante relativo al vettore y di output ogni vettore di input che consenta di produrre y e sia minimo rispetto ad una riduzione radiale.

Sebbene la definizione tradizionale faccia riferimento alla funzione di produzione, in generale si può affermare che ad un isoquante appartengono tutte le combinazioni di input che consentono di produrre esattamente y unità di output.

Un ulteriore sottoinsieme dell'insieme di fabbisogno di input, che, unitamente all'isoquante, risulta di particolare importanza nella misurazione dell'efficienza, è costituito dal cosiddetto **sottoinsieme efficiente**, espresso da:

$$EffL(y) = \left\{ x \mid x \in L(y), x' \leq x \Rightarrow x' \notin L(y) \right\}$$

È evidente dalle espressioni precedenti che il sottoinsieme efficiente è incluso nell'isoquante, in quanto la definizione di sottoinsieme efficiente coinvolge la relazione d'ordine tra due vettori, mentre la definizione di isoquante esprime una riduzione radiale del vettore considerato.

Dall'esame della Figura 2.3, dove, nel caso semplificato di due soli input, sono rappresentati sia l'isoquante che il sottoinsieme efficiente dell'insieme di fabbisogno di input, emergono chiaramente le caratteristiche distintive dei due insiemi considerati.

In riferimento a due piani di produzione x^A e x^B , si osserva che, sebbene entrambi i processi siano posizionati sull'isoquante dell'insieme di fabbisogno di input, il processo rappresentato dal punto x^A è caratterizzato da un uso "eccessivo" dell'input x_2 . Si può, quindi, affermare che quest'ultimo processo produttivo appartiene all'isoquante ma non al sottoinsieme efficiente dell'insieme $L(y)$.

In modo analogo, è possibile definire l'isoquante dell'insieme degli output producibili, utilizzando l'espressione seguente (Deprins e Simar, 1983):

$$IsoqP(x) = \left\{ y \mid y \in P(x), \delta y \notin P(x), \forall \delta \in [1, \infty) \right\}$$

Similmente, il sottoinsieme efficiente di $P(x)$ è essere espresso da:

$$EffP(x) = \left\{ y \mid y \in P(x), y' \geq y \Rightarrow y' \notin P(x) \right\}$$

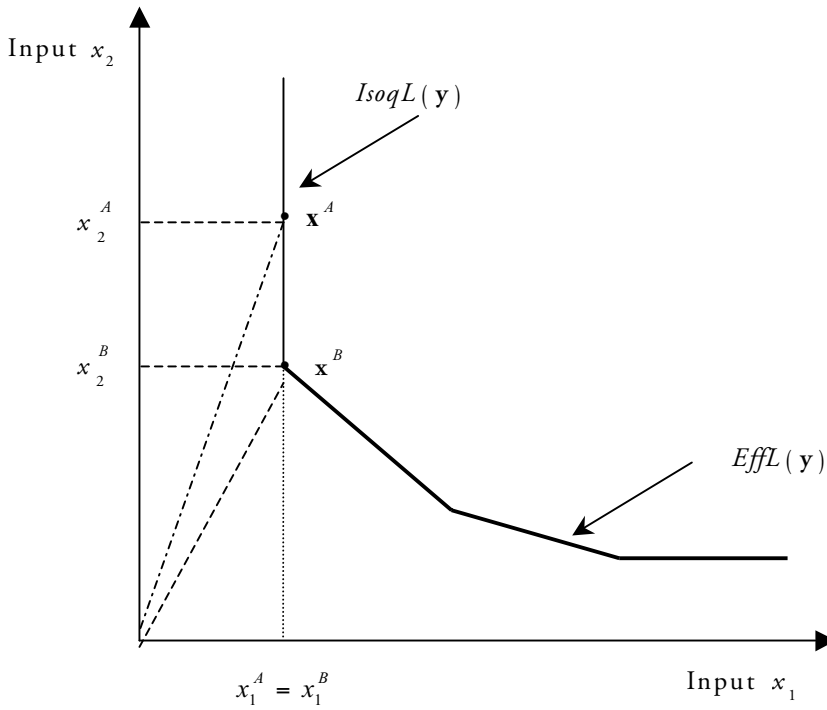


Figura 2.3 – Isoquante e Sottoinsieme efficiente di $L(x)$.

2.3 LE CARATTERISTICHE STRUTTURALI DELLA TECNOLOGIA DI PRODUZIONE

2.3.1 La descrizione della tecnologia di produzione

La descrizione della tecnologia rappresenta il passo fondamentale per lo sviluppo di una procedura per la misurazione dell'efficienza.

Le molteplici proprietà, che specificano le caratteristiche principali della tecnologia, possono essere riferite a diverse rappresentazioni formali.

Nell'analisi empirica possono sorgere particolari condizioni e determinati obiettivi da soddisfare, che intervengono nella specificazione di un modello di produzione e quindi nell'individuazione di una particolare tecnologia.

La scelta di una particolare rappresentazione della tecnologia, che soddisfa specifiche proprietà, individuandone la tipologia, ossia, ad esempio, *regolare*, *monotona* e *convessa*, condiziona fortemente la successiva individuazione delle

metodologie utilizzabili per l'identificazione della frontiera e, conseguentemente, delle misure di efficienza che ne derivano.

2.3.2 Tecnologie regolari

Tra le principali proprietà che possono essere attribuite all'insieme di produzione, si possono considerare, dapprima, tre proprietà che specificano la condizione di **regolarità** della tecnologia.

La proprietà di *chiusura* di \tilde{Z} in R^r è soddisfatta nel caso in cui l'insieme di produzione sia costituito da un numero finito di processi, o da un numero infinito di processi isolati.

L'ipotesi della *produttività* richiede l'esistenza di almeno un processo, che impiega qualche input e dia luogo ad almeno un output positivo, e viene espressa imponendo che in \tilde{Z} esista almeno un elemento non confrontabile con il vettore nullo.

La cosiddetta ipotesi di *inesistenza della terra di cuccagna* (Tani, 1986) è riferita all'insieme di produzione \tilde{Z} attraverso: $\mathbf{z} \in \tilde{Z}, \mathbf{z} \geq 0 \Rightarrow \mathbf{z} = 0$. In tal modo si impone che la produzione di un output positivo richieda sempre l'impiego di qualche input, ossia che non si possa produrre qualcosa dal nulla.

Proprietà equivalenti alle precedenti, ma in un certo senso più deboli², possono essere specificate in relazione all'insieme di fabbisogno di input.

Una **tecnologia** è **input-regolare** se l'insieme $L(\mathbf{y})$ è **regolare**, ossia se, $\forall \mathbf{y} \geq 0$ si ha: i) $L(\mathbf{y}) \neq \emptyset$; ii) $L(\mathbf{y})$ è un insieme chiuso; iii) se $\mathbf{0} \in L(\mathbf{y}) \Rightarrow \mathbf{y} = \mathbf{0}$.

L'ipotesi che $L(\mathbf{y})$ non sia un insieme vuoto richiede che sia sempre possibile produrre un qualsiasi livello di output. Inoltre, come osserva Varian (1978), l'ipotesi che l'insieme di fabbisogno di input sia chiuso viene assunta per ragioni tecniche ed è irrilevante nella maggior parte dei casi.

Un'implicazione della suddetta ipotesi è racchiusa nella seguente considerazione. Se si suppone che \mathbf{x}^i sia una successione di combinazioni di input, ciascuna delle quali consente di produrre \mathbf{y} , e che questa successione converga alla combinazione di input \mathbf{x}^0 , allora l'ipotesi di chiusura assicura che la combinazione limite dovrà consentire di produrre \mathbf{y} .

In altri termini, l'insieme di fabbisogno di input deve contenere la sua frontiera.

² Va osservato, infatti, che mentre le proprietà di chiusura, produttività ed inesistenza della terra di cuccagna dell'insieme Z implicano che l'insieme sia regolare, l'insieme Z di una tecnologia input-regolare può non essere chiuso.

2.3.3 Tecnologie monotone

Nelle applicazioni empiriche si assume, generalmente, che la tecnologia sia **monotona**, ossia che aumentando la quantità impiegata di uno o più input sia possibile produrre una quantità di output almeno uguale a quella prodotta inizialmente (Varian, 1999).

In altre parole, se è possibile disporre liberamente dei beni, si può ragionevolmente ritenere che, in relazione ad una combinazione di input \mathbf{x} , che consente di produrre il vettore di output \mathbf{y} , è sempre possibile individuare una combinazione di input \mathbf{x}' , in cui ciascuna componente rappresenta una quantità di input superiore o al limite uguale alla componente corrispondente nel vettore \mathbf{x} , attraverso la quale produrre la medesima quantità di output \mathbf{y} .

Questa caratteristica della tecnologia viene descritta attraverso la cosiddetta proprietà di *possibilità di distruzione gratuita dei beni* o, per utilizzare la terminologia anglosassone, *free disposal*.

In riferimento all'insieme di produzione \tilde{Z} , la *free disposal* può essere espressa da: $\mathbf{z} \in \tilde{Z}, \mathbf{z}' \leq \mathbf{z} \Rightarrow \mathbf{z}' \in \tilde{Z}$.

La proprietà afferma che, se un processo di produzione è possibile, allora lo sono anche tutti quei processi che impiegano quantità di input non minori per ottenere quantità non superiori di output. Questa proprietà si basa sull'ipotesi che ci si possa disfare senza difficoltà, ossia senza sostenere costi aggiuntivi, dei beni in eccesso, rappresentati da input oppure da output del processo.

Una tale caratteristica della tecnologia può essere espressa attraverso due distinte proprietà, una relativa agli input e una relativa agli output. In particolare, la *possibilità di distruzione gratuita degli input* (*free disposal* o *strong disposability of input*) può essere espressa in termini di *monotonicità* dell'insieme di fabbisogno di input, ossia, in termini formali: $\mathbf{x} \in L(\mathbf{y}), \mathbf{x}' \geq \mathbf{x} \Rightarrow \mathbf{x}' \in L(\mathbf{y})$.

La Figura 2.4 riproduce l'insieme di fabbisogno di input, raffigurato in precedenza nella Figura 2.1, nel caso in cui si introduca l'ipotesi di *monotonicità*. L'insieme $L(\mathbf{y})$ comprende ora anche tutte le combinazioni di input posizionati nell'area grigia a destra dei vettori \mathbf{x}^A , \mathbf{x}^B , e \mathbf{x}^C .

La possibilità di distruzione gratuita degli output (*strong disposability of output*) può essere definita (Färe, Grosskopf e Lovell, 1985) mediante la seguente espressione:

$$\mathbf{y}' \leq \mathbf{y} \in P(\mathbf{x}) \Rightarrow \mathbf{y}' \in P(\mathbf{x})$$

L'ipotesi di monotonicità è frequentemente assunta nella costruzione di frontiere di produzione, sebbene, in ambito non parametrico, Färe e Lovell (1978) e Färe e Grosskopf (1983a) sviluppano analisi di efficienza rispetto a tecnologie in cui la disposizione o disponibilità degli input e degli output non è libera. Un

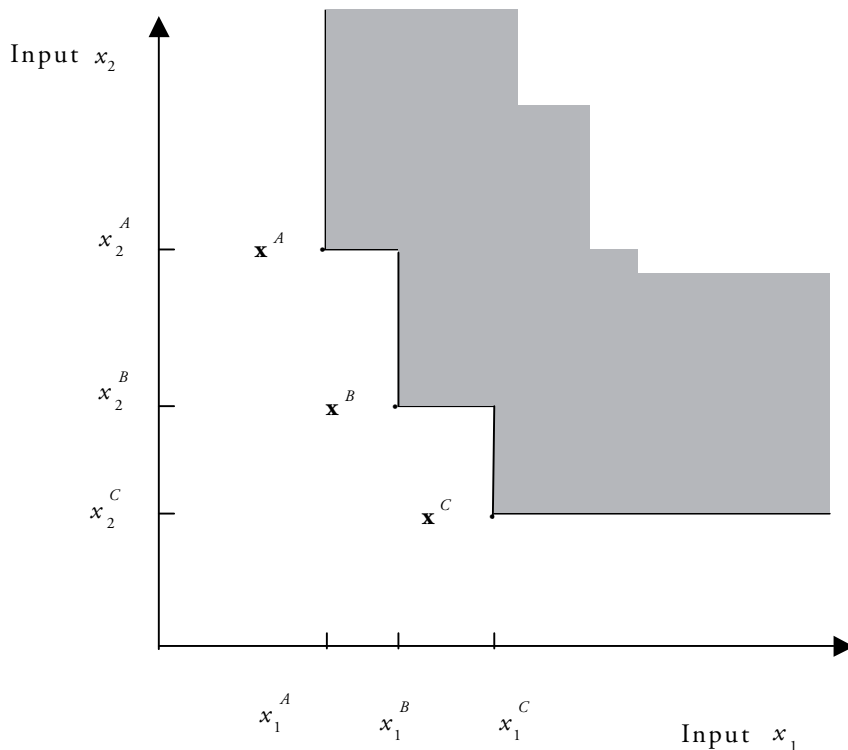


Figura 2.4 – Insieme di fabbisogno di input con ipotesi di *monotonicità*.

simile tentativo di estendere l'analisi anche ad insiemi di fabbisogno di input non monotoni si trova anche nel contesto dell'analisi parametrica in Borts e Mishan (1962).

2.3.4 Tecnologie convesse

Nella moderna teoria economica della produzione, il riferimento a tecnologie **convesse**, ossia caratterizzate dall'assunzione dell'ipotesi di *convessità* dell'insieme di produzione e dell'insieme di fabbisogno di input, rappresenta una delle condizioni necessarie per raggiungere un'allocazione delle risorse che sia efficiente in senso paretiano.

La stretta convessità di \tilde{Z} comporta l'unicità del punto di massimo profitto per ogni vettore di prezzi e rappresenta il caso di una tecnologia con rendimenti decrescenti ovunque.

L'ipotesi di convessità può, quindi, essere introdotta in riferimento all'insieme di produzione, definendo \tilde{Z} come un **insieme convesso**, ossia assumendo che:

$$z, z' \in \tilde{Z} \Rightarrow (\lambda z + (1 - \lambda)z') \in \tilde{Z} \quad \forall \lambda \in [0, 1]$$

Va osservato che la convessità dell'insieme di produzione implica la convessità dell'insieme di fabbisogno di input e dell'insieme degli output.

Nel primo caso, si ha che l'insieme $L(y)$ è convesso se:

$$x', x \in L(y) \Rightarrow (\lambda x + (1 - \lambda)x') \in L(y) \quad \forall \lambda \in [0, 1]$$

Come emerge dalla Figura 2.5, che estende la rappresentazione di un insieme di fabbisogno di input monotono della Figura 2.4, ipotizzare che \tilde{Z} sia convesso significa affermare che, se x^A e x^B consentono di produrre y unità di output, allora una qualsiasi media ponderata dei due, $(\lambda x^A + (1 - \lambda)x^B)$, consentirà di produrre la medesima quantità di output.

Nel secondo caso, l'insieme $P(x)$ è convesso se:

$$y^0, y \in P(x) \Rightarrow (\theta y + (1 - \theta)y^0) \in P(x) \quad \forall \lambda \in [0, 1]$$

Le precedenti proprietà di convessità degli insiemi $L(y)$ e $P(x)$ sono indipendenti, ossia l'una non implica l'altra.

Della proprietà in esame può essere introdotta anche una versione forte, denominata *convessità stretta*. In riferimento, ad esempio, all'insieme di produzione si ha:

$$z, z' \in \tilde{Z}, z \neq z', \lambda \in (0, 1) \Rightarrow \exists z^0 \in \tilde{Z} : z^0 \geq \lambda z + (1 - \lambda)z'$$

L'ipotesi di stretta convessità è generalmente introdotta per garantire unicità di soluzione a problemi di ottimizzazione vincolata.

Del pari, l'isoquanto dell'insieme di fabbisogno di input si definisce *convesso* quando:

$$x', x \in IsoqL(y) \Rightarrow \lambda x + (1 - \lambda)x' \in L(y), \quad \forall \lambda \in [0, 1]$$

e *strettamente convesso* quando oltre ad essere convesso soddisfa la seguente proprietà:

$$x', x \in IsoqL(y) \Rightarrow \lambda x + (1 - \lambda)x' \notin IsoqL(y), \quad \forall \lambda \in (0, 1)$$

L'ipotesi di convessità è sottesa alla stragrande maggioranza delle rappresentazioni della tecnologia di riferimento per la costruzione della frontiera di produzione.

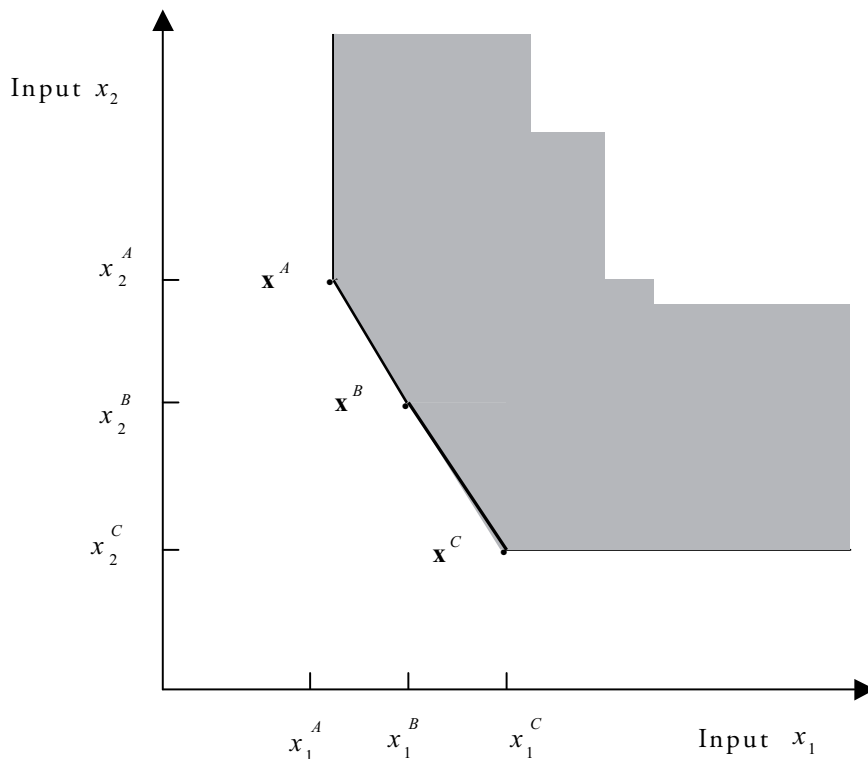


Figura 2.5 – Insieme di fabbisogno di input convesso.

2.4 IL MODELLO DI ANALISI DELLE ATTIVITÀ

Nel modello di rappresentazione della tecnologia di produzione denominato **analisi delle attività** (*activity analysis*), introdotto da von Neumann (1937) e sviluppato successivamente dai neo-walrasiani, in particolare da Koopmans (1951, 1953, 1987), il produttore può scegliere tra un numero ridotto e finito di tecniche di produzione o attività. L'intento, come sottolinea Tani (1986), è quello di fornire uno schema più realistico e finalizzato anche alle applicazioni, uno schema che consenta, in particolare, l'utilizzazione dei metodi di programmazione lineare in alternativa ai più tradizionali metodi di calcolo differenziale nei problemi di ottimizzazione vincolata.

Koopmans stesso, nell'Introduzione ad *Activity Analysis of Production and Allocation*, riconosce l'enorme influenza esercitata sul piano metodologico dagli studi relativi alla programmazione lineare, condotti durante la seconda guerra mondiale.

Sul piano economico però, come emerso in precedenza, il modello di analisi delle attività si sviluppa a partire dall'insoddisfazione esistente in relazione alla rappresentazione analitica della tecnologia produttiva, basata sulla "tradizionale" funzione di produzione.

Sebbene Koopmans non intenda rifiutare del tutto il concetto di funzione di produzione, di cui farà largo impiego in altri contesti, l'Autore comprende la necessità di presentare un'analisi più generale, basata sulla rappresentazione per processi, che possa anche offrire alla stessa funzione di produzione più solide fondamenta.

Riprendendo le parole di Koopmans (1953) "...*Activity analysis is concerned with the construction of conceptual models to study and appraise criteria, rules, and practices for the allocation of resources...*"

In un tale contesto assume una rilevanza fondamentale la definizione del concetto di base della tecnologia, ossia dell'*attività* o processo elementare, in termini di una combinazione di merci qualitativamente specificate, in rapporti quantitativi fissi, come input utilizzati per produrre altre merci, espresse dagli output, in rapporti quantitativamente determinati rispetto agli input.

Nel caso di attività caratterizzate da due input che producono un solo output è possibile procedere ad una rappresentazione grafica nello spazio a due dimensioni. Nella Figura 2.6 sono rappresentate tre possibili attività mediante i tre raggi uscenti dall'origine, A_1 , A_2 e A_3 , denominati, raggio di attività o raggio del processo produttivo (*activity ray* o *production process ray*), la cui pendenza caratterizza la proporzione fissa dei due fattori produttivi utilizzati.

Nei modelli di analisi delle attività, quindi, il produttore può scegliere di utilizzare i processi base oppure processi misti, ossia tutti quelli che risultano da una combinazione convessa dei processi base. Dalla Figura 2.6 emerge, ad esempio, che per produrre l'output y in modo efficiente, si può utilizzare una qualsiasi combinazione delle attività A_2 e A_3 o delle attività A_1 e A_2 che appartengono all'isoquanto $a_1 a_2 a_3$.

Invece, le combinazioni di A_1 e A_3 sono inefficienti, in quanto, da essi si otterrebbe un'attività che non giace sull'isoquanto, che comporta, quindi, l'impiego di una quantità maggiore di input per produrre il medesimo livello di output (Koopmans, 1951).

Il criterio di efficienza adottato è quello paretiano per cui un processo è efficiente se non esiste nessun altro processo nell'insieme di produzione con cui è possibile produrre una quantità maggiore di almeno un output con gli stessi input, o la stessa quantità di output con minori input.

Nella descrizione della tecnologia, Koopmans segue un'impostazione assiomatica introducendo alcune proprietà che si riveleranno di grandissima

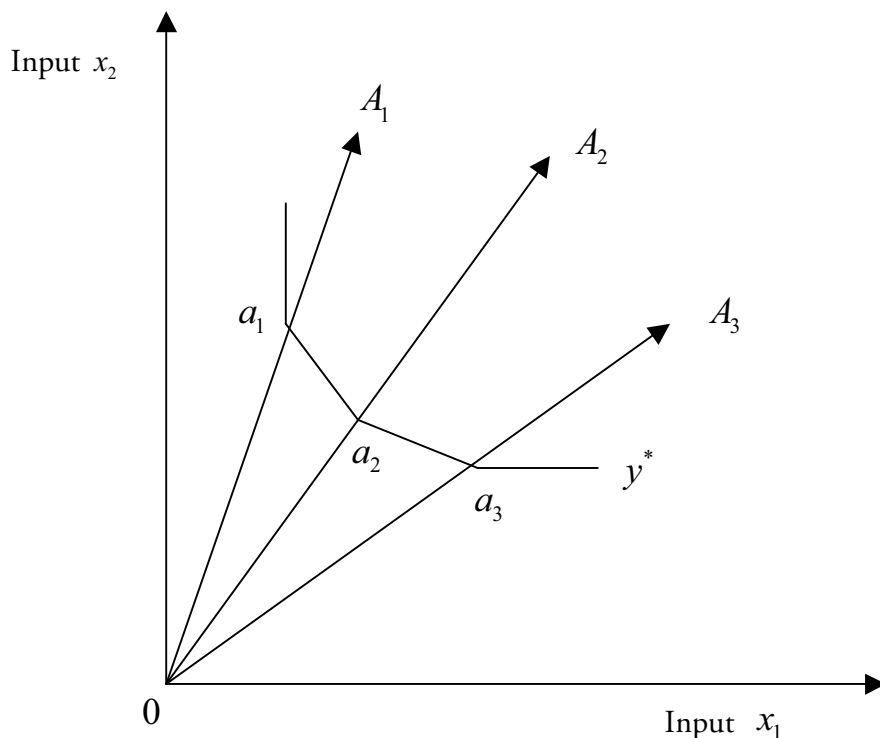


Figura 2.6 – Isoquanti nel modello di analisi delle attività.

importanza, quale l'assioma di irreversibilità, secondo cui non è possibile trovare un insieme di livelli possibili di attività il cui effetto congiunto sia un output netto nullo per tutte le merci, e l'assioma di inesistenza della Terra di Cuccagna.

L'ipotesi di convessità e di rendimenti costanti di scala, introdotte da Koopmans per l'individuazione dell'insieme dei punti efficienti, sebbene siano alquanto restrittive, possono facilmente essere rimosse o generalizzate, per cui non incidono sull'estrema flessibilità che caratterizza il metodo proposto.

Introducendo una notazione più generale (Färe e Grosskopf, 1996), si possono considerare K attività o processi produttivi ($k= 1, \dots, K$). Ciascuna attività è definita attraverso un insieme di coefficienti w_k , rappresentati da numeri reali non negativi, che indicano l'intensità con la quale una particolare attività è coinvolta nella produzione degli output.

Il modello base, espresso in termini di insieme delle possibilità produttive $P(\mathbf{x})$ è il seguente:

$$P(\mathbf{x}) = \left\{ \begin{array}{l} (y_1, y_2, \dots, y_M) : y_m \leq \sum_{k=1}^K w_k y_{km}, \quad m = 1, \dots, M \\ \sum_{k=1}^K w_k y_{km} \leq x_{ks}, \quad s = 1, \dots, S \\ w_k \geq 0, \quad k = 1, \dots, K \end{array} \right\}$$

Questa tecnologia soddisfa le proprietà di chiusura, di inesistenza della terra di Cuccagna e della convessità e presenta, infine, rendimenti costanti di scala.

Si può, infine, sottolineare che il modello di analisi delle attività si è rivelato particolarmente utile nella misurazione dell'efficienza nel contesto dell'approccio non parametrico, ed, in particolare, il suo uso caratterizza l'impostazione metodologica seguita in molti studi (ad esempio, in Färe, Grosskopf e Lovell, 1985).

2.5 LA RAPPRESENTAZIONE DELLA TECNOLOGIA MEDIANTE LA FUNZIONE DI PRODUZIONE

La tecnologia può essere rappresentata tramite una funzione caratterizzata da un numero limitato di parametri. Varian (1978) osserva che le rappresentazioni parametriche della tecnologia non dovrebbero essere considerate delle riproduzioni fedeli delle possibilità di produzione, piuttosto dovrebbero descrivere, da un punto di vista ingegneristico, i piani di produzione che sono fisicamente possibili.

La **funzione di produzione**, che costituisce una rappresentazione della tecnologia che privilegia i processi tecnicamente efficienti trascurando quasi tutti gli altri, ha ricoperto, come si è già accennato, un posto centrale nella letteratura sulla teoria della produzione. Considerando il caso di produzione singola, per ogni possibile vettore di input è identificata la massima quantità di output producibile. Questo legame funzionale prende il nome di funzione di produzione.

A partire da un insieme di produzione, nel quale i processi sono espressi nella forma (\mathbf{x}, y) , si definisce una funzione $f: R_+^s \rightarrow R_+$ nel modo seguente:

$$\forall \mathbf{x} \in R_+^s : f(\mathbf{x}) = \max \{ y \mid \mathbf{x} \in L(y) \}$$

Al fine di garantire che la funzione risulti ben definita, ovvero che il problema di massimo ammetta sempre una soluzione per ogni $\mathbf{x} \in \mathcal{R}_+^s$, l'insieme di produzione deve godere delle proprietà di chiusura, limitazione superiore degli input, possibilità di inazione e *free disposal* degli input.

È importante sottolineare che la funzione di produzione, descrivendo solo le relazioni tecnologiche output efficienti, rappresenta il più alto livello tecnologicamente possibile di y , per il vettore di fattori \mathbf{x} . Una rappresentazione grafica per una funzione di produzione generale è riportata nella Figura 2.7 nel caso di due soli input³.

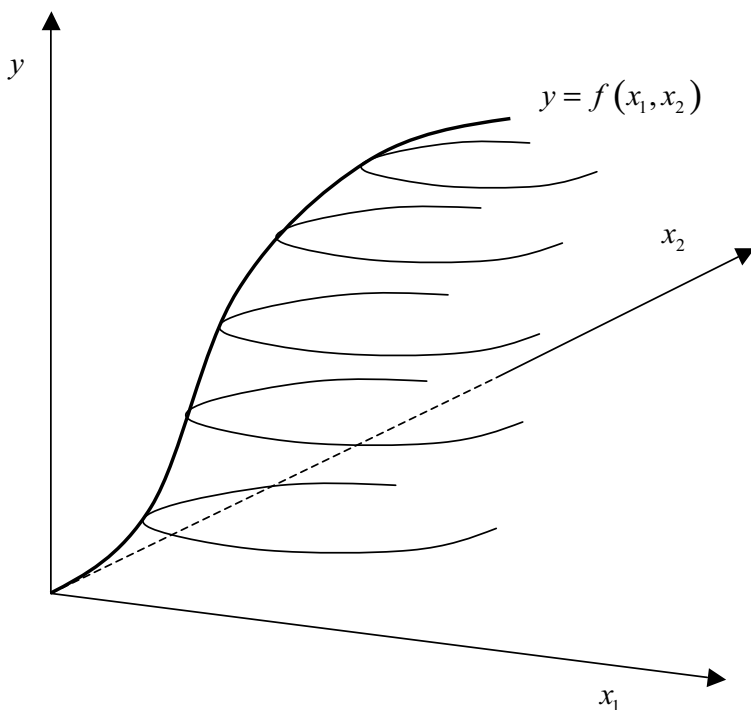


Figura 2.7 – La funzione di produzione.

³ Si noti che da qui in avanti nelle rappresentazioni grafiche gli input e gli output che specificano il processo di produzione, concetti divenuti ormai familiari a questo punto della trattazione, saranno indicati utilizzando solamente i simboli corrispondenti.

Alla funzione di produzione possono essere attribuite varie proprietà, in parte corrispondenti a quelle dell'insieme di produzione, in parte proprie (Tani, 1986).

Dalla monotonicità dell'insieme di fabbisogno degli input, e quindi dalla proprietà della *free disposal*, discende che la funzione di produzione è non decrescente, ossia un incremento negli input non può determinare una diminuzione dell'output, per cui $\mathbf{x}' \geq \mathbf{x} \Rightarrow f(\mathbf{x}') \geq f(\mathbf{x})$.

Nel caso di un solo output, la convessità dell'insieme di fabbisogno di input è equivalente alla quasi concavità della funzione di produzione.

Infatti $L(y)$ è un insieme convesso se e solo se la funzione di produzione $f(\mathbf{x})$ è una funzione quasi concava, così $L(y) = \{\mathbf{x} | f(\mathbf{x}) \geq y\}$ che è proprio l'insieme di livello superiore di $f(\mathbf{x})$, ma la funzione è quasi concava se e solo se ha un insieme di livello superiore convesso.

Sebbene non sia strettamente necessario, si suppone, solitamente, che la funzione di produzione goda delle proprietà di continuità e differenziabilità, che implicano un più marcato carattere di sostituibilità generalizzata fra gli input.

L'assunto di derivabilità consente di introdurre il concetto di *produttività marginale* o prodotto marginale del fattore i -esimo, ossia la derivata parziale:

$\frac{\delta f(\mathbf{x})}{\delta x_i}$, che misura il contributo marginale del fattore x_i alla produzione, lasciando inalterate le quantità impiegate degli altri fattori.

Dovendo esprimere processi produttivi efficienti, le produttività marginali sono sempre positive (Giusti, 1994).

Le *elasticità del prodotto* rispetto ai fattori, sono espresse dalle quantità positive:

$$\alpha_i = \frac{\delta f(\mathbf{x})}{\delta x_i} \frac{x_i}{y} = \frac{\delta f(\mathbf{x})/y}{\delta x_i/x_i},$$

in quanto, generalmente, l'incremento di un input non determina la diminuzione della quantità prodotta.

Attraverso l'ipotesi di continuità e derivabilità della funzione di produzione è possibile tradurre in termini analitici un'ipotesi che ricopre notevole importanza nella teoria neoclassica della produzione, ossia quella della *sostituibilità degli input*. Il riferimento è alla possibilità che, data una certa combinazione di input in grado di produrre una data quantità di output, sia sempre possibile sostituire una piccola quantità di un input qualsiasi con un'adeguata quantità

di uno qualunque degli altri input. In termini formali, il *saggio marginale di sostituzione (SMS)* tra l'input x_i e l'input x_j , a parità degli altri fattori, ossia il grado di sostituibilità tra gli input ai fini dell'ottenimento di un dato un livello di output, è espresso da:

$$SMS_{ij} = \frac{\delta f(\mathbf{x})/\delta x_i}{\delta f(\mathbf{x})/\delta x_j}$$

Assumendo, quindi, di produrre in corrispondenza di un punto sulla frontiera di produzione, nel caso a due dimensioni, ossia nel caso di due soli fattori, il saggio marginale di sostituzione corrisponde all'inclinazione dell'isoquanto, mentre nel caso multi-dimensionale, corrisponde all'inclinazione di una superficie di isoquanto, misurata rispetto a una direzione particolare (Varian, 1978).

L'**elasticità di sostituzione**, introdotta da Hicks (1932) come una misura della facilità con cui il fattore che varia può essere sostituito dagli altri, esprime, invece, la curvatura di un isoquanto. In termini formali, l'elasticità di sostituzione è il rapporto tra la variazione percentuale del rapporto tra i fattori e la variazione percentuale del loro saggio marginale di sostituzione, mantenendo costante il livello dell'output. In simboli, nel caso di due soli fattori x_1 e x_2 , si ha:

$$\sigma_{12} = \frac{d(x_2/x_1)/(x_2/x_1)}{d(SMS_{12})/SMS_{12}}$$

La teoria neoclassica considera spesso *funzioni di produzione omotetiche* e, quindi, tecnologie *omotetiche*. La proprietà dell'*omotetia* assicura che, data una mappa di isoquanti, il saggio marginale di sostituzione tra fattori sia costante lungo ogni raggio che parte dall'origine (Vaggi, 1987).

La classe principale di funzioni di produzione caratterizzate da omotetia è quella delle funzione omogenee⁴ e la più nota fra esse è la funzione di produzione introdotta da Cobb e Douglas (1928), che, nel caso di due input, assume la forma $f(x_1, x_2) = Ax_1^\alpha x_2^\beta$, dove A è un parametro che dipende dal grado di efficienza nel modo di organizzare l'attività produttiva (Zamagni, 1994), α e β

⁴ In generale, una funzione di produzione è omogenea di grado μ , se, $\forall \lambda \in \mathbb{R}_+$, risulta: $f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^\mu f(x_1, x_2, \dots, x_n)$. I rendimenti di scala, ossia la relazione esistente tra variazioni di output e variazioni proporzionali di tutti gli input, sono specificati dai valori assunti dal parametro μ . Si deriva che il processo manifesta rendimenti di scala crescenti se $\mu > 1$, rendimenti di scala decrescenti se $\mu < 1$ ed, infine, rendimenti costanti se $\mu = 1$ (Giusti, 1994).

sono costanti positive⁵. Come sottolineato da Varian (1978), questa funzione è la frontiera di un insieme delle produzioni possibili del tipo:

$$Z = \left\{ (y, -x_1, -x_2) \in R^3 : y \leq x_1^\alpha x_2^\beta \right\}$$

L'insieme di fabbisogno di input assume la forma:

$$L(y) = \left\{ (x_1, x_2) \in R_+^2 : y \leq x_1^\alpha x_2^\beta \right\}$$

mentre l'isoquante è espresso da:

$$IsoqL(y) = \left\{ (x_1, x_2) \in R_+^2 : y = x_1^\alpha x_2^\beta \right\}$$

La funzione di produzione Cobb-Douglas, come è emerso nel capitolo precedente, ha formato oggetto di numerose applicazioni in diversi campi, dallo studio della crescita all'analisi dell'efficienza, grazie alle proprietà di cui gode e all'elevato grado di adattamento mostrato. Una prima rassegna sui risultati ottenuti nelle applicazioni si trova in Douglas (1976).

Riprendendo le osservazioni di Koopmans (1951), attraverso un'ulteriore elaborazione del concetto di funzione di produzione si sviluppa la distinzione tra situazioni produttive nelle quali l'insieme delle combinazioni dei fattori che sono possibili da un punto di vista tecnico non è vincolato, consentendo l'esistenza di sostituibilità continua tra i fattori, e situazioni in cui alcuni fattori possono essere combinati solo in rapporti fissi l'uno con l'altro, nel rispetto dei principi tecnologici coinvolti.

Nel secondo tipo di situazione, le corrispondenti funzioni di produzione presentano dei gomiti nei punti in cui i rapporti tra le quantità dei fattori disponibili coincidono con i rapporti tecnici caratteristici del processo in questione.

In un tale contesto, Arrow, Chenery, Minhas e Solow (1961) evidenziano le limitazioni connesse all'utilizzazione della funzione Cobb-Douglas, che impone a priori restrizioni sulla sostituibilità tra gli input, risultando caratterizzata da elasticità di sostituzione costante ed uguale a uno.

La famiglia delle *funzioni di produzione ad elasticità costante di sostituzione* (*Constant Elasticity of Substitution*, CES), cui ha dato origine la funzione introdotta da Arrow, Chenery, Minhas e Solow (ACMS), consente di superare, almeno parzialmente, questo inconveniente, aggiungendo flessibilità all'approc-

⁵ Nel caso della funzione Cobb-Douglas, la variazione subita dall'output al solo variare della scala di esercizio del processo è misurata dalla somma dei coefficienti ($\alpha + \beta$). Se $(\alpha + \beta) = 1$ si avranno rendimenti di scala costanti, mentre se $(\alpha + \beta) > 1$ la funzione presenta rendimenti crescenti ed, infine, se $(\alpha + \beta) < 1$ i rendimenti saranno decrescenti.

cio tradizionale grazie alla considerazione dell'elasticità di sostituzione come di un parametro da stimare.

La funzione CES, nel caso di due soli input, assume (Varian, 1978) la forma seguente:

$$y = \left[a_0 + a_1 x_1^\rho + a_2 x_2^\rho \right]^{1/\rho}$$

Se $a_0 = 0$ la funzione presenta rendimenti di scala costanti e può assumere la forma di altre funzioni di produzione a seconda dei valori assunti dal parametro ρ .

In particolare, al tendere di ρ a zero⁶, gli isoquanti della funzione CES tendono ad assumere la forma di quelli della funzione di Cobb-Douglas (Figura 2.8a), mentre al tendere di ρ a $-\infty$ l'isoquanto della funzione di produzione CES⁷ prende la forma dell'isoquanto associato ad una tecnologia di Leontief, che è caratterizzata dall'esistenza di coefficienti di produzione fissi e, quindi, di non sostituibilità fra i fattori, come evidenziato nella Figura 2.8b.

La tecnologia di Leontief può essere rappresentata attraverso la specificazione dell'insieme di produzione seguente:

$$Z = \left\{ (y, -x_1, -x_2) \in R^3 : y \leq \min(ax_1, bx_2) \right\}$$

ovvero dall'insieme di fabbisogno di input espresso da:

$$L(y) = \left\{ (x_1, x_2) \in R_+^2 : y \leq \min(ax_1, bx_2) \right\}$$

In questo caso, definito il livello dell'output, le quantità di tutti gli input risultano esattamente determinate, a meno di "sprechi".

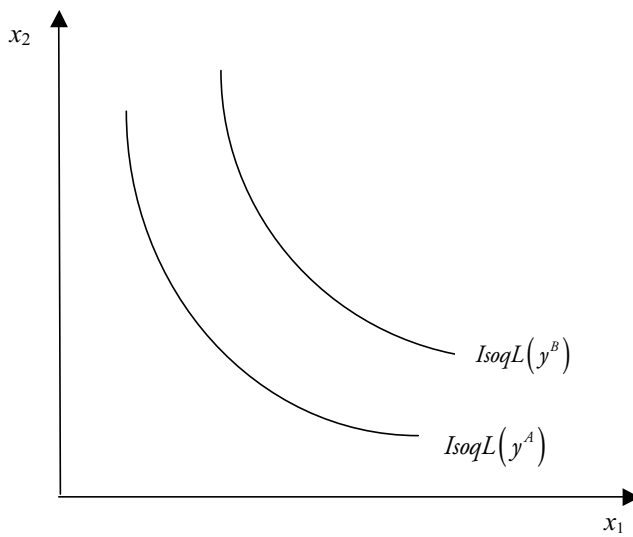
Come sottolinea Giusti (1994), nell'ambito della teoria neoclassica, le ricerche empiriche sulla produzione hanno interessato, prevalentemente, la specificazione, sia a livello aggregato sia a livello di singola impresa, di forme funzionali ben determinate, vincolate al soddisfacimento di particolari proprietà. Il comportamento regolare della funzione di produzione presuppone che la funzione sia monotona positiva e quasi concava rispetto agli input, condizione che si

⁶ Varian (1978), assumendo rendimenti di scala costanti, considera:

$$\lim_{\rho \rightarrow 0} b(\rho) = \frac{y^\rho - a_1 x_1^\rho - a_2 x_2^\rho}{\rho}$$

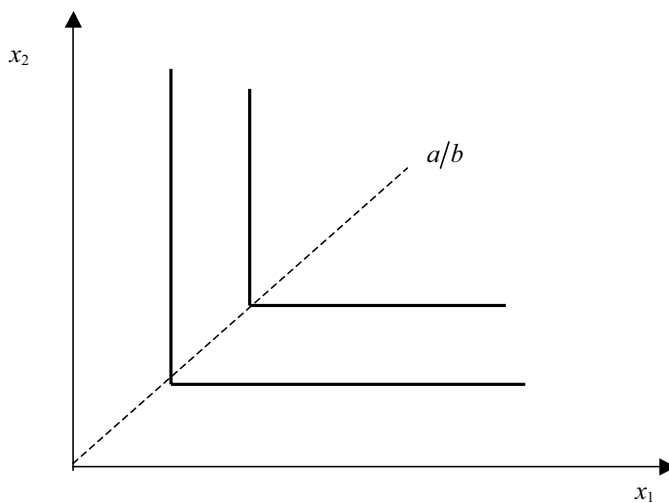
⁷ Assumendo che $a_1 = a_2$ e considerando $x_1 = \min(x_1, x_2)$, Varian (1978) dimostra che:

$$x_1 = \lim_{\rho \rightarrow -\infty} \left[x_1^\rho + x_2^\rho \right]^{1/\rho}$$



$$IsoqL(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y = x_1^\alpha x_2^\beta\}$$

Figura 2.8a – Isoquanti di produzione: Tecnologia Cobb-Douglas.



$$IsoqL(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y = \min(ax_1, bx_2)\}$$

Figura 2.8b – Isoquanti di produzione: Tecnologia Leontief.

traduce nella proprietà di convessità degli isoquanti. Ulteriori restrizioni sono relative all'imposizione o meno di rendimenti di scala costanti e alla considerazione di funzioni omogenee e additive⁸.

Gli studi degli ultimi decenni hanno rivelato l'inadeguatezza dei modelli tradizionali, in quanto caratterizzati da eccessive restrizioni, espresse attraverso ipotesi a priori sulla forma della tecnologia non verificabili statisticamente, sottolineando che le specificazioni formali dovrebbero poter esprimere e rappresentare il più ampio spettro tecnologico possibile, e i relativi effetti economici, al fine di ridurre al minimo le assunzioni imposte a priori.

Si assiste, quindi, all'introduzione di *funzioni flessibili*, appartenenti per lo più alla classe delle funzioni non omotetiche, in cui non sono imposti vincoli restrittivi a priori. Tali funzioni dovrebbero contenere un numero sufficientemente elevato di parametri, da sottoporre a verifica statistica, e includere, come casi particolari, i modelli più semplici, quali la Cobb-Douglas, le funzioni CES e la funzione di Leontief.

Tra le più comuni specificazioni funzionali flessibili⁹, si può citare la *famiglia delle funzioni translogaritmiche*, la cui caratteristica è quella di essere espressa come approssimazione locale, almeno del secondo ordine, di un'arbitraria funzione di produzione e, quindi, di una qualsiasi tecnologia produttiva rappresentabile da una funzione almeno due volte derivabile, essendo desunta

⁸ Una funzione di produzione è additiva se e solo se può rappresentarsi nella forma:

$$f(x_1, x_2, \dots, x_s) = f(x_1) + f(x_2) + \dots + f(x_s) = 0$$

dove le funzioni $f(x_i)$ sono strettamente monotone e dipendono da una sola variabile.

⁹ Diewert (1971) introduce la funzione lineare generalizzata che assume la forma:

$$y = \Phi \left(\sum_i \sum_j a_{ij} \sqrt{x_i x_j} \right)$$

dove i parametri sono non negativi e verificano la relazione di simmetria ($\alpha_{ij} = \alpha_{ji}$), mentre Φ è una qualsiasi funzione continua e crescente, che verifica l'uguaglianza $\Phi(0) = 0$. Nel caso di rendimenti costanti di scala la Φ è la funzione identità per cui l'espressione precedente si semplifica. Successivamente Diewert (1973) propose una generalizzazione della funzione di produzione Cobb-Douglas con n fattori, definita da:

$$y = k \prod_{i=1}^s \prod_{j=1}^s \left(\frac{1}{2} x_i + \frac{1}{2} x_j \right)^{\alpha_{ij}}$$

dove i parametri α_{ij} soddisfano le seguenti restrizioni nel caso di rendimenti costanti di scala:

$$\alpha_{ij} = \alpha_{ji}, \quad 1 \leq i, j \leq s \quad \text{e} \quad \sum_i \sum_j \alpha_{ij} = 1.$$

Nel caso in cui $\alpha_{ij} = 0 \quad \forall i \neq j$ la precedente forma funzionale si riduce alla funzione di Cobb-Douglas.

da un particolare sviluppo in serie di Taylor troncato ai termini del secondo o terzo ordine (Giusti, 1994).

Christensen, Jorgenson e Lau (1971,1973) introducono quella che può definirsi la più utilizzata tra le forme flessibili, ossia la funzione translogaritmica, abbreviata spesso in *translog*, caratterizzata dalla presenza di termini sia lineari che quadratici nei logaritmi delle quantità di un determinato numero di input.

La separabilità e le altre restrizioni, come ad esempio sui rendimenti di scala, possono essere imposte parametricamente e verificate statisticamente. La funzione *translog* può essere rappresentata, nel caso di un solo prodotto e di s fattori, mediante l'espressione:

$$\log y = \log \alpha_0 + \sum_{i=1}^s \alpha_i \log x_i + \frac{1}{2} \sum_{i=1}^s \sum_{j=1}^s \gamma_{ij} \log x_i \log x_j$$

Nell'adozione di una funzione di tipo *translog* non si assume alcuna ipotesi restrittiva sulla tecnologia, tuttavia i parametri devono verificare alcune condizioni affinché la *translog* possa essere considerata una funzione di produzione a comportamento regolare. In particolare, l'output deve crescere in modo monotono rispetto agli input, gli isoquanti devono realizzare una condizione di convessità in senso stretto (Vitali, 1999) e i parametri devono essere simmetrici, ossia $\gamma_{ij} = \gamma_{ji}$. La prima condizione richiede che:

$$\frac{\delta y}{\delta x_i} > 0 \Rightarrow \frac{\delta \ln y}{\delta \ln x_i} = \frac{\delta y}{\delta x_i} \cdot \frac{x_i}{y} > 0.$$

La seconda è verificata se la matrice hessiana orlata delle derivate parziali del primo e secondo ordine è definita negativa.

Il problema della scelta della forma funzionale più idonea a rappresentare una specifica situazione reale, che vada oltre il puro adattamento ai dati osservati, non è di facile soluzione ed ha costituito l'oggetto di numerosi studi.

La modellizzazione econometrica dei processi produttivi, attraverso l'utilizzazione di funzioni flessibili, ha determinato lo sviluppo di una ricchissima letteratura, con riguardo principalmente a ricerche globali e settoriali concernenti diversi paesi.

Fra i numerosi studi si può citare, solo per fare un esempio, quello di Berndt, Darrough e Diewert (1977) nel quale l'attenzione è concentrata su diverse funzioni flessibili, tra le quali si ritrova la *translog*, la cui adattabilità viene sottoposta a verifica empirica attraverso l'utilizzazione di dati relativi a particolari aggregati economici che caratterizzano il Canada.

2.6 LA RAPPRESENTAZIONE DELLA TECNOLOGIA A MEZZO DELLA FUNZIONE DI TRASFORMAZIONE E DELLA FUNZIONE DI DISTANZA

Nei casi in cui le unità produttive siano caratterizzate dalla produzione di molteplici output, la tecnologia può essere rappresentata mediante una generalizzazione a m dimensioni della funzione di produzione, utilizzata in particolar modo nel contesto teorico dell'equilibrio economico generale, denominata **funzione di trasformazione** (*transformation function*), consistente in un legame implicito tra input e output, espresso da $T(\mathbf{x}, \mathbf{y}): R^m \rightarrow R$, dove $T(\mathbf{x}, \mathbf{y}) = 0$ se e solo se (\mathbf{x}, \mathbf{y}) è un processo efficiente in Z .

La funzione di trasformazione può essere definita, per l'output generico, y_r , come la massima quantità di y_r che può essere prodotta, nell'ipotesi di prefissate quantità degli altri beni prodotti, con assegnate quantità degli input dei fattori (Giusti, 1994) ed assume, indicando con \mathbf{y}^* il vettore dei prodotti, la forma:

$$T(\mathbf{y}^*, \mathbf{x}) = \max_{y_r} \left\{ y_r \mid (y_r, \mathbf{y}^*, \mathbf{x}) \in Z \right\}$$

Le proprietà che la caratterizzano sono, in particolare, quella di essere una funzione reale limitata per ogni insieme $(\mathbf{y}^*, \mathbf{x})$, non crescente rispetto a \mathbf{y}^* e non decrescente rispetto a \mathbf{x} ed, infine, di essere continua a destra e concava.

Una seconda e più importante estensione, che interviene in modo significativo nella teoria della dualità, è rappresentata dalla **funzione di distanza** (*distance function*), introdotta da Shephard (1953, 1970), che, non solo caratterizza la struttura della tecnologia di produzione, sia nel senso degli input che in quello degli output, ma è strettamente legata alla misura dell'efficienza tecnica.

A partire da un insieme di produzione e dai suoi insiemi di fabbisogno di input regolari, monotonici e convessi, la funzione di distanza, nel senso degli input, detta anche **input distanza** (*input distance function*) può essere definita (Kumbhakar e Lovell, 2000) attraverso:

$$d_I(\mathbf{x}, \mathbf{y}) = \max \left\{ \lambda \mid \frac{\mathbf{x}}{\lambda} \in L(\mathbf{y}) \right\}$$

Una definizione analoga, fornita da Färe e Lovell (1978) e Deprins e Simar (1983), è la seguente:

$$d_I(\mathbf{x}, \mathbf{y}) = \frac{1}{\min \left\{ \lambda \mid \lambda \mathbf{x} \in L(\mathbf{y}) \right\}}$$

La Figura 2.9 illustra la funzione di input distanza nel caso semplificato in cui il processo di produzione sia caratterizzato dalla presenza di due soli input (Kumbhakar e Lovell, 2000).

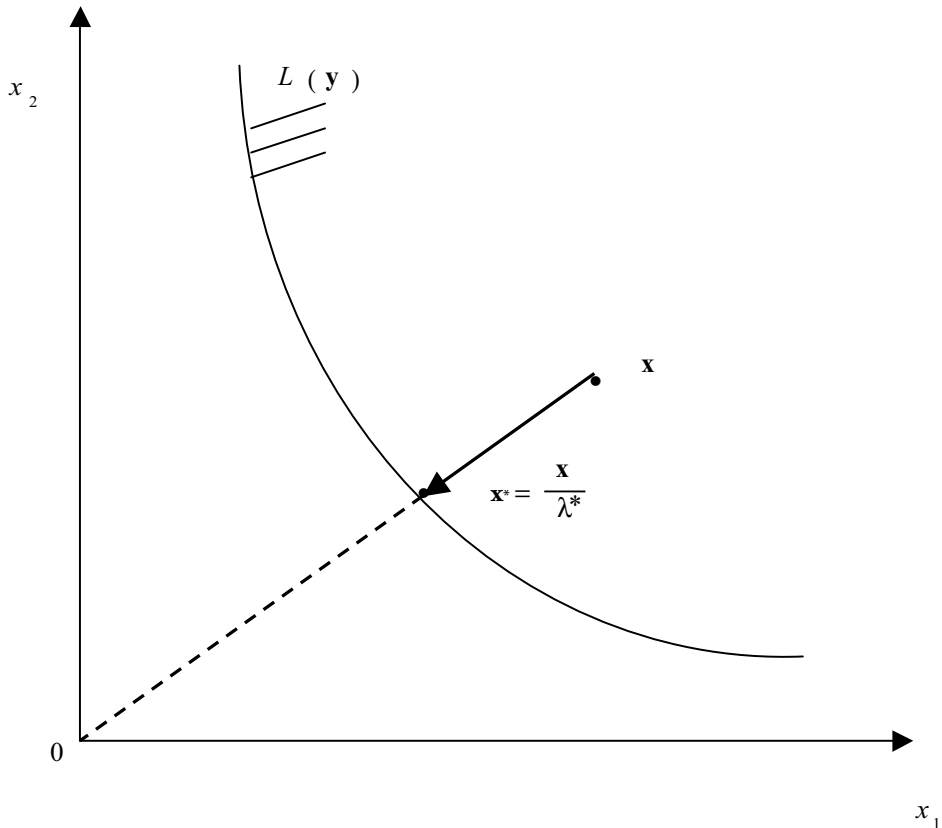


Figura 2.9 – La funzione di input distanza.

La funzione di input distanza esprime l'ammontare massimo, che può essere indicato con λ^* , in relazione al quale un vettore di input x può essere ridotto, in modo radiale, cioè tale da mantenere inalterate le proporzioni tra i vari fattori impiegati, ed essere ancora in grado di produrre il vettore di output y .

Geometricamente dato un processo di produzione (x, y) , se si immagina di tracciare la semiretta uscente dall'origine che passa per il punto x , in virtù del fatto che l'insieme di fabbisogno di input gode delle proprietà di regolarità, monotonicità e convessità, questa semiretta attraverserà la frontiera di $L(y)$ in un punto, che

identifica un vettore di input \hat{x} , in grado di produrre y e nel quale le proporzioni tra gli input sono uguali a quelle del vettore x dato. Il vettore x può, quindi, essere contratto radialmente attraverso uno scalare positivo, $\lambda^* > 1$, che può essere definito mediante il rapporto tra le lunghezze dei due vettori, ossia $\frac{\|x\|}{\|x^*\|}$ dove $\|\cdot\|$ indica il modulo di un vettore e x^* il vettore input efficiente.

La funzione di input distanza, essendo definita in riferimento all'insieme di fabbisogno di input, soddisfa le seguenti proprietà:

1. $d_I(x, 0) = +\infty$ e $d_I(0, y) = 0$;
2. $d_I(x, y)$ è una funzione semicontinua a destra;
3. $d_I(\delta x, y) = \delta d_I(x, y)$ per $\delta > 0$;
4. $d_I(\delta x, y) \geq d_I(x, y)$ per $\delta \geq 1$;
5. $d_I(x, \delta y) \leq d_I(x, y)$ per $\delta \geq 1$;

Se l'insieme di fabbisogno di input $L(y)$ gode della proprietà di monotonicità forte, la quarta e la quinta proprietà possono essere sostituite dall'unica proprietà seguente:

$$6. d_I(x', y) \geq d_I(x, y) \text{ per } x' \geq x \text{ e } d_I(x, y') \leq d_I(x, y) \text{ per } y' \geq y$$

Nel caso in cui l'insieme $L(y)$ è convesso, la funzione di input distanza è una funzione concava in x .

La funzione di **output distanza** (*output distance function*), definita, invece, in relazione all'insieme degli output producibili, esprime l'ammontare minimo, indicato con μ^* , del quale il vettore di output può essere ridotto e rimanere ancora producibile con il vettore di input dato e può esprimersi (Kumbhakar e Lovell, 2000) attraverso:

$$d_O(x, y) = \min \left\{ \mu \left| \frac{y}{\mu} \in P(x) \right. \right\}$$

Färe e Lovell (1978) e Deprins e Simar (1983) suggeriscono la seguente definizione:

$$d_O(x, y) = \frac{1}{\max \left\{ \mu \mid \mu y \in P(x) \right\}}$$

Nella Figura 2.10, che riproduce la funzione di output distanza nel caso semplificato di due soli output, il vettore y è producibile utilizzando la quantità

di input x . D'altra parte, il vettore degli output y può essere espanso in senso radiale attraverso uno scalare positivo $\mu^* > 1$ fino a raggiungere il vettore di output efficiente y^* .

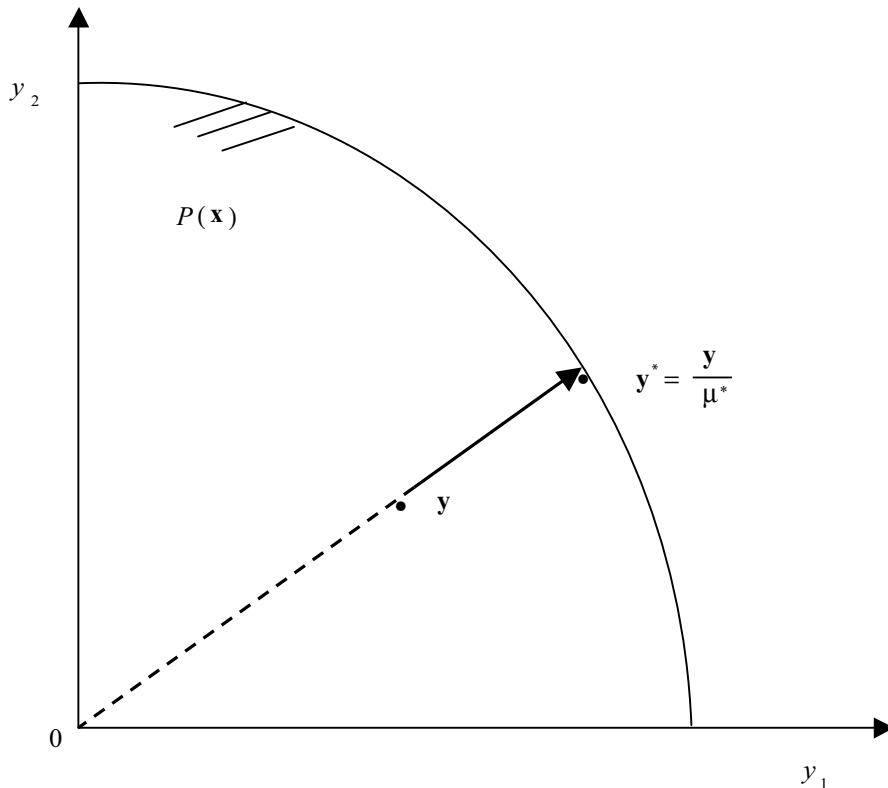


Figura 2.10 – La funzione di output distanza.

In altre parole, come osservano Färe e Grosskopf (1996), la funzione di output distanza esprime la massima espansione radiale possibile di y , posizionando il vettore stesso, attraverso la trasformazione $\frac{y}{\mu}$ sull'isoquante di $P(x)$, ossia sull'insieme definito da:

$$IsoqP(x) = \left\{ y : y \in P(x), \text{ e } y' \geq y \Rightarrow y' \notin P(x) \right\}$$

La funzione di output distanza soddisfa le seguenti proprietà, derivanti dalle corrispondenti proprietà dell'insieme $P(x)$:

1. $d_o(x, 0) = 0$ e $d_o(0, y) = +\infty$;

2. $d_o(\mathbf{x}, \mathbf{y})$ è una funzione semicontinua a sinistra;
3. $d_o(\mathbf{x}, \delta \mathbf{y}) = \delta d_o(\mathbf{x}, \mathbf{y})$ per $\delta > 0$;
4. $d_o(\delta \mathbf{x}, \mathbf{y}) \leq d_o(\mathbf{x}, \mathbf{y})$ per $\delta \geq 1$;
5. $d_o(\mathbf{x}, \delta \mathbf{y}) \leq d_o(\mathbf{x}, \mathbf{y})$ per $0 \leq \delta \leq 1$;

Nel caso in cui l'insieme soddisfi la proprietà di monotonicità forte, vale la proprietà seguente, che sostituisce la quarta e la quinta proprietà, ossia

$$6. d_o(\mathbf{x}', \mathbf{y}) \leq d_o(\mathbf{x}, \mathbf{y}) \text{ per } \mathbf{x}' \geq \mathbf{x} \text{ e } d_o(\mathbf{x}, \mathbf{y}') \leq d_o(\mathbf{x}, \mathbf{y}) \text{ per } \mathbf{y}' \leq \mathbf{y}$$

Se l'insieme degli output producibili è convesso, la funzione di output distanza è convessa in \mathbf{y} .

Färe e Grosskopf (1996) osservano che, sotto l'ipotesi di disponibilità debole degli output, definita, come si vedrà più avanti, *weak disposability*, la funzione di output distanza caratterizza completamente la tecnologia di produzione.

Quindi, formalmente risulta che le funzioni di input e output distanza caratterizzano interamente la tecnologia di produzione, come si evince, ad esempio, dalle seguenti espressioni:

$$L(\mathbf{y}) = \left\{ \mathbf{x} \mid d_I(\mathbf{x}, \mathbf{y}) \geq 1 \right\} \quad IsoqL(\mathbf{x}) = \left\{ \mathbf{y} \mid d_I(\mathbf{x}, \mathbf{y}) = 1 \right\}$$

$$P(\mathbf{x}) = \left\{ \mathbf{y} \mid d_o(\mathbf{x}, \mathbf{y}) \leq 1 \right\} \quad IsoqP(\mathbf{x}) = \left\{ \mathbf{y} \mid d_o(\mathbf{x}, \mathbf{y}) = 1 \right\}$$

I suddetti Autori evidenziano inoltre che, nel caso di rendimenti di scala costanti, la funzione di output distanza è il reciproco della funzione di input

distanza, ossia $d_I(\mathbf{x}, \mathbf{y}) = \frac{1}{d_o(\mathbf{x}, \mathbf{y})}$.

2.7 I RENDIMENTI DI SCALA

Nell'ottica che s'intende conferire alla trattazione in questo lavoro, è opportuno, a questo punto, riprendere il concetto di **rendimenti di scala**.

Con questa espressione ci si riferisce alla relazione tra variazioni di output e variazioni equiporzionali di tutti gli input. Il termine di "scala" sta ad indicare che ciò che cambia è soltanto la "dimensione" del processo, ovvero la scala

di esercizio del processo, mentre la struttura e le caratteristiche dello stesso non mutano (Zamagni, 1994).

In termini formali, si può affermare che una tecnologia di produzione presenta *rendimenti costanti di scala* (*Constant Return to Scale, CRS*) se è soddisfatta una qualsiasi delle seguenti condizioni che si riferiscono a diverse rappresentazioni della tecnologia:

a) $z \in \tilde{Z} \Rightarrow \lambda z \in \tilde{Z} \quad \forall \lambda \in R_+$, ossia l'insieme di produzione \tilde{Z} è un cono con vertice nell'origine;

b) $x \in L(y) \Rightarrow \lambda x \in L(\lambda y) \quad \forall \lambda \in R_+$;

c) $f(\lambda x) = \lambda f(x) \quad \forall \lambda \in R_+$, cioè se la funzione di produzione è omogenea di grado 1.

La tecnologia esibisce *rendimenti di scala non decrescenti* (*NonDecreasing Return to Scale NDRS*) se, in riferimento all'insieme di produzione è soddisfatta la condizione seguente: $z \in \tilde{Z} \Rightarrow \lambda z \in \tilde{Z} \quad \forall \lambda > 1$.

Secondo una versione forte della proprietà, la tecnologia presenta *rendimenti crescenti di scala* se $z \in \bar{Z}, \lambda > 1 \Rightarrow \exists z' \in \tilde{Z} : z' \geq \lambda z$, in cui \bar{Z} indica la chiusura di \tilde{Z} . In questo caso, dato un processo possibile, anche ogni processo che, conservando le proporzioni, ne aumenti la scala, è possibile.

In riferimento alla descrizione della tecnologia tramite la funzione di produzione si hanno rendimenti crescenti di scala se $f(\lambda x) > \lambda f(x) \quad \forall \lambda > 1$, ovvero se l'aumento dell'output è superiore a λ .

Una tecnologia presenta *rendimenti di scala non crescenti* (*NonIncreasing Return to Scale NIRS*) se: $z \in \tilde{Z} \Rightarrow \lambda z \in \tilde{Z} \quad \forall \lambda \in [0,1]$. Secondo questa proprietà, come afferma Tani (1986), ogni processo, e quindi ogni elemento che lo definisce, è considerato divisibile, nel senso che, dato un processo possibile, si possono attivare processi che abbiano le stesse proporzioni degli input e degli output ma in scala ridotta.

La versione forte di questa proprietà considera *rendimenti decrescenti di scala* se $z \in \bar{Z}, \lambda \in (0,1) \Rightarrow \exists z' \in \tilde{Z} : z' \geq \lambda z$. Rispetto ad un tecnologia descritta mediante una funzione di produzione si hanno rendimenti decrescenti di scala se $f(\lambda x) < \lambda f(x) \quad \forall \lambda > 1$, ovvero se l'aumento degli output è inferiore a λ .

La considerazione dei rendimenti di scala è molto importante nella teoria della produzione, in quanto con essi si intende formalizzare tutte quelle situazioni nelle quali il variare della scala, o dimensione, produttiva altera le condizioni di produzione nel senso di renderle più convenienti (rendimenti crescen-

ti), meno convenienti (rendimenti decrescenti) o invariate (rendimenti costanti), indipendentemente dai prezzi degli input.

Le definizioni di rendimenti di scala crescenti e di rendimenti decrescenti risultano significative solo per insiemi di produzione per i quali vale la proprietà di limitazione superiore degli output, ossia quando $\forall \bar{z} \in \tilde{Z}$, l'insieme dei vettori $z \in \tilde{Z}$, tali che $z \geq \bar{z}$, è limitato (la proprietà afferma, quindi, che vi è un limite alla possibilità di produrre beni e servizi a partire da una quantità prefissata di input).

Come osserva Varian (1999), una tecnologia può esibire rendimenti di scala diversi per diversi livelli di produzione. Può accadere, ad esempio, che per bassi livelli di produzione, la tecnologia manifesti rendimenti di scala crescenti, mentre per livelli più elevati si possono presentare rendimenti di scala costanti o anche decrescenti.

Per illustrare il legame concettuale esistente tra tipo di tecnologia e rendimenti di scala (Ray, 2004), si può considerare la Figura 2.11, che si riferisce, per semplicità di esposizione, a processi produttivi con un solo input e un unico output e ad una tecnologia di produzione descritta dall'insieme delle possibilità produttive $Z = \{(x, y) : y \leq f(x); x \geq a\}$, in cui $y^* = f(x)$ è la funzione di produzione e a è il livello minimo, o scala, per cui la funzione di produzione non è definita¹⁰.

Nell'esempio mostrato nella parte a) della Figura 2.11, i rendimenti di scala crescenti valgono per livelli di input da 0 fino a x_0 . In questa regione la produttività marginale di x è crescente e la funzione di produzione è convessa. Oltre il livello di input x_0 i rendimenti di scala sono decrescenti, mentre, localmente, nel punto x_0 valgono rendimenti costanti di scala. Nella parte b) della Figura 2.11 la funzione di produzione è globalmente concava, ma i rendimenti di scala crescenti valgono esclusivamente in corrispondenza dei livelli di input compresi tra x_m , che rappresenta la scala minima di attivazione del processo, e x_0 . Di nuovo, in corrispondenza del livello di input x_0 , si hanno rendimenti di scala costanti, mentre, per livelli superiori di utilizzazione del fattore x , i rendimenti di scala divengono decrescenti.

Emerge quindi che la presenza di rendimenti di scala crescenti non è vincolata alla convessità della tecnologia di produzione. Dagli esempi illustrati nella Figura 2.11, si evince, infatti, che nel primo caso l'insieme di produzione non è convesso, mentre nel secondo caso, la concavità della funzione di produzione assicura la convessità dell'insieme Z .

¹⁰ Nel caso in cui non esista una scala minima, a è uguale a zero.

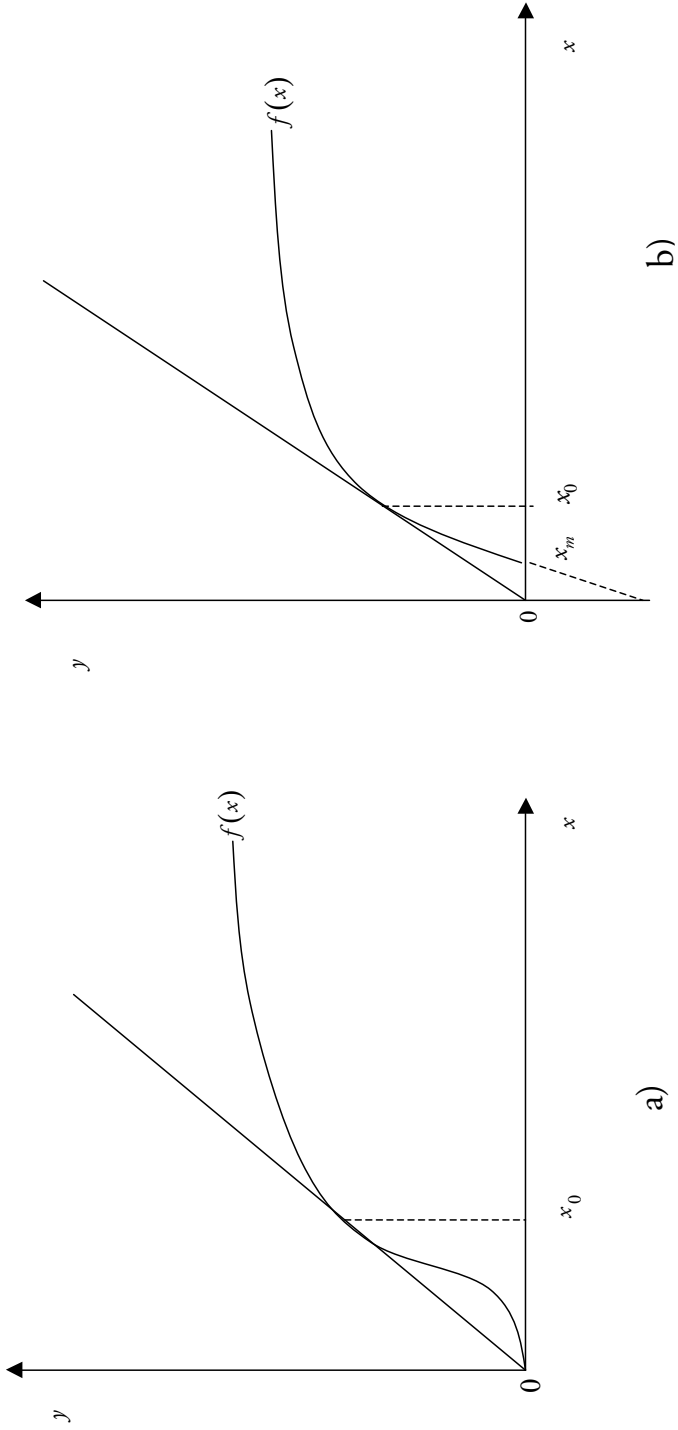


Figura 2.11 – Funzione di produzione e rendimenti di scala.

L'analisi dei rendimenti di scala ha rappresentato l'oggetto di numerosi studi empirici. Tra i molteplici settori di attività coinvolti, si possono considerare, in particolare per l'Italia, il settore dei trasporti pubblici e quello dell'assistenza sanitaria.

Nel contesto dei trasporti pubblici locali, Levaggi (1994) rileva la presenza di economie di scala di breve periodo per le imprese minori, utilizzando sia un'analisi parametrica che un'applicazione non parametrica. Dagli studi di Cambini e Filippini (2003) e di Fazioli, Filippini e Künzle (2003) emergono persistenti economie di scala per le imprese maggiori, da cui consegue che numerosi operatori svolgono il processo produttivo ad un livello dimensionale troppo ridotto.

Nell'ambito dell'assistenza sanitaria pubblica, occorre notare che molti interventi di razionalizzazione della struttura dell'offerta ospedaliera presuppongono l'esistenza di vantaggi derivanti dalla concentrazione della produzione in pochi luoghi di cura. Fabbri (2002) esamina la produzione ospedaliera di sei regioni italiane con l'obiettivo di analizzare, tra l'altro, l'efficienza di scala delle diverse unità produttive rappresentate dagli ospedali pubblici.

2.8 LA FRONTIERA DELL'INSIEME DI PRODUZIONE E LA MISURA DI EFFICIENZA

Nelle pagine precedenti sono state illustrate diverse rappresentazioni della tecnologia di produzione, che descrive l'insieme di tutti i processi produttivi realizzabili. L'obiettivo di un'impresa o unità produttiva è quello di individuare, tra i processi possibili, il *processo efficiente*, che, in considerazione di quanto esposto fino ad ora, si colloca sulla frontiera delle possibilità produttive. A questo punto appare essenziale fornire una descrizione del tentativo di Farrell (1957), più volte richiamato, di tradurre in termini "operativi" la definizione di efficienza in senso paretiano.

Per illustrare graficamente le misure di efficienza, di cui intende suggerire l'utilizzazione per la valutazione di processi produttivi caratterizzati da unico output, y , e due input, x_1 e x_2 , Farrell assume la presenza di una tecnologia caratterizzata da rendimenti costanti di scala. Tali ipotesi consente di utilizzare il concetto di isoquanto unitario, esprime la combinazione dei due fattori che un'unità produttiva efficiente può impiegare per produrre un'unità di output.

La Figura 2.12 riproduce il diagramma utilizzato da Farrell nel quale l'isoquanto unitario è rappresentato dalla curva SS' .

L'*efficienza tecnica* (*technical efficiency*) dell'unità produttiva P è definita da Farrell mediante il rapporto OQ/OP ed esprime la capacità dell'organizzazione

produttiva di minimizzare le quantità di input impiegate per la produzione di una data quantità di output.

Il punto Q corrisponde, infatti, a un processo che impiega i due fattori di produzione nelle stesse proporzioni di P ma in quantità inferiori. L'esistenza del punto Q dimostra, quindi, che la quantità di output ottenuta dal processo P potrebbe essere prodotta impiegando solo una frazione OQ/OP delle quantità utilizzate per ciascuno dei due fattori.

Secondo il suddetto Autore il rapporto OQ/OP presenta alcune proprietà desiderabili per una misura di efficienza. Innanzitutto, assume il valore 1, (o 100 se espresso in forma percentuale) per un'unità produttiva perfettamente efficiente, e assume valori sempre più piccoli al crescere dell'ammontare degli input impiegati per unità di output. Inoltre, fintantoché l'isoquanto SS' presenta una pendenza negativa, l'incremento nell'input per un'unità di output di un fattore determina, a parità di condizioni, il raggiungimento di una minore efficienza tecnica dell'unità produttiva.

Questa misura di efficienza esprime, dunque, la quantità di input che potrebbe essere "risparmiata" per produrre la stessa quantità di output del processo osservato, utilizzando al meglio la tecnologia disponibile.

Atteso che la linea AA' presenta una pendenza uguale al rapporto dei prezzi dei due fattori, l'*efficienza allocativa (price efficiency)* può essere definita mediante la relazione OR/OQ , che esprime la capacità dell'unità produttiva di scegliere combinazioni di input ottimali rispetto ai prezzi.

In questo contesto, il metodo di produzione ottimale è rappresentato da Q' mentre il processo R è scelto come standard di riferimento poiché, pur essendo un processo di produzione impossibile, presenta un costo complessivo uguale a quello minimo che, dati tecnologia e prezzi, deve essere sostenuto per produrre la quantità di input considerata. La scelta di Q e non di P , come processo da confrontare con quello ottimale, dipende, invece, dall'esigenza di valutare l'inefficienza allocativa indipendentemente da quella tecnica (Nisticò e Prosperetti, 1991).

Infine, il rapporto OR/OQ , esprimibile anche come prodotto delle due misure di efficienza, tecnica e allocativa, ossia mediante l'espressione $OR/OP = OQ/OP \cdot OR/OQ$, definisce l'efficienza produttiva totale (*overall efficiency*) e rappresenta la capacità dell'unità di minimizzare i costi totali di produzione.

Secondo questa impostazione, l'efficienza tecnica può, quindi, essere definita come l'abilità di una organizzazione produttiva di operare sulla frontiera dell'insieme di produzione cui appartiene.

Per l'individuazione della frontiera di produzione efficiente, Farrell adotta un approccio operativo, suggerendo di utilizzare le migliori *performance* osservate tra le unità considerate nella specifica situazione analizzata. In tal modo si

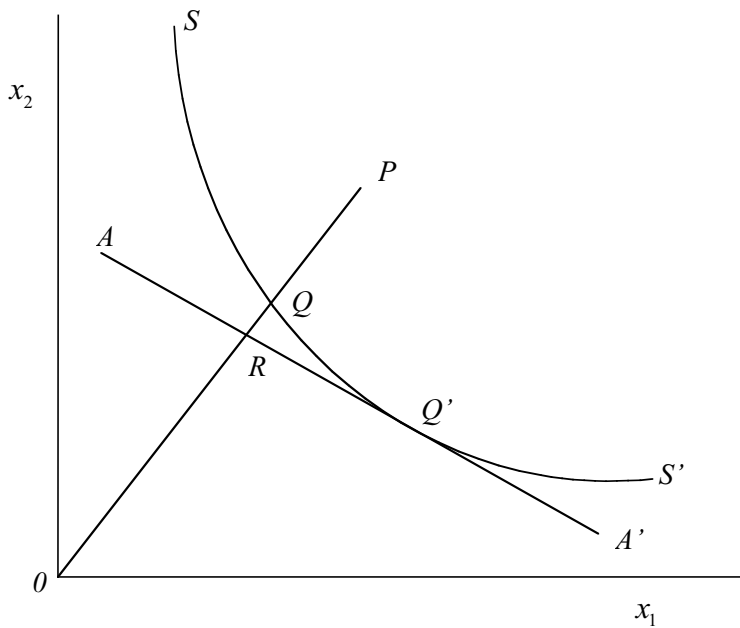


Figura 2.12 – Le misure di efficienza di Farrell.

giunge alla definizione di misure che esprimono l'*efficienza relativa* piuttosto che l'*efficienza assoluta*, altrimenti basate su considerazioni di tipo ingegneristico.

Utilizzando l'ipotesi di convessità e di rendimenti di scala costanti, Farrell stima l'isoquante attraverso l'impiego di sistemi di equazioni lineari, definendo la frontiera così ottenuta, rappresentata nella Figura 2.13, come la specificazione più pessimistica o conservativa, nel senso che la frontiera risulta quanto più possibile vicina ai processi produttivi realmente osservati.

Successivamente, Färe e Lovell (1978) sottolineano che la misura di efficienza di Farrell può essere oggetto di due distinte interpretazioni. In altre parole, il concetto di efficienza tecnica può essere scomposto in due concetti più deboli, quello di *input efficienza* e quello di *output efficienza*. Dal calcolo delle due misure derivano risultati identici solo in riferimento alla tecnologia presa in esame da Farrell (1957), caratterizzata dalla proprietà di distruzione senza costo degli input e degli output e dalla presenza di rendimenti di scala costanti.

Deprins e Simar (1983) estendono le osservazioni di Färe e Lovell (1978), che fanno riferimento al caso di produzione singola, al caso più generale, che interessa unità produttive multi-input e multi-output.

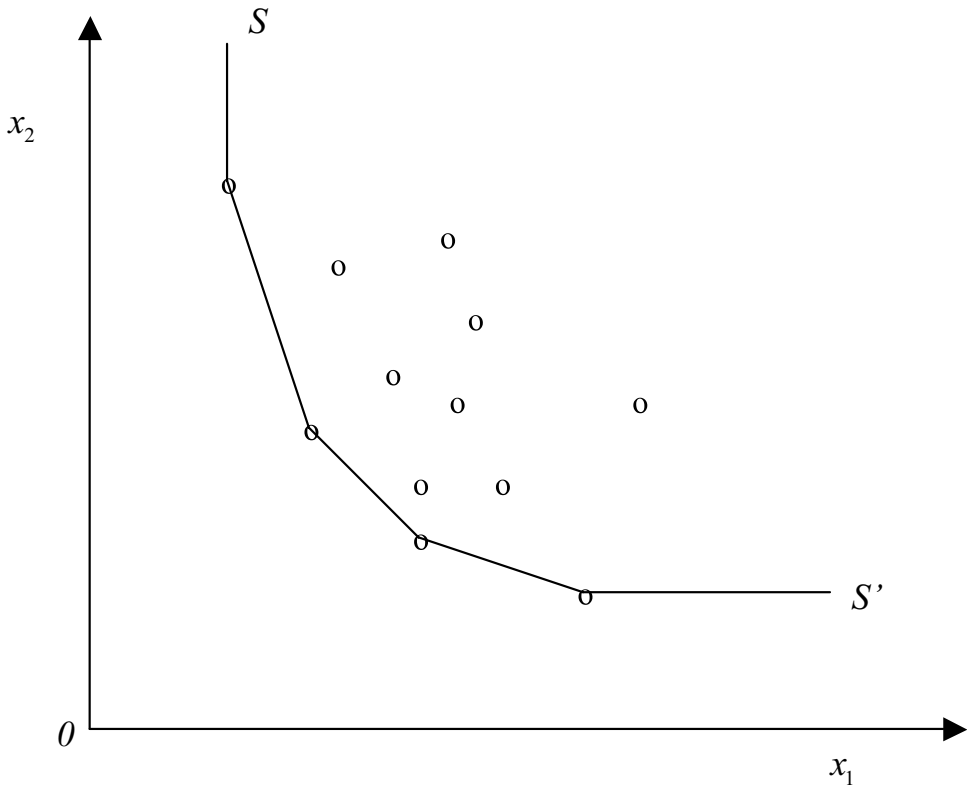


Figura 2.13 – La frontiera di produzione –isoquante unitario.

In particolare, la misura di input efficienza di Farrell, in riferimento al processo produttivo (\mathbf{x}, \mathbf{y}) , definita mediante l'espressione:

$$F_I(\mathbf{x}, \mathbf{y}) = \min \left\{ \theta \mid \theta \mathbf{x} \in L(\mathbf{y}) \right\},$$

per $\mathbf{x} \in L(\mathbf{y})$ e $F_I(\mathbf{x}, \mathbf{y}) = +\infty$ per $\mathbf{x} \notin L(\mathbf{y})$

La misura di output efficienza di Farrell assume, invece, la seguente specificazione:

$$F_O(\mathbf{x}, \mathbf{y}) = \left[\max \left\{ \phi \mid \phi \mathbf{y} \in P(\mathbf{x}) \right\} \right]^{-1},$$

per $\mathbf{y} \in P(\mathbf{x})$ e $F_O(\mathbf{x}, \mathbf{y}) = +\infty$ per $\mathbf{y} \notin P(\mathbf{x})$.

Deprins e Simar (1983) sottolineano che la misura di input efficienza di Farrell è equivalente al reciproco della funzione di input distanza $d_I(\mathbf{x}, \mathbf{y})$,

ossia $F_I(\mathbf{x}, \mathbf{y}) = \frac{1}{d_I(\mathbf{x}, \mathbf{y})}$. La distanza è quindi anche una misura di inefficienza.

È evidente quindi che se il processo (\mathbf{x}, \mathbf{y}) è input efficiente allora $\mathbf{x} = \mathbf{x}^*$ e la misura suddetta è uguale a 1 (Figura 2.9). Ne consegue che $\forall (\mathbf{x}, \mathbf{y}) \in Z$ input-inefficienti si avrà $F_I(\mathbf{x}, \mathbf{y}) < 1$.

Se, viceversa, $F_I(\mathbf{x}, \mathbf{y}) > 1$ allora (\mathbf{x}, \mathbf{y}) non è un processo possibile.

La misura di input inefficienza radiale soddisfa quindi le proprietà seguenti:

$F_I(\mathbf{x}, \mathbf{y}) \leq 1$, $F_I(\mathbf{x}, \mathbf{y}) = 1 \Leftrightarrow IsoqL(\mathbf{y})$; $F_I(\mathbf{x}, \mathbf{y})$ è non crescente in \mathbf{x} ; $F_I(\mathbf{x}, \mathbf{y})$ è omogenea di grado -1 in \mathbf{x} ; $F_I(\mathbf{x}, \mathbf{y})$ è invariante rispetto all'unità di misura di \mathbf{x} e \mathbf{y} . Proprietà analoghe valgono per la misura di output efficienza che coincide con la funzione di output distanza¹¹.

I suddetti Autori, dimostrano che le due misure di efficienza assumono il medesimo valore solo nel caso in cui la tecnologia sia omogenea di grado 1.

La misura di Farrell valuta l'efficienza in riferimento all'isoquanto piuttosto che al sottoinsieme efficiente, concetto più coerente con la nozione di efficienza paretiana.

Il rischio è perciò quello di giungere ad una valutazione errata dell'efficienza dell'unità produttiva considerata.

Per cui un processo produttivo sarà input efficiente, raggiungendo un valore uguale a 1, se e solo se $\mathbf{x} \in IsoqL(\mathbf{y})$. Ma, mentre $\mathbf{x} \in EffL(\mathbf{y})$ implica che $F_I(\mathbf{x}, \mathbf{y}) = 1$, il viceversa non è vero.

Färe e Lovell (1978) dimostrano che la misura di Farrell identifica correttamente le unità efficienti nel senso di Koopmans solo nel caso in cui:

$$IsoqL(\mathbf{y}) = EffL(\mathbf{y})$$

¹¹ A tale proposito, va osservato che Färe e Lovell (1978) dimostrano che la funzione di output efficienza di Farrell è l'inversa della funzione di output distanza. Gli Autori definiscono la funzione di output efficienza attraverso l'espressione:

$$F_o(\mathbf{x}, \mathbf{y}) = \min \{ \phi \mid \mathbf{x} \in L(\mathbf{y}/\phi) \}$$

mentre l'espressione della funzione di distanza coincide con la forma suggerita da Deprins e Simar (1983):

$$d_o(\mathbf{x}, \mathbf{y}) = \left[\max \{ \mu \mid \mu \mathbf{y} \in P(\mathbf{x}) \} \right]^{-1}$$

Tale condizione è verificata, ad esempio, nel caso di rappresentazioni parametriche della tecnologia, per le funzioni di produzione Cobb Douglas e CES, ma non per altre forme parametriche, come la Leontief e la sottofamiglia di funzione di produzione con elasticità di sostituzione variabile e disponibilità debole degli input.

Färe e Lovell (1978) propongono, quindi, una generalizzazione non radiale della misura di Farrell, la cosiddetta *misura di Russell*, che può essere espressa mediante:

$$\mathfrak{R}(\mathbf{x}, \mathbf{y}) = \min \left\{ \sum_{i=1}^k \theta_i / k \mid \theta_1 x_1, \theta_2 x_2, \dots, \theta_k x_k; \dots \in L(\mathbf{y}), \theta_i \in (0, 1] \right\}$$

Nell'espressione precedente la media aritmetica semplice degli scalari $\theta_i \in (0, 1]$, può essere sostituita da altre funzioni, come la media geometrica semplice, o una media aritmetica ponderata suggerita successivamente da Ruggiero e Bretschneider (1995).

La natura non radiale della misura di Russell deriva dal fatto che ogni elemento del vettore $\mathbf{x} \in L(\mathbf{y})$ può essere ridotto di un ammontare diverso, espresso dallo specifico scalare θ_i .

La misura di Farrell può essere derivata come caso particolare della misura di Russell, ossia quando tutti gli elementi sono uguali ad una costante, per cui $\theta_i = \theta \quad \forall i = 1, \dots, k$.

Färe e Lovell (1978) dimostrano che la misura di Russell soddisfa le seguenti proprietà:

1. Se $\mathbf{x} \in L(\mathbf{y}), y > 0$ allora $\mathfrak{R}(\mathbf{x}, \mathbf{y}) = 1 \Leftrightarrow \mathbf{x} \in \text{EffL}(\mathbf{y})$
2. Se $\mathbf{x} \in L(\mathbf{y}), y > 0$ e $\mathbf{x} \notin \text{EffL}(\mathbf{y})$ allora $\mathfrak{R}(\mathbf{x}, \mathbf{y})$ dovrebbe confrontare \mathbf{x} con qualche $\mathbf{x}^* \in \text{EffL}(\mathbf{y})$.
3. Se $\mathbf{x} \in L(\mathbf{y}), \lambda \mathbf{x} \in L(\mathbf{y}), y > 0$ allora $\mathfrak{R}(\lambda \mathbf{x}, \mathbf{y}) = \lambda^{-1} \mathfrak{R}(\mathbf{x}, \mathbf{y})$ per ogni $\lambda \in [\lambda^0, +\infty)$ dove $\lambda^0 \mathbf{x} \in \text{IsoqL}(\mathbf{y})$.
4. Se $\mathbf{x} \in L(\mathbf{y}), y > 0$ e se $\mathbf{x}' \geq \mathbf{x}$ allora $\mathfrak{R}(\mathbf{x}, \mathbf{y}) > \mathfrak{R}(\mathbf{x}', \mathbf{y})$

Attraverso la prima proprietà si richiede che un processo sia definito efficiente se e solo se il vettore degli input appartiene al sottoinsieme efficiente di $L(\mathbf{y})$, mentre, con la seconda, si intende individuare il vettore \mathbf{x}^* rispetto al quale il vettore realmente osservato dovrebbe essere confrontato nella valutazione del-

l'efficienza. La terza proprietà è l'unica soddisfatta anche dalla misura di Farrell e specifica una proprietà di omogeneità. L'ultima espressione identifica la proprietà di monotonicità, attraverso la quale si esige che l'efficienza vari in modo inverso rispetto all'incremento degli input.

La misura di efficienza radiale di Farrell è stata oggetto di numerosi studi.

La natura restrittiva dell'assunzione di rendimenti costanti di scala viene riconosciuta da Farrell e Fieldhouse (1962) i quali suggeriscono un'appropriata trasformazione dei dati che dovrebbe consentire la presenza di rendimenti di scala non costanti nel contesto del modello di analisi delle attività.

In riferimento ad una rappresentazione parametrica della tecnologia, Førsund e Hjalmarsson (1979) propongono una generalizzazione della misura di efficienza di Farrell, che consente la separazione dell'efficienza di scala dall'efficienza puramente tecnica, utilizzando una funzione di produzione.

Ray (1998) estende l'approccio dei suddetti Autori allo scopo di fornire una misura dell'efficienza di scala nel caso in cui la tecnologia sia descritta da una funzione flessibile del tipo *translog*.

Infine, Färe, Grosskopf e Lovell (1994), come si vedrà più avanti, propongono una scomposizione della misura di efficienza tecnica in due distinte misure, rappresentate, rispettivamente, da una misura di congestione degli input (*input congestion*) e da una misura di efficienza puramente tecnica.

Le applicazioni empiriche che hanno coinvolto la misura di efficienza, sia tecnica che economica, rispetto alla frontiera sono innumerevoli, sia in ambito privato che in ambito pubblico. Benché nel corso del presente lavoro si avrà occasione di considerare diverse applicazioni molto interessanti, si possono citare, qui, solo per fare qualche esempio per l'Italia, nel privato, per il settore bancario, Bianchi (1983) e Cardani, Castagna e Galeotti (1991) e per il settore tessile, Ferrari e Ganugi (1993). Nel pubblico, per i trasporti urbani, Petretto e Viviani (1984) e Selmi (1991) e per la riforma della pubblica amministrazione, Petretto e Pisauo (1995), per le poste Gazzei (1993) e Ferrari e Maltagliati (1997).

Capitolo Terzo

L'EFFICIENZA RISPETTO ALLA FRONTIERA DELLA POSSIBILITÀ PRODUTTIVE: L'APPROCCIO PARAMETRICO

3.1 L'EVOLUZIONE DELL'APPROCCIO ECONOMETRICO: DALLA FUNZIONE MEDIA ALLA FRONTIERA DI PRODUZIONE

L'approccio econometrico tradizionale alla stima delle funzioni di produzione (Fuss, Mc Fadden e Mundlack, 1978), originato dal lavoro di Cobb e Douglas (1928), ha generalmente seguito il paradigma teorico neoclassico, secondo il quale il comportamento del produttore, di minimizzazione dei costi o di massimizzazione dei profitti, dati i prezzi di mercato dei fattori e del prodotto, e data la tecnologia di produzione esistente, consente il raggiungimento del pieno equilibrio (Samuelson, 1947).

Tale approccio si basa sulla rappresentazione della tecnologia attraverso la specificazione della funzione di produzione come una relazione, spesso lineare nei parametri o nei logaritmi dei parametri, espressa da $y_i = f(\mathbf{x}_i; \boldsymbol{\beta})$ o da $\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta})$ che lega l'output a un termine costante e a un insieme di variabili, costituito dagli input del processo di produzione. Alla precedente relazione, come anticipato nel primo capitolo, si associa anche una variabile casuale (v.c.) con valore atteso nullo, che, rappresentando un termine di disturbo o errore, consente di includere nel modello l'agire di errori e fattori aleatori che intervengono nelle azioni poste in essere dal produttore nel tentativo di soddisfare le necessarie condizioni per la massimizzazione del profitto.

In termini formali, il **modello "tradizionale"** per la stima della funzione di produzione può essere scritto nel modo seguente:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) \cdot \exp\{v_i\} \quad [3.1]$$

ovvero, in forma del tutto analoga, secondo l'espressione:

$$\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i \quad [3.2]$$

Nelle formulazioni [3.1] e [3.2], si ipotizza che l'output y ottenuto da ciascun processo osservato, per $i = 1, \dots, n$, sia uguale a quello espresso dalla funzione di produzione, aumentato o diminuito di una quantità dipendente da fattori puramente aleatori, rappresentati, ad esempio, da avverse condizioni meteorologiche o da imprevedibili variazioni nella *performance* del lavoro o dei macchinari, inclusi nel termine stocastico¹, v_i , usualmente descritto da una v.c. normale con media nulla e varianza costante.

In relazione alle modalità di introduzione del termine stocastico nei modelli "classici", va precisato che le diverse formulazioni suggerite si basano su ipotesi diverse in relazione agli effetti esercitati dai fattori aleatori sul livello degli input impiegati. In riferimento ad una tale problematica Zellner, Kmenta e Drèze (1966), seguendo quanto già evidenziato da Mundlak e Hoch (1965), suggeriscono la specificazione $y_i = A \log L_i^\alpha + \log K_i^\beta e^{v_i}$ intendendo in tal modo indicare che i disturbi stocastici non si "trasmettono" agli input, come invece si ipotizza nella specificazione seguita da Marschack e Andrews (1944). In altre parole, i fattori aleatori non condizionano la determinazione degli input del processo ma intervengono successivamente alla definizione delle quantità da impiegare, modificando la quantità di output prodotta.

Si può, quindi, affermare che, nell'ambito dei modelli tradizionali, l'allontanamento dalla situazione ottimale, descritta dalla funzione di produzione, è attribuito esclusivamente alla presenza di disturbi ed errori puramente casuali che possono determinare un posizionamento del processo osservato al di sopra o al di sotto della funzione stessa, con uguale probabilità.

Secondo una tale impostazione la funzione di produzione stimata non si identifica, quindi, con il concetto teorico di funzione di produzione, ma assume, come già anticipato, la configurazione di **funzione media**.

In relazione all'analisi dell'efficienza tecnica di unità produttive omogenee, per la stima dei parametri della funzione di produzione, che rappresenta una relazione media tra le osservazioni, e per la determinazione della misura di efficienza per ciascuna unità considerata, sono state utilizzate tecniche di regressione basate, prevalentemente, sul metodo OLS.

¹ Per facilitare la lettura si è deciso di utilizzare una notazione uniforme, in particolare per quanto riguarda l'indicazione dei termini di errore nel modello, sebbene i diversi Autori che hanno contribuito allo sviluppo dell'approccio parametrico, adottino talvolta simboli diversi.

Tra le diverse applicazioni empiriche esistenti in letteratura, si può citare, ad esempio, quella di Petretto e Viviani (1984) che stimano la funzione di produzione per il servizio dei trasporti pubblici utilizzando gli OLS con le ipotesi classiche sull'errore casuale.

La misura di efficienza viene definita attraverso il rapporto tra il livello dell'output osservato per ciascuna unità produttiva, y_i , e la stima ottenuta dalla funzione di produzione, \hat{y}_i , che rappresenta, in tale ambito, la produzione media dell'insieme delle unità produttive considerate, dati i valori degli input.

Con riferimento alla specificazione della v.c. v_i nel modello [3.1], come una *variabile rappresentativa dell'efficienza produttiva*, il rapporto $\pi_i = y_i / \hat{y}_i$ per $i = 1, \dots, n$, dove n è il numero delle unità considerate nell'analisi, costituisce un'indicazione attendibile dell'efficienza produttiva per le diverse unità di produzione. Tale indicatore può assumere valori superiori o inferiori all'unità² esprimendo una maggiore o minore efficienza dell'unità produttiva rispetto alla "media".

I suddetti Autori sottolineano, tuttavia, che un indicatore più interessante, sebbene più difficile da ottenere, potrebbe essere costruito attraverso l'individuazione di valori standard ottimi, da un punto di vista tecnico, dell'output per ciascuna unità, che rappresenterebbero un livello di produzione "desiderabile" da raggiungere per i dati valori delle variabili indipendenti. Evidenziando che un simile risultato potrebbe essere ottenuto grazie alla costruzione di una frontiera di produzione "ideale", mediante una particolare modificazione del modello "classico" in precedenza adottato, Petretto e Viviani ottengono due "stime" ideali dell'output, che denominano *valori obiettivi standard*, aggiungendo all'equazione di regressione, rispettivamente, la quantità 2δ e δ , dove il parametro δ rappresenta la deviazione standard stimata della v.c. v_i . Assumendo che il disturbo stocastico sia distribuito normalmente, nel primo caso considerato solo il 2,5% delle osservazioni rimarrebbe sopra la frontiera "ideale".

L'insoddisfazione verso misure di efficienza determinate in riferimento ad una funzione media diviene sempre più evidente in letteratura, come risulta anche dal lavoro di Nisticò e Viviani (1987) nel quale, riprendendo il modello suggerito da Petretto e Viviani (1984), viene valutata la *performance* delle imprese operanti nel trasporto pubblico.

Come osservano i suddetti Autori, la qualità dei risultati ottenuti attraverso la stima "diretta" della funzione di produzione, attuata sulla base dei modelli

² Petretto e Viviani (1984) ottengono (attraverso la trasformazione logaritmica) un indicatore centrato rispetto a zero piuttosto che a uno.

tradizionali, mediante i metodi di stima di uso più comune, dipende dalle ipotesi assunte sul comportamento dell'impresa, che, in relazione alla scelta del processo di produzione, può tenere in esplicita considerazione anche l'azione dei fattori di inefficienza.

In quest'ultimo caso le stime dei parametri risulterebbero evidentemente distorte in quanto non sarebbe sostenibile l'ipotesi di incorrelazione tra le variabili esplicative ed il termine stocastico. Se, invece, la scelta dell'imprenditore in relazione alle quantità di input da utilizzare, conformemente al principio della minimizzazione dei costi, fosse tale da ignorare l'agire dei fattori di inefficienza, i risultati, in termini di proprietà statistiche degli stimatori, sarebbero, sebbene in parte, più soddisfacenti.

Dalle riflessioni precedenti, emerge, chiaramente, che l'unico modo di pervenire a modelli empirici coerenti con la teoria economica, come già osservato da Varian (1978), consiste nell'esame attento del processo decisionale dell'impresa.

Si era assistito, infatti, come già anticipato, alla progressiva diffusione in letteratura di un orientamento diretto allo sviluppo di una teoria del comportamento del produttore nella quale le motivazioni del singolo rimanevano immutate, ma il successo, nel raggiungimento degli obiettivi, non era affatto garantito. In altre parole, le relazioni produttive dovevano prevedere la possibilità di fallire nel tentativo di ottimizzare.

Riprendendo le parole di Marschack e Andrews (1944), il valore della componente stocastica della funzione di produzione dovrebbe dipendere "... *on the technical knowledge, the will, effort and luck of a given entrepreneur in a given year, as can be summarized in the word technical efficiency*".

Da un punto di vista concettuale, poiché la funzione di produzione teorica esprime il massimo output ottenibile da diverse combinazioni di input, ed è quindi definibile come frontiera delle possibilità produttive, appare più corretto ipotizzare che i valori teorici dell'output, per ciascuna unità produttiva, possano collocarsi con *maggior probabilità sotto la frontiera*.

Gli scostamenti dalla frontiera sono derivanti dall'agire di fattori di inefficienza che, per definizione, agiscono in modo unilaterale, per cui sarebbe più coerente specificare la componente di errore o disturbo casuale come distribuita unilateralmente, se è possibile ritenere che nella particolare situazione osservata gli errori di osservazione esercitino un'influenza irrilevante, o almeno, agiscano in modo asimmetrico con media negativa (Nisticò e Viviani, 1987).

In un tale contesto, l'attività produttiva di un'impresa qualsiasi, in riferimento ad un dato periodo di tempo, è definita come efficiente se gli input impiegati e gli output ottenuti corrispondono ad un punto situato sulla funzione di produzione intesa come frontiera (Gazzei, Lemmi, Viviani, 1997). Il gra-

do di efficienza tecnica è dunque descritto, come anticipato nel paragrafo 1.1, dal confronto tra la quantità osservata dell'output e la quantità massimale di prodotto che l'unità avrebbe potuto produrre utilizzando i medesimi livelli osservati di input. In tal modo, l'output ottenuto da ciascuna unità è esprimibile attraverso un modello del tipo $y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) \cdot ET_i$.

Un primo tentativo di estendere la frontiera di produzione in questa direzione si deve ad Aigner e Chu (1968), che come già detto, sviluppano in senso parametrico l'idea originale di Farrell (1957), introducendo quelle che saranno note in letteratura come **frontiere deterministiche non statistiche**.

I suddetti Autori riconoscono che l'output delle imprese considerate possa giacere sotto la frontiera di produzione a causa dell'agire di una molteplicità di fattori, tra i quali si possono considerare gli effetti derivanti da *shock* puramente casuali che possono interessare il processo produttivo, l'agire di fattori di inefficienza tecnica e, infine, l'eventuale presenza di inefficienza economica.

Al fine di illustrare i metodi di stima suggeriti, Aigner e Chu, considerano, per semplicità, una funzione di produzione di tipo Cobb-Douglas con un solo output e due input, espressa da $y = Ax_1^\alpha x_2^\beta$, valida per n unità produttive³.

Per ottenere una frontiera di produzione, piuttosto che una funzione media, i suddetti Autori suggeriscono di stimare i parametri della funzione, resa lineare attraverso i logaritmi, mediante un sistema di programmazione quadratica, risolvibile con l'algoritmo di Wolfe, ed esprimibile nel modo seguente:

$$\min \left\{ \sum_{i=1}^n \left[A + \alpha \log x_{1i} + \beta \log x_{2i} - \ln y_i \right]^2 \right\} \quad [3.3]$$

soggetto a :

$$\left[A + \alpha \log x_{1i} + \beta \log x_{2i} \right] \geq \ln y_i$$

$$i = 1, \dots, n$$

³ È necessario osservare che Aigner e Chu (1968) aggiungono alla funzione di produzione un termine, indicato con u , che dovrebbe riflettere l'effetto sui valori dell'output derivante da fattori puramente casuali. Il trattamento di questa variabile è implicito e non ne viene specificata una particolare distribuzione aleatoria. Il termine è utilizzato solo allo scopo di evidenziare il concetto di frontiera di produzione in quanto, essendo $y = Ax_1^\alpha x_2^\beta u$, i suddetti Autori impongono che $y \leq Ax_1^\alpha x_2^\beta$, ossia che tutti i residui rispetto alla frontiera siano negativi. Al fine di evidenziare i successivi sviluppi metodologici del concetto di frontiera di produzione si è deciso di non riportare il termine u nell'esposizione dei metodi di programmazione.

Un ulteriore metodo di stima proposto consiste nella soluzione del problema di programmazione lineare che segue:

$$\min \left\{ \sum_{i=1}^n \left[A + \alpha \log x_{1i} + \beta \log x_{2i} - \ln y_i \right] \right\}$$

soggetto a :

$$\left[A + \alpha \log x_{1i} + \beta \log x_{2i} \right] \geq \ln y_i$$

$$i = 1, \dots, n$$
[3.4]

Seguendo uno dei due metodi precedenti, si ottiene però una stima della funzione di produzione complessiva, che Aigner e Chu definiscono *funzione di produzione dell'industria* ("industry production function"), per cui, al fine di identificare le funzioni di produzioni individuali per ciascuna impresa, è necessario introdurre nel modello uno o più parametri che esprimano l'efficienza tecnica di ciascuna unità produttiva.

Aigner e Chu seguono l'approccio di Hildebrand e Liu (1965) che si basa sulla separazione, su basi teoriche, degli effetti dell'efficienza tecnica dai disturbi casuali, con l'obiettivo di consentire solo a quest'ultimi di influenzare la stima dei parametri.

Secondo l'impostazione suggerita, la specificazione esplicita delle differenze nell'efficienza tecnica delle unità considerate può avvenire attraverso l'introduzione, in una funzione di produzione Cobb-Douglas, con lavoro e capitale come unici fattori della produzione, di due specifici parametri, R e r , che possono essere impiegati come *proxy* per descrivere il livello tecnico delle imprese nell'utilizzazione degli input, rispettivamente, di capitale e lavoro.

In particolare, R esprime il rapporto tra il valore delle attrezzature disponibili e il valore dell'intero stabilimento, mentre r è il rapporto tra le unità di personale tecnico e il numero di lavoratori complessivamente impiegati.

La funzione di produzione Cobb-Douglas si trasforma, perciò, nell'espressione:

$$y = Ax_1^{\alpha \log r} \cdot x_2^{\beta \log R}$$
[3.5]

Poiché l'output effettivo raggiunto dal produttore è influenzato anche dal processo di aggiustamento effettuato in conseguenza del verificarsi di particolari situazioni economiche, che caratterizzano i mercati dei fattori e dei prodotti, la stima dei parametri avviene attraverso la soluzione di un sistema di equazioni costituito dalla funzione di produzione e dalle funzioni di domanda dei due fattori, derivate ipotizzando situazioni di concorrenza perfetta e comportamento di massimizzazione del profitto per il produttore.

Su queste basi, Aigner e Chu illustrano l'uso dei metodi di programmazione per la determinazione della frontiera di produzione con i dati dell'industria metallifera statunitense nel periodo 1957-58 per stati.

Va osservato che i parametri della funzione di produzione vengono sia calcolati, attraverso i suddetti metodi, sia stimati utilizzando gli OLS, in riferimento alla seguente specificazione:

$$\log y = \log A + b_0 \log L + e_0 \log R_{-1} \log K_{-1},$$

dove L indica il fattore lavoro, K_{-1} la variabile ritardata, di un anno, che esprime il valore del capitale e y il valore aggiunto.

Le stime ottenute dai metodi di programmazione lineare (PL) e quadratica (PQ), e dagli OLS, non differiscono sostanzialmente, assumendo comunque valori plausibili nel caso dei parametri relativi al fattore lavoro, ossia per l'elasticità e il prodotto marginale.

I due approcci, quello basato sulla programmazione e quello basato sulle stime statistiche, conducono, invece, a risultati molto diversi nel caso dell'elasticità e del prodotto marginale del capitale.

In particolare, rispetto al valore derivante dalla funzione di produzione media, e, quindi, ottenuta applicando il metodo OLS, l'elasticità del lavoro, stimata attraverso la frontiera di produzione, è più elevata se viene applicata la PQ e più bassa se la scelta ricade sul metodo PL (Tabella 3.1). Le differenze appaiono più pronunciate nel caso del prodotto marginale del lavoro e dei parametri legati al capitale⁴.

Utilizzando la variabile R come indicatore dell'efficienza tecnica all'interno delle imprese dell'industria dei metalli primari, i suddetti Autori specificano la *funzione di produzione dell'industria* sostituendo il valore massimo ottenuto per R (nel caso empirico quello osservato per lo stato della Louisiana) nella forma generale della funzione. Si ottiene⁵:

$$\log y = 0,8221 \log L + 0,0219 \cdot 4,2673 \log K_{-1} + \text{costante} \quad [3.6]$$

Tale funzione esprime il massimo output ottenibile dal capitale e dal lavoro, dato il livello tecnologico nell'industria. La funzione di produzione individuale,

⁴ Va osservato che Aigner e Chu si riferiscono ad un modello completo, che coinvolge anche la funzione di domanda del lavoro. La stima dei parametri viene effettuata sia utilizzando il metodo PQ stimando simultaneamente le due equazioni, sia attraverso gli OLS a due stati. Nella Tabella 3.1 sono riportati esclusivamente i risultati ottenuti per la funzione di produzione.

⁵ Si osservi che i parametri nell'espressione [3.6] derivano dalla stima simultanea effettuata con il metodo PQ.

| <i>Metodo</i> | Elasticità del lavoro | Coefficiente e_0 | Prodotto marginale lavoro | Elasticità del capitale | Elasticità tecnologica |
|---------------|-----------------------|--------------------|---------------------------|-------------------------|------------------------|
| <i>OLS</i> | 0,908 | 0,0333 | 1,546 | 0,1278 | 0,5115 |
| <i>PL</i> | 0,873 | 0,0031 | 1,485 | 0,0132 | 0,0165 |
| <i>PQ</i> | 1,071 | 0,0269 | 1,822 | 0,1148 | 0,1441 |

Tabella 3.1–Risultati empirici in Aigner e Chu (1968).

per ciascuno stato, può essere ottenuta sostituendo il valore corrispondente di R nell'espressione generale [3.6].

L'utilizzazione delle tecniche di programmazione lineare viene suggerita anche da Seitz (1971), che, estendendo il modello suggerito da Farrell (1957) e Farrell e Fieldhouse (1962), stima una *frontiera di produzione dell'industria*, proponendo, inoltre, misure di efficienza produttiva, tecnica, di prezzo, e di scala.

L'indice di efficienza tecnica, definito da Seitz "*Technical efficiency given scale*" (*TES*), viene ottenuto come distanza relativa di un punto, rappresentante il processo che deve essere valutato, rispetto al punto corrispondente sulla funzione di produzione stimata che avviluppa, "*envelopes*", tutti i processi osservati.

Utilizzando la funzione stimata e i *TES* calcolati, per aggiustare il costo di produzione unitario tenendo conto dello spreco di fattori, si può determinare una misura dell'efficienza di prezzo. Il prodotto dell'efficienza di prezzo e dell'efficienza tecnica esprime la misura dell'efficienza produttiva. La natura e il grado delle economie di scala, sono oggetto di ulteriori analisi basate su procedimenti analoghi a quelli che hanno condotto all'individuazione dell'efficienza tecnica e dell'efficienza produttiva. In particolare, l'efficienza di scala viene stimata utilizzando il medesimo problema di programmazione lineare con la dimensione di scala eliminata.

Seitz dimostra, quindi, che non solo è possibile ottenere una stima ragionevole di una frontiera di produzione, ma che i risultati ottenuti possono essere utilizzati come base di calcolo per la definizione degli indici di efficienza multidimensionali.

L'applicazione empirica riguarda 181 impianti di generazione del vapore costruiti negli Stati Uniti durante gli anni 1947-63, rispetto ai quali, non solo vengono ottenute misure di efficienza tecnica, ma viene anche impostata un'analisi di regressione che include tra le variabili indipendenti una serie di caratte-

ristiche degli impianti stessi (come la temperatura del vapore, la pressione, ecc.) per individuare possibili determinanti degli indicatori *TES*.

3.2 ALCUNI SVILUPPI DELLE FRONTIERE DETERMINISTICHE: APPROCCIO PROBABILISTICO E APPROCCIO STATISTICO

Un interessante studio che ha contribuito allo sviluppo delle frontiere di produzione si deve, come già anticipato, a Timmer (1971), che tenta di recepire il suggerimento formulato molto brevemente da Aigner e Chu, secondo i quali “...we have attempted to interpret the traditional theory strictly, in that the frontier we construct is truly a surface of maximum points. Under a different goal one may pursue less than 100 per cent frontiers using the change constrained programming ideas of Charnes-Cooper where $Ax_1^\alpha x_2^\beta \geq y$ would be translated into a probability statement, $P(Ax_1^\alpha x_2^\beta \geq y) \geq \tau$, with τ a specific minimum probability with which the statement is to hold”.

L'idea, quindi, è quella di costruire una frontiera probabilistica attraverso l'espressione $P(Ax_1^\alpha x_2^\beta \geq y) \geq \tau$ dove τ è una probabilità fissata esogenamente, per la quale vale la disuguaglianza.

Si può affermare che l'obiettivo di Timmer sia quello di superare le difficoltà connesse al fatto che le frontiere di produzione, secondo l'approccio di Aigner e Chu, risultano estremamente sensibili ai valori anomali, in quanto non considerano la presenza di disturbi casuali.

A tal fine, egli suggerisce di stimare inizialmente la frontiera, denominata **frontiera di produzione probabilistica** (*probabilistic production frontier*), minimizzando la somma dei residui, come in [3.4], ottenendo in tal modo le imprese efficienti, e di eliminare poi una percentuale uguale a $(1-\tau)$ delle prime unità efficienti, fino a che non venga raggiunto il livello specificato di τ .

Alternativamente, le osservazioni efficienti potrebbero essere eliminate una alla volta fino a quando i coefficienti stimati non si stabilizzino.

L'applicazione empirica è relativa alle aziende agricole statunitensi i cui processi sono osservati durante il periodo 1960-1967. La stima della funzione di produzione, che si configura come una funzione di produzione aggregata dell'agricoltura negli Stati Uniti, viene effettuata utilizzando l'*impresa media* in ogni stato e in ciascun anno come osservazione individuale, e quindi utilizzando una matrice dei dati di dimensione 8x48.

I risultati ottenuti, applicando gli OLS, confermano le precedenti stime di Griliches (1963), e vengono considerate la base di confronto⁶ per valutare l'attendibilità dei valori dei coefficienti derivanti dall'utilizzazione delle frontiere di produzione, ottenute applicando sugli stessi dati il metodo PL (Tabella 3.2).

| | <i>Metodo</i> | | | |
|---------------|---------------|-------------------|------------------|------------------|
| | OLS | LP ₁₀₀ | LP ₉₈ | LP ₉₇ |
| Costante | 1,7350 (53,8) | 1,6693 | 1,8578 | 1,8828 |
| Lavoro | 0,1919 (6,7) | 0,6015 | 0,3287 | 0,2679 |
| Capitale | 0,3726 (11,7) | 0,4887 | 0,3689 | 0,4842 |
| Terra | 0,0458 (4,2) | ... | 0,0298 | 0,0090 |
| Fertilizzanti | 0,1484 (16,0) | 0,1334 | 0,1428 | 0,1693 |
| Bestiame | 0,2510 (19,5) | 0,2347 | 0,2045 | 0,1885 |
| Semente | 0,1579 (5,4) | 0,1043 | 0,2243 | 0,1712 |

Tabella 3.2 – Risultati empirici in Timmer (1971) – stime della funzione di produzione.

Si può osservare che le stime della frontiera di produzione, ottenute attraverso il modello di programmazione lineare che coinvolge tutte le osservazioni, indicato con LP_{100} , sembrano produrre risultati privi di significato interpretativo, in particolare, in riferimento al valore molto grande del coefficiente del fattore lavoro (0,6015).

L'applicazione della frontiera di produzione probabilistica produce, invece, buoni risultati, determinando un elevato avvicinamento dei valori dei coefficienti stimati, dapprima, mediante l'eliminazione del 2% delle imprese efficienti (LP_{98}), e, successivamente, di un ulteriore 1% (LP_{97}), rispetto ai coefficienti della funzione media, ottenuta attraverso gli OLS. Differenze più elevate permangono nel valore della costante, che, come atteso, con il metodo LP_{97} risulta superiore di circa il 14% rispetto a quanto ottenuto applicando gli OLS. Per quanto riguarda il fattore lavoro, sebbene le elasticità assumano dei valori quasi identici, l'interpretazione non può essere sottoposta a verifica statistica.

⁶ In riferimento ai coefficienti ottenuti con gli OLS, i valori tra parentesi rappresentano le stime degli errori standard.

Come sottolinea Timmer, le stime dei coefficienti della funzione di produzione, pur avendo un significato in sé, rappresentano lo strumento necessario per la derminazione dell'efficienza tecnica per ciascuna impresa. A tal fine, vengono utilizzati i residui ottenuti dai modelli precedenti.

In particolare, i residui della funzione media, stimata applicando gli OLS, vengono spiegati, mediante un modello di regressione lineare, utilizzando un insieme di 48 variabili *dummy* che identificano gli stati coinvolti nell'analisi. Le misure di efficienza più soddisfacenti si dimostrano quelle derivanti dal modello LP_{98} , che risultano, come prevedibile, altamente correlate ($r=0,89$) con quelle derivanti dagli OLS.

A conferma di quanto detto e a fini illustrativi, nella Tabella 3.3, vengono riportati, per un limitato insieme di stati, alcune tra le misure ottenute.

| | Metodo | |
|--------------|-------------|-------------------|
| | Residui OLS | Residui LP_{98} |
| South Dakota | 0,1423 | 0,991 |
| Florida | 0,2117 | 0,978 |
| Illinois | 0,0362 | 0,963 |
| Nebraska | -0,0270 | 0,945 |
| Texas | -0,0296 | 0,921 |
| Ohio | -0,1250 | 0,880 |

Tabella 3.3 – Risultati empirici in Timmer (1971): Misure di efficienza tecnica.

Un aspetto interessante del lavoro di Timmer, comune al già citato studio di Seitz (1971), è rappresentato dal fatto che l'efficienza tecnica, a livello di singola impresa, viene analizzata attraverso un modello di regressione lineare al fine di individuare le possibili determinanti e trarne, di conseguenza, le dovute indicazioni per apportare miglioramenti nello svolgimento dei processi produttivi.

Si può affermare che il metodo applicato da Timmer, che costituisce una variante del "*change-constrained programming*" di Charnes e Cooper (1963), è diretto al superamento di uno dei limiti che caratterizzano i metodi di programmazione lineare, ossia l'estrema sensibilità alle osservazioni estreme. I risultati ottenuti da Timmer, ma anche quelli derivanti dall'applicazione di Aigner e Chu, dimostrano, infatti, l'elevata sensibilità delle stime ai dati osservati, ed in particolare alla presenza di osservazioni anomale. La mancanza di una possibile interpretazione economica della percentuale di osservazioni

che si posizionerebbero “fuori la frontiera”, come già anticipato, rappresenta però il maggior punto di debolezza dell’approccio delle frontiere probabilistiche.

Tale approccio lascia comunque irrisolta la principale limitazione connessa all’utilizzazione dei metodi di programmazione lineare per la stima delle funzioni frontiera. Non è possibile, infatti, effettuare le usuali procedure di inferenza statistica sul valore dei parametri stessi, in quanto, con essi, i parametri vengono *calcolati* piuttosto che *stimati*⁷, come accade, invece, nel caso delle funzioni medie attraverso l’impiego di tecniche di regressione (Tabella 3.2).

Come soluzione ai diversi problemi derivanti dall’utilizzazione delle frontiere deterministiche non statistiche, ottenute attraverso l’impiego dei metodi di programmazione, si sviluppa un approccio che utilizza tecniche statistiche per la stima della frontiera di produzione. Tale orientamento, suggerito inizialmente da Afriat (1972), e successivamente sviluppato, tra gli altri, da Richmond (1974) e Greene (1980), conduce alle cosiddette (Lovell e Schmidt, 1988) **frontiere deterministico statistiche**.

La caratteristica principale di quest’approccio è quella di esprimere, in termini generali, per ciascuna unità produttiva, l’output ottenibile da un dato vettore di input attraverso il seguente modello:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) \cdot \exp\{-u_i\} \quad [3.7]$$

che può essere, equivalentemente, formulato⁸ mediante l’espressione:

$$\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) - u_i \quad [3.8]$$

⁷ A tale critica risponde, in seguito, Schmidt (1976), sottolineando la circostanza per cui, se si ipotizza una determinata distribuzione probabilistica per u , l’approccio di Aigner e Chu può avere un’interpretazione statistica. In particolare, Schmidt afferma che nel caso in cui si ipotizzi che il termine u segua una distribuzione esponenziale, e si adotti un problema di programmazione lineare, ovvero si ipotizzi una distribuzione metà normale (*half normal*) e il problema impostato sia di programmazione quadratica, allora i valori calcolati rappresentano, in realtà, stime di massima verosimiglianza dei parametri della frontiera di produzione. In entrambi i casi, però, come sottolineato dallo stesso Autore, le proprietà statistiche degli stimatori di massima verosimiglianza non possono essere ottenute nei modi tradizionali in quanto risulta violata una delle condizioni di regolarità. In particolare, se il termine stocastico è distribuito unilateralmente non è rispettata la condizione che il campo di variazione della variabile casuale debba essere indipendente dai parametri da stimare.

⁸ Va osservato che Afriat (1972) e Richmond (1974) utilizzano una formulazione analoga esprime la frontiera di produzione come: $y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) \cdot u_i$

In relazione alle precedenti specificazioni, il termine di disturbo u_i , non negativo, rappresenta l'*inefficienza tecnica* di ciascuna unità produttiva considerata rispetto alla frontiera di produzione, mentre l'efficienza tecnica, che assume valori compresi tra zero e uno, è definita dalla seguente espressione:

$$ET_i = \exp\{-u_i\}.$$

Ipotizzando, ad esempio, una funzione di produzione Cobb-Douglas con s input, si ottiene la seguente specificazione della frontiera di produzione (Lovell e Schmidt, 1988):

$$y_i = A \prod_{j=1}^s x_{ij}^{\beta_j} \exp\{-u_i\} \quad [3.9]$$

dove il termine di disturbo u_i si distribuisce secondo una determinata v.c. di tipo unilaterale, come, ad esempio, una normale troncata o una esponenziale.

In particolare, al fine di individuare la corretta specificazione della frontiera di produzione, Afriat (1972) sottolinea l'opportunità di riconoscere la possibilità che i processi di produzione realmente osservati non raggiungano la perfetta efficienza a causa di fattori diversi, rappresentati, ad esempio, dalla mancanza di sufficienti informazioni o da improvvise modificazioni nei mercati dei fattori.

Attraverso sviluppi successivi basati su alcune proprietà connesse alla funzione di produzione, il suddetto Autore suggerisce, in assenza di ulteriori informazioni teoriche o evidenze empiriche, che facciano luce sul meccanismo probabilistico connesso alla forma della distribuzione del termine di efficienza tecnica, di considerare un semplice modello in cui la componente u viene descritta da una v.c. beta.

Afriat, inoltre, mette in risalto che nel caso in cui la generica forma funzionale $f(\mathbf{x}_i; \boldsymbol{\beta})$, utilizzata per la descrizione della tecnologia, venga specificata come una funzione Cobb-Douglas, risulta naturale considerare una distribu-

zione gamma per il termine $\log \frac{1}{u}$.

Per quanto riguarda la stima dei parametri del modello, il suddetto Autore suggerisce di utilizzare il metodo della massima verosimiglianza⁹.

⁹ È interessante sottolineare che Afriat introduce anche un approccio per verificare la coerenza dei dati analizzati con le ipotesi legate all'efficienza produttiva, che sarà ulteriormente sviluppato da Hanoch e Rothschild (1972). In particolare, osservando che i modelli parametrici per la descrizione dei processi di produzione impongono diverse restrizioni sulle caratteristiche

L'ipotesi avanzata da Afriat sulla possibilità di associare al termine di errore unilaterale, una particolare v.c., venne sviluppata ulteriormente da Richmond (1974), con riferimento ad una v.c. gamma, $G(u, n)$ con parametro n , in un modello in cui la tecnologia di produzione è descritta da una funzione Cobb-Douglas.

Richmond propose una misura del livello di efficienza media per il complesso delle unità produttive considerate, espressa da $z = \int_0^{\infty} \exp(-u) G(u, n) du$, che si riduce attraverso successive semplificazioni a $z = 2^{-n}$, e suggerì, al fine di evidenziare le caratteristiche della distribuzione dell'efficienza tra le diverse unità produttive, di utilizzare come indicatore la percentuale di unità con un livello di efficienza almeno uguale a c , dove c è una costante ($0 < c < 1$).

Il suddetto Autore, inoltre, riprendendo un concetto anticipato da Winsten (1957) nel suo commento al lavoro di Farrell, evidenzia che per stimare i parametri di una funzione di produzione di tipo Cobb-Douglas, con le ipotesi distributive precedenti sul termine di errore, possono essere utilizzati gli OLS, dai quali si ottengono stime corrette dei parametri ad eccezione dell'intercetta, la cui stima può, comunque, essere ottenuta utilizzando i residui del modello.

Tale metodo, denominato, come già anticipato, *COLS*, costituisce la tecnica più utilizzata per stimare i parametri delle frontiere deterministico - statistiche.

Il metodo COLS consiste nello stimare dapprima la funzione di produzione mediante gli OLS, ottenendo una funzione media, per poi passare alla frontiera di produzione, correggendo il termine costante attraverso l'aggiunta della media del termine u . In termini formali, il modello [3.9] può essere riscritto, utilizzando i logaritmi, nel modo seguente:

$$\ln y_i = \alpha + \sum_{j=1}^s \beta_j \ln x_{ij} - u_i$$

Indicando con μ la media di u_i , si ottiene:

di base della funzione di produzione, come la concavità e la monotonicità, che non possono essere sottoposte a verifica all'interno del contesto delle procedure di stima, Hanoch e Rothschild propongono l'utilizzazione di test basati sulla programmazione lineare e che non derivano da ipotesi parametriche sulla forma della funzione di produzione. Lo studio rappresenta, quindi, un tentativo di fornire strumenti per comprendere la natura e le potenzialità dei dati in una situazione di completa assenza di formali procedure statistiche per la verifica delle ipotesi, che caratterizza, appunto, l'approccio parametrico deterministico.

$$\ln y_i = (\alpha - \mu) + \sum_{j=1}^s \beta_j \ln x_{ij} - (u_i - \mu) \quad [3.10]$$

dove il nuovo termine $(\mu - u_i)$ ha media zero. L'equazione [3.10] può essere stimata utilizzando gli OLS per ottenere le migliori stime lineari non distorte di $(\alpha - \mu)$ e dei parametri β_j . La media μ può essere stimata in modo non distorto e consistente a partire dai momenti centrali dei residui degli OLS, dopo aver ipotizzato una distribuzione specifica per il termine u . Quindi, aggiungendo il valore stimato per μ al termine $(\alpha - \mu)$ si procede alla "correzione" della stima del termine costante.

Il metodo COLS presenta però un grave inconveniente, ossia non è in grado di garantire che, dopo la correzione, la funzione di produzione si sia spostata sufficientemente affinché tutte le osservazioni siano sotto o sulla frontiera. Un ulteriore limite di tale metodo è rappresentato dal fatto che la stima del termine μ dipende dal tipo di v.c. ipotizzata per la componente di inefficienza.

Per superare tali inconvenienti si può ricorrere al metodo suggerito da Greene (1980), denominato, come anticipato, MOLS, nel quale, per la correzione del termine costante, si sostituisce alla stima della media μ , il residuo OLS positivo più elevato, cosicché il termine costante viene aumentato fino a quando nessun residuo risulta positivo, ed almeno uno è uguale a zero.

In particolare, sottolineando la valenza teorica delle frontiere di produzione¹⁰, in quanto riproducono il concetto di funzione di produzione neoclassico consentendo l'analisi di situazioni non efficienti che possono verificarsi nello svolgimento del processo produttivo, il suddetto Autore concentra la propria attenzione sui metodi di stima di tipo statistico. In relazione al metodo della massima verosimiglianza (*Maximum Likelihood, ML*), ritiene che i problemi teorici relativi alle proprietà statistiche degli stimatori ML possano essere superati nel caso in cui la distribuzione del termine di efficienza soddisfi determinati criteri.

Ipotizzando una distribuzione gamma per u , che consente di garantire il soddisfacimento delle usuali proprietà di regolarità, Greene suggerisce, quindi, anche l'utilizzazione del metodo ML, che conduce a stime consistenti e asintoticamente efficienti dei parametri delle frontiere di produzione deterministico-statistiche.

¹⁰ Greene (1980) si riferisce all'approccio delle frontiere di produzione, introdotte da Aigner e Chu, con il termine di modelli a frontiera piena (*full frontier*) per evidenziare i vincoli imposti sul termine di errore o disturbo.

Da quanto finora esposto, l'adozione di quest'ultimo approccio all'analisi dell'efficienza rispetto alla frontiera, sembra configurarsi come la soluzione "ideale" ai diversi problemi in precedenza emersi in letteratura, legati sia al concetto economico di funzione di produzione sia alla stima statistica del modello.

Va osservato, invece, che, già da tempo, grazie al lavoro di Aigner, Amemiya e Poirier (1976), si rifletteva sul fatto che le frontiere di produzione deterministico statistiche presentassero una grave limitazione, ossia quella di attribuire la totalità della variabilità dell'output delle unità produttive al diverso grado di efficienza che le stesse raggiungono nello svolgimento del processo di produzione. Una tale limitazione è diretta conseguenza della specificazione di un'unica v.c. per gli effetti aleatori, relativa al termine di efficienza u .

Aigner, Amemiya e Poirier, sottolineano, infatti, che, nel contesto della funzione di produzione dell'industria, una tale procedura può risultare corretta solo nel caso in cui la fonte delle differenze tra l'output prodotto dalle imprese sia pienamente riconducibile alla diversa abilità dei produttori nell'utilizzare la tecnologia migliore (*best practice*).

Se, invece, come è ragionevole ritenere, agiscono sulla produzione anche fattori casuali simmetrici, rappresentati, ad esempio, da errori di misurazione sul prodotto o da fattori ambientali esterni, allora la distribuzione del termine di errore non può essere unilaterale.

I suddetti Autori propongono un modello che prevede la possibilità di considerare sia errori con segno negativo sia errori con segno positivo, attraverso l'adozione di una struttura di pesi sulle osservazioni. In particolare, in riferimento al modello, espresso in forma matriciale, mediante $y = X\beta + \varepsilon$, dove y è il vettore di n osservazioni sulla variabile dipendente ed X è una matrice di osservazioni relative a s regressori, gli Autori ipotizzano che ogni elemento del vettore di disturbo, ε , sia espresso da:

$$\varepsilon_i = \begin{cases} \varepsilon_i^* / \sqrt{(1-\theta)} & \text{se } \varepsilon_i^* > 0 \quad i = 1, \dots, n \\ \varepsilon_i^* / \sqrt{\theta} & \text{se } \varepsilon_i^* \leq 0 \end{cases}$$

dove gli errori, ε_i^* , sono distribuiti indipendentemente secondo una v.c. normale con media nulla e varianza costante per $0 < \theta < 1$; altrimenti ε_i^* si distribuisce secondo una v.c. normale troncata negativa, se $\theta = 1$, o secondo una v.c. normale troncata positiva se $\theta = 0$.

Il parametro θ rappresenta una misura della variabilità relativa nelle due fonti di errore, ossia dell'incapacità ad utilizzare la migliore tecnologia disponibile e dell'agire di fattori casuali.

I problemi connessi alla stima di massima verosimiglianza dei parametri della frontiera di produzione che deriva dal modello sono superati attraverso la soluzione di un problema equivalente, che consiste nel minimizzare l'espressione:

$$S = \theta \sum_{\varepsilon_i \leq 0} \varepsilon_i^2 + (1 - \theta) \sum_{\varepsilon_i > 0} \varepsilon_i^2 .$$

Il fatto che, nella funzione precedente, si ritrovino come casi particolari, il modello della funzione media, se $\theta = \frac{1}{2}$, e il modello delle frontiere di produzione deterministiche, se $\theta = 0$ o $\theta = 1$, evidenzia il tentativo dei suddetti Autori di trovare un giusto compromesso tra le due situazioni estreme, superando in tal modo i limiti dell'uno e dell'altro approccio.

Aigner, Amemiya e Poirier ricavano gli stimatori dei parametri nel caso in cui i pesi, e quindi i termini θ , siano noti, ma anche per il caso, notevolmente più complesso, nel quale il valore di θ non è noto e deve, perciò, essere stimato insieme agli altri parametri del modello.

3.3 LE FRONTIERE DI PRODUZIONE STOCASTICHE, OVVERO COME SUPERARE I LIMITI DELLE FRONTIERE DETERMINISTICHE

Si può affermare che nelle **frontiere di produzione stocastiche** confluiscono, trovando un ragionevole equilibrio, due aspetti essenziali che caratterizzavano gli sviluppi metodologici precedenti relativi all'individuazione e alla stima della frontiera di produzione.

Al pari dell'approccio delle frontiere deterministico-statistiche, l'approccio delle frontiere stocastiche utilizza tecniche statistiche per stimare una frontiera di produzione, ottenuta aggiungendo ad una funzione di produzione, che descrive la tecnologia di riferimento, un termine di disturbo, che esprime l'efficienza tecnica raggiunta da ciascuna unità produttiva nello svolgimento del processo, rappresentato da una v.c. con caratteristiche particolari, ossia tali da garantire che tutte le osservazioni giacciono sotto o, al più, sulla frontiera.

Analogamente all'approccio delle funzioni medie, e a differenza di quello delle frontiere deterministiche, l'approccio stocastico, introduce esplicitamente un termine di disturbo simmetrico che cattura gli effetti dovuti ad errori di misurazione e ad altri fattori casuali non controllabili dall'unità produttiva.

In tal modo, i due diversi aspetti dei fattori di "disturbo" che esercitano la propria azione sulla relazione che lega gli input impiegati agli output prodotti,

evidenziati da Marschak e Andrews (1944) come “*error of measurement*” e “*technical efficiency*”, possono essere tenuti in adeguata considerazione, consentendo, al tempo stesso, di coniugare la teoria economica con il comportamento dell’impresa.

L’introduzione dell’approccio stocastico alle frontiere di produzione si deve ad Aigner, Lovell e Schmidt (1977) e Meeusen e van den Broeck (1977). Un ulteriore contributo è rappresentato da Battese e Corra (1977) i quali intervengono, in particolare, suggerendo una specifica parametrizzazione che dovrebbe facilitare l’interpretazione dei risultati.

Sebbene i suddetti Autori utilizzino una diversa notazione e facciano riferimento ad ipotesi diverse in relazione alla v.c. che descrive il termine di inefficienza u , l’essenza dell’approccio proposto è la medesima, per cui si procederà ad un’unica esposizione, salvo evidenziare le peculiarità che caratterizzano i diversi lavori nel caso specifico.

In termini formali, la frontiera di produzione stocastica può essere espressa da:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) \cdot \exp\{v_i\} \cdot \exp\{-u_i\} \quad [3.11]$$

ovvero, ricorrendo ai logaritmi, mediante l’espressione:

$$\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i - u_i \quad [3.12]$$

Il termine di errore complessivo, espresso per ciascuna unità produttiva i , ($i = 1, \dots, n$) da $\varepsilon_i = v_i - u_i$, è composto da due elementi supposti indipendenti: un termine di inefficienza u_i , distribuito unilateralmente, ed un termine di errore aleatorio v_i che, oltre agli effetti derivanti sia da *shock* di carattere esogeno che da eventuali errori di misurazione, riproduce anche i disturbi contenuti in qualsiasi relazione empirica. In altre parole, alle relazioni [3.7] e [3.8] si aggiunge un termine di errore aleatorio v simmetrico, indipendentemente ed identicamente distribuito secondo una v.c. normale $N(0, \sigma_v^2)$.

L’elemento innovativo nel modello precedente consiste nel fatto che, essendo $y_i \leq f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i$, si consente alla frontiera di produzione stessa di variare casualmente, sia tra le unità, in un dato momento temporale (*cross-section*), sia nel tempo, per la stessa unità produttiva (*time series*). In base a quest’ultima interpretazione, quindi, la frontiera è stocastica con disturbi casuali positivi o negativi che sono il risultato di eventi esterni favorevoli o sfavorevoli, che vanno oltre il controllo delle imprese.

Dopo aver specificato una forma funzionale per la relazione che lega gli input all’output, e ipotizzato una particolare v.c. per il termine u , le stime dei parametri del modello [3.12] possono essere ottenute sia con il metodo COLS che con

il metodo ML. Sebbene il metodo COLS sia più facile da implementare¹¹, le stime ottenute dalla ML risultano asintoticamente più efficienti.

Aigner, Lovell e Schmidt ipotizzano¹² che la componente di errore u_i sia distribuita indipendentemente da v_i e soddisfi la condizione $u_i \geq 0$. Pur sottolineando la possibilità di impiegare diverse v.c. per la descrizione di u , tra cui anche la distribuzione esponenziale:

$$f(u) = \frac{2}{\sigma_u} \exp\left\{-\frac{u}{\sigma_u}\right\},$$

i suddetti Autori concentrano l'attenzione sulla v.c. metà normale (*Half Normal*), espressa da:

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\}$$

con media e varianza rappresentate, rispettivamente da:

$$E(u) = \sqrt{\frac{2}{\pi}}\sigma_u \quad \text{e} \quad \text{Var}(u) = \left[\frac{\pi-2}{\pi}\right]\sigma_u^2.$$

A partire dalla funzione di ripartizione che caratterizza la somma di una v.c. normale e di una v.c. metà normale, ottenuta da Weinstein (1964), si deriva agevolmente la funzione di densità del termine di errore complessivo ε .

Infatti, grazie all'ipotesi di indipendenza, la funzione di densità congiunta di u e v , $f(u, v)$, è facilmente ottenibile come prodotto delle due densità individuali. La densità marginale di ε è ottenibile per integrazione in u della densità congiunta $f(u, \varepsilon)$, ottenuta sostituendo a v , l'espressione $v = \varepsilon + u$. Si ha quindi:

$$f(\varepsilon) = \int_0^{\infty} f(u, \varepsilon) du = \frac{2}{\sigma} \phi\left(\frac{\varepsilon}{\sigma}\right) \left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \quad -\infty < \varepsilon < \infty \quad [3.13]$$

dove $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = \sigma_u/\sigma_v$ e ϕ e Φ sono rispettivamente la funzione di densità e di ripartizione della v.c. normale standardizzata.

¹¹ La disponibilità di software econometrici, come Limdep, Frontier 4.1 e Stata, che consentono una rapida massimizzazione numerica della funzione di verosimiglianza, ha ridotto i vantaggi del metodo COLS rispetto alla ML.

¹² Va osservato che Aigner, Lovell e Schmidt esprimono il modello della frontiera stocastica secondo la formulazione equivalente: $y_i = f(x_i; \beta) + \varepsilon_i$, dove $\varepsilon_i = v_i + u_i$ con $u_i \leq 0$.

Tale densità è asimmetrica con media e varianza espresse, rispettivamente, da:

$$E(\varepsilon) = -E(u) = -\frac{\sqrt{2}}{\sqrt{\pi}}\sigma_u$$

e

$$\text{Var}(\varepsilon) = \text{Var}(u) + \text{Var}(v) = \left(\frac{\pi - 2}{\pi}\right)\sigma_u^2 + \sigma_v^2$$

È interessante osservare che la densità [3.13] può essere ricondotta alla v.c. *skew normal* introdotta da Azzalini (1985, 1986) ed espressa da:

$$f(x) = 2\phi(x)\Phi(\delta x) \quad \delta \in \Re.$$

Ponendo, infatti, $\varepsilon = \alpha x$ dove x è una v.c. *skew normal* con $\delta = -\lambda$, Dominguez-Molina *et al.* (2004) evidenziano il legame esistente tra le due v.c. e le potenzialità applicative della distribuzione suggerita da Azzalini nell'analisi dell'efficienza.

Aigner, Lovell e Schmidt¹³ sottolineano che la parametrizzazione del termine di errore complessivo espressa dalla [3.13], risulta particolarmente conveniente in quanto λ può essere interpretato come un indicatore della variabilità relativa delle due fonti di errore, sebbene σ_u^2 non rappresenti, esattamente, la varianza di u .

Dal comportamento del parametro λ si può dedurre il peso relativo delle due fonti di errore nella determinazione della frontiera di produzione. Si ha, infatti, che se $\lambda \rightarrow 0$ allora $\sigma_v^2 \rightarrow \infty$ e/o $\sigma_u^2 \rightarrow 0$, per cui la componente di errore simmetrico domina nella determinazione dell'errore complessivo ε . In tal caso la funzione di densità di ε si trasforma una v.c. $N(0, \sigma^2)$. D'altra parte, quando $\sigma_v^2 \rightarrow 0$ il termine di errore unilaterale diventa la fonte dominante di variazione casuale nel modello e l'errore complessivo, per $\varepsilon \leq 0$ segue una distribuzione metà normale.

Un'ulteriore riflessione concerne il fatto per cui il modello proposto si riduce al modello della frontiera deterministica nel caso in cui $\sigma_v^2 = 0$ e al modello introdotto da Drèze, Kmenta e Zellner (1966) se $\sigma_u^2 = 0$.

Considerando un campione casuale di n osservazioni, in riferimento ad un modello lineare del tipo $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$, con $\varepsilon_i = v_i - u_i$, si ottiene la funzione di log-verosimiglianza, espressa da:

¹³ Aigner, Lovell e Schmidt suggeriscono inoltre che un'ulteriore parametrizzazione della frontiera stocastica sarebbe ottenibile a partire dalla densità di ε utilizzando σ^2 insieme al parametro $\alpha = \sigma_u^2 / \sigma^2$.

$$\ln L(y|\boldsymbol{\beta}, \lambda, \sigma^2) = \sum_{i=1}^n \ln [f_i(\varepsilon_i)],$$

che assume una forma quasi identica a quella derivata da Amemiya (1973), ossia:

$$\ln L(y|\boldsymbol{\beta}, \lambda, \sigma^2) = \frac{n}{2} \ln \frac{2}{\pi} - n \ln \sigma + \sum_{i=1}^n \ln \left[1 - \Phi \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right] - \frac{1}{2\sigma^2} \sum_{i=1}^n \varepsilon_i^2 \quad [3.14]$$

e che può essere massimizzata allo scopo di ottenere stime di $\boldsymbol{\beta}$, σ_u^2 e σ_v^2 .

I vantaggi apportati dall'utilizzazione delle nuove frontiere stocastiche vengono posti in evidenza da Aigner, Lovell e Schmidt, attraverso due applicazioni empiriche su dati che avevano costituito l'oggetto, rispettivamente, del lavoro di Aigner e Chu (1968) e di Timmer (1971).

La specificazione della funzione di produzione stimata dai tre suddetti Autori, è la seguente: $\ln Y = \beta_0 + \beta_1 \ln L + \beta_2 (\ln R \cdot \ln K) + (v - u)$.

Nella Tabella 3.4 vengono riportati i risultati delle stime della frontiera stocastica¹⁴, secondo due diverse specificazioni della forma distributiva del termine di errore unilaterale (esponenziale e metà normale), che possono essere confrontati con i valori risultanti dall'applicazione della funzione media, stimata attraverso gli OLS, e con i risultati ottenuti da Aigner e Chu (Tabella 3.1). Va osservato che Aigner, Lovell e Schmidt riportano anche i risultati ottenuti in precedenza da Hildebrand-Liu, sebbene i dati su cui sono basate le stime, pur relativi al medesimo settore industriale, siano riferiti al periodo 1956-57.

L'elemento interessante da rilevare è che i parametri della frontiera stocastica, stimati attraverso il metodo ML, in entrambe le formulazioni, ottenute specificando due distinte v.c. per il termine u , risultano molto più vicini ai valori ottenuti per la funzione media, stimata mediante gli OLS, piuttosto che ai valori dei parametri della frontiera deterministica, calcolati applicando i metodi di programmazione.

L'analisi delle stime dei parametri delle distribuzioni dei termini di errore consente di concludere che, nel caso esaminato, la componente di errore simmetrico sovrasta la componente unilaterale che esprime l'efficienza tecnica. In altre parole, si può affermare che esiste una sostanziale variabilità della frontiera tra gli stati ma la variabilità dei punti osservati al di sotto della frontiera è piut-

¹⁴ I numeri tra parentesi esprimono i valori del t ratio, nel caso delle stime OLS, e dei t ratio asintotici, nel caso di stime ML.

| Metodo | β_0 | β_1 (Elasticità del lavoro) | β_2 |
|--|------------------|--------------------------------------|-------------------|
| Funzione media | 0,9146 (2,04) | 0,9168 (7,31) | 0,04164 (2,19) |
| Frontiera stocastica <i>u metà normale</i> | 0,9600 (2,06) | 0,9105 (7,68) | 0,04125 (2,29) |
| Frontiera stocastica <i>u</i> <i>esponenziale</i> | 0,9601 (2,20) | 0,9144 (7,71) | 0,04208 |
| Hildebrand-Liu OLS | | 0,988 | 0,0421 |
| Aigner e Chu OLS | | 0,908 | 0,0333 |
| Aigner e Chu PL | | 0,873 | 0,0031 |
| Aigner e Chu PQ | | 1,071 | 0,0269 |

Tabella 3.4 – Risultati empirici in Aigner, Lovell e Schmidt (1977):
prima applicazione.

tosto limitata. Infatti, rispetto alla stima della varianza del termine di errore casuale nel caso OLS, che è uguale a 0,07764, nel caso metà normale, le stime della varianza dei due termini di errore, simmetrico e unilaterale, assumono rispettivamente i valori $\hat{\sigma}_v^2 = 0,0692$ e $\hat{\sigma}_u^2 = 0,000686$ mentre nel caso esponenziale, si ottiene $\hat{\sigma}_u^2 = 0,0180$ e $\hat{\sigma}_v^2 = 0,0691$.

La seconda applicazione delle frontiere stocastiche effettuata da Aigner, Lovell e Schmidt riguarda i medesimi dati utilizzati da Timmer (1971) relativi al settore agricolo negli Stati Uniti¹⁵. La funzione di produzione da stimare assume la specificazione seguente:

$$\ln Y = \beta_0 + \beta_1 \ln \text{Lavoro} + \beta_2 \text{Capitale} + \beta_3 \text{Terra} + \beta_4 \text{Fertilizzante} + \beta_5 \text{Bestiame} + \beta_6 \text{Semente} + (v - u)$$

La Tabella 3.5 presenta le stime ottenute da Aigner, Lovell e Schmidt utilizzando un modello in cui il termine di disturbo unilaterale u segue una distribuzione esponenziale. Sono riportati, inoltre, i risultati derivanti dall'applicazione della funzione media, stimata attraverso gli OLS e quelli ottenuti dall'utilizzazione della frontiera di produzione probabilistica introdotta da Timmer (Tabella 3.2).

¹⁵ Rispetto al lavoro di Timmer (1971), Aigner, Lovell e Schmidt escludono dal data set gli anni 1966 e 1967 e, quindi, complessivamente 96 osservazioni.

| | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|---|------------------|------------------|------------------|-------------------|-------------------|-------------------|------------------|
| OLS | 1,8072 | 0,1149 (3,67) | 0,2976 (8,85) | 0,06061 (4,81) | 0,1411 (12,40) | 0,2581 (16,72) | 0,1956 (6,44) |
| Frontiera stocastica <i>u esponenziale</i> | 1,8143 (42,9) | 0,1148 (3,71) | 0,2776 (8,96) | 0,06061 (4,87) | 0,1411 (12,56) | 0,2581 (16,93) | 0,1956 (6,52) |
| Timmer OLS | 1,7350 (53,8) | 0,1919 (6,7) | 0,3726 (11,7) | 0,0458 (4,2) | 0,1484 (16,0) | 0,2510 (19,5) | 0,1579 (5,4) |
| Timmer LP100 | 1,6693 | 0,6015 | 0,4887 | ... | 0,1334 | 0,2347 | 0,1043 |
| Timmer LP98 | 1,8578 | 0,3287 | 0,3689 | 0,0298 | 0,1428 | 0,2045 | 0,2243 |
| Timmer LP97 | 1,8828 | 0,2679 | 0,4842 | 0,0090 | 0,1693 | 0,1885 | 0,1712 |

Tabella 3.5 – Risultati empirici in Aigner, Lovell e Schmidt (1977):
seconda applicazione.

I risultati ottenuti appaiono estremamente interessanti. I valori stimati dei parametri del modello stocastico sono molto simili a quelli derivanti dalla stima della funzione di produzione media, suggerendo la sostanziale assenza, confermata peraltro dalle stime dei parametri delle distribuzioni dei termini di errore, dell'effetto esercitato dal termine di disturbo unilaterale, rappresentante l'efficienza tecnica.

Si può osservare, quindi, che nei casi in cui forte è la presenza di fattori di disturbo puramente casuali, provocati da *shock* esterni al processo o da errori di misurazione, la stima di una frontiera deterministica conduce a risultati fuorvianti, e, talvolta, privi di un'interpretazione economica coerente. L'effetto dei disturbi puramente aleatori ricade completamente sul termine unilaterale, l'unico presente nei modelli della frontiera deterministica, e si traduce, non solo in un'errata definizione della misura dell'efficienza tecnica, ma anche, talora, in una funzione di produzione stimata i cui parametri appaiono privi di significato.

L'applicazione della frontiera di produzione probabilistica di Timmer sembra risolvere in parte il problema, presumibilmente legato alla presenza di valori anomali, e produce stime alquanto vicine a quelle ottenute con l'applicazione della frontiera stocastica.

Riprendendo la trattazione teorica dell'approccio stocastico alle frontiere di produzione, appare interessante a questo punto sottolineare che Meeusen e van den Broeck (1977), diversamente da Aigner, Lovell e Schmidt, suggeriscono il modello [3.11] specificando una funzione di produzione Cobb-Douglas, e prediligendo, per il termine di inefficienza, la v.c. esponenziale, in quanto dotata di proprietà notevoli sia dal punto di vista teorico che computazionale.

La funzione di densità marginale dell'errore complessivo¹⁶, considerando che $v \sim N(0, \sigma_v^2)$, assume la forma seguente:

$$f(\varepsilon) = \int_0^{\infty} f(u, v) du = \left(\frac{1}{\sigma_u} \right) \Phi \left(-\frac{\varepsilon}{\sigma_v} - \frac{\sigma_u}{\sigma_v} \right) \exp \left\{ \frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\}$$

dove Φ è la funzione di ripartizione della v.c. normale standardizzata. La v.c. $f(\varepsilon)$ è distribuita asimmetricamente con media e varianza date, rispettivamente, da: $E(\varepsilon) = -E(u) = -\sigma_u$ e $Var(\varepsilon) = \sigma_u^2 + \sigma_v^2$.

La funzione di log-verosimiglianza per un campione di n unità produttive può essere scritta (Kumbhakar e Lovell, 2000) nel modo seguente:

$$\ln L = \text{costante} - n \ln \sigma_u + n \left(\frac{\sigma_v^2}{2\sigma_u^2} \right) + \sum_i \Phi(-A) + \sum_i \frac{\varepsilon_i}{\sigma_u}$$

dove $A = -\frac{\tilde{\mu}}{\sigma_v}$ e $\tilde{\mu} = -\varepsilon - \left(\frac{\sigma_v^2}{\sigma_u} \right)$

Un ulteriore aspetto distintivo del lavoro di Meeusen e van den Broeck è rappresentato dalla particolare attenzione dedicata dagli Autori alla problematica connessa all'identificazione della misura di efficienza. In particolare, per fornire informazioni relative alla distribuzione dei livelli di efficienza tra le imprese, i suddetti Autori suggeriscono di utilizzare una misura di efficienza media¹⁷ e di determinare la proporzione di unità con efficienza produttiva almeno uguale a qualche valore dato κ .

L'applicazione empirica è relativa ai dati del censimento francese delle industrie manifatturiere, da cui vengono selezionate 10 industrie al cui interno sono

¹⁶ Va osservato che Meeusen e van den Broeck (1977) specificano l'espressione seguente:

$$f_w(w) = \int_{-\infty}^{\infty} f_z(z) f_v(w-z) dz = \frac{\lambda}{2} \exp \left(\frac{\sigma^2 \lambda^2}{2} - \lambda w \right) \operatorname{erfc} \left(\frac{\sigma^2 \lambda^2 - w}{\sigma \sqrt{2}} \right)$$

nella quale $W=Z+V$ dove z è il termine di efficienza, distribuito secondo una v.c. esponenziale e v è l'errore casuale distribuito normalmente con media nulla e varianza costante.

¹⁷ In riferimento alla formulazione adottata da Meeusen e van den Broeck, la misura di efficienza media è ottenuta attraverso: $E(K) = \frac{\lambda}{\lambda+1}$, dove $K = \exp(-z)$ con distribuzione espressa da $f(K) = \lambda k^{\lambda-1}$ per $0 < k \leq 1$.

presenti un numero variabile di imprese (si va dalle 953 imprese per macchine e attrezzature meccaniche alle 135 per i prodotti del vetro).

Meeusen e van den Broeck sottolineano che, malgrado la distribuzione del termine di errore composto non sia normale, i coefficienti della funzione di produzione Cobb-Douglas non differiscono significativamente da quelli ottenuti stimando la funzione stessa con le procedure standard (OLS). L'unica differenza sostanziale, come atteso, consiste in uno spostamento verso l'alto dell'intercetta.

Il risultato fondamentale che deve essere sottolineato e che ricalca quanto già emerso dall'applicazione di Aigner, Lovell e Schmidt è che la maggior parte della varianza della distribuzione del termine di errore composto può essere attribuita ad errori puramente casuali e, perciò, non può essere attribuita all'inefficienza. Per illustrare la dimensione della "distorsione" introdotta nella misura di efficienza tecnica dall'inclusione dei fattori puramente casuali nell'unico termine di errore considerato nel modello [3.7] o [3.8], la Tabella 3.6 riporta, per i dieci settori considerati da Meeusen e van den Broeck, i valori dell'efficienza media del modello dell'errore composto e i valori delle efficienze calcolate secondo il metodo suggerito da Richmond (1974).

Emerge chiaramente che i valori delle efficienze ottenute secondo l'impostazione di Richmond, ossia adottando una frontiera deterministico statistica, producono risultati che non possono essere accettati come plausibili. Infatti, le efficienze medie assumono dei valori troppo bassi, non ammissibili dal punto di vista economico, in quanto si può ragionevolmente presumere che, in mercati concorrenziali, le imprese meno efficienti vengano automaticamente eliminate.

Come già anticipato, Battese e Corra (1977) propongono una diversa parametrizzazione della funzione di log verosimiglianza [3.14] del modello della frontiera stocastica che può essere utile ai fini computazionali.

I suddetti Autori suggeriscono l'utilizzazione del parametro $\gamma = \sigma_u^2 / \sigma^2$, che assumendo valori compresi tra zero e uno, consente un'interpretazione più immediata dei risultati ottenuti dalla stima della frontiera stocastica, con riferimento alla prevalenza degli effetti di un tipo di errore rispetto all'altro.

In particolare, in riferimento alla frontiera di produzione $y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$, dove le due componenti di errore, v e u sono descritti, rispettivamente, da una v.c. normale e da una v.c. metà normale, ossia $v \sim N(0, \sigma_v^2)$ e $u \sim N^+(0, \sigma_u^2)$, la funzione di log-verosimiglianza per un campione di n osservazioni assume la forma seguente:

$$\ln L = -\frac{n}{2} \ln \frac{\pi}{2} - \frac{n}{2} \ln \sigma^2 + \sum_{i=1}^N \ln [1 - \Phi(z_i)] - \frac{1}{2\sigma^2} \sum_{i=1}^N (\ln y_i - x_i \boldsymbol{\beta})^2$$

| <i>Industria</i> | Afriat –Richmond Gamma | Frontiera stocastica <i>u</i> esponenziale |
|---|---------------------------|---|
| 1.Prodotti del vetro | 0,5697 | 0,9435 |
| 2.Prodotti del latte | 0,6333 | 0,9265 |
| 3.Tessile | 0,5183 | 0,9089 |
| 4.Costruzione di macchine e strumenti meccanici | 0,6068 | 0,9036 |
| 5.Macchinari elettrici | 0,5565 | 0,8595 |
| 6.Veicoli e ciclomotori | 0,5789 | 0,8249 |
| 7.Industria chimica | 0,4987 | 0,7901 |
| 8.Carta | 0,5864 | 0,7888 |
| 9.Calzature | 0,7173 | 0,7560 |
| 10.Distillerie e bevande | 0,4686 | 0,7069 |

Tabella 3.6 – Risultati empirici in Meeusen e van den Broeck (1977): efficienza produttiva.

dove: $z_i = \frac{(\ln y_i - x_i \beta)}{\sigma} \sqrt{\frac{\gamma}{1-\gamma}}$, $\gamma = \frac{\sigma_u^2}{\sigma^2}$ e $\Phi(\cdot)$ è la funzione di ripartizione della v.c. normale standardizzata.

Le deviazioni dalla frontiera possono essere attribuite interamente a disturbi casuali fuori il controllo dell'impresa nel caso in cui $\gamma = 0$, mentre derivano completamente dal mancato raggiungimento dell'efficienza tecnica nello svolgimento del processo nel caso in cui $\gamma = 1$.

È opportuno osservare che in nessuno dei lavori precedentemente esaminati, che possono considerarsi le pietre miliari nella letteratura delle frontiere stocastiche, si è affrontato il problema di determinare, in forma rigorosa, una misura di efficienza a livello di singola impresa. L'obiettivo essendo solo quello di evidenziare i limiti dei modelli di frontiera piena e fornire una soluzione per il loro superamento.

3.4 LA MISURA DI EFFICIENZA TECNICA NEI MODELLI STOCASTICI

Da quanto finora esposto, è evidente che un aspetto fondamentale, più o meno esplicito, che caratterizza la letteratura delle frontiere di produzione, a partire dal lavoro di Farrell (1957), risiede proprio nell'identificazione di una misura di efficienza per ogni singola impresa rispetto ad uno standard ottimo

che rispecchi il più possibile le effettive condizioni tecnologiche ed ambientali in cui si svolge il processo di produzione.

Nel contesto dell'approccio stocastico, una tale esigenza si traduce nell'analisi del termine di disturbo u , che esprime esclusivamente l'inefficienza tecnica¹⁸. Infatti, come sottolineato da Aigner, Lovell e Schmidt, grazie all'approccio delle frontiere stocastiche, l'inefficienza può, in linea di principio, essere determinata mediante il rapporto $y_i / [f(\mathbf{x}_i; \boldsymbol{\beta}) \exp\{v_i\}]$, invece che attraverso il rapporto $y_i / [f(\mathbf{x}_i; \boldsymbol{\beta})]$ in precedenza utilizzato, consentendo in tal modo di distinguere il mancato raggiungimento dell'efficienza tecnica da altre fonti di disturbo che, essendo non controllabili, non possono essere attribuite all'agire del produttore nello svolgimento del processo.

Sebbene l'interesse verta sull'errore unilaterale, la difficoltà che caratterizza l'approccio stocastico è rappresentata dall'impossibilità di osservare direttamente la componente u , in quanto i residui ottenuti dalla stima del modello [3.12], ossia $\ln y_i - \ln f(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$, possono essere considerati come stime del termine di errore complessivo ε e non di u .

È evidente che il residuo complessivo contiene informazioni sull'errore unilaterale, ma la questione è *come estrarre queste informazioni*. In altre parole, il problema consiste nello scomporre i residui complessivi in due stime separate della componente di errore u_i dovuta all'inefficienza tecnica, e della componente v_i , dovuta a disturbi puramente casuali.

Una soluzione a questo problema viene fornita da Jondrow, Lovell, Materov e Schmidt (1982).

Riflettendo sul fatto che i residui stimati $\hat{\varepsilon}_i$ contengono informazioni su u , i suddetti Autori concentrano l'attenzione sulla distribuzione condizionata di u dato ε , $f(u|\varepsilon)$, dimostrando che, nel caso $u_i \sim N^+(0, \sigma_u^2)$, ossia nel caso si ipotizzi una v.c. metà normale per il termine unilaterale u , tale distribuzione assume la forma seguente:

$$f(u|\varepsilon) = \frac{f(u, \varepsilon)}{f(\varepsilon)} = \frac{1}{1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)} \cdot \frac{1}{\sqrt{2\pi\sigma_*}} \exp\left[-\frac{1}{2\sigma_*^2} \left(u + \frac{\sigma_u^2 \varepsilon}{\sigma^2}\right)^2\right] \quad [3.15]$$

¹⁸ Appare opportuno sottolineare, nuovamente, che in questo lavoro l'attenzione si concentra, in particolare, sulla definizione e sui metodi di stima dell'efficienza tecnica. Quest'ultimo concet-

dove $\sigma_*^2 = \frac{\sigma_u^2 \sigma_v^2}{\sigma^2}$ e $u \geq 0$. Ad eccezione del termine che coinvolge il fattore $1 - \Phi$,

l'espressione [3.15] somiglia alla densità di una v.c. $N(\mu_*, \sigma_*^2)$, con $\mu_* = -\frac{\sigma_u^2 \varepsilon}{\sigma^2}$.

Osservando che la funzione di ripartizione Φ è valutata in $\frac{\varepsilon \lambda}{\sigma} = -\frac{\mu_*}{\sigma_*}$ il termine $1 - \Phi\left(\frac{\varepsilon \lambda}{\sigma}\right)$ non è altro che la probabilità che una v.c. $N(\mu_*, \sigma_*^2)$ sia positiva. Quindi, nel caso considerato la distribuzione condizionata di u dato ε è una v.c. $N(\mu_*, \sigma_*^2)$ troncata a zero¹⁹ con media $\mu_* = -\sigma_u^2 \varepsilon / \sigma^2$ e varianza $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$.

Una stima puntuale del termine di inefficienza u può essere ottenuta utilizzando sia la media che la moda della distribuzione condizionata. In particolare, il valore atteso è espresso da:

$$E(u_i | \varepsilon_i) = \mu_{*i} + \sigma_* \frac{\phi(-\mu_{*i} / \sigma_*)}{1 - \Phi(-\mu_{*i} / \sigma_*)}$$

che, sfruttando l'uguaglianza $-\mu_* / \sigma_* = \varepsilon \lambda / \sigma$, dove $\lambda = \frac{\sigma_u}{\sigma_v}$, può essere anche scritto:

$$E(u_i | \varepsilon_i) = \sigma_* \left[\frac{\phi(\varepsilon_i \lambda / \sigma)}{1 - \Phi(\varepsilon_i \lambda / \sigma)} - \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right] \quad [3.16]$$

to si contrappone a quello di efficienza economica, rispetto al quale la frontiera viene definita in termini di comportamento economico dell'unità produttiva, vale a dire, in riferimento alla sua capacità di minimizzare i costi e/o massimizzare le entrate ed i profitti (Lovell, 1993).

¹⁹ Seguendo Johnson, Kotz e Balakrishnan (1994) una v.c. X presenta una distribuzione normale doppiamente troncata se la sua funzione di densità assume la forma:

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \left[\frac{1}{\sqrt{2\pi}\sigma} \int_A^B \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt \right]^{-1}$$

dove, nel caso considerato in [3.15], il punto di troncamento inferiore A è uguale a 0, mentre B , che rappresenta il punto di troncamento superiore, coincide con ∞ .

Il secondo stimatore puntuale, rappresentato dalla moda della distribuzione condizionata, è esprimibile mediante:

$$M(u_i | \varepsilon_i) = -\varepsilon_i \left(\frac{\sigma_u^2}{\sigma^2} \right) \quad \text{se } \varepsilon_i \leq 0$$

$$M(u_i | \varepsilon_i) = 0 \quad \text{se } \varepsilon_i > 0$$
[3.17]

Va osservato che nelle applicazioni empiriche²⁰ il valore atteso viene utilizzato più frequentemente del valore modale, nonostante quest'ultimo abbia una interessante interpretazione come stimatore di massima verosimiglianza.

Materov (1981) dimostrò, infatti, che $M(u_i | \varepsilon_i)$ può essere derivato massimizzando la densità congiunta di u_i e v_i , rispetto a u_i e v_i , soggetto al vincolo che $v_i - u_i = \varepsilon_i$.

Dopo aver ottenuto stime puntuali di u è possibile stimare l'efficienza tecnica di ciascun produttore attraverso l'espressione:

$$ET_i = \exp\{-\hat{u}_i\}$$
[3.18]

nella quale ad \hat{u}_i si sostituisce l'espressione [3.16] o [3.17].

In relazione alle proprietà statistiche dei precedenti stimatori puntuali, Waldman (1984) dimostra²¹, in riferimento alle ipotesi considerate da Jondrow, Lovell, Materov e Schmidt (1982), che il valore atteso condizionale di u dato ε è una funzione non lineare in ε , e che esso rappresenta uno stimatore corretto di u . Infatti, considerando la legge del valore atteso iterato si ha (Green, 2000):

$$E \left[E(u_i | \varepsilon_i) \right] = E(u)$$

²⁰ Nel caso in cui si ipotizzi che il termine u segua una distribuzione esponenziale si ottengono risultati simili ai precedenti. In particolare la distribuzione condizionata di dato ε è una v.c. normale con media uguale a $-\varepsilon/\sigma_v + \sigma_v/\sigma_u$ e varianza uguale a σ_v^2 , per cui la media e la moda assumono, rispettivamente la forma:

$$E(u_i | \varepsilon_i) = \sigma_v \left[\frac{\phi(A)}{1 - \Phi(A)} - A \right] \quad \text{e} \quad \begin{cases} M(u_i | \varepsilon_i) = -\varepsilon - \sigma_v^2/\sigma_u & \text{se } \varepsilon \leq -\sigma_v^2/\sigma_u \\ M(u_i | \varepsilon_i) = 0 & \text{se } \varepsilon > -\sigma_v^2/\sigma_u \end{cases}$$

²¹ Va osservato che Waldman (1984) effettua un confronto tra lo stimatore suggerito da Jondrow, Lovell, Materov e Schmidt (1982) e due stimatori lineari, rappresentati, rispettivamente da $-\varepsilon$, giustificato dal fatto che $E(-\varepsilon) = E(-v + u)$, e da $\alpha + \beta\varepsilon$, dove $\beta = -V(u)/[V(u) + V(v)]$ e $\alpha = E(u) - \beta E(\varepsilon) = E(u)(1 + \beta)$. L'autore conclude affermando che il valore atteso condizionato è da preferire ai precedenti stimatori lineari.

Sfortunatamente, le stime dell'efficienza tecnica, ottenute dalla [3.18], sia utilizzando la [3.16] che la [3.17], non sono consistenti in quanto la variabilità associata alla distribuzione di $(u_i | \varepsilon_i)$ è indipendente da i .

Battese e Coelli (1988) suggeriscono l'utilizzazione del seguente stimatore puntuale dell'efficienza tecnica a livello di unità produttiva:

$$ET_i = E\left(\exp\{-\hat{u}_i\} | \varepsilon_i\right) = \left[\frac{1 - \Phi(\sigma_* - \mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)} \right] \cdot \exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\}$$

Va osservato che, non solo quest'ultimo stimatore può condurre a risultati diversi, in quanto $E\{-u_i | \varepsilon_i\} \neq E\{\exp\{-\hat{u}_i\} | \varepsilon_i\}$, ma soffre della medesima limitazione del precedente [3.18], ossia non è consistente, come sottolineano Kumbhakar and Lovell (2000).

È possibile ottenere intervalli di fiducia per le stime puntuali dell'efficienza tecnica, sfruttando, ad esempio, il fatto che la densità di $(u_i | \varepsilon_i)$ è una v.c. $N^+(\mu_*, \sigma_*^2)$ nel caso in cui u segua una v.c. metà normale. In particolare, Horrance e Schmidt (1996) derivano intervalli fiduciali $(1 - \alpha)$ 100% per $u_i | \varepsilon_i$ espressi da:

$$\left(\mu_{*i} + \sigma_* \Phi^{-1} \left[1 - \frac{\alpha}{2} \Phi \left(\frac{\mu_{*i}}{\sigma_*} \right) \right], \mu_{*i} + \sigma_* \Phi^{-1} \left[1 - \left(1 - \frac{\alpha}{2} \right) \Phi \left(\frac{\mu_{*i}}{\sigma_*} \right) \right] \right)$$

Grazie alle proprietà della funzione esponenziale è possibile ottenere agevolmente intervalli fiduciali per l'efficienza tecnica $ET_i = E\{\exp\{-u_i\} | \varepsilon_i\}$.

Va osservato che tali intervalli non presentano buone proprietà di copertura essendo basati sui percentili della distribuzione di $\exp\{-u_i\} | \varepsilon_i$ e non sulla distribuzione campionaria degli stimatori.

A tale proposito Hjalmarsson, Kumbhakar e Heshmati (1996) derivano intervalli di fiducia per lo stimatore $E(u_i | \varepsilon_i)$.

3.5 LA SPECIFICAZIONE DI FORME DISTRIBUTIVE PER IL TERMINE DI ERRORE UNILATERALE

Nei modelli che specificano la frontiera di produzione stocastica, il termine di errore unilaterale viene solitamente descritto utilizzando la distribuzione metà

normale o la distribuzione esponenziale. Entrambe le distribuzioni, come evidenziato in Figura 3.1, presentano il valore modale nel punto $u = 0$, che si traduce nell'ipotesi che la probabilità del comportamento non efficiente del produttore decresca monotonicamente all'aumentare del grado di inefficienza.

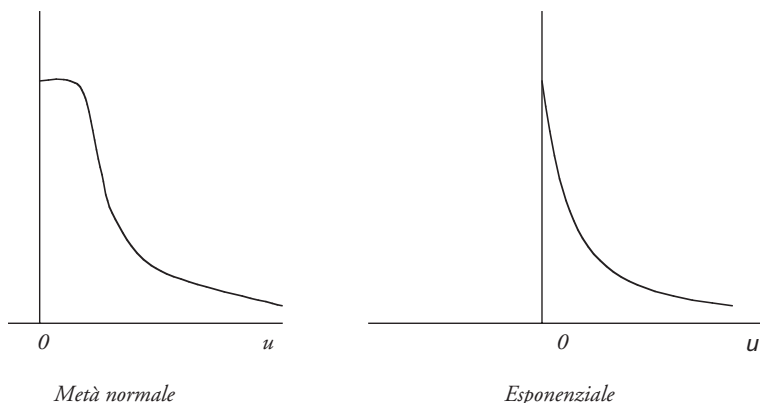


Figura 3.1 – Distribuzioni per il termine di inefficienza: v.c. metà normale ed esponenziale.

Poiché alcune caratteristiche dei processi produttivi, legate all'abilità del produttore, contrastano con tale assunzione, si rende utile una generalizzazione di tali forme distributive. Di conseguenza si è assistito, in letteratura, all'introduzione di v.c. più flessibili per la descrizione del termine di errore che descrive l'inefficienza tecnica delle unità produttive, in modo da riprodurre più fedelmente il comportamento dei produttori nelle situazioni reali.

In particolare, Stevenson (1980) generalizza le ipotesi avanzate da Aigner, Lovell e Schmidt sulla forma distributiva di u , suggerendo il modello di frontiera stocastica, in riferimento ad una funzione di costo, in cui il termine di errore bilaterale v_i segue, come usuale, una distribuzione normale $N(0, \sigma^2)$, mentre il termine u_i che descrive l'inefficienza tecnica segue una distribuzione normale troncata²² (Figura 3.2), che essendo caratterizzata dai due parametri μ e σ assicura una maggiore flessibilità al modello.

²² Va osservato che Stevenson (1980) suggerisce anche l'utilizzazione di una v.c. gamma per il termine u , accanto alla distribuzione normale $N(0, \sigma^2)$ per l'errore simmetrico, sebbene non proceda ad una verifica empirica del modello.

La funzione di densità dell'errore complessivo $\varepsilon = v - u$ assume, quindi, la forma seguente:

$$f(\varepsilon_i) = \frac{1}{\sigma} \phi\left(\frac{\varepsilon + \mu}{\sigma}\right) \left[1 - \Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right)\right] \left[\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}$$

dove $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$, $\lambda = \sigma_u / \sigma_v$ e ϕ è la densità della v.c. normale standardizzata.

Va osservato che nel caso in cui $\mu = 0$ il modello si riduce a quello introdotto da Aigner, Lovell e Schmidt.

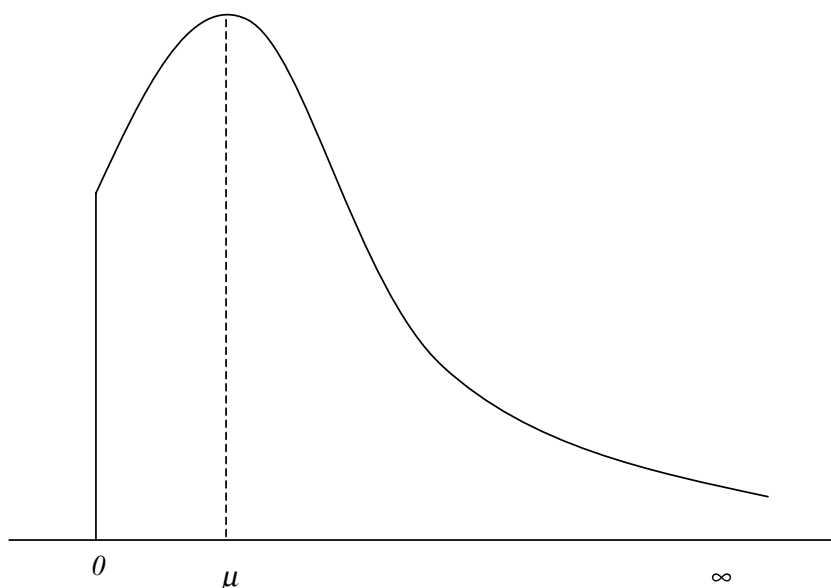


Figura 3.2 – La distribuzione del termine di inefficienza: la v.c. normale troncata.

Il suddetto Autore procede nella verifica empirica della significatività del modello proposto, che definisce generalizzato, considerando sia la stima della frontiera di produzione, ipotizzando due diverse specificazioni per la funzione di produzione, sia la stima della funzione di costo.

Nel primo esempio di applicazione, l'Autore considera gli stessi dati utilizzati da Hildebrand e Liu, che, come già visto, sono successivamente ripresi da

Aigner e Chu e da Aigner, Lovell e Schmidt, relativi all'industria dei metalli primari negli Stati Uniti per il periodo 1957-58.

La Tabella 3.7 riporta i risultati ottenuti²³ nella stima del modello

$\ln Y = \beta_0 + \beta_1 \ln L + \beta_2 (\ln R \cdot \ln K) + (v - u)$, in cui le variabili assumono il significato già esposto in precedenza. Le stime confermano il buon adattamento della v.c. normale troncata al comportamento del produttore, in quanto il parametro μ assume un valore negativo, e statisticamente significativo, sulla base del test t asintotico.

| | Funzione media (OLS) | Frontiera di produzione u metà normale | Modello generalizzato u normale troncata |
|------------------|----------------------|---|---|
| β_0 | 1,212* | 1,030 | 1,380* |
| β_1 | 0,862* | 0,862* | 0,805* |
| β_2 | 0,052* | 0,052* | 0,063* |
| σ | | 0,168* | 0,189* |
| $\sqrt{\lambda}$ | | 0,014 | 66,5 |
| μ | | | -0,318* |

* Significatività al 99%

Tabella 3.7 – Risultati empirici in Stevenson (1980).

Sottolineando che la specificazione di una particolare v.c. per il termine di errore unilaterale dovrebbe basarsi su informazioni relative ai meccanismi economici che generano inefficienza, Lee (1983) riconosce che, in genere, nelle applicazioni empiriche non si dispone di tali informazioni e che, quindi, la scelta di una particolare funzione è in parte arbitraria.

Differenti specificazioni possono condurre non solo a differenti misure di efficienza, ma anche a diverse stime dei coefficienti degli input, come evidenziato nello studio di Van den Broeck, Forsund, Hjalmarsson e Meeusen (1980).

²³ È opportuno osservare che i risultati presentati da Stevenson, relativi al modello di Aigner, Lovell e Schmidt, in cui u è descritto da una v.c. metà normale, differiscono da quelli riportati dagli stessi Autori (Tabella 3.4).

Lee propone un test che consente la valutazione dell'adattamento della distribuzione unilaterale specificata per la componente di inefficienza nella frontiera stocastica, derivando l'espressione analitica in particolare per la v.c. metà normale e normale troncata, che rappresentano le ipotesi più frequentemente adottate nelle applicazioni.

Va ricordato che, nelle diverse formulazioni precedentemente esposte, che caratterizzano l'approccio parametrico stocastico per l'analisi dell'efficienza rispetto alla frontiera di produzione, si ipotizza sempre che il termine di errore bilaterale, v , segua una distribuzione normale con media nulla e varianza costante.

A tale proposito, un interessante tentativo di proporre un modello in cui l'ipotesi di normalità viene abbandonata è quello di Kopp e Mullah (1990), nel quale viene, inoltre, suggerito l'impiego del metodo dei momenti generalizzato per la stima dei parametri della frontiera.

Un aspetto estremamente interessante è rappresentato dalla sensibilità delle misure di efficienza a diverse specificazione della forma funzionale utilizzata per la descrizione della tecnologia di riferimento.

Per la valutazione dell'efficienza produttiva degli ospedali pubblici italiani, Venanzoni (1996) analizza la sensibilità dei risultati ottenuti stimando diverse specificazioni della funzione di costo *translog* ed utilizzando differenti indicatori di output. I risultati mostrano un forte grado di correlazione tra le misure di efficienza, nonostante le notevoli diversità esistenti tra le diverse specificazioni utilizzate.

3.6 RECENTI SVILUPPI DELLE FRONTIERE STOCASTICHE

3.6.1 Qualche cenno all'analisi dei dati panel

Tra i molteplici sviluppi che hanno interessato l'approccio stocastico, non si può non citare, brevemente, l'introduzione dei modelli per dati *panel*, secondo l'espressione, che costituisce una generalizzazione del modello di Aigner, Lovell e Schmidt, suggerita da Pitt e Lee (1981): $\ln y_{it} = \ln f(\mathbf{x}_{it}; \boldsymbol{\beta}) + v_{it} - u_{it}$, dove $i = 1, \dots, n$ e $t = 1, \dots, T$.

Nel contesto dell'approccio di base, che prevede l'assunzione di *invarianza nel tempo dell'inefficienza*, secondo la quale $u_{it} = u_i$, si possono individuare due diverse specificazioni, ossia il modello a effetti fissi (*fixed effects model*) e il modello a effetti variabili (*random-effects model*).

Nel primo caso gli errori unilaterali u_i si traducono in parametri di intercetta specifici per ogni unità produttiva, in conseguenza anche della mancata specificazione di una particolare forma distributiva.

Nel secondo caso si assume che gli errori u_i siano descritti da v.c. con media e varianza costante. Si ipotizza inoltre che i termini u_i siano incorrelati con i regressori e con i termini v_{it} .

Tra i principali contributi che si sono susseguiti in letteratura, si possono richiamare, ad esempio, lo studio di Battese e Coelli (1988), che generalizzano il modello ad efficienza invariante nel tempo suggerendo la v.c. normale troncata. Successivamente, Battese, Coelli e Colby (1989) estendono ulteriormente l'approccio, consentendo l'analisi di dati *panel* non bilanciati.

In relazione ai metodi di stima, Schmidt e Sickles (1984) osservano che, nel contesto precedente, non risulta necessario specificare una particolare v.c. per u , in quanto i parametri del modello possono essere stimati utilizzando i metodi tradizionali per dati *panel*, che si basano su modificazioni particolari degli OLS.

La limitazione principale dell'approccio precedente consiste nell'assumere che l'inefficienza tecnica sia costante nel tempo, che si traduce, in particolare quando i periodi di osservazione sono numerosi, in un'ipotesi fortemente contrastante con la realtà di un mercato concorrenziale.

L'introduzione di modelli ad *efficienza variabile nel tempo* si deve, inizialmente, agli studi di Kumbhakar (1990), Battese e Coelli (1992), Cornwell, Schmidt e Sickles (1990) e Lee e Schmidt (1993) che propongono diverse generalizzazioni del modello di Schmidt e Sickles al fine di considerare la circostanza per cui l'efficienza può variare nel tempo, anche in conseguenza delle modificazioni delle capacità del produttore nell'organizzazione del processo produttivo.

In riferimento alle diverse forme funzionali suggerite, si può considerare, ad esempio, quella proposta da Battese e Coelli, secondo la quale:

$$u_{it} = \left\{ \exp \left[-\eta(t - T) \right] \right\} u_i,$$

dove le u_i sono *i.i.d* secondo una v.c. normale troncata e η è un parametro da stimare. Nel modello suggerito da Cornwell, Schmidt e Sickles la variabilità nel tempo dell'efficienza è introdotta ipotizzando che i parametri di intercetta per le diverse unità produttive, in diversi periodi di tempo, siano espresse da una funzione quadratica del tempo.

Tra le molteplici applicazioni empiriche che hanno coinvolto l'approccio stocastico utilizzando dati *panel*, si possono citare, ad esempio, lo studio di Viviani, Mealli e Pierotti (1999), che analizzano il caso della distribuzione commerciale attraverso la specificazione e stima di una funzione frontiera dinamica, nella quale si tiene conto delle innovazioni gestionali del settore.

I suddetti Autori confrontano i risultati ottenuti tramite l'approccio parametrico, impiegando dati relativi alla grande distribuzione in Toscana, con quelli derivanti dall'applicazione di una frontiera non parametrica, specificata attraverso il cosiddetto²⁴ *Free Disposal Hull*.

Molto interessante è anche il lavoro di Fabbri (1998), che stima l'efficienza produttiva e di costo nel trasporto pubblico locale, con una particolare applicazione al caso dell'Emilia Romagna.

3.6.2 L'introduzione dell'effetto di variabili esogene

L'analisi dell'influenza esercitata da fattori esogeni sulla misura di efficienza delle unità produttive risulta di particolare rilevanza, in quanto consente, non solo di definire la misura di efficienza raggiunta dalle unità produttive nello svolgimento del processo, ma anche di identificare eventuali fattori che contribuiscono alla determinazione della stessa.

Kumbhakar e Lovell (2000), riprendendo quanto già affermato da Simar, Lovell e Vanden Eackaut (1994) rilevano che l'analisi dell'efficienza produttiva deve, o almeno dovrebbe, svolgersi in due fasi.

La prima consiste nella stima di una frontiera di produzione che rappresenti un "*benchmark*" rispetto al quale stimare l'efficienza tecnica delle unità produttive.

La seconda fase concerne l'analisi di variabili esogene, che caratterizzano l'ambiente nel quale si svolge la produzione e che, pur non rappresentando né input né output del processo, esercitano un'influenza sulla *performance* dell'unità produttiva.

Va osservato che tali variabili possono produrre i loro effetti nei confronti della tecnologia di produzione, che descrive le condizioni strutturali in cui gli input sono trasformati in output, oppure nei confronti dell'efficienza con la quale si svolge il processo di trasformazione.

In conseguenza di ciò, l'introduzione delle variabili esogene, nei modelli per la misurazione dell'efficienza rispetto alla frontiera, può avvenire seguendo diversi approcci, che si basano su ipotesi specifiche in relazione agli effetti esercitati da questi fattori.

Numerosi studi seguono un primo approccio, suggerito inizialmente da Pitt e Lee (1981), che si basa sull'ipotesi per cui le variabili esogene **agiscono sulla**

²⁴ Tale metodologia sarà oggetto di analisi nel quarto capitolo.

struttura della tecnologia di produzione, intervenendo direttamente nel processo produttivo e modificando il risultato della produzione y_i .

Secondo tale impostazione, i fattori esogeni rappresentati, ad esempio, da specifiche caratteristiche aziendali, sono inclusi, ammesso che siano quantificabili, direttamente nella funzione di produzione, insieme agli input utilizzati nel processo.

Considerando le variabili esogene attraverso il vettore $\mathbf{z} = (z_1, z_2, \dots, z_k)$, la frontiera di produzione, espressa in termini logaritmici, assume la forma seguente: $\ln y_i = \ln f(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\beta}) + v_i - u_i$, $i = 1, \dots, n$, dove il vettore $\boldsymbol{\beta}$ include sia i parametri legati alla tecnologia di produzione sia quelli legati alle variabili esogene.

Bauer e Hancock (1993), ad esempio, utilizzando una funzione di produzione translogaritmica, includono tra gli input del processo produttivo anche un vettore di variabili di tipo ambientale.

Sulla base dell'idea originariamente suggerita da Timmer (1971), si sviluppa il cosiddetto **approccio a due stadi**, formulato tra gli altri da Ali e Flinn (1989) e Kalirajan (1990), nel quale si assume che le variabili esogene influenzino il risultato della produzione, y_i , in modo indiretto, attraverso gli effetti prodotti sull'efficienza con cui il processo viene posto in essere, per cui si ipotizza che gli elementi del vettore \mathbf{z} siano in qualche modo correlati con quelli di \mathbf{u} , o almeno con lo stimatore $E(u_i | \boldsymbol{\varepsilon}_i)$. Il *primo stadio* consiste nella stima della frontiera di produzione, che conduce poi all'ottenimento della misura di efficienza tecnica per ciascuna unità produttiva, mediante la [3.18]. Nel *secondo stadio* della procedura, le misure di efficienza sono associate ai fattori esterni attraverso un modello di regressione, che può assumere la forma seguente: $ET_i = g(\mathbf{z}_i; \boldsymbol{\gamma}) + w_i$, dove $w_i \sim i.i.d.N(0, \sigma_w^2)$, mentre $\boldsymbol{\gamma}$ è un vettore di parametri da stimare.

In particolare, Ali e Flinn, analizzando un campione di aziende produttrici di riso in Pakistan, associano le misure di efficienza ottenute, stimando una frontiera di profitto stocastica, a tre gruppi di fattori: fattori istituzionali, rappresentati, ad esempio, dalle date fissate per il raccolto o la diffusione di fertilizzanti; fattori socioeconomici, come gli anni di formazione scolastica del capofamiglia; e fattori legati alle risorse di base, come la dimensione dell'azienda o l'uso di particolari macchinari. La stima dei parametri della frontiera stocastica del primo stadio è effettuata mediante il metodo ML utilizzando la formulazione di Battese e Corra (1977), mentre nel secondo stadio il modello viene stimato utilizzando gli OLS.

Kalirajan, al fine di analizzare in modo più completo e funzionale l'efficienza economica delle unità produttive, propone di stimare simultaneamente, utiliz-

zando il metodo ML, la frontiera di produzione, descritta da una funzione di produzione translog, e l'equazione esprimente la produttività marginale. Nonostante i problemi derivanti dall'adozione della funzione translog, Kalirajan ottiene, dai residui del modello, oltre a misure di efficienza tecnica anche misure di efficienza allocativa.

Riprendendo la teoria dello sviluppo economico di Schumpeter, il suddetto Autore sottolinea che l'efficienza tecnica dipende da fattori che determinano la conoscenza e le capacità tecniche individuali, oltre che da fattori di tipo socio economico legati all'ambiente nel quale il produttore svolge la sua attività. Suggerisce quindi un modello semi-logaritmico esplicativo delle misure di efficienza tecnica, introducendo alcune variabili che rappresentano i due insiemi di fattori suddetti, ossia il livello di conoscenza tecnica e lo status socio economico del produttore.

Per quanto riguarda le diverse *metodologie utilizzate nel secondo stadio dell'analisi*, attraverso il quale si vuole stimare l'effetto delle variabili esterne sulla misura di efficienza, che rappresenta l'elemento di originalità dell'approccio, sono state suggerite in letteratura numerose forme funzionali²⁵ e tecniche di stima.

Poiché la variabile dipendente assume sempre valori compresi tra zero e uno, gli OLS potrebbero risultare inappropriati, per cui si può considerare una preliminare trasformazione della variabile dipendente, oppure si può applicare una tecnica di stima per variabile dipendente limitata come, ad esempio, il modello Tobit.

Dall'approccio in due stadi emerge però la presenza di una distorsione, in quanto, mentre nel primo stadio si ipotizza l'identica distribuzione dei termini di errore unilaterale, nel secondo stadio, si assume che l'efficienza stimata sia funzione di un insieme di variabili esplicative, contraddicendo in tal modo l'assunzione di identica distribuzione.

²⁵ Mester (1993), ad esempio, propose di utilizzare una forma funzionale logistica, per lo stimatore di u_i , del tipo:

$$E(u_i | \varepsilon_i) = \frac{\exp(\mathbf{z}_i' \boldsymbol{\gamma}^K)}{1 + \exp(\mathbf{z}_i' \boldsymbol{\gamma}^K)} + \xi_i^K \quad K = M, S$$

dove \mathbf{z}_i è un vettore di variabili indipendenti per l' i -esima unità produttiva, mentre $\boldsymbol{\gamma}^k$ rappresenta un vettore di parametri per i due gruppi, M e S ("Mutual" e "Stock" nel settore del risparmio e prestito), di imprese considerate nell'analisi, per le quali Mester propone due diversi modelli di frontiera di costo stocastica in modo da rendere l'analisi più flessibile. Infine, il termine di disturbo casuale ξ_i^K , è una v.c. normale.

Tra i diversi studi sull'argomento, si può citare, ad esempio, Berger e Mester (1997) e Wang e Schmidt (2002). In particolare, quest'ultimi Autori, evidenziano, attraverso simulazioni Monte Carlo, la presenza di distorsione in entrambe le fasi di analisi, dimostrando che se nel primo stadio le stime dell'efficienza sono ottenute non includendo le variabili esogene, nel secondo stadio, la regressione produrrà stime distorte verso il basso dell'effetto di \mathbf{z} su u .

Il superamento dei limiti connessi all'approccio a due stadi costituisce l'oggetto di numerosi studi che conducono allo sviluppo del cosiddetto **approccio a uno stadio (o della stima simultanea)**, attraverso il quale è possibile ottenere, contemporaneamente, sia la stima della misura dell'efficienza tecnica che la determinazione dell'effetto delle variabili esogene considerate nell'analisi.

Un primo tentativo, dovuto a Deprins e Simar (1989), è sviluppato *nel contesto dell'approccio delle frontiere deterministiche* del tipo [3.7].

La relazione attraverso la quale i termini di efficienza vengono spiegati dalle variabili esogene considerate rilevanti, viene specificata dai suddetti Autori mediante una funzione esponenziale, grazie alla quale può essere assicurata la positività di u_i . In tal modo si introduce una struttura log-lineare per la regressione di u_i sulle variabili \mathbf{z}_i , ossia:

$$E(u_i | \mathbf{z}_i) = \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\} \quad i = 1, \dots, n$$

Il modello completo può scriversi: $\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) - \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\} + w_i$ dove il termine di disturbo w_i ha media nulla e rappresenta la parte della produzione che non è spiegata dagli altri due termini dell'espressione. Va osservato che w_i deve verificare la relazione $w_i \leq \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\}$. Considerando \mathbf{b} e \mathbf{g} stimatori, rispettivamente, di $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, ed ipotizzando che per la funzione di produzione sia stato scelto un modello log-lineare, si può definire una misura di efficienza per la i -esima unità produttiva nel modo seguente:

$$ET_i = \frac{y_i}{\hat{y}_i} = \exp(\hat{w}_i) \quad \text{dove} \quad \hat{w}_i = y_i - \mathbf{x}_i' \mathbf{b} + \exp\{\mathbf{g}' \mathbf{z}_i\}$$

Gli Autori in questione rilevano che la misura precedente può essere scomposta in: $ET_i = \exp(y_i - \mathbf{x}_i' \mathbf{b}) \exp\{\exp(\mathbf{g}' \mathbf{z}_i)\}$

Il primo fattore è minore o uguale a 1 e corrisponde alla misura di efficienza tecnica in un modello che non tiene conto dell'influenza delle variabili esogene; il secondo fattore rappresenta un termine di aggiustamento che costituisce una stima del contributo delle variabili esogene alla performance di ogni produttore ed assume valori superiori a 1. Valori relativamente più grandi di quest'ultimo

termine esprimono situazioni ambientali non favorevoli nello svolgimento del processo. In tal modo determinano un incremento della misura dell'efficienza tecnica grezza, in conseguenza del fatto che il processo produttivo è stato svolto in condizioni avverse. Poiché ET_i può assumere valori più grandi di 1, gli Autori suggeriscono, al fine di facilitare i confronti, di dividere tutte le misure per il $\max(ET_i)$, in modo da ottenere grandezze inferiori o uguali a 1.

Poiché il modello non è lineare nei parametri, la stima può essere effettuata mediante i minimi quadrati non lineari, oppure utilizzando gli stimatori di ML, dopo aver specificato un'opportuna distribuzione casuale per il termine u_i , come, ad esempio, la distribuzione Weibull.

L'approccio descritto, pur apportando notevoli elementi migliorativi sia rispetto alla prima formulazione, che considera l'inclusione delle variabili esogene nella determinazione delle caratteristiche della tecnologia, perché include un modello esplicativo dell'efficienza, sia rispetto all'approccio a due fasi, perché il problema dell'omissione di variabili è superato, è basato su un modello di tipo deterministico per la frontiera di produzione, che rappresenta una notevole limitazione, a causa della non inclusione di un termine di disturbo aleatorio simmetrico.

La *reformulazione della problematica in un contesto stocastico* ha rappresentato l'oggetto dei lavori di Kumbhakar, Ghosh e McGuskin (1991), Reifschneider e Stevenson (1991), Huang e Liu (1994) e Battese e Coelli (1995).

In particolare, Kumbhakar, Ghosh e McGuskin considerano la funzione di produzione Zellner-Revankar per la rappresentazione della tecnologia del latte nelle fattorie statunitensi, che conduce alla specificazione della frontiera di produzione stocastica seguente: $y \exp(\lambda y) = \beta_0 \prod x_i^{\alpha_i} \exp(v_i - u_i)$

Il termine di errore u_i , che descrive l'inefficienza tecnica²⁶ delle unità produttive è costituito da una componente deterministica, che dipende da un vettore di fattori esogeni osservabili, \mathbf{z} , e da una componente puramente casuale non osservabile, \mathbf{w} , ossia:

$$\mathbf{u} = \mathbf{z}' \boldsymbol{\gamma} + \mathbf{w} \quad [3.19]$$

dove $\boldsymbol{\gamma}$ è il vettore dei parametri dei fattori esogeni.

Alternativamente, si può ipotizzare che \mathbf{u} sia distribuito come una v.c. normale troncata a zero da sopra, $N^+(\mathbf{z}' \boldsymbol{\gamma}, \sigma_u^2)$.

²⁶ I suddetti Autori considerano, inoltre, la possibilità che un'unità produttiva sia anche allocativamente non efficiente, attraverso l'introduzione nel modello delle condizioni di primo ordine per la massimizzazione del profitto.

Tale ipotesi estende la formulazione di Stevenson (1980), nella quale si assume che la media della v.c. normale troncata che descrive il termine \mathbf{u} sia costante e diversa da zero. Infatti, la media della normale troncata può assumere diversi valori tra le unità produttive, esprimendo in tal modo le diversità esistenti tra le unità rispetto a fattori specifici, come, ad esempio, la qualità degli input utilizzati o la formazione del produttore, che non possono essere inclusi nella funzione di produzione direttamente, ma che sono in grado di spiegare le differenze nell'efficienza tecnica tra le unità produttive.

Nel caso in cui $\mathbf{z}'\boldsymbol{\gamma}$ sia uguale ad una costante, si ottiene la specificazione proposta da Stevenson, che può essere sottoposta a verifica statistica ipotizzando che $\boldsymbol{\gamma} = \mathbf{0}$, ad eccezione della costante. Se anche il termine costante è posto uguale a zero, il modello suddetto si riduce a quello introdotto da Aigner, Lovell, e Schimdt (1977), poiché la v.c. per u diviene una metà normale.

Nell'applicazione empirica, utilizzando il metodo ML, i suddetti Autori derivano le stime dei parametri di tre diversi modelli in cui specifici fattori sono inclusi tra le determinanti della misura di efficienza.

Considerando la stima di una frontiera di costo stocastica, Reifschneider e Stevenson (1991) suggeriscono di considerare la seguente formulazione: $\mathbf{u} = g(\mathbf{z}) + \mathbf{w}$, che non si differenzia sostanzialmente dal modello [3.19], se non per la considerazione di diverse ipotesi sulla forma distributiva del termine di errore puramente stocastico \mathbf{w} , che viene descritto dalle distribuzioni gamma, metà normale ed esponenziale. La definizione di una specifica forma funzionale $g(\mathbf{z})$ per la componente sistematica del termine di efficienza assicura che u sia non negativo.

Il modello specificato da Huang e Liu (1994), pur riproducendo l'impostazione di base dei modelli precedenti, introduce due elementi innovativi. In particolare, il modello è specificato inizialmente mediante due equazioni, esprimenti, rispettivamente, la frontiera stocastica, $y = f(\mathbf{x}; \boldsymbol{\beta}) + v$, e il termine di efficienza, $\mathbf{u} = g(\mathbf{z}; \boldsymbol{\gamma}) + \mathbf{w}$, le quali combinate insieme²⁷, danno luogo alla seguente espressione: $\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i - [g(\mathbf{z}_i; \boldsymbol{\gamma}) + \mathbf{w}]$.

Le condizioni sull'errore unilaterale sono soddisfatte troncando il termine \mathbf{w} in modo tale che $\mathbf{w} \geq -g(\mathbf{z}; \boldsymbol{\gamma})$. Perciò, un primo elemento innovativo consiste nel fatto che, invece di troncare a zero una distribuzione normale con

²⁷ Va osservato che Huang e Liu (1994) utilizzano una diversa simbologia, ma al fine di agevolare la lettura si è deciso di uniformare la notazione.

media variabile, Huang e Liu (1994) si riferiscono ad una distribuzione normale con media nulla nel punto di troncamento variabile $g(\mathbf{z}; \boldsymbol{\gamma})$.

Un'ulteriore peculiarità del modello suddetto consiste nel fatto che la funzione $g(\mathbf{z}; \boldsymbol{\gamma})$ può includere interazioni tra gli elementi di \mathbf{z} e quelli del vettore dei fattori produttivi \mathbf{x} .

L'applicazione empirica del modello suggerito, riguarda le industrie elettroniche a Taiwan e adotta una funzione di produzione translogaritmica con quattro input, mentre tra gli elementi del vettore \mathbf{z} sono inclusi l'età dell'azienda, la percentuale di esportazioni sul fatturato totale e la spesa relativa in ricerca e sviluppo.

Battese e Coelli (1995), riferendosi a dati *panel*, suggerirono l'utilizzazione di un modello simile ai precedenti, in cui le variabili esogene sono incluse nell'analisi contemporaneamente alle variabili che descrivono il processo produttivo. In particolare, il modello suggerito dagli Autori suddetti differisce da quello di Reifschneider e Stevenson (1991), giacché si assume che le v.c. w_i siano indipendentemente ma non identicamente distribuite.

Battese e Coelli stimano i parametri utilizzando il metodo ML, per l'analisi dell'efficienza di 14 aziende indiane produttrici di riso osservate su un periodo di dieci anni. Tra le variabili esplicative della misura di efficienza tecnica sono incluse alcune caratteristiche del produttore, rappresentate dall'età e dal livello di istruzione.

Il modello suggerito da Battese e Coelli è stato applicato da Audibert (1997), sempre in ambito agrario, ed ha costituito l'oggetto del lavoro di Battese e Broca (1997), nel quale i risultati sono stati confrontati con quelli ottenuti dal modello di Huang e Liu.

Simar, Lovell e Vanden Eeckaut (1994) proposero un modello di tipo moltiplicativo per il termine di errore unilaterale.

In riferimento alla frontiera stocastica, in cui il termine di errore simmetrico v_i , indipendente da u_i , si assume distribuito normalmente, gli Autori suggeriscono per u_i , che si ipotizza sia distribuito secondo una v.c. casuale unilaterale, la relazione $u_i = \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\} \cdot \eta_i$ nella quale $\eta_i \geq 0$, identicamente e indipendentemente distribuiti, $E(\eta_i) = 1$, $Var(\eta_i) = \sigma_{\eta_i}^2$. Da queste ultime assunzioni segue che: $E(u_i) = \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\}$ e $Var(u_i) = \exp\{2\boldsymbol{\gamma}' \mathbf{z}_i\} \sigma_{\eta_i}^2$.

Sostituendo l'espressione $u_i = \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\} \cdot \eta_i$ per il termine unilaterale, che risulta maggiore o uguale a zero, si ottiene il modello seguente:

$$\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) - \exp\{\boldsymbol{\gamma}' \mathbf{z}_i\} \cdot \eta_i + v_i \cdot$$

I suddetti Autori sottolineano che nella specificazione proposta, u_i rappresenta una trasformazione di scala di un processo sottostante η_i , riprendendo in tal modo il modello presentato in Deprins e Simar (1989), ma superandone le difficoltà connesse alla definizione deterministica della frontiera di produzione.

La struttura moltiplicativa dovrebbe consentire, inoltre, di superare i problemi statistici legati alla specificazione additiva del termine unilaterale, che utilizzando la simbologia degli Autori, può essere espressa da $u_i = \exp\{\boldsymbol{\gamma}'\mathbf{z}_i\} + \eta_i$. In questo caso, infatti, dovendo essere $u_i \geq 0$, si richiede che $\eta_i \geq -\exp\{\boldsymbol{\gamma}'\mathbf{z}_i\}$, ma ciò implica che η_i non soddisfa la condizione di distribuzione identica ed indipendente, in quanto risulta legato alle variabili esogene.

Simar, Lovell e Vanden Eeckaut suggeriscono diverse soluzioni per la stima dei parametri, ipotizzando diverse specificazioni per la v.c. u_i , in particolare, le distribuzioni gamma, esponenziale, metà normale e normale troncata, e due diversi metodi di stima, ossia il metodo dei minimi quadrati non lineari e della massima verosimiglianza, ma non forniscono un'applicazione dei metodi suggeriti.

Nel contesto del trattamento delle variabili esogene nelle frontiere stocastiche di produzione risulta di estrema importanza, a questo punto, introdurre i cosiddetti **modelli eteroschedastici**.

Nel tentativo di risolvere i problemi derivanti dalla presenza di *eteroschedasticità* nelle stime dei parametri della frontiera di produzione stocastica Caudill e Ford (1993) and Caudill, Ford e Gropper (1995) propongono un modello stocastico nel quale la varianza del termine di disturbo unilaterale non è costante, ma risulta espressa come funzione di un insieme di variabili legate, in particolare, alle caratteristiche dell'impresa.

Tali modelli possono essere considerati come una strategia alternativa per l'introduzione dell'effetto delle variabili esogene nella determinazione dell'efficienza attraverso lo studio e il trattamento dell'eterogeneità presente nella componente di errore unilaterale che dipende, appunto, da fattori esterni.

In particolare, Caudill and Ford dimostrarono, attraverso simulazioni Monte Carlo, che gli effetti della presenza di eteroschedasticità sulle stime dei parametri dei modelli di frontiera sono più dannosi che nel caso della stima di funzioni di produzione medie. In particolare, l'eteroschedasticità provocherebbe una sottostima del termine di intercetta e una sovrastima degli altri parametri.

Considerando che il termine di errore unilaterale presenta eteroschedasticità moltiplicativa (Greene, 2000), ossia: $\sigma_{u_i} = \sigma \exp(\mathbf{z}_i\boldsymbol{\gamma})$ che può semplificarsi in

$\sigma_{ui} = \exp(\mathbf{z}_i \boldsymbol{\gamma})$ se il vettore di variabili esogene \mathbf{z} include l'intercetta, la funzione di densità del termine di errore complessivo nella parametrizzazione di Aigner, Lovell e Schmidt (1977) diventa:

$$f_i(\varepsilon_i) = \frac{2}{\sigma_i} \phi\left(\frac{\varepsilon_i}{\sigma_i}\right) \left[1 - \Phi\left(\frac{\lambda_i \varepsilon_i}{\sigma_i}\right)\right] \quad -\infty < \varepsilon_i < \infty$$

dove $\sigma_i^2 = \sigma_{ui}^2 + \sigma_v^2$, $\lambda_i = \sigma_{ui} / \sigma_v$, e ϕ e Φ sono rispettivamente la funzione di densità e di ripartizione della v.c. normale standardizzata.

I suddetti Autori presentano un'interessante applicazione del modello alla stima della frontiera di costo per il settore bancario, utilizzando l'algoritmo di Berndt, Hall, Hall e Hausman (1974), per la massimizzazione della funzione di verosimiglianza.

È necessario osservare che già in precedenza Reifschneider e Stevenson (1991) avevano formulato una simile specificazione, senza però procedere nella stima, ipotizzando che il parametro della distribuzione metà normale che descrive il termine u , ossia σ_u^2 , fosse funzione delle caratteristiche delle unità produttive, a loro volta legate ai fattori esterni, attraverso l'espressione $\sigma_{ui} = \sigma_{ui} + h(\mathbf{z}_i)$ con $h(\mathbf{z}_i) \geq 0$.

Anche il modello di Simar, Lovell e Vanden Eeckaut anticipa in qualche modo il contributo di Caudill e Ford poiché la varianza di u_i considerata dagli Autori, dipende da i , e quindi, varia al variare delle osservazioni.

Si ritiene interessante a questo punto riportare l'espressione della funzione di log verosimiglianza nel caso in cui gli errori u_i siano eteroschedastici. Per un campione di n unità produttive si ha:

$$\ln L(y|\boldsymbol{\beta}, \lambda, \sigma_v^2, \boldsymbol{\delta}) = \sum_{i=1}^n \ln [f_i(\varepsilon_i)] \text{ e quindi:}$$

$$\begin{aligned} \ln L(y|\boldsymbol{\beta}, \lambda, \sigma_v^2, \boldsymbol{\delta}) &= \text{costante} - \frac{1}{2} \sum_i \ln [g(\mathbf{z}_i; \boldsymbol{\delta}) + \sigma_v^2] + \\ &+ \sum_i \ln \Phi\left(-\frac{\varepsilon_i \lambda_i}{\sigma_i}\right) - \frac{1}{2} \sum_i \frac{\varepsilon_i^2}{\sigma_i^2} \end{aligned}$$

dove $\sigma_i^2 = \sigma_v^2 + g(\mathbf{z}_i; \boldsymbol{\delta})$, $\lambda_i = \frac{\sqrt{g(\mathbf{z}_i; \boldsymbol{\delta})}}{\sigma_v}$ e $g(\mathbf{z}_i; \boldsymbol{\delta})$ è la funzione generica che esprime la dipendenza dalla variabili esogene.

L'efficienza tecnica per ciascuna unità produttiva è derivabile attraverso l'espressione [3.18] nella quale la stima del termine di inefficienza può essere ottenuta utilizzando lo stimatore:

$$E(u_i | \varepsilon_i) = \frac{\sigma_{ui} \sigma_v}{\sigma_i} \left[\frac{\phi(\varepsilon_i \lambda_i / \sigma_i)}{1 - \Phi(\varepsilon_i \lambda_i / \sigma_i)} - \left(\frac{\varepsilon_i \lambda_i}{\sigma_i} \right) \right]$$

che rappresenta una generalizzazione del valore atteso della distribuzione condizionata per il caso omoschedastico²⁸. Un'applicazione empirica della frontiera con errori unilaterali eteroschedastici sarà analizzata nel sesto capitolo.

In un recente articolo, Wang (2002) tenta di riunire in un unico modello le caratteristiche principali dei due approcci che, in un unico stadio, introducono l'effetto delle variabili esogene z , attraverso l'inserimento di modelli esplicativi dei parametri della distribuzione normale troncata, in particolare della media, come nei modelli di Huang e Liu (1994) Kumbhakar, Ghosh e McGuskin (1991), Battese e Coelli (1995) e della varianza, come nei modelli di Simar, Lovell e Vanden Eeckaut (1994) e Caudill, Ford and Gropper (1995).

Wang intende misurare l'effetto non monotono che alcune variabili esogene possono esercitare sull'efficienza, nel senso di considerare effetti di segno opposto, ossia d'incremento o di diminuzione dell'efficienza, registrati per alcune unità nel campione.

Precedentemente, Bhattacharyya, Kumbhakar e Bhattacharyya (1995), nel contesto dell'analisi di dati *panel*, avevano stimato una frontiera stocastica di costo, utilizzando una funzione translog, in cui sia la media che la varianza del termine di errore unilaterale $u_i \geq 0$ erano funzioni del tempo e di alcune caratteristiche individuali delle unità produttive.

Attraverso il modello proposto, i suddetti Autori, erano in grado di considerare, nella valutazione dell'efficienza, non solo l'effetto di fattori legati al tempo o espressioni caratteristiche specifiche delle unità produttive, ma anche l'eventuale presenza di eterogeneità.

Un vantaggio ulteriore derivante da una tale generalizzazione, è rappresentato dalla possibilità di sottoporre a verifica il modello standard caratterizzato da un termine di errore che descrive l'efficienza con media e varianza costante.

3.6.3 La meta-frontiera di produzione

Battese e Rao (2002), allo scopo di consentire il confronto tra l'efficienza di unità produttive facenti parte di diversi gruppi, caratterizzati dalla presen-

²⁸ Si noti che anche il termine v_i , che descrive l'agire di fattori puramente aleatori, può essere eteroschedastico. In tal caso è possibile considerare una funzione $\sigma_{vi}^2 = h(w_i; \gamma)$ attraverso

za di diverse tecnologie, utilizzano il concetto di meta-frontiera in ambito stocastico.

L'idea di individuare una funzione di produzione comune, in relazione ad unità produttive scarsamente omogenee, definita *metafrontiera* di produzione (*metaproduction function*), è stata introdotta inizialmente da Hayami (1969) e quindi sviluppata da Hayami e Ruttan (1970, 1971).

Analizzando i differenziali di produttività tra paesi sviluppati e in via di sviluppo, Hayami e Ruttan (1970) sottolineano la circostanza per cui i produttori agricoli in diversi paesi, o in diverse regioni di uno stesso paese, non si trovano tutti sulla medesima funzione di produzione. Tale situazione può essere il risultato di una diversa abilità dei produttori nell'adottare nuove tecnologie, ma anche di una diversa diffusione della tecnologia agraria e della tecnologia in generale. La meta-funzione di produzione descrive l'insieme di tutte le tecnologie di produzione alternative possibili, che sono solo parzialmente disponibili per i produttori in un dato paese o in una data regione agricola in un particolare periodo di tempo, ed involupa tutte le osservazioni.

I suddetti Autori stimano una funzione di produzione Cobb-Douglas tra i diversi paesi, utilizzando gli OLS, per tre distinti periodi di tempo, ossia per gli anni 1955, 1960 e 1965.

I risultati ottenuti, che confermano quelli di studi precedenti, sono analizzati anche a livello di singolo paese, allo scopo di fornire utili informazioni sui fattori che determinano l'esistenza di differenziali di produttività.

Successivamente, Mundlak e Hellinghausen (1982) stimano una funzione di produzione comune per diversi paesi, sempre in riferimento all'agricoltura, utilizzando un approccio che considera i fattori che intervengono nella scelta di una particolare tecnologia, specificati dalle cosiddette variabili di stato.

Il modello della meta-frontiera stocastica viene esteso da Battese, Rao e O'Donnell (2004).

Rispetto al modello presentato in Battese e Rao (2002), i suddetti Autori, che considerano dati *panel*, assumono che il meccanismo generatore dei dati sia unico per tutte le unità produttive che operano in una data tecnologia, definendo la meta-frontiera di produzione come una funzione caratterizzata da una determinata forma matematica, che comprende le componenti deterministiche delle funzioni di produzione stocastiche per le unità che operano sotto le diverse tecnologie coinvolte nell'analisi.

so la quale si introducono gli effetti di variabili, spesso collegate alla dimensione dell'unità produttiva, che determinano la presenza di eteroschedasticità negli errori.

Considerando dati *cross-section*, e supponendo di poter individuare R diversi gruppi all'interno di una medesima industria e che per il gruppo j -esimo sia disponibile un campione di N_j unità produttive che produce un output da un insieme di input, la frontiera di produzione stocastica (Aigner, Lovell e Schmidt, 1977) per questo gruppo è definita da:

$$y_{ij} = f(\mathbf{x}_{ij}; \boldsymbol{\beta}_j) e^{v_j - u_{ij}}$$

dove le variabili coinvolte assumono l'usuale significato.

La meta-frontiera di produzione per il complesso delle unità produttive nell'industria è espressa da:

$$y_i^* = f(\mathbf{x}_i; \boldsymbol{\beta}^*) e^{x_i \boldsymbol{\beta}^*} \quad i = 1, \dots, N = \sum_{j=1}^R N_j$$

dove $\boldsymbol{\beta}^*$ indica il vettore dei parametri per la funzione meta-frontiera tale che:

$$\mathbf{x}_i \boldsymbol{\beta}^* \geq \mathbf{x}_{ij} \boldsymbol{\beta}_j.$$

Da ciò si evince che Battese, Rao e O'Donnell (2004) definiscono la funzione meta-frontiera come una funzione parametrica deterministica, di una qualche forma specificata, tale che i suoi valori siano non inferiori ai valori corrispondenti alle diverse componenti deterministiche delle funzioni di frontiera stocastiche dei diversi gruppi.

La figura 3.3 illustra il modello della meta-frontiera nel caso in cui si abbiano tre diversi gruppi all'interno della medesima industria. Per ciascuna frontiera di produzione, i valori osservati sono indicati con il numero corrispondente mentre i valori non osservabili, sulla frontiera stocastica di ciascun gruppo, sono indicati con i numeri asteriscati. I valori sulla meta-frontiera sono sempre superiori o uguali alle diverse funzioni deterministiche associate alle frontiere stocastiche dei diversi gruppi.

L'output dell'unità produttiva i -esima, definito dalla frontiera stocastica per il gruppo j -esimo, può essere alternativamente espresso in termini della funzione meta-frontiera attraverso:

$$y_{ij} = e^{-u_{ij}} \times \frac{e^{x_{ij} \boldsymbol{\beta}_j}}{e^{x_i \boldsymbol{\beta}^*}} \times e^{x_i \boldsymbol{\beta}^* + v_j}$$

dove, il primo termine nella parte destra dell'equazione, rappresenta l'efficienza tecnica relativa alla frontiera stocastica del gruppo j -esimo, ossia:

$$ET_{ij} = \frac{y_{ij}}{e^{x_{ij} \boldsymbol{\beta}_j + v_j}} = e^{-u_{ij}};$$

mentre il secondo termine indica il cosiddetto rapporto di divergenza tecnologica (*Technology Gap Ratio, TGR*) per l'osservazione i -esima, ossia:

$$TGR_{ij} = \frac{e^{x_{ij}\beta_j}}{e^{x_i\beta^*}}.$$

Quest'ultimo rapporto esprime, per ciascuna unità produttiva, dato il livello di input impiegati, la distanza tra l'output sulla frontiera di produzione del gruppo j -esimo e l'output potenzialmente raggiungibile, definito dalla funzione meta-frontiera.

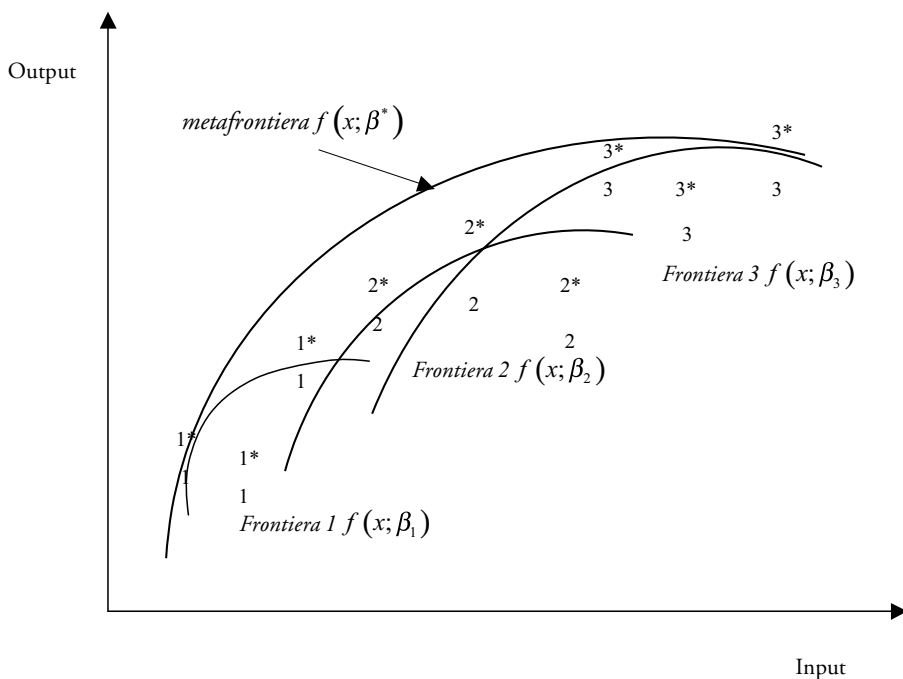


Figura 3.3 – Il modello meta-frontiera.

L'efficienza tecnica dell'unità produttiva i -esima rispetto alla meta-frontiera è definita dall'espressione:

$$ET_i^* = \frac{y_{ij}}{e^{x_{ij}\beta^* + v_{ij}}} = e^{-u_{ij}}.$$

Dalle espressioni precedenti deriva che una formulazione alternativa della misura di efficienza rispetto alla meta-frontiera è la seguente:

$$ET_i^* = ET_{ij} \times TGR_{ij}$$

dalla quale emerge una proprietà di scomposizione simile a quella evidenziata per il caso non parametrico. Infatti, la misura di efficienza rispetto alla meta-frontiera, definibile come efficienza complessiva, viene ottenuta come prodotto tra la misura di efficienza, che essendo definita rispetto alla frontiera di ciascun gruppo può denominarsi efficienza interna (*within*), e il *TGR* che può essere ricondotta ad una misura di efficienza esterna (*between*).

Per la stima dei parametri e delle misure associate alla meta-frontiera, Battese, Rao e O'Donnell (2004) suggeriscono di utilizzare la procedura seguente.

Dopo aver ottenuto le stime ML dei parametri della frontiera stocastica per ciascuno degli R gruppi, si stimano i parametri β^* della meta-frontiera in modo tale che la funzione stimata sia la migliore copertura delle componenti deterministiche delle frontiere stocastiche stimate per i diversi gruppi. Per identificare la migliore copertura, i suddetti Autori suggeriscono di utilizzare due criteri.

Secondo il primo criterio, basato sulla somma delle deviazioni assolute, le stime dei parametri della metafrontiera sono ottenute risolvendo il seguente problema di ottimizzazione:

$$\min L = \sum_{i=1}^N \left| \ln f(\mathbf{x}_i; \beta^*) - \ln f(\mathbf{x}_i; \hat{\beta}_j) \right|$$

soggetto a :

$$\ln f(\mathbf{x}_i; \beta^*) \geq \ln f(\mathbf{x}_i; \hat{\beta}_j)$$

Tra i diversi vantaggi derivanti dall'adozione del criterio precedente, gli Autori sottolineano che se $f(\mathbf{x}_i; \beta)$ è log-lineare nei parametri, il problema di ottimizzazione si semplifica nel seguente sistema di programmazione lineare:

$$\min L = \sum_{i=1}^N \left(\mathbf{x}_i \beta^* - \mathbf{x}_{ij} \hat{\beta}_j \right)$$

soggetto a :

$$\mathbf{x}_i \beta^* \geq \mathbf{x}_{ij} \hat{\beta}_j$$

Poiché le stime delle frontiere stocastiche per i diversi gruppi, $\hat{\beta}_j$, sono date nel problema di programmazione lineare, la soluzione può essere ottenuta in modo equivalente minimizzando la funzione obiettivo $L^* = \mathbf{x} \beta^*$, soggetta alla restrizione lineare:

$$\ln f(\mathbf{x}_i; \beta^*) \geq \ln f(\mathbf{x}_i; \hat{\beta}_j),$$

dove \bar{x} è il vettore riga di medie degli elementi dei vettori x_i per tutte le osservazioni.

Il secondo criterio suggerito dagli Autori, basato sulla somma dei quadrati delle deviazioni dei valori della meta-frontiera da quelli delle frontiere di ciascun gruppo, conduce al seguente problema di ottimizzazione vincolata:

$$\min L = \sum_{i=1}^N \left(x_i \beta^* - x_{ij} \beta_j \right)^2$$

soggetto a :

$$x_i \beta^* \geq x_{ij} \beta_j$$

A differenza del primo metodo che attribuisce un identico peso a tutte le distanze radiali per le unità considerate, questo metodo attribuisce un peso maggiore alle deviazioni associate alle imprese che hanno un più elevato TGR.

Le stime dell'efficienza tecnica di ciascuna unità produttiva rispetto alla meta-frontiera possono essere ottenute mediante: $\hat{ET}_i^* = \hat{ET}_{ij} \times \hat{TGR}_{ij}$, dove \hat{ET}_{ij} è la stima della misura di efficienza a livello individuale ottenuta utilizzando ad esempio lo stimatore proposto da Jondrow, Lovell, Materov e Schmidt (1982), mentre \hat{TRG}_{ij} è ottenuto utilizzando le stime dei parametri coinvolti.

È interessante osservare, in un tale contesto che, recentemente, Gori, Grassetti e Minotti (2004), analizzando l'efficienza degli ospedali in Lombardia, allo scopo di tenere in considerazione la struttura gerarchica dei dati, tipica nel settore della sanità, propongono un modello multilivello per la funzione di produzione, descritta da una funzione traslogaritmica, in cui non viene considerato il termine di errore unilaterale che caratterizza le frontiere di produzione.

Capitolo Quarto

L'EFFICIENZA RISPETTO ALLA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE: L'APPROCCIO NON PARAMETRICO

4.1 PREMESSA

L'approccio non parametrico raggruppa molteplici metodologie, le quali, pur avendo in comune il fatto di adottare una rappresentazione della tecnologia che privilegia l'insieme di produzione, o l'insieme di fabbisogno di input, presentano tratti distintivi in relazione alle caratteristiche della tecnologia di riferimento per la misurazione dell'efficienza.

La metodologia DEA, secondo l'approccio di Charnes, Cooper e Rhodes (1978) ma anche di Burley (1980), si riferisce, in genere, a tecnologie convenzionali, ossia tecnologie regolari, convesse e monotone, mentre i modelli suggeriti da Färe, Grosskopf e Lovell (1985) e da Deprins, Simar e Tulkens (1984) considerano tecnologie "non convenzionali" in cui viene rimossa, rispettivamente, la proprietà di monotonicità, ossia la libera disponibilità degli input e degli output, e la proprietà di convessità.

L'opportunità di considerare una tecnologia che soddisfi determinate ipotesi e la possibilità di verificarne la coerenza rispetto ai dati osservati, hanno rappresentato un importante oggetto di ricerca nell'ambito della teoria della produzione.

McFadden (1966), mettendo in evidenza come la specificazione dei limiti tecnologici alle possibili azioni dell'impresa possa essere descritta dall'insieme delle possibilità di produzione, determinato dallo stato della conoscenza e dalle leggi fisiche, rileva che l'importanza ricoperta, nella teoria della produzione tradizionale, dalle ipotesi convenzionali di disponibilità forte e di convessità deriva dalla convenienza analitica che ne consegue, piuttosto che dal loro realismo economico. Il suddetto Autore dimostra che tali ipotesi non risultano strettamente necessarie nello sviluppo e nell'applicazione di molteplici analisi economiche. Lo studio dei

casi in cui le due proprietà non risultano verificate è seguito, inoltre, da un'interessante illustrazione grafica.

Come già anticipato in precedenza, lo sviluppo di specifici test non parametrici per la verifica delle ipotesi su particolari proprietà della tecnologia risale ai lavori di Afriat (1972), Hanoch e Rothschild (1972) e prosegue con Varian (1984) e Färe, Grosskopf e Lovell (1987).

È importante osservare che le tecniche non parametriche per la determinazione di misure di input e output efficienza, rispetto ad una frontiera che inviluppa i dati osservati, costruita ipotizzando sia tecnologie convenzionali sia tecnologie non convenzionali, sono estremamente numerose. Tutte condividono però un obiettivo fondamentale: individuare quali tra le n unità produttive considerate nell'analisi formano la superficie di copertura (*envelopment surface*), ovvero costruire la funzione di produzione empirica o frontiera efficiente.

I due modelli DEA più frequentemente utilizzati nelle applicazioni empiriche¹, ossia il modello a rendimenti di scala costanti (CRS), introdotto da Charnes, Cooper e Rhodes (1978) e quello a rendimenti di scala variabili (*Variable Return to Scale*, VRS), sviluppato da Banker, Charnes e Cooper (1984), consentono l'identificazione di frontiere lineari, definite, secondo la terminologia anglosassone², "*piecewise linear envelopment surfaces*" (Grosskopf, 1986). Come rilevano Ali e Seiford (1993), è possibile determinare superfici di tipo non lineare, come, ad esempio, la "*piecewise log linear*" o la "*piecewise Cobb-Douglas*", se si utilizzano, nella determinazione dell'efficienza, combinazioni moltiplicative degli input e degli output relativi alle unità analizzate.

La determinazione della misura di efficienza per ciascuna delle n unità produttive costituenti il campione o la popolazione oggetto di analisi, implica la soluzione di n sistemi di programmazione lineare, costruiti secondo un particolare *orientamento*, che può riflettere il concetto di input o quello di output distanza.

Un diverso approccio, che non si basa sulla scelta di una particolare direzione rispetto alla quale definire la distanza, ma che fonda la valutazione dell'efficienza sulla contemporanea espansione degli input e contrazione degli output,

¹ Il modello a rendimenti di scala costanti e quello a rendimenti variabili sono frequentemente denominati utilizzando le iniziali degli Autori che ne hanno suggerito l'introduzione. Quindi, il modello a rendimenti costanti di scala è indicato anche con l'acronimo CCR, che deriva da Charnes, Cooper e Rhodes (1978) mentre il modello a rendimenti variabili di scala è definito anche BCC, dal lavoro di Banker, Charnes e Cooper (1984).

² È opportuno a questo punto sottolineare nuovamente che durante l'esposizione dei diversi concetti si farà ricorso molto spesso alla terminologia anglosassone, in modo particolare in questo capitolo, in quanto il suo uso è ormai invalso nella letteratura delle frontiere di produzione.

conduce alla formulazione di due misure di efficienza tecnica alternative a quelle derivanti dai cosiddetti modelli orientati. In particolare, Chambers, Chung e Färe (1996) introdussero la funzione di distanza direzionale (*directional distance function*), basata sulla “*benefit function*” di Luenberger (1992), mentre Färe, Grosskopf e Lovell (1985,1994) svilupparono la cosiddetta “*graph hyperbolic distance function*”. Entrambe le misure si basano sulla ricerca della proiezione efficiente di un processo produttivo osservato, (x_i, y_i) in modo tale da incrementare gli output prodotti e, simultaneamente, ridurre gli input utilizzati.

Il rapido sviluppo delle metodologie non parametriche, che si sono estese a tematiche concernenti la valutazione dell'efficienza rispetto alla scala di produzione attivata, l'inclusione di variabili categoriali, l'introduzione di criteri di giudizio esterni, l'analisi longitudinale, lo studio di problematiche connesse alle caratteristiche della copertura a cui si perviene (*proper envelopment*), e così via, è stato alimentato, come evidenziano Charnes, Cooper, Lewin e Seiford (1995) anche dalla necessità di risolvere problemi pratici che sorgevano durante il processo di applicazione empirica del metodo (Dyson, R.G., Allen, R., Camanho, A.S., Podinoski, V.V., Sarrico, C.S. e Shale, E.A., 2001).

Continuamente si aggiungono ad una già voluminosa letteratura importanti contributi nella forma di nuovi modelli, estensioni creative di modelli esistenti, applicazioni empiriche innovative in nuove aree ancora “inesplorate”.

In questo capitolo non si intende fornire una rassegna esaustiva dei diversi avanzamenti metodologici, per i quali si rinvia, ad esempio, ai testi di Coelli, Rao Prasada e Battese (1999), Cooper, Seiford e Tone, (2000), Thanassoulis (2001) e Ray (2004), ma si vuole evidenziare come lo sviluppo di tecniche specifiche, in contesti diversi e ad opera di diversi studiosi, sia confluito nella creazione di un insieme di strumenti metodologici per la costruzione di frontiere non parametriche, finalizzate alla misurazione dell'efficienza, estremamente flessibile ed in continuo mutamento ed adattamento alle più disparate situazioni reali.

Nell'esposizione dei concetti fondamentali che caratterizzano il suddetto approccio si seguirà un'impostazione basata sulle caratteristiche della tecnologia di produzione, espresse in termini di proprietà formali, che andranno a specificare una particolare “forma” della frontiera delle possibilità produttive rispetto alla quale, in termini di distanza, viene poi misurata l'efficienza di ciascuna unità produttiva considerata.

I riferimenti ad applicazioni empiriche saranno piuttosto limitati per due ragioni principali. In primo luogo, molti degli Autori che hanno contribuito allo sviluppo delle frontiere non parametriche si sono concentrati sull'aspetto metodologico non procedendo, nei lavori principali che si esporranno, a sperimentazioni degli strumenti suggeriti. In secondo luogo, poiché le metodologie non parametriche hanno interessato applicazioni nei campi più disparati, si ritie-

ne più interessante, in questo capitolo, dare maggiore spazio agli sviluppi metodologici, e considerare invece alcuni studi empirici in modo più omogeneo nel quinto capitolo, in relazione al settore dell'istruzione universitaria.

4.2 L'ANALISI DELL'EFFICIENZA RISPETTO A TECNOLOGIE CONVESSE:

LA DATA ENVELOPMENT ANALYSIS

4.2.1 Il modello di base con rendimenti di scala costanti

L'evoluzione dei metodi di programmazione lineare e la conseguente diffusione degli stessi nelle applicazioni, grazie ai lavori di Farrell e Fieldhouse (1962) e Boles (1967,1971), pone le basi per lo sviluppo di un nuovo approccio all'analisi dell'efficienza produttiva, quello non parametrico.

È il lavoro di Charnes, Cooper e Rhodes (1978) che, in letteratura, segna la nascita del metodo che sarà successivamente denominato *Data Envelopment Analysis* (DEA)³.

Appare doveroso osservare che, in realtà, con "*Measuring the efficiency of decision making units*", Charnes, Cooper e Rhodes (1978) si inseriscono in un ambito di ricerca che era stato precedentemente l'oggetto degli studi di diversi Autori.

In "*Discussion to Mr. Farrell Paper*", Hoffman (1957) osserva, infatti, che il metodo di programmazione lineare poteva essere adatto agli scopi di Farrell (1957) per l'individuazione di tutte le "superfici efficienti" (*efficient facets*) per la determinazione di una misura di *performance* per ogni singola unità di produzione.

Successivamente, Farrell e Fieldhouse (1962) recepiscono il suggerimento di Hoffman, fornendo anche un'applicazione empirica e proponendo una possibile estensione che considerava tecnologie caratterizzate da rendimenti crescenti di scala.

Ai lavori precedenti, vanno aggiunti gli sforzi dei cosiddetti "*The Berkeley Agricultural Economists*", tra cui, in particolare, va ricordato il già citato contributo di Boles (1971), tesi a migliorare e sviluppare il metodo di Farrell nell'ottica suggerita da Hoffman. Come osservano Førsund e Sarafoglou (2002), tali studiosi non hanno ricevuto però il meritato riconoscimento nella letteratura delle frontiere non parametriche, anche a causa della ridotta diffusione delle riviste su cui gli studi erano pubblicati.

³ Il termine Data Envelopment Analysis viene introdotto per la prima volta da Charnes, Cooper e Rhodes (1981).

Infine, il legame esistente tra il calcolo dell'efficienza di Farrell e l'applicazione dei problemi di programmazione lineare era stato adeguatamente evidenziato da Burley (1976, 1980), anche attraverso un'interessante applicazione empirica.

In una prima versione del lavoro, Charnes, Cooper e Rhodes dedicano ampio spazio al tentativo di collocare l'approccio proposto nel contesto della teoria neoclassica della produzione. In conseguenza delle difficoltà incontrate per la pubblicazione, gli Autori ridussero drasticamente i collegamenti esistenti con la teoria della produzione e si rivolsero verso un pubblico di studiosi di ricerca operativa (*Operational Research, OR*).

Un contributo fondamentale dello studio di Charnes, Cooper e Rhodes è rappresentato dal legame esplicito, che gli Autori evidenziano fortemente, tra un indice di produttività, che nel caso multi-input e multi-output aggrega fattori e prodotti in un'unica formula espressa dal rapporto tra una somma ponderata degli output e una somma ponderata degli input (*weighted outputs to weighted inputs*), e la misura di efficienza di Farrell, nel caso di rendimenti costanti di scala.

Si può ragionevolmente affermare che l'elemento decisivo per lo sviluppo della metodologia DEA, è rappresentato dalla necessità di trovare i pesi più "adatti", attraverso la massimizzazione del rapporto di produttività, soggetto ai vincoli derivanti dalla frontiera e da quello di normalizzazione. Si ottiene così la cosiddetta forma "frazionaria" (*ratio form*) del modello di CCR, che corrisponde ad una definizione di efficienza più ingegneristica che economica.

Charnes, Cooper e Rhodes coniano il termine *Decision Making Unit (DMU)*, per indicare l'unità produttiva generica⁴, allo scopo di includere anche organizzazioni *non profit* come, ad esempio, scuole e ospedali, il cui obiettivo non si riduce alla massimizzazione del profitto e la cui produzione riguarda un insieme di output, identificabili e misurabili, ottenuti attraverso l'impiego di specifiche quantità di input, ma per i quali, spesso, non sono disponibili i prezzi di mercato.

Ipotizzando di considerare n unità ($h = 1, \dots, i, \dots, n$), ciascuna delle quali produce m output impiegando s input, per cui l' i -esima unità impiega il vettore di input:

$$\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{si}) \in R_+^s$$

per produrre il vettore di output:

⁴ In questo capitolo, il termine DMU sarà utilizzato come sinonimo di unità produttiva.

$$y_i = (y_{1i}, y_{2i}, \dots, y_{mi}) \in R_+^m,$$

per ottenere una misura della produttività media per ciascuna *DMU* è necessario introdurre uno specifico criterio di aggregazione.

A tal fine, non disponendo di prezzi per i rispettivi output e input, si possono utilizzare i cosiddetti “prezzi ombra” (*shadow price*), specificando, per ciascuna *DMU* considerata:

$$v_i = (v_{1i}, v_{2i}, \dots, v_{si})$$

come il vettore di prezzi ombra per gli input e:

$$u_i = (u_{1i}, u_{2i}, \dots, u_{mi})$$

come il vettore di prezzi ombra per gli output. In altre parole, si ipotizza di disporre, per ciascuna unità produttiva, di un sistema di pesi, espresso dai vettori v_i e u_i , attraverso i quali misurare l'importanza relativa dei fattori e dei prodotti.

Utilizzando tale sistema di pesi, si può ottenere la seguente misura di produttività media per l'unità *i*-esima:

$$P_i = \frac{\sum_{r=1}^m u_{ri} y_{ri}}{\sum_{j=1}^s v_{ji} x_{ji}}$$

Va osservato che i prezzi ombra non solo devono essere non negativi, ma devono anche essere specificati in modo tale che nessuna unità produttiva presenti un indicatore di produttività superiore all'unità, ossia $P_i \leq 1 \forall i$.

Le condizioni precedenti si traducono formalmente nel seguente problema:

$$P_i = \frac{\sum_{r=1}^m u_{ri} y_{ri}}{\sum_{j=1}^s v_{ji} x_{ji}} \leq 1$$

$$u_{ri} \geq 0 \quad r = 1, \dots, m \quad v_{ji} \geq 0 \quad j = 1, \dots, s$$

Poiché, in generale, esistono molteplici vettori dei prezzi (u_i, v_i) che soddisfano l'espressione precedente, il criterio di selezione si basa sulla ricerca della coppia di vettori che massimizza la produttività media di ciascuna unità produttiva sotto il vincolo che tutte le misure ottenute debbano essere inferiori o uguali a 1.

Si ottiene, quindi, il seguente sistema di programmazione frazionaria o frazionale:

$$\max P_i = \frac{\sum_{r=1}^m u_{ri} y_{ri}}{\sum_{j=1}^s v_{ji} x_{ji}}$$

soggetto a :

$$\frac{\sum_{r=1}^m u_{ri} y_{ri}}{\sum_{j=1}^s v_{ji} x_{ji}} \leq 1; \quad b = 1, \dots, i, \dots, n \quad [4.1]$$

$$u_r, v_j \geq 0; \quad r = 1, \dots, m; \quad j = 1, \dots, s$$

Le informazioni sulle quantità di input impiegati e di output prodotti da ciascuna delle DMU considerate sono utilizzate come insieme di riferimento (*reference set*) per effettuare la valutazione della *performance*.

In questo senso, quindi, l'efficienza tecnica di ciascuna unità, rappresenta una misura di efficienza relativa, in quanto determinata in riferimento a tutte le altre unità facenti parte dell'insieme di produzione. In particolare, se $P_i = 1$ l'unità considerata opera in modo pienamente efficiente, mentre se $P_i \leq 1$ essa realizza il proprio processo di produzione in modo relativamente meno efficiente di quanto riesca a fare un'opportuna combinazione di altre unità simili.

I suddetti Autori osservano che il processo di massimizzazione è tale da attribuire a ciascuna DMU il sistema di pesi più favorevole. L'intento della metodologia è, infatti, quello di determinare un sistema di prezzi ombra, non noto in precedenza, consentendo alla singola unità produttiva di "scegliere" la coppia di vettori v_i e u_i in modo da massimizzare la propria misura di efficienza, sotto la condizione che nessuna tra le DMU considerate ottenga, con i medesimi pesi, un indice superiore a 1.

Utilizzando i risultati derivanti da precedenti studi (Charnes e Cooper, 1962, 1973), Charnes, Cooper e Rhodes (1978) traducono il problema di programmazione frazionaria precedente, di non facile soluzione, in una **forma lineare**.

Considerando dapprima il reciproco del problema [4.1], Charnes, Cooper e Rhodes (1978) ottengono il seguente modello per la determinazione della misura di efficienza:

$$\min f_i = \frac{\sum_{j=1}^s v_{ji} x_{ji}}{\sum_{r=1}^m u_{ri} y_{ri}}$$

soggetto a :

$$\frac{\sum_{j=1}^s v_{ji} x_{ji}}{\sum_{r=1}^m u_{ri} y_{ri}} \geq 1; \quad h = 1, \dots, i, \dots, n \quad [4.2]$$

$$u_r, v_j \geq 0; \quad r = 1, \dots, m; \quad j = 1, \dots, s$$

I suddetti Autori suggeriscono, quindi, di sostituire la precedente formulazione con un problema di programmazione lineare ordinario del tipo:

$$\max \phi_i$$

soggetto a :

$$\begin{aligned} \sum_{h=1}^n \lambda_h y_{rh} &\geq y_{ri} \phi_i & r = 1, \dots, m \\ \sum_{h=1}^n \lambda_h x_{jh} &\leq x_{ji} & j = 1, \dots, s \\ \lambda_h &\geq 0; & h = 1, \dots, i, \dots, n \end{aligned} \quad [4.3]$$

Il duale del problema [4.3] assume una forma assolutamente equivalente al modello [4.2].

Si può osservare, infatti, che né la funzione obiettivo né i vincoli vengono alterati se tutti i pesi sono moltiplicati per un fattore di scala non negativo k , ossia se:

$$\mu_{ri} = k u_{ri} \quad \omega_{ji} = k v_{ji}.$$

Ponendo:

$$k = \frac{1}{\sum_j v_{ji} x_{ji}}$$

si ottiene:

$$\sum_{j=1}^s \omega_j x_{ji} = \sum_{j=1}^s kv_j x_{ji} = \frac{1}{\sum_j v_{ji} x_{ji}} \sum_j v_j x_{ji} = 1$$

per cui il problema [4.2] si traduce nel problema di programmazione lineare ordinario duale al problema [4.3], ovvero assume la formulazione seguente:

$$\min g_i = \sum_{j=1}^s \omega_j x_{ji}$$

soggetto a :

$$-\sum_{r=1}^m \mu_r y_{rh} + \sum_{j=1}^s \omega_j x_{jh} \geq 0 \quad h = 1, \dots, i, \dots, n$$

$$\sum_{r=1}^m \mu_r y_{ri} = 1$$

$$\mu_r, \omega_j \geq 0 \quad j = 1, \dots, s \quad i = 1, \dots, n$$

[4.4]

nel quale, come osservano Charnes e Cooper (1985), si introduce la condizione

$$\sum_{r=1}^m \mu_r y_{ri} = 1.$$

Charnes, Cooper e Rhodes dimostrano che si può utilizzare la formulazione lineare per risolvere quella frazionaria; in altre parole, i problemi sono equivalenti. Dalla soluzione del problema [4.3], che rappresenta la cosiddetta *forma di avvolgimento* (*envelopment form*) si ricava una *misura di output efficienza tecnica*, espressa da:

$$ET_i^{\circ} = \frac{1}{\phi_i}$$

in quanto il problema è orientato nel senso degli output, ossia, come affermano Charnes e Cooper (1985) “...can be interpreted as a problem in which one maximizes the “intensity” ϕ_i of the output vector y_i subject to envelopment from above and below.”

In modo del tutto speculare è possibile considerare un problema di programmazione lineare che conduce alla determinazione della misura di input efficienza tecnica, rispetto ad una tecnologia caratterizzata da rendimenti costanti di scala e disponibilità forte. Il modello CCR orientato nel senso degli input è sviluppato, in particolare, nel lavoro di Banker, Charnes e Cooper (1984). In termini formali si ha:

$$\begin{aligned}
 & \min \theta_i \\
 & \text{soggetto a :} \\
 & \sum_{h=1}^n \lambda_h x_{jh} \leq \theta_i x_{ji} \quad j = 1, \dots, s \\
 & \sum_{h=1}^n y_{rh} \lambda_h \geq y_{ri} \quad r = 1, \dots, m \\
 & \lambda_h \geq 0; \quad h = 1, \dots, i, \dots, n
 \end{aligned} \tag{4.5}$$

dal quale si ottiene la misura di efficienza tecnica radiale, espressa dallo scalare θ .

Charnes, Cooper e Rhodes, riferendosi alla misura di efficienza di Pareto-Koopmans e definendo alcune variabili che esprimono la riduzione possibile nell'impiego di qualche input o l'incremento di un output a parità di input, denominate, rispettivamente, **input e output slack**, riconoscono che una DMU può essere definita efficiente solo nel caso in cui soddisfi contemporaneamente due condizioni. In primo luogo si deve avere: $g_i^* = \phi_i^* = 1$, per cui le unità sono efficienti nel senso di Farrell e nessuna espansione radiale è possibile, e, in secondo luogo, tutte le variabili *slack* devono uguali a zero⁵.

Le difficoltà sorgono a causa del fatto che la contrazione radiale di un vettore di input (o l'espansione radiale di un vettore di output), su cui si basa la misura di Farrell, produce un vettore che sicuramente appartiene all'isoquanto dell'insieme di fabbisogno di input (o dell'insieme degli output producibili) ma che può non appartenere al sottoinsieme efficiente, come illustrato nel secondo capitolo, e anche attraverso la Figura 2.3.

In riferimento al caso semplificato di processi caratterizzati dall'impiego di due input per la produzione di un unico output, il problema è facilmente esplicabile mediante la seguente rappresentazione grafica (Figura 4.1).

La frontiera efficiente risulta definita dalle unità produttive *B* e *C*. La misura di Farrell delle unità *A* e *D* è espressa, rispettivamente, dal rapporto $0A^1 / 0A$ e $0D^1 / 0D$. Una differenza sostanziale intercorre però tra i due punti sulla frontiera efficiente considerati come termine di paragone per *A* e *D*. Mentre il pun-

⁵ Sulla medesima problematica si è soffermata anche l'attenzione di Färe (1975), che riconosce il ruolo assunto dagli *slack factors* nella determinazione dell'efficienza tecnica in senso paretiano. Successivamente Banker, Charnes e Cooper (1984) suggeriscono l'utilizzazione di uno specifico problema di programmazione lineare per la determinazione degli *slack* che sarà esposto più avanti.

to D^1 appartiene al sottoinsieme efficiente, e quindi, risulta pienamente efficiente anche nel senso di Pareto-Koopmans, il punto A^1 giace sull'isoquanto e presenta un eccesso nell'uso del secondo input, espresso da A^1B e denominato, come già anticipato, input *slack*. Infatti, l'unità "virtuale" A^1 potrebbe ridurre ulteriormente la quantità impiegata dell'input x_2 producendo sempre la medesima quantità di output.

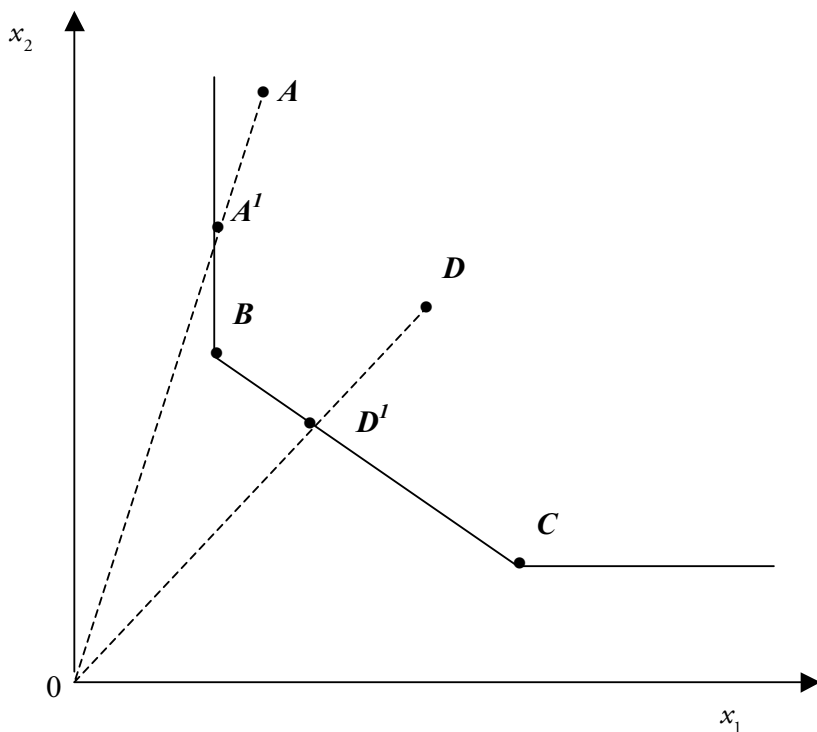


Figura 4.1 – La misura di efficienza e il problema dell'uso eccessivo degli input.

La valutazione della *performance* è affrontata da Deprins e Simar (1983) attraverso l'identificazione dell'insieme delle possibilità produttive, che, nel caso di rendimenti costanti di scala e disposizione forte, assume la configurazione geometrica di un cono, espresso da:

$$Z = \left\{ \begin{array}{l} (\mathbf{x}, \mathbf{y}) \in R^{s+m} \left| (\mathbf{x}, \mathbf{y}) = \sum_{h=1}^n \gamma_h (\mathbf{x}_h, \mathbf{y}_h) + \sum_{j=1}^s \mu_j (\mathbf{0}_s, \mathbf{e}_j^s) - \sum_{r=1}^m v_r (\mathbf{e}_r^m, \mathbf{0}_m) \right. \\ \left. \gamma_h \geq 0, \mu_j \geq 0, v_r \geq 0 \right. \end{array} \right\}$$

nel quale sono inclusi gli ortanti positivi, per $\mathbf{x} \in R^s$, e gli ortanti negativi, per $\mathbf{y} \in R^m$, attraverso i vettori $\mathbf{0}_s, \mathbf{0}_m$ (i cui elementi sono uguali a zero) e $\mathbf{e}_j^s, \mathbf{e}_r^m$ (che esprimono la j -esima e la r -esima colonna della matrice identità).

La misura di input efficienza, per l' i -esima unità osservata, è calcolata mediante la soluzione del problema di programmazione lineare seguente:

$$\min \theta_i$$

soggetto a :

$$y_i = \sum_{b=1}^n \gamma_b y_b - \sum_{r=1}^m v_r \mathbf{e}_r^m$$

$$\theta \mathbf{x}_i = \sum_{b=1}^n \gamma_b \mathbf{x}_b - \sum_{j=1}^s \mu_j \mathbf{e}_j^s$$

$$\gamma_i, \mu_j, v_r \geq 0, \quad \theta_i \geq 0$$

Considerando gli sviluppi metodologici precedenti, si può affermare che il merito fondamentale di Charnes, Cooper e Rhodes è quello di aver convertito il problema consistente nella massimizzazione della produttività di un'impresa in un problema ordinario di programmazione lineare, facilmente trattabile dal punto di vista informatico, che consiste nella massimizzazione della misura di efficienza di Farrell, soggetto al vincolo derivante dalla frontiera.

Va osservato, inoltre, che mentre Farrell (1957) presenta le sue idee in modo molto semplice per diffonderle, come lui stesso afferma, a "*statisticians, businessmen and civil servants, many of whom have little knowledge of economic theory or mathematics*", in Charnes, Cooper e Rhodes (1978) il problema legato al calcolo della misura di efficienza è presentato in una forma matematicamente più stringente ma anche comprensibile dalla comunità scientifica.

Una prima applicazione della metodologia DEA viene effettuata da Charnes, Cooper e Rhodes (1981) in relazione alla valutazione dell'efficienza di unità produttive operanti nell'ambito dell'istruzione pubblica.

Nel loro "*Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to Program Follow Through*" i suddetti Autori, non solo dimostrano l'applicabilità del nuovo metodo di programmazione matematica al settore dell'istruzione scolastica pubblica e i vantaggi che ne derivano, ma introducono per la prima volta il concetto di "valutazione del programma" (*program evaluation*), per indicare la presenza di diversi gruppi di unità omogenee all'interno dell'intero campione, rispetto ai quali è importante definire una specifica misura di efficienza tecnica.

Riferendosi solo ad alcune delle variabili disponibili, proprio al fine di illustrare le potenzialità applicative della DEA, i suddetti Autori propongono di descrivere il processo specificando gli input e gli output riprodotti nella Tabella 4.1.

| Unità produttive | Input | Output |
|------------------------------------|---|--|
| Scuole pubbliche negli Stati Uniti | 1. Livello di istruzione della madre (<i>percentage of high school graduates among female parents</i>) 2. Posizione lavorativa migliore dei membri della famiglia (<i>Highest occupation of a family member</i>) 3. Indicatore del numero di visite dei genitori nella scuola (<i>parent visit index</i>) 4. Indicatore di partecipazione dei genitori nell'attività scolastica (<i>parent counseling index</i>) 5. Numero di insegnanti (<i>Number of Teacher</i>) | 1. Abilità nella lettura (<i>Total Reading Score</i>) 2. Abilità in Matematica (<i>Total Mathematics Score</i>) 3. Misura dell'autostima (<i>Coopersmith Self-Eseem Inventory</i>) |

Tabella 4.1 – La rappresentazione del processo di produzione in Charnes, Cooper e Rhodes (1981): la specificazione degli input e degli output.

Suddividendo le DMU considerate in due gruppi, in funzione della loro partecipazione al progetto *Program Follow Through* (PFT), gli Autori ottengono misure di efficienza interna ed esterna, che non evidenziano una particolare superiorità in termini di performance per le scuole che partecipano al programma (PFT) rispetto a quelle che non vi partecipano (NFT). Nella Tabella 4.2, si riportano le misure di efficienza ottenute per alcune unità a titolo puramente esemplificativo.

| <i>Program Envelope</i> | | | | <i>Inter-Envelope</i> | | | |
|-------------------------|------|-------|------|-----------------------|------|-------|------|
| Unità | PTF | Unità | NFT | Unità | PTF | Unità | NFT |
| 1 | 1,00 | 50 | 0,95 | 1 | 0,92 | 50 | 1,00 |
| 2 | 0,90 | 51 | 0,92 | 2 | 1,00 | 51 | 1,00 |
| 3 | 0,98 | 52 | 1,00 | 3 | 0,94 | 52 | 1,00 |

Tabella 4.2 – Risultati empirici in Charnes, Cooper e Rhodes (1981): misure di efficienza.

Dall'esposizione precedente emerge che le prime applicazioni empiriche della DEA interessano in particolare il settore pubblico ed analizzano imprese multi-output, caratterizzate da una funzione obiettivo che non risulta rappresentata dalla massimizzazione del profitto o dalla minimizzazione dei costi.

Su tale argomento, Bessent e Bessent (1980) e Bessent, Bessent, Kennington e Reagan (1982) sottolineano che, per la valutazione dell'efficienza tecnica di una scuola, non può individuarsi metodologia migliore della DEA, in quanto consente di superare le difficoltà di specificazione degli input e degli output e di interpretazione dei risultati, che caratterizzavano i precedenti approcci basati sulla stima di una funzione di produzione.

In particolare, Bessent, Bessent, Kennington e Reagan, applicando la DEA a 167 scuole elementari statunitensi, descritte come unità produttive che impiegano 12 input (rappresentati da misure che esprimono le caratteristiche degli scolari e misure che esprimono le risorse a disposizione delle scuole) e 2 output (che riflettono diversi punteggi ottenuti da ciascun scolaro in relazione a particolari test cognitivi), mostrano che il 46,7% delle scuole non risultano, in realtà, efficienti.

I suddetti Autori presentano, quindi, un'analisi dei processi che caratterizzano le scuole più efficienti e quelle meno efficienti, in termini delle quantità impiegate di alcuni input, il cui uso risulta eccessivo, e che potrebbero essere riallocate o gestite dalle scuole stesse in maniera più efficiente. In tal modo vengono suggerite anche alcune indicazioni di *policy*.

4.2.2 Il modello a rendimenti di scala variabili

Nello sviluppo della metodologia DEA assume un ruolo centrale lo studio di Banker, Charnes e Cooper (1984) che segna l'introduzione del modello DEA a rendimenti di scala variabili.

In "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis" i suddetti Autori osservano, tra l'altro, che la formulazione espressa in Charnes, Cooper e Rhodes (1978, 1981) non rispecchia fedelmente l'usuale concetto di funzione di produzione sia a livello di singola unità produttiva sia a livello aggregato.

Infatti, il riferimento è ad una funzione che avviluppa (*envelope*), i dati osservati per il complesso delle n unità considerate, e che rappresenta la frontiera efficiente rispetto alla quale ciascuna unità deve essere valutata.

Con l'obiettivo di superare le difficoltà interpretative di una tale formulazione, Banker, Charnes e Cooper (1984) focalizzano l'attenzione sulla dimostra-

zione del legame esistente tra la funzione di distanza di Shephard (1970) e la misura di efficienza calcolata nel modello CCR.

In particolare, i suddetti Autori ipotizzano che l'insieme delle possibilità produttive Z soddisfi le seguenti proprietà:

1) convessità, secondo la quale:

$$\text{se } (\mathbf{x}_i, \mathbf{y}_i) \in Z, (i = 1, \dots, n), \mu_i \geq 0 \text{ e } \sum_{b=1}^n \mu_b = 1 \Rightarrow \left(\sum_{b=1}^n \mu_b \mathbf{x}_b, \sum_{b=1}^n \mu_b \mathbf{y}_b \right) \in Z$$

2) proprietà di disponibilità forte, che gli Autori denominano “*Inefficiency Postulate*”, secondo la quale:

$$\text{se } (\mathbf{x}_i, \mathbf{y}_i) \in Z \text{ e } \bar{\mathbf{x}} \geq \mathbf{x} \Rightarrow (\bar{\mathbf{x}}, \mathbf{y}) \in Z$$

e, in riferimento agli output:

$$\text{se } (\mathbf{x}_i, \mathbf{y}_i) \in Z \text{ e } \bar{\mathbf{y}} \leq \mathbf{y} \Rightarrow (\mathbf{x}, \bar{\mathbf{y}}) \in Z$$

3) proprietà di rendimenti costanti di scala, denominata “*Ray Unboundedness*”, per cui:

$$\text{se } (\mathbf{x}, \mathbf{y}) \in Z, k > 0 \Rightarrow (k\mathbf{x}, k\mathbf{y}) \in Z$$

4) proprietà di estrapolazione minima (*minimum extrapolation*), secondo la quale Z è l'intersezione di tutti gli insiemi di produzione, che soddisfano le proprietà precedenti e sono tali che ogni processo ne faccia parte, ossia:

$$(\mathbf{x}_i, \mathbf{y}_i) \in Z, i = 1, \dots, n.$$

L'insieme delle possibilità produttive, Z , a cui fare riferimento per effettuare la valutazione è quindi il più piccolo insieme coerente con i dati osservati e con le precedenti proprietà.

I suddetti Autori evidenziano, inoltre, come sia possibile caratterizzare ogni processo produttivo appartenente all'insieme delle possibilità produttive che soddisfi le proprietà suddette. In particolare, dalla proprietà di convessità e di rendimenti costanti di scala discende che ogni processo produttivo della forma

$$\left(k \sum_{b=1}^n \mu_b \mathbf{x}_b, k \sum_{b=1}^n \mu_b \mathbf{y}_b \right)$$

$$\text{con } k > 0, \mu_b \geq 0 \text{ e } \sum_{b=1}^n \mu_b = 1$$

si trova nell'insieme delle possibilità produttive Z . Dalla proprietà di disponibilità degli input e dalla minima estrapolazione deriva che $(\mathbf{x}_i, \mathbf{y}_i) \in Z$ se e solo se:

$$x_i \geq k \sum_{b=1}^n \mu_b x_{ib} \text{ e } y_i \leq k \sum_{b=1}^n \mu_b y_{ib} \text{ per } k > 0 \text{ e } \mu_b \geq 0$$

che soddisfa inoltre la condizione $\sum_{b=1}^n \mu_b = 1$.

Riprendendo il concetto di insieme di fabbisogno di input $L(y)$, Banker, Charnes e Cooper esprimono la funzione di input distanza, espressa da: $d_I(x, y) = 1 / \min \{ \lambda \mid \lambda x \in L(y) \}$, in relazione alla precedente caratterizzazione dei processi produttivi attraverso il sistema:

$$\min \lambda$$

soggetto a :

$$\lambda x \geq k \sum_{b=1}^n \mu_b x_b$$

$$y \leq k \sum_{b=1}^n \mu_b y_b \tag{4.6}$$

$$k > 0, \mu_b \geq 0 \text{ e } \sum_{b=1}^n \mu_b = 1$$

Sostituendo l'espressione $\lambda_b = k\mu_b$ e considerando che $F_I(x, y) = 1/d_I(x, y)$ si ottiene il problema di programmazione lineare che identifica il modello a rendimenti costanti orientato nel senso degli input [4.5].

Il duale al problema [4.6] rappresenta l'equivalente della metodologia introdotta da Charnes, Cooper e Rhodes (1978), che può essere esposta secondo la formulazione *non-Archimedean* (Charnes, Cooper e Rhodes, 1979) attraverso la quale la condizione di non negatività delle variabili u_r e v_j , che rappresentano i cosiddetti "pesi virtuali", viene sostituita dalla condizione di positività stretta, che nel contesto della analisi dell'attività risulta più aderente all'ipotesi per cui tutte le risorse impiegate e gli output prodotti debbano assumere valori positivi⁶.

⁶ L'espressione, dalla cui massimizzazione si ricava la misura di output efficienza, è la seguente:

$$\max z_i = \sum_{r=1}^m u_r y_{ri}$$

$$\text{soggetto a : } \sum_{r=1}^m u_r y_{rb} - \sum_{j=1}^s v_j x_{jb} \leq 0 \quad \sum_{j=1}^s v_j x_{j0} = 1 \quad u_r, v_j \geq \varepsilon \quad \forall r, j$$

dove $\varepsilon > 0$ è una quantità piccola a piacere (*non-archimedean*).

Il modello CCR, viene esteso da Banker, Charnes e Cooper (1984), i quali considerando situazioni meno restrittive sviluppano un nuovo strumento per la misurazione dell'efficienza delle unità produttive, in riferimento a tecnologie di produzione caratterizzate da **rendimenti variabili di scala**. Gli Autori non precedono però ad un'applicazione empirica della metodologia.

In tal modo è possibile distinguere la misura di efficienza puramente tecnica dalla misura di efficienza dovuta alla dimensione o scala di svolgimento del processo.

Considerando per l'insieme di produzione Z le proprietà di convessità, di disponibilità degli input e di minima estrapolazione, deriva che un processo (x, y) è nell'insieme Z , se e solo se:

$$x \geq \sum_{b=1}^n \lambda_b x_b \text{ e } y \leq \sum_{b=1}^n \lambda_b y_b \text{ per } \lambda_b \geq 0 ,$$

a cui si aggiunge la condizione:

$$\sum_{b=1}^n \lambda_b = 1$$

La funzione di distanza di Shephard riferita all'insieme di fabbisogno di input può essere espressa attraverso il precedente problema di programmazione lineare da cui si deriva il seguente problema duale, espresso, per l'unità i , nella seguente formulazione:

$$\begin{aligned} & \max \sum_{r=1}^m u_r y_{ri} - u_i \\ & \text{soggetto a :} \\ & \sum_{r=1}^m u_r y_{rb} - \sum_{j=1}^s v_j x_{jb} - u_i \leq 0 \quad b = 1, \dots, n \\ & \sum_{j=1}^s v_j x_{ji} = 1 \\ & u_r, v_j \geq 0 \end{aligned} \tag{4.7}$$

nella quale u_i può assumere qualsiasi segno.

Banker, Charnes e Cooper (1984), evidenziano che il valore di u_i , ottenuto dalla soluzione del problema precedente, indica la presenza di rendimenti crescenti di scala, se $u_i^* < 0$, di rendimenti costanti di scala nel caso in cui $u_i^* = 0$, e infine, di rendimenti decrescenti se $u_i^* > 0$.

La formulazione del modello VRS nella *envelopment form* per la determinazione della misura di efficienza, riproduce esattamente il problema [4.3], se orientato nel senso degli output, oppure il problema [4.5], se l'orientamento è di contrazione degli input, a cui si aggiunge un ulteriore vincolo sulla somma dei pesi, espresso da $\sum_{i=1}^n \lambda_i = 1$. In definitiva, per ottenere la misura di efficienza *output-oriented* con rendimenti di scala variabili è necessario risolvere, per ciascuna unità considerata, il seguente problema:

$$\begin{aligned} & \max \phi_i \\ & \text{soggetto a :} \\ & \sum_{h=1}^n \lambda_h y_{rh} \geq y_{ri} \phi_i \quad r = 1, \dots, m \\ & \sum_{h=1}^n \lambda_h x_{jh} \leq x_{ji} \quad j = 1, \dots, s \\ & \lambda_h \geq 0; \quad h = 1, \dots, i, \dots, n \\ & \sum_{h=1}^n \lambda_h = 1 \end{aligned}$$

Per poter identificare le situazioni caratterizzate dalla presenza di *slack* nel caso di processi con rendimenti di scala variabili, Banker, Charnes e Cooper suggeriscono di utilizzare il seguente problema, espresso nella formulazione input-oriented:

$$\begin{aligned} & \min \theta_i - \varepsilon \left[\sum_{j=1}^s s_j + \sum_{r=1}^m s'_r \right] \\ & \text{soggetto a :} \\ & \theta_i x_{ji} - \sum_{h=1}^n x_{jh} \lambda_h - s_j = 0 \quad j = 1, \dots, s \\ & y_{ri} = \sum_{h=1}^n y_{rh} \lambda_h - s'_r \quad r = 1, \dots, m \\ & \sum_{h=1}^n \lambda_h = 1 \\ & \lambda_i, s'_r, s_j \geq 0; \quad \forall i, j, r \end{aligned} \tag{4.8}$$

nel quale s_j e s_r indicano, rispettivamente, gli *slack* negli input e negli output che caratterizzano un determinato processo produttivo.

Va osservato che la misura di output efficienza $ET_i^\theta = 1/\phi_i$ e quella di input efficienza $ET_i^l = \theta_i$ assumeranno lo stesso valore solo nel caso in cui la tecnologia sia caratterizzata da rendimenti costanti di scala, e quindi le due misure siano ottenute risolvendo rispettivamente, il sistema [4.3] oppure il problema [4.5].

Poiché, sotto l'ipotesi di rendimenti variabili di scala, la misura di output efficienza non coincide con quella di input efficienza, la scelta del tipo di orientamento, ossia l'espansione degli output o la contrazione degli input, assume un'importanza determinante ai fini dell'interpretazione dei risultati.

Mediante una rappresentazione grafica, riprodotta nella Figura 4.2, Banker, Charnes e Cooper illustrano i concetti di efficienza tecnica ed efficienza di scala.

La frontiera efficiente nell'ipotesi di rendimenti di scala variabili è rappresentata dalla spezzata che congiunge i punti *FGEC*, mentre nel caso si ipotizzi la presenza di rendimenti di scala costanti la frontiera è costituita dalla linea *OS* passante per il punto *E*. Infine, la linea spezzata *OEC* identifica la frontiera caratterizzata da rendimenti di scala non crescenti (*NIRS*) rispetto alla frontiera *CRS*.

In riferimento al punto *A*, che rappresenta l'unità da valutare, si può determinare una misura complessiva di input efficienza tecnica (*overall technical efficiency*), espressa dal rapporto MN/MA ottenuta confrontando il punto *A* con il punto *N* che riflette la produttività media raggiungibile alla dimensione di scala più produttiva (*most productive scale size*, *MPSS*) rappresentata dal punto *E*.

L'efficienza puramente tecnica (*pure technical efficiency*) di *A*, nel senso degli input, ottenuta dal confronto con il punto *B* sulla frontiera efficiente *VRS*, caratterizzato dalla stessa dimensione di scala di *A*, è misurata dal rapporto MB/MA . Quindi, l'efficienza di scala di *A* rispetto agli input è misurata dal rapporto MN/MB , per cui l'efficienza complessiva è esprimibile mediante il prodotto tra l'efficienza di scala e la pura efficienza tecnica, ossia $MN/MA = MB/MA \cdot MN/MB$.

Si osservi che la misura di efficienza complessiva, ossia che comprende sia la misura di efficienza tecnica sia quella di scala, derivabile dalla soluzione del problema di programmazione lineare che esprime un modello *CRS*, risulta sempre inferiore o uguale alla misura di efficienza tecnica pura, ottenibile mediante la soluzione del modello *VRS*. Una misura di efficienza di scala (*ES*), per ciascuna *DMU*, può quindi ottenersi mediante il rapporto tra le due precedenti misure. In termini formali si ha:

$$ES = \frac{ET_{CRS}}{ET_{VRS}}.$$

Per stabilire se l'inefficienza di scala sia da attribuire ad un sottodimensionamento o a un sovradimensionamento dell'unità produttiva si può procedere confrontando le misure di efficienza di Farrell determinate rispetto alle frontiere caratterizzate da rendimenti di scala costanti e da rendimenti di scala non crescenti. Dunque, nell'esempio della Figura 4.2 per il punto A si possono determinare le misure di efficienza seguenti: $TE_{VRS} = MB/MA$, $TE_{CRS} = TE_{NIRS} = MN/MA$.

Poiché $TE_{CRS} = TE_{NIRS} < TE_{VRS}$ si è in presenza di rendimenti di scala crescenti, in altre parole l'unità è sottodimensionata. Considerando, invece, il punto D in Figura 4.2 si verifica che: $TE_{VRS} = TE_{NIRS} > TE_{CRS}$ per cui la proiezione del punto D, nel senso degli input, si trova nella regione caratterizzata da rendimenti decrescenti di scala, che manifestano un sovradimensionamento dell'unità produttiva rispetto alla scala di produzione efficiente.

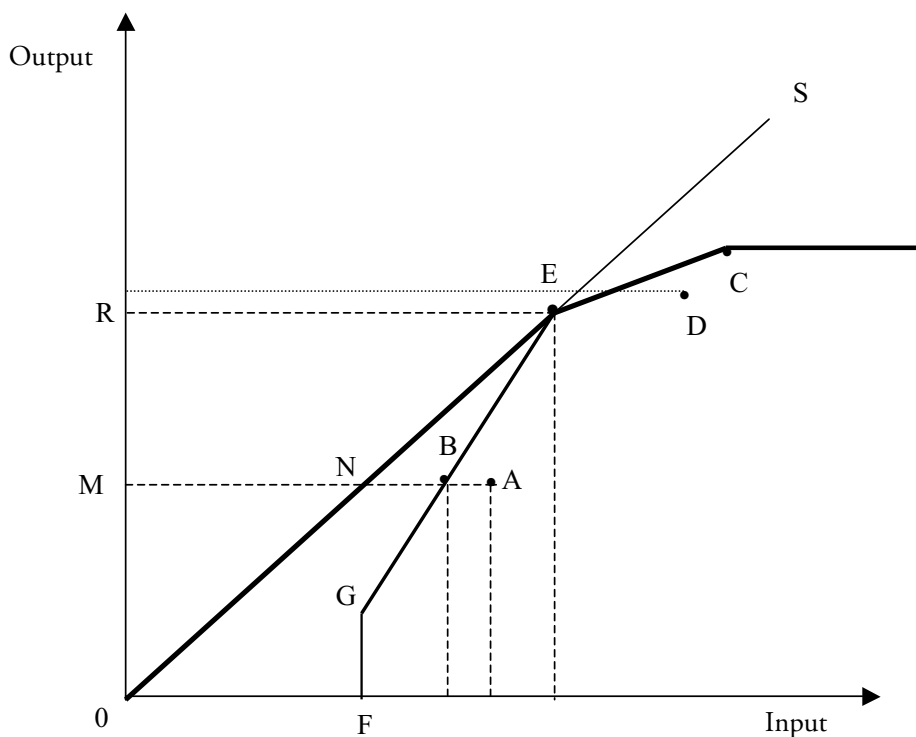


Figura 4.2 – Efficienza tecnica e di scala.

È d'obbligo rilevare, brevemente, che per l'identificazione della natura dei rendimenti di scala in qualunque punto della frontiera, sono stati sviluppati in letteratura due approcci alternativi rispetto al suddetto criterio di Banker, Charnes e Cooper, definito anche approccio duale (Ray, 2004).

Il primo metodo, denominato, approccio primario (*primal approach*), introdotto da Banker (1984) si ricollega all'individuazione della MPSS in un contesto multi input e multi output (Starrett, 1977). In particolare, Banker (1984) dimostra la relazione esistente tra un processo che si trova in una MPSS all'interno di un insieme delle possibilità produttive con rendimenti variabili e la soluzione ottima di un problema DEA-CCR.

Il secondo metodo, come già anticipato, è stato introdotto da Färe, Grosskopf e Lovell (1985) ed è denominato, ricorrendo alla terminologia inglese, "*nesting approach*". Tale approccio si basa sull'analisi della relazione gerarchica tra insiemi di produzione sotto ipotesi alternative concernenti i rendimenti di scala.

Tra le prime applicazioni del modello a rendimenti variabili si può considerare lo studio di Färe, Grosskopf e Weber (1989), nel quale l'approccio non parametrico è utilizzato per la misurazione della *performance* nel settore pubblico, ed in particolare, di un campione di scuole del Missouri. In riferimento al concetto di funzione di produzione della formazione (*educational production function*) i suddetti Autori evidenziano i vantaggi derivanti dal poter includere molteplici output nell'analisi di efficienza. L'intento è di introdurre maggiore flessibilità, attraverso la considerazione di rendimenti variabile di scala e all'utilizzazione di test statistici ottenuti dall'applicazione della procedura *jackknife*.

La DEA si è rivelata una metodologia particolarmente adatta ad analizzare l'efficienza del settore creditizio italiano⁷ a causa delle peculiarità che lo caratterizzano (Resti, 1994).

Bruschini e Viviani (2002) utilizzano il metodo DEA per la misurazione della performance delle filiali di una banca operante su un distretto industriale territorialmente individuato. L'analisi dell'efficienza, condotta per tre gruppi dimensionali, viene effettuata in riferimento a diverse funzioni obiettivo che caratterizzano la gestione interna.

Una tale impostazione, di tipo "micro", era stata già suggerita da Gazzei e Viviani (2001) in uno studio sull'efficienza interna delle diverse filiali di 22 filiali di un'azienda di credito operante nell'Italia centrale condotto utilizzando un modello DEA a rendimenti di scala variabili.

⁷ Le applicazioni della DEA al settore creditizio sono molto numerose anche in altri Paesi. Si può citare, ad esempio, Hartman e Storbeck (1996).

I suddetti Autori individuano un sistema di indicatori ordinati ad albero, esplicativi della performance complessiva, la cui interpretazione viene arricchita attraverso una modellizzazione statistica, superando la semplice impostazione caratterizzata l'approccio basato sui singoli rapporti (*ratio*).

Come già anticipato, le frontiere non parametriche sono proficuamente utilizzate per valutare l'efficienza di unità produttive operanti in particolari settori del trasporto, come, ad esempio, il trasporto pubblico, con i già citati lavori di Levaggi (1994), Cambini e Filippini (2003) e Fazioli, Filippini e Künzle (2003) e la gestione dei servizi portuali, con il lavoro di Marchese, Ferrari e Benacchio (2000). Quest'ultimi Autori analizzano l'efficienza tecnica dei terminal portuali dedicati alla movimentazione di container utilizzando un modello DEA con rendimenti di scala variabili orientato nel senso degli input.

4.2.3 Misure di efficienza non radiali: il modello DEA additivo e il modello per la determinazione della misura di Russell

Il problema derivante dalla presenza di *slack*, nella soluzione ottima dei modelli DEA a rendimenti costanti e a rendimenti variabili, sorge perché l'obiettivo consiste nell'espandere, contemporaneamente, tutti gli output o di contrarre tutti gli input secondo la medesima proporzione. La misura di efficienza che deriva dalla soluzione del problema di programmazione lineare, ad esempio, del problema [4.5] è, infatti, come sottolineato, una misura di efficienza radiale.

Nei modelli non radiali si consente, invece, ad ogni singolo output di incrementarsi e ad ogni input di contrarsi, secondo proporzioni diverse.

Charnes, Cooper, Golany, Seiford e Steetz (1985) svilupparono il cosiddetto modello additivo (*additive model*), che, ad esempio, nella formulazione più semplice, orientata nel senso degli output, massimizza la somma totale degli *slack* che esistono nei diversi output dei processi produttivi osservati⁸, proiettando in tal modo il processo considerato sul sottoinsieme efficiente dell'insieme degli output producibili.

In particolare, sotto l'ipotesi di rendimenti di scala variabili, per l'unità i , si ha il seguente problema:

⁸ In modo analogo, nel modello additivo orientato nel senso degli input si massimizza la somma degli *slack* negli input. Esiste, inoltre, anche un modello in cui si massimizza la somma complessiva degli *slack*, sia negli input che negli output.

$$\max S = \sum_{r=1}^m s'_r$$

soggetto a :

$$\sum_{h=1}^n y_{rh} \lambda_h + s'_r = y_{ri} \quad r = 1, \dots, m$$

$$\sum_{h=1}^n x_{jh} \lambda_h \leq x_{ji} \quad j = 1, \dots, s \quad [4.9]$$

$$\sum_{h=1}^n \lambda_h = 1 \quad \lambda_h, s'_r \geq 0; \quad \forall i, j, r$$

Come sottolinea Ray (2004), l'utilità principale del modello additivo consiste nella circostanza per cui è possibile identificare con certezza l'appartenenza del vettore di output osservato y_i al sottoinsieme efficiente dell'insieme $P(\mathbf{x})$. Si può concludere, infatti, che $y_i \notin P(\mathbf{x})$ a meno che la funzione obiettivo S non sia uguale a 0 nella soluzione ottima.

Un aspetto controverso legato all'impiego del modello additivo è rappresentato dal fatto che la funzione obiettivo S non ha un'interpretazione chiara, essendo la somma degli *slack* in relazione ai diversi output, valutati utilizzando diverse unità di misura. Inoltre, la dimensione complessiva di S dipende dalla scala di misurazione degli output.

Una soluzione ai problemi derivanti dall'utilizzazione di una misura di efficienza di tipo radiale consiste nell'impiego della misura di Russell (*RM, Russell Measure*), già introdotta nel secondo capitolo.

Tale misura, definita da $RM_y = \frac{1}{\rho_y}$ può essere agevolmente determinata, risolvendo il seguente sistema (orientato nel senso degli output):

$$\rho_y = \max \frac{1}{m} \sum_r \phi_r$$

$$\text{soggetto a : } \sum_{h=1}^n \lambda_h y_{rh} = \phi_r y_{ri} \quad r = 1, \dots, m$$

$$\sum_{h=1}^n \lambda_h x_{jh} \leq x_{ji} \quad j = 1, \dots, s \quad [4.10]$$

$$\sum_{h=1}^n \lambda_h = 1 \quad \phi_r \geq 1$$

$$\lambda_h \geq 0; \quad h = 1, \dots, n$$

Va osservato che la misura di Russell può essere considerata una versione invariante nella scala del semplice modello additivo (si veda la Formula 4.9). Infatti, definendo:

$$y_{ri} + s'_r = y_{ri} \left(1 + \frac{s'_r}{y_{ri}} \right) \equiv \phi_r y_{ri} \quad r = 1, \dots, m$$

si ottiene

$$\rho_y = 1 + \frac{1}{m} \sum_r \frac{s'_r}{y_{ri}},$$

da cui si evince che, essendo gli *slack* per i singoli output divisi per le rispettive quantità osservate, i fattori ρ_y , e, quindi, la misura di Russell è invariante per cambiamenti di scala.

Ray (2004) osserva che, nel caso in cui esistano degli *slack* negli output in corrispondenza di una soluzione ottima in un modello DEA radiale, allora la misura non radiale di Russell sarà più piccola della misura convenzionale ottenuta dalla soluzione di un modello BCC orientato nel senso degli output.

La precedente misura, proposta da Färe e Lovell (1978), come un indicatore di efficienza “migliore” di quello suggerito da Farrell, è stata oggetto dell’attenzione di diversi studiosi.

In particolare, Russell (1985) dimostra che la misura non radiale di Färe e Lovell non soddisfa un insieme di proprietà desiderabili per un indicatore di efficienza, tra cui la proprietà di monotonicità stretta e la proprietà di omogeneità, già analizzata da Zieschang (1985).

Quest’ultimo Autore suggerisce una misura di Russell estesa, ottenibile attraverso la soluzione di una procedura costituita da due fasi. Nella prima fase si ottiene una misura radiale mentre nella seconda si massimizza la somma degli *slack*.

La limitazione principale della procedura precedente consiste nel fatto che, massimizzando la somma degli *slack*, si perviene all’individuazione del punto sulla frontiera più lontano (*furthest efficient point*) da quello osservato. In tal modo la valutazione della *performance* potrebbe non essere effettuata in relazione ad unità simili a quella osservata.

A tale proposito, Coelli (1998) suggerisce una procedura multi-stadio (*multi-stage*) consistente in una sequenza di problemi di programmazione lineare di tipo radiale⁹.

4.3 L'ANALISI DELL'EFFICIENZA RISPETTO A TECNOLOGIE NON MONOTONE

4.3.1 Il concetto di congestione degli input e degli output

Come osservano Borts e Mishan (1962), fin dall'introduzione in letteratura del concetto di isoquanto o *production indifference curve* dovuta a Johnson (1913), l'attenzione degli economisti si è concentrata, in modo prevalente, nell'esaminare unicamente la "*economic range*", ossia la parte dell'isoquanto rispetto alla quale il prodotto marginale di entrambi i fattori è positivo e nella quale può essere individuato il punto di minimo costo, dato un qualunque vettore di prezzi.

Considerare solo la regione economica dell'isoquanto potrebbe essere sufficiente se si ritiene che il produttore sia completamente libero di scegliere la combinazione che minimizza i costi. Possono però sorgere situazioni nelle quali la scelta dei fattori da impiegare risulta vincolata ed è, quindi, necessario estendere l'analisi anche alle aree esterne all'area economicamente vantaggiosa.

L'ampliamento dell'analisi fino ad includere anche le cosiddette regioni non economiche (*uneconomic region*), definite come regioni nelle quali il prodotto marginale di uno o più fattori è negativo, era stata precedentemente discussa in letteratura da Allen (1947) ma, a detta degli Autori, senza l'accuratezza necessaria per la trattazione dell'argomento.

Contrastando l'opinione prevalente nella letteratura neoclassica che vede la funzione di produzione definita solo per la regione economica, Borts e Mishan (1962), modellano tecnologie con isoquanti caratterizzati da inclinazioni positive, riferendosi per convenienza espositiva a funzioni omogenee di primo grado.

Successivamente, come già anticipato, Mc Fadden (1978) sviluppa un'interessante analisi dell'ipotesi di disposizione debole (*weak disposability*) degli input, sottolineando che l'assunzione standard di *free disposal* è usualmente invocata per convenienza analitica piuttosto che per il suo realismo economico.

In riferimento alla funzione di produzione, il concetto di **congestione** fu introdotto da Färe e Svensson (1980) per caratterizzare le tecnologie nelle quali non è possibile disporre liberamente dei fattori produttivi.

Secondo i suddetti Autori esempi di tecnologie di produzione che presentano congestione si ritrovano frequentemente in agricoltura e nei trasporti.

⁹ Tale procedura è implementata nel software DEAP ideato da Coelli (1996) che sarà oggetto di attenzione nel sesto capitolo.

In tali casi si verifica che, mantenendo costante l'impiego di un sottoinsieme di fattori produttivi, l'incremento di altri input determina un "ingorgo" nella produzione degli output. Färe e Svensson propongono, quindi, un trattamento assiomatico generale di queste situazioni tecnologiche in riferimento alla funzione di produzione, specificando diverse forme di congestione di diversa intensità¹⁰.

In "*Measuring Congestion in production*" Färe e Grosskopf (1983a), allo scopo di includere il concetto di congestione, generalizzano la misurazione dell'efficienza, seguendo l'impostazione di Farrell (1957).

Si può ragionevolmente affermare che il mancato interesse verso la definizione di una misura della congestione nei primi lavori sull'efficienza derivi, in parte, dall'ipotesi sottostante l'approccio di Farrell e di altri studiosi. Si ritiene infatti che gli isoquant non siano mai positivamente inclinati.

In altre parole, ripetendo l'aspetto centrale del problema, si assume solitamente che la produzione avvenga esclusivamente all'interno della regione economica, dove tutti i prodotti marginali sono non negativi, e mai in regioni non economiche, dove qualche fattore presenta un prodotto marginale negativo. Tutto ciò equivale ad imporre la proprietà di disponibilità forte degli input, secondo la quale l'incremento di alcuni input non dovrebbe causare la diminuzione dell'output, o, viceversa, la diminuzione di alcuni input non dovrebbe determinare l'incremento dell'output.

In tal modo si preclude la possibilità di far emergere il fenomeno della congestione, che, invece, può presentarsi nelle situazioni reali e può condizionare lo svolgimento del processo di produzione. In alcuni casi la produzione può avvenire in regioni non economicamente vantaggiose a causa dell'indivisibilità degli input (Borts e Mishan, 1962), o a causa di qualsiasi altro fenomeno esogeno che ostacola il produttore nella scelta delle quantità ottime da impiegare.

Allo scopo di estendere la misurazione dell'efficienza per includere la congestione degli input, Färe e Grosskopf specificano una tecnologia regolare che non impone la disponibilità forte degli input, ma solo la disponibilità debole espressa da:

$$\mathbf{x} \in L(\mathbf{y}) \Rightarrow \lambda \mathbf{x} \in L(\mathbf{y}), \quad \lambda \geq 1.$$

Determinante per la definizione della misura di congestione, come parte integrante della misura di efficienza tecnica, risulta la distinzione, consentita

¹⁰ È inoltre sottolineata l'importanza della congestione nel provare la legge delle proporzioni variabili.

dall'assenza della proprietà di disponibilità forte, tra l'isoquante dell'insieme di fabbisogno di input, $IsoqL(y)$, e il sottoinsieme efficiente debole (*weak efficient subset*), espresso da:

$$WEffL(y) = \left\{ x \mid x \in L(y) \text{ e } u < x \Rightarrow u \notin L(y) \right\}$$

dove la relazione $<$ tra i due vettori comprende anche il caso in cui alcuni degli elementi dei vettori posti a confronto siano uguali a zero (Färe, Grosskopf e Lovell, 1985). Il sottoinsieme $WEffL(y)$ si trova sempre nella regione economica della produzione, a differenza dell'isoquante.

Sostituendo l'ipotesi di *free disposal*, con una versione debole della proprietà, si possono considerare tutte quelle situazioni nelle quali la disposizione dei fattori in eccesso non è "libera" ma comporta dei costi per l'unità produttiva.

Le situazioni appena descritte possono essere illustrate graficamente attraverso la Figura 4.3, nella quale l'insieme di fabbisogno di input che soddisfa la proprietà di disposizione forte, indicato con $L^S(y)$, consiste nell'insieme dei vettori di input racchiusi dall'area¹¹ A^*BCDF^* , mentre l'insieme più ristretto che soddisfa la disposizione debole, $L^W(y)$, è rappresentato dall'insieme dei vettori compresi nell'area $ABCDF$.

L'isoquante dell'insieme $L^S(y)$, ossia $IsoqL^S(y)$, è costituito dalla spezzata che congiunge i punti A^*BCDF^* , mentre il sottoinsieme $WEffL(y)$ è rappresentato dai segmenti BC e CD .

Considerando la congestione come un tipo particolare di inefficienza tecnica, caratterizzata da spreco di risorse o da perdita di output, Färe e Grosskopf (1983a) specificano le varie forme di efficienza riferendosi al diagramma riprodotto in Figura 4.4, nel quale il segmento AA^1 rappresenta il piano dei prezzi, che corrisponde alla minimizzazione dei costi, mentre $L^W(y)$ è l'insieme di fabbisogno di input caratterizzato dalla proprietà debole di disponibilità degli input. In riferimento al processo produttivo P e relativamente al punto S , i suddetti Autori definiscono tre componenti di efficienza.

In primo luogo, considerando le medesime assunzioni di Farrell, l'*efficienza tecnica* è espressa dal rapporto OR/OP . Questa misura viene poi scomposta in due componenti: i) una **misura di efficienza "puramente" tecnica**, ottenuta dal rapporto OQ/OP e ii) una **misura di congestione** rappresentata da $C = OR/OQ$. Il terzo tipo di efficienza, allocativa o di prezzo (*allocative efficiency*) è rappresentata dal rapporto $EA = OS/OR$ e coincide con la definizione originale di Farrell.

¹¹ Si considerano tutti i punti, rappresentanti vettori possibili, posizionati sulla spezzata e a destra della stessa.

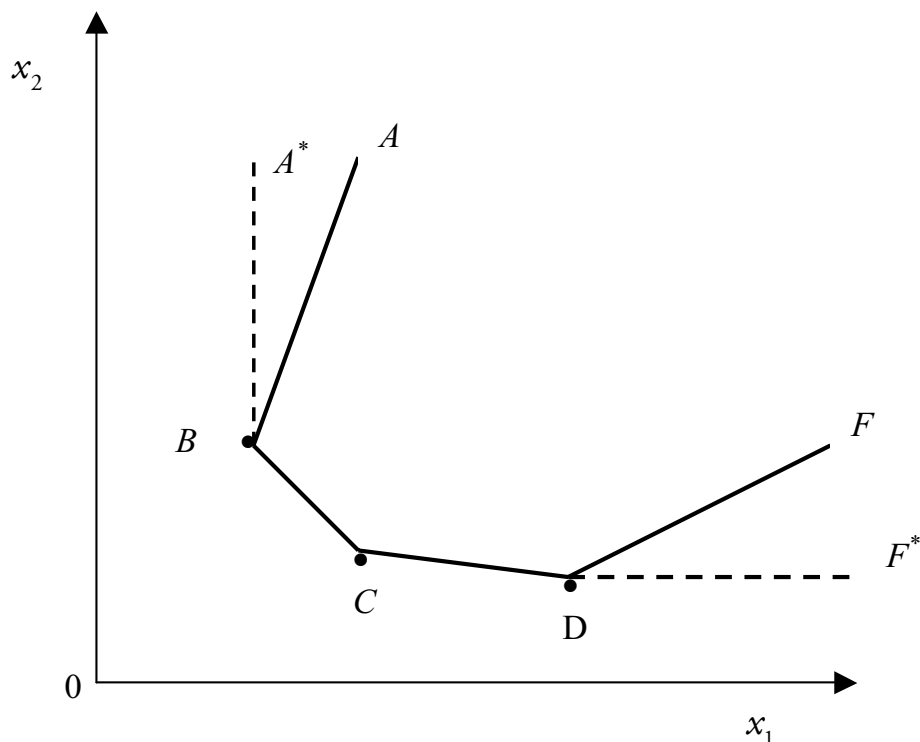


Figura 4.3 – Disponibilità forte e disponibilità debole degli input.

Le tre misure radiali definiscono insieme la misura di efficienza complessiva EC (*overall efficiency*) espressa dal rapporto $EC = OS/OP$.

In relazione ai precedenti sviluppi si nota che per misurare la congestione, utilizzando una misura radiale, è necessario riferirsi anche ad una tecnologia monotona, per la quale vale la *free disposal*, e quindi al punto R . Ovviamente l'output y non può essere prodotto nel punto R ma, diminuendo l'ammontare impiegato del secondo input, potrebbe essere prodotto nel punto V .

A causa della presenza di congestione, la diminuzione di un input non determina la diminuzione dell'output. In altre parole, il processo produttivo osservato può collocarsi sull'isoquanto, come, ad esempio accade per il punto Q , ma nell'area non economica della produzione, ossia dove l'isoquanto si ripiega all'indietro assumendo inclinazione positiva.

Poiché Farrell si riferì ad una tecnologia monotona, il concetto di congestione, e quindi la sua misura, risultava “nascosto”, ovvero incluso, in quello di efficienza tecnica.

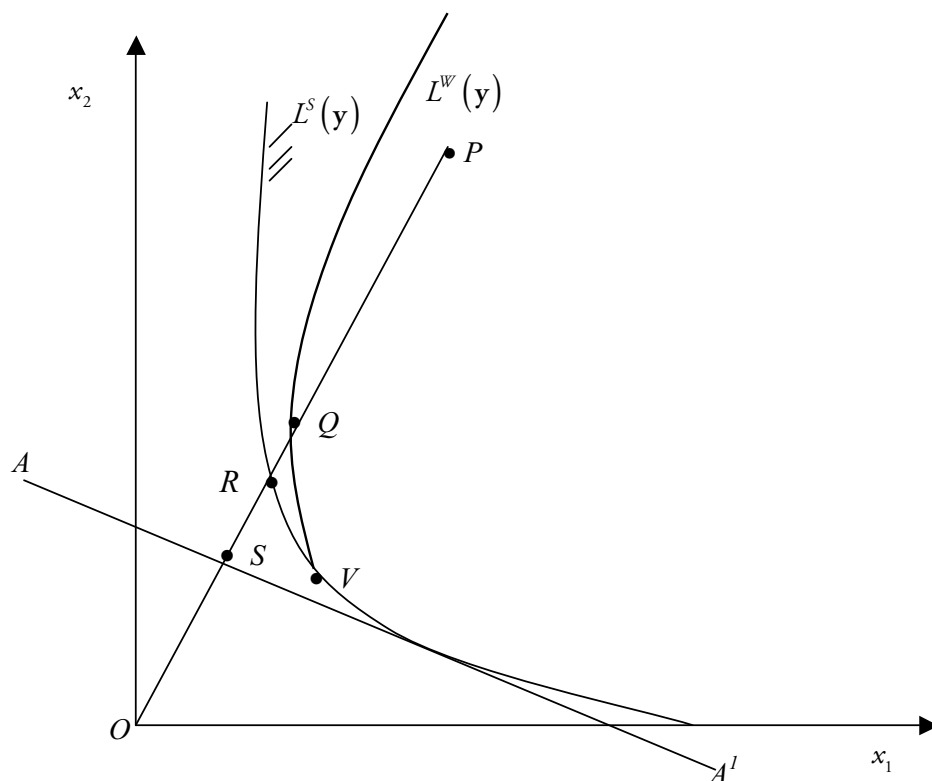


Figura 4.4 – Le misure di efficienza.

I suddetti Autori presentano una definizione formale di congestione specificando che una tecnologia di produzione può definirsi congestionata se per qualche $y \geq 0$ e $x \in L(y)$ esiste un $x' \geq x$ tale che:

$$x' \in L(y') \text{ per } 0 \leq y' \leq y, \quad x' \notin L(y).$$

Tale definizione coincide con la nozione generalmente accettata secondo cui la congestione degli input si verifica quando l'incremento di uno o più fattori comporta la diminuzione degli output, o quando la diminuzione degli input determina un incremento degli output.

Da quanto riportato in precedenza, ogni misura della congestione deve confrontare il processo produttivo appartenente ad una tecnologia flessibile, in cui vale la disposizione debole degli input, con quello definibile in una tecnologia monotona. A tale scopo, è necessario introdurre una misura di input efficienza debole (*weak efficiency measure*) definita da:

$$W^I(\mathbf{x}, \mathbf{y}) = \begin{cases} \min [\lambda \geq 0 : (\lambda K(\mathbf{x}) \cap L(\mathbf{y})) \neq \emptyset], \\ (\mathbf{x}, \mathbf{y}) \in D(W) \\ +\infty, (\mathbf{x}, \mathbf{y}) \in \text{Complementare } D(W) \end{cases}$$

in cui le quantità $K(\mathbf{x})$ e $D(W)$ sono rappresentate da:

$$K(\mathbf{x}) = \{ \mathbf{x}' \in R_+^s : \mathbf{x}' \leq \mathbf{x} \}$$

$$D(W) = \left\{ (\mathbf{x}, \mathbf{y}) \in R_+^m \times R_+^s : \exists \lambda \geq 0 \text{ tale che } (\lambda K(\mathbf{x}) \cap L(\mathbf{y})) \neq \emptyset \right\}.$$

La figura 4.5 illustra graficamente il meccanismo connesso alla definizione della misura di efficienza debole.

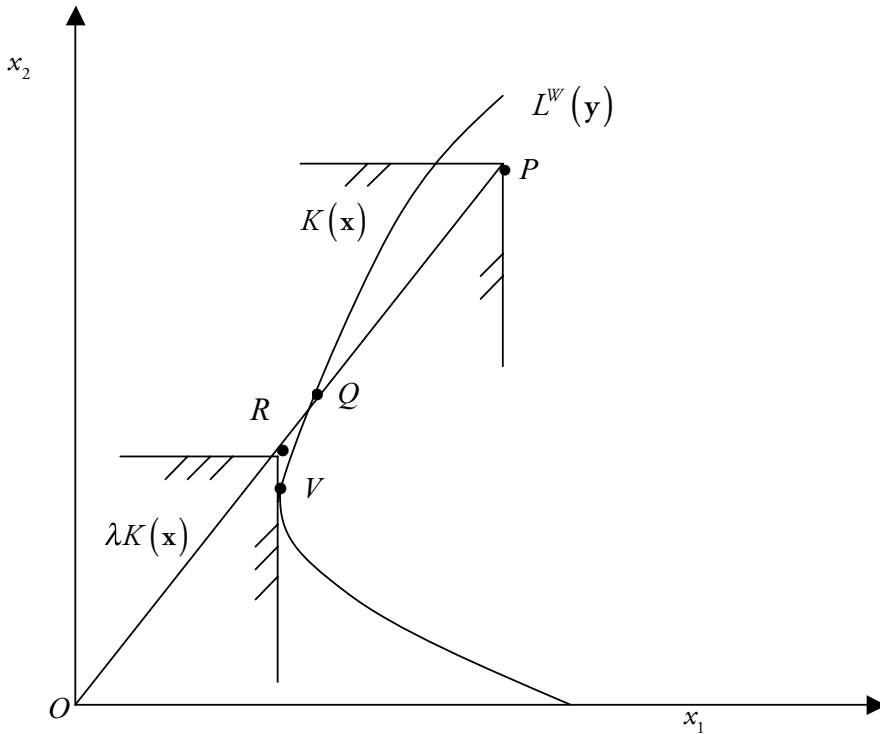


Figura 4.5 – La misura di efficienza debole.

Per il punto P , che rappresenta il vettore $\mathbf{x} \in L(\mathbf{y})$, la misura di efficienza debole è $W^I(\mathbf{x}, \mathbf{y}) = \lambda$ dove $V = \lambda K(\mathbf{x}) \cap L(\mathbf{y})$.

In altre parole, la misura $W^I(\mathbf{x}, \mathbf{y})$ è tale da determinare una spinta verso il basso dell'ortante non positivo $K(\mathbf{x})$ lungo la linea che identifica il vettore \mathbf{x} , per l'ammontare massimo possibile nel rispetto della condizione:

$$(\lambda K(\mathbf{x}) \cap L(\mathbf{y})) \neq \emptyset.$$

In riferimento alla Figura 4.5, la misura di efficienza debole è espressa dal rapporto $W^I(\mathbf{x}, \mathbf{y}) = OR/OP$. Evidentemente l'output \mathbf{y} non può essere ottenuto utilizzando la combinazione di input espressa dal punto R ma, attraverso una riduzione del secondo fattore, può essere ottenuto utilizzando la combinazione V .

Dalla misura di efficienza debole si ottiene una misura di congestione attraverso: $C(\mathbf{x}, \mathbf{y}) = OR/OQ = W^I(\mathbf{x}, \mathbf{y})/F^I(\mathbf{x}, \mathbf{y})$, dove $F^I(\mathbf{x}, \mathbf{y})$ è la misura di efficienza tecnica di Farrell¹² espressa da $F^I(\mathbf{x}, \mathbf{y}) = OQ/OP$.

Essendo la misura di input efficienza tecnica di Farrell sempre maggiore o uguale alla misura di efficienza puramente tecnica, ossia $F^I(\mathbf{x}, \mathbf{y}) \geq W^I(\mathbf{x}, \mathbf{y})$, verificandosi l'uguaglianza solo nel caso in cui la tecnologia presenti ovunque disponibilità forte degli input, deriva che $0 < C(\mathbf{x}, \mathbf{y}) \leq 1$.

Färe e Grosskopf sottolineano, dopo aver dimostrato le proprietà che la $W^I(\mathbf{x}, \mathbf{y})$ soddisfa, che un vettore di input $\mathbf{x} \in L(\mathbf{y})$ può essere tecnicamente efficiente se e solo se $\mathbf{x} \in WEff(\mathbf{y})$, dove il sottoinsieme efficiente debole è espresso da $WEffL(\mathbf{y}) = \left\{ \mathbf{x} \in R_+^s \mid \mathbf{x} \in L(\mathbf{y}) \ W(\mathbf{x}, \mathbf{y}) = 1 \right\}$.

In *Measuring Output efficiency*, Färe e Grosskopf (1983b), analizzano le caratteristiche delle misure di efficienza orientate nel senso degli output, determinate in riferimento ad una tecnologia descritta attraverso l'insieme degli output producibili $P(\mathbf{x})$ e caratterizzata da diverse proprietà, concernenti, in particolare, la proprietà di disponibilità degli output, centrale per l'identificazione dell'eventuale presenza di congestione (*output congestion*).

¹² Färe e Grosskopf (1983a) definiscono la misura di input efficienza di Farrell nel modo seguente:

$$F^I(\mathbf{x}, \mathbf{y}) = \begin{cases} \min \{ \lambda \geq 0 : \lambda \mathbf{x} \in L(\mathbf{y}) \}, & (\mathbf{x}, \mathbf{y}) \in D(F) \\ +\infty, & (\mathbf{x}, \mathbf{y}) \in \text{Complementare } D(F) \end{cases}$$

I concetti precedenti sono estesi ulteriormente in Färe, Grosskopf e Lovell (1985). In generale non si può assumere che gli output soddisfino la proprietà della disponibilità forte, poiché la tecnologia può produrre anche *bad* output, rappresentati, ad esempio, dai rifiuti di produzione, di cui l'unità produttiva non può liberamente disporre in conseguenza degli obblighi relativi alla tutela dell'ambiente.

In tali casi, la misura di efficienza tecnica di Farrell, definita rispetto alla tecnologia caratterizzata da rendimenti costanti di scala e disponibilità forte degli output, espressa dai suddetti Autori¹³ nella forma: $F^o(\mathbf{x}, \mathbf{y}) = \max \{ \phi \geq 0 : \phi \mathbf{y} \in P(\mathbf{x}) \}$, può differire dalla misura di efficienza tecnica debole, ottenuta rilassando la condizione di disponibilità forte degli output, ossia:

$$W^o(\mathbf{x}, \mathbf{y}) = \begin{cases} \max [\phi \geq 0 : (\phi M(\mathbf{y}) \cap P(\mathbf{x})) \neq \emptyset], (\mathbf{x}, \mathbf{y}) \in D(W^o) \\ +\infty, (\mathbf{x}, \mathbf{y}) \in \text{Complementare } D(W^o) \end{cases}$$

dove:

$$M(\mathbf{y}) = \{ \mathbf{y}' : \mathbf{y}' \geq \mathbf{y} \}$$

e

$$D(W^o) = \{ (\mathbf{x}, \mathbf{y}) \in R_+^m \times R_+^s : \exists \phi \geq 0 \text{ tale che } (\phi M(\mathbf{y}) \cap P(\mathbf{x})) \neq \emptyset \}.$$

Le due misure di output efficienza, ossia la misura di Farrell e quella di efficienza debole, possono differire solo sotto l'ipotesi di disposizione debole degli output. In tal caso, una quantità di output ridotta proporzionalmente continua ad essere producibile senza cambiamenti nella combinazione degli input.

Una misura appropriata della perdita di output dovuta a restrizioni sulla libera disponibilità degli output può essere ottenuta attraverso il rapporto:

$$C^o(\mathbf{x}, \mathbf{y}) = F^o(\mathbf{x}, \mathbf{y}) / W^o(\mathbf{x}, \mathbf{y}).$$

L'analisi degli output non desiderabili, o *bad* output, ha rappresentato l'oggetto di numerosi studi, tra i quali, si possono citare Färe, Grosskopf, Lovell e Pasurka (1989), Scheel, H. (2001) e Seiford e Zhu (2002).

L'introduzione della **misura di efficienza di scala**, nel contesto dell'approccio sviluppato da Färe, Grosskopf e Lovell (1983), avviene secondo una prospettiva *input-oriented* attraverso la definizione di una misura di efficienza tecnica debole, in riferimento ad una tecnologia che esibisce rendimenti costanti di scala $W^C(\mathbf{x}, \mathbf{y})$, che viene poi posta a confronto con l'efficienza tecnica debole

¹³ Si osservi che Färe, Grosskopf e Lovell (1985) definiscono la misura di output efficienza come funzione inversa della funzione di output distanza.

della medesima unità produttiva in riferimento ad una tecnologia in cui non vi siano restrizioni relative ai rendimenti di scala.

Si ottiene quindi la misura di efficienza di scala mediante l'espressione:

$$S(\mathbf{x}, \mathbf{y}) = W^C(\mathbf{x}, \mathbf{y}) / W(\mathbf{x}, \mathbf{y}).$$

L'analisi della fonte dell'inefficienza di scala, ossia delle situazioni per cui $S(\mathbf{x}, \mathbf{y}) < 1$, causate dal fatto di svolgere il processo di produzione in regioni di rendimenti crescenti o di rendimenti decrescenti, può essere effettuata confrontando la $W^C(\mathbf{x}, \mathbf{y})$ con un'ulteriore misura di efficienza tecnica debole definita in riferimento ad una tecnologia che presenta rendimenti di scala non crescenti.

I suddetti Autori giungono, quindi, alla seguente scomposizione della misura di efficienza tecnica debole rispetto ad una tecnologia con rendimenti di scala costanti:

$$W^C(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}, \mathbf{y}) \cdot C(\mathbf{x}, \mathbf{y}) \cdot S(\mathbf{x}, \mathbf{y})$$

in cui

$$C(\mathbf{x}, \mathbf{y}) = W(\mathbf{x}, \mathbf{y}) / F(\mathbf{x}, \mathbf{y}).$$

L'espressione precedente descrive la struttura dell'efficienza tecnica, affermando che un'unità produttiva può essere *tecnicamente inefficiente* a causa del verificarsi di tre diverse situazioni. Innanzitutto, il vettore degli input \mathbf{x} potrebbe essere un punto interno di $L(\mathbf{y})$, e quindi $F(\mathbf{x}, \mathbf{y}) < 1$, in secondo luogo \mathbf{x} può congestionare la produzione \mathbf{y} , ossia si può avere $C(\mathbf{x}, \mathbf{y}) < 1$. Infine, il processo produttivo (\mathbf{x}, \mathbf{y}) può non essere efficiente nella scala di produzione attivata e perciò $S(\mathbf{x}, \mathbf{y}) < 1$.

4.3.2 La determinazione delle misure di efficienza debole e della congestione

Per calcolare le diverse misure di efficienza sviluppate utilizzando tecnologie di produzione definite "*piecewise linear technology*", Färe, Grosskopf e Lovell (1985) suggeriscono di ricorrere alle tecniche di programmazione lineare.

Riprendendo il modello di analisi delle attività per la rappresentazione della tecnologia di produzione, i suddetti Autori considerano un insieme di dati composto da n attività o osservazioni¹⁴ riguardanti la trasformazione dei vettori di input:

¹⁴ Si osservi che Färe, Grosskopf e Lovell (1985) utilizzano una notazione in parte diversa da quella adottata nel presente lavoro.

$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{is}) \in R_+^s$
 nei vettori di output:

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im}) \in R_+^m.$$

L'insieme delle attività produttive possibili può essere definito considerando, per la totalità delle unità osservate, la matrice Y degli output di dimensione $n \times m$ degli output, la matrice X ($n \times s$) degli input e il vettore $\mathbf{w} = (w_1, w_2, \dots, w_n) \in R_+^n$ che esprime l'intensità con la quale ciascuna attività può essere attivata.

La tecnologia più restrittiva che soddisfa la disponibilità forte degli input e degli output e la proprietà di rendimenti costanti di scala può essere espressa attraverso la costruzione dell'insieme di produzione Z che involuppa i dati osservati. In termini formali si ha:

$$Z = \left\{ (\mathbf{x}, \mathbf{y}) : \mathbf{y} \leq \mathbf{w}Y, \mathbf{w}X \leq \mathbf{x} \right\}$$

In riferimento all'insieme Z , per quanto esposto nel secondo capitolo, si può considerare una relazione di input corrispondenza:

$$L : R_+^m \rightarrow L(\mathbf{y}) = \left\{ \mathbf{x} \mid \mathbf{x} \in R_+^s \text{ e } (\mathbf{x}, \mathbf{y}) \in Z \right\}$$

per cui l'insieme di fabbisogno di input presenta isoquanti descritti mediante l'espressione:

$$IsoqL(\mathbf{y}) = \left\{ \mathbf{x} \mid \mathbf{x} \in L(\mathbf{y}) \text{ e } \lambda \mathbf{x} \notin L(\mathbf{y}) \quad \forall \lambda \in [0, 1] \right\},$$

e una relazione di output corrispondenza:

$$P : R_+^s \rightarrow P(\mathbf{x}) = \left\{ \mathbf{y} \mid \mathbf{y} \in R_+^m \text{ e } (\mathbf{x}, \mathbf{y}) \in Z \right\}$$

rispetto alla quale l'insieme $P(\mathbf{x})$ è caratterizzato da isoquanti espressi da:

$$IsoqP(\mathbf{x}) = \left\{ \mathbf{y} \mid \mathbf{y} \in P(\mathbf{x}), \lambda \mathbf{y} \notin P(\mathbf{x}), \lambda > 1 \right\}$$

Färe, Grosskopf e Lovell (1987) sottolineano che l'efficienza tecnica del processo produttivo $(\mathbf{x}_i, \mathbf{y}_i) \in Z$, rispetto alla tecnologia rappresentata dall'insieme di produzione Z , può assumere quattro diverse interpretazioni.

Si può considerare l'efficienza tecnica del vettore degli input \mathbf{x}_i nella produzione del vettore degli output \mathbf{y}_i , mediante la quale si esprime la riduzione proporzionale di tutti gli input che consente di produrre il medesimo vettore di output. In tal caso si può ottenere una misura radiale utilizzando l'input corrispondenza attraverso: $F(\mathbf{x}_i, \mathbf{y}_i) = \min \left\{ \theta : \theta \mathbf{x}_i \in L(\mathbf{y}_i) \right\}$ o impiegando l'output corrispondenza mediante $F^o(\mathbf{x}_i, \mathbf{y}_i) = \min \left\{ \theta : \mathbf{y}_i \in P(\theta \mathbf{x}_i) \right\}$.

L'efficienza tecnica del vettore y_i ottenuto dal vettore x_i e, quindi, l'espansione radiale degli output ottenibile utilizzando il medesimo livello di input, può essere misurata utilizzando l'output corrispondenza ottenendo:

$$F^o(x_i, y_i) = \max \{ \phi : \phi y_i \in P(x_i) \}$$

oppure l'input corrispondenza per cui si ha:

$$F(x_i, y_i) = \max \{ \phi : x_i \in L(\phi y_i) \}.$$

Considerando una tecnologia di riferimento meno restrittiva, che considera la versione debole della disponibilità degli input, mantenendo la versione forte della proprietà in relazione agli output e i rendimenti costanti di scala, si ottiene l'insieme di produzione seguente:

$$Z^w = \left\{ (x, y) : y \leq wY, wX = \sigma x, 0 \leq \sigma \leq 1 \right\}$$

Come osservano Färe e Grosskopf (1998), poiché la tecnologia soddisfa rendimenti costanti di scala, il parametro σ , può essere posto uguale a uno per cui l'insieme di produzione, che presenta ancora la disponibilità debole degli input, diviene $Z^w = \left\{ (x, y) : y \leq wY, wX = x \right\}$, rispetto al quale la misura di input efficienza può essere determinata risolvendo il problema di programmazione lineare seguente:

$$W^C(x_i, y_i) = \min \theta_i$$

soggetto a :

$$\theta_i x_{ij} = \sum_{h=1}^n w_h x_{hj} \quad j = 1, \dots, s$$

$$y_{ir} \leq \sum_{h=1}^n w_h y_{hr} \quad r = 1, \dots, m$$

$$w_h \geq 0 \quad h = 1, \dots, n$$

[4.11]

A questo punto è chiaro che, seguendo l'approccio precedente, la determinazione dell'efficienza di ciascuna unità produttiva può essere effettuata in riferimento ad una frontiera costruita ipotizzando specifiche proprietà. Ad esempio, si può considerare una tecnologia caratterizzata da disponibilità forte degli input, disponibilità debole degli output e rendimenti di scala costanti espressa dall'insieme di produzione seguente:

$$Z^o = \left\{ (x, y) : y\mu = wY, wX \leq x, 0 \leq \mu \leq 1 \right\}$$

Rimuovendo il vincolo dei rendimenti di scala costanti, mediante l'introduzione di condizioni sulla somma degli elementi del vettore di intensità w , si possono considerare tecnologie ancora più flessibili che considerano rendimenti di sca-

la variabili e versioni deboli della disponibilità degli input o degli output. Ad esempio, l'insieme di produzione caratterizzato da disponibilità debole degli input, disponibilità forte degli output e rendimenti di scala variabili è espresso da:

$$Z^w = \left\{ (x, y) : y \leq wY, wX = \sigma x, 0 \leq \sigma \leq 1, \sum_{b=1}^n w_b = 1 \right\}$$

Dalle espressioni precedenti per l'insieme di produzione, possono essere derivati specifici problemi di programmazione matematica, la cui soluzione consente la determinazione della misura di efficienza.

Nell'ambito della procedura di scomposizione dell'efficienza tecnica suggerita da Färe, Grosskopf e Lovell (1983) e oggetto di applicazione empirica in Byrnes, Färe e Grosskopf (1984) e Byrnes, Färe, Grosskopf e Lovell (1988), va osservato che i risultati possono essere sensibili all'ordine secondo cui le due componenti sono calcolate (McDonald, 1996).

In entrambe le applicazioni empiriche citate, per la determinazione delle misure di output efficienza viene definita inizialmente la componente di scala, $S(x, y)$, rapportando la misura di efficienza di Farrell, calcolata assumendo una tecnologia con rendimenti costanti di scala e disponibilità forte degli input, che può essere indicata dall'insieme $L(y|C, S)$, alla misura determinata ipotizzando una tecnologia con rendimenti di scala variabili e disponibilità forte degli input, espressa attraverso $L(y|V, S)$.

Successivamente, viene determinata la componente di congestione, $C(x, y)$, attraverso il rapporto tra la misura riferita alla tecnologia con rendimenti variabili di scala e disponibilità forte degli input, $L(y|V, S)$, e la misura di efficienza tecnica debole, calcolata in riferimento alla tecnologia caratterizzata da rendimenti variabili di scala e disponibilità debole degli input, $L(y|W)$.

Secondo tale impostazione, quindi, in riferimento alla tecnologia considerata nel lavoro di Farrell (1957), viene prima rilassata la condizione di rendimenti costanti di scala e successivamente quella di disposizione forte.

McDonald (1996) mette in evidenza la possibilità di seguire un diverso ordine nella scomposizione della misura di efficienza tecnica di Farrell. Si può, infatti, determinare inizialmente una misura di efficienza radiale rispetto ad una tecnologia caratterizzata da rendimenti costanti di scala, ma per la quale si considera solo la condizione debole di disponibilità degli input, indicata da $L(y|C, W)$, e quindi ottenere la componente di congestione attraverso il rapporto con la misura di Farrell, determinata rispetto alla tecnologia

$L(\mathbf{y}|C, S)$. Successivamente, può essere rilassata anche la condizione di rendimenti costanti di scala, al fine di determinare la componente di scala come rapporto tra la misura calcolata rispetto alla tecnologia con rendimenti costanti di scala e disponibilità debole, $L(\mathbf{y}|C, W)$ e la misura di efficienza rispetto alla tecnologia con rendimenti variabili e disponibilità debole degli input, rappresentata da $L(\mathbf{y}|V, W)$.

Riconoscendo la diversità di risultati a cui si può giungere scegliendo una strada piuttosto che un'altra, Färe e Grosskopf (2000) suggeriscono di utilizzare la seconda scomposizione, determinando dapprima la componente di congestione, se la congestione degli input sembra essere un fenomeno rilevante nei processi sotto osservazione¹⁵.

Färe e Grosskopf (2000) evidenziano l'esistenza di specifiche relazioni tra le diverse tecnologie coinvolte nella scomposizione. Seguendo il primo procedimento si ha:

$$L(\mathbf{y}|C, S) \supseteq L(\mathbf{y}|V, S) \supseteq L(\mathbf{y}|V, W)$$

Il secondo procedimento conduce a:

$$L(\mathbf{y}|C, S) \supseteq L(\mathbf{y}|C, W) \supseteq L(\mathbf{y}|V, W)$$

Si osservi che gli insiemi coinvolti sono solo parzialmente ordinati.

La prima scomposizione, definendo la componente di scala come differenza tra $L(\mathbf{y}|C, S)$ e $L(\mathbf{y}|V, S)$, potrebbe non rilevare la presenza di congestione a causa dell'ipotesi di disponibilità forte degli input.

Appare interessante osservare che in Färe, Grosskopf e Lovell (1983) la tecnologia di riferimento iniziale è caratterizzata dalla presenza di rendimenti variabili di scala e disponibilità debole rispetto alla quale viene determinata la misura di input efficienza tecnica debole, $W(\mathbf{x}, \mathbf{y})$. Imponendo le restrizioni relative alla disponibilità degli input, ossia considerando la tecnologia che soddisfa la *free disposal*, viene determinata la misura che considera ancora rendimenti di scala variabili, $F(\mathbf{x}, \mathbf{y})$ e quindi la componente di congestione. Considerando, invece, la tecnologia caratterizzata da rendimenti di scala costanti, mantenendo la versione debole della proprietà di disponibilità degli input,

¹⁵ Va osservato che risulta possibile identificare anche lo specifico fattore che determina la congestione attraverso la soluzione di specifici problemi di programmazione (Färe, Grosskopf e Lovell, 1994)

viene determinata la misura indicata con $W^C(\mathbf{x}, \mathbf{y})$, attraverso la quale può essere derivata la componente di scala.

In definitiva, mentre nei lavori successivi l'impostazione suggerita è quella di procedere verso la considerazione di una tecnologia di riferimento sempre più flessibile fino a giungere alla tecnologia $L(\mathbf{y}|V, W)$ con rendimenti variabili e disponibilità debole (Byrnes, Färe e Grosskopf, 1984, Färe, Grosskopf e Lovell, 1994) in "The Structure of Technical Efficiency" Färe, Grosskopf e Lovell (1983) procedono in modo inverso, partendo dalla tecnologia con meno restrizioni e alternando l'inserimento delle proprietà sui rendimenti di scala e sulla disponibilità degli input.

Nel contesto dell'analisi e della misurazione del fenomeno della congestione dei fattori e dei prodotti, si ritiene doveroso osservare che, in contrapposizione all'approccio richiamato in precedenza, sviluppato grazie agli studi di Färe, Grosskopf e Lovell, esiste in letteratura un diverso approccio basato sull'analisi degli *slack* introdotto da Brockett, Cooper, Wang e Shin (1998), Cooper, Seiford e Zhu (2000, 2001) e Cooper, Deng, Huang e Li (2002).

Senza entrare nello specifico dei diversi argomenti che hanno caratterizzato l'accesso dibattito intercorso tra i suddetti Autori negli ultimi anni¹⁶, appare interessante evidenziare le linee distintive dei due approcci, utilizzabili per l'identificazione e la misurazione della congestione.

Sinteticamente si può affermare che mentre Färe, Grosskopf e Lovell, come già ampiamente esposto, utilizzano una misura radiale di Farrell per misurare la congestione come differenza nella tecnologia sotto l'ipotesi di disposizione forte e debole, Brockett, Cooper, Wang e Shin procedono secondo un approccio in due fasi basato sull'identificazione degli *slack*.

La prima fase consiste nella soluzione di modello che identifica la misura di efficienza e l'ammontare degli *slack* negli input e negli output. Nella seconda fase viene massimizzata la somma degli eccessi negli input definiti rispetto al punto sulla frontiera efficiente. La differenza tra gli *slack*, per ciascun input, ottenuti attraverso le due fasi precedenti, rappresenta una misura di congestione e consente l'identificazione della fonte, ossia del fattore responsabile del fenomeno oggetto di analisi. L'ammontare totale che esprime l'eccesso di input, calcolato nella prima fase, è scomposto quindi in due componenti che esprimono l'efficienza tecnica e la congestione.

¹⁶ A tale proposito si può citare, ad esempio, la successione di interventi: Färe e Grosskopf (2000), Cooper, Seiford e Zhu (2001), Färe e Grosskopf (2001).

4.4 L'ANALISI DELL'EFFICIENZA RISPETTO A TECNOLOGIE NON CONVESSE: IL *FREE DISPOSAL HULL*

Deprins, Simar e Tulkens (1984) propongono un metodo non parametrico per la misurazione dell'efficienza di un insieme di unità produttive basato sulla costruzione, a partire dai dati osservati, di un insieme di produzione che descrive una tecnologia non convessa, in cui sono mantenute le ipotesi di libera disponibilità degli input e degli output.

Il fatto di non assumere l'ipotesi di convessità comporta che nella determinazione della misura dell'efficienza tecnica vengano coinvolti nel confronto, in quanto processi produttivi efficienti che costituiscono la frontiera delle possibilità produttive, esclusivamente processi produttivi realmente osservati e non processi "ipotetici" rappresentati da combinazioni convesse di processi esistenti. L'insieme di produzione è ottenuto dall'unione di tutti gli ortanti, positivi negli output y , e negativi negli input x , la cui origine coincide con un punto osservato. In termini formali, riprendendo la notazione degli Autori, si ha:

$$Z = \left\{ \begin{array}{l} (\mathbf{x}, \mathbf{y}) \in R_+^{s+m} \left| (\mathbf{x}, \mathbf{y}) = (\mathbf{x}_i, \mathbf{y}_i) + \sum_{r=1}^s \mu_r (\mathbf{e}_r^s, \mathbf{0}^m) + \sum_{j=1}^m \nu_j (\mathbf{0}^s, \mathbf{e}_j^m) \right. \\ \left. (\mathbf{x}_i, \mathbf{y}_i) \in Z, \quad \mu_r \geq 0, \quad \nu_j \geq 0, \quad r = 1, \dots, s, \quad j = 1, \dots, m \right. \end{array} \right\}$$

dove $\mathbf{0}^s$ e $\mathbf{0}^m$ sono vettori nulli di dimensione s e m , per cui $(\mathbf{0}^s, \mathbf{0}^m)$ rappresenta l'origine dello spazio input-output, mentre \mathbf{e}_r^s denota un vettore di dimensione s , i cui elementi sono tutti nulli ad eccezione della r -esima componente uguale a 1, e \mathbf{e}_j^m rappresenta il vettore di dimensione m , i cui elementi, ad eccezione della componente j -esima, che assume valore 1, sono uguali a zero. Le variabili μ_r e ν_j esprimono la proprietà di *free disposal*, rispettivamente degli input e degli output, da cui deriva la denominazione attribuita alla metodologia.

La Figura 4.6 illustra il *Free Disposal Hull (FDH)* ottenuto da una data combinazione input-output, per il caso semplificato di processi che impiegano un solo input per produrre un solo output.

I punti rappresentati con il simbolo A_i ($i = 1, \dots, 5$) rappresentano processi produttivi realmente osservati. A partire dai processi osservati, in virtù della proprietà di *free disposal*, qualsiasi altro processo che produce quel medesimo livello di output utilizzando una quantità non inferiore di input è un processo possibile.

L'area tratteggiata alla destra della spezzata $x_1 A_1 B A_3 C A_4$ rappresenta, quindi, l'insieme delle possibilità produttive derivato dal *FDH* dei dati osservati, la cui frontiera è rappresentata da una funzione a gradini.

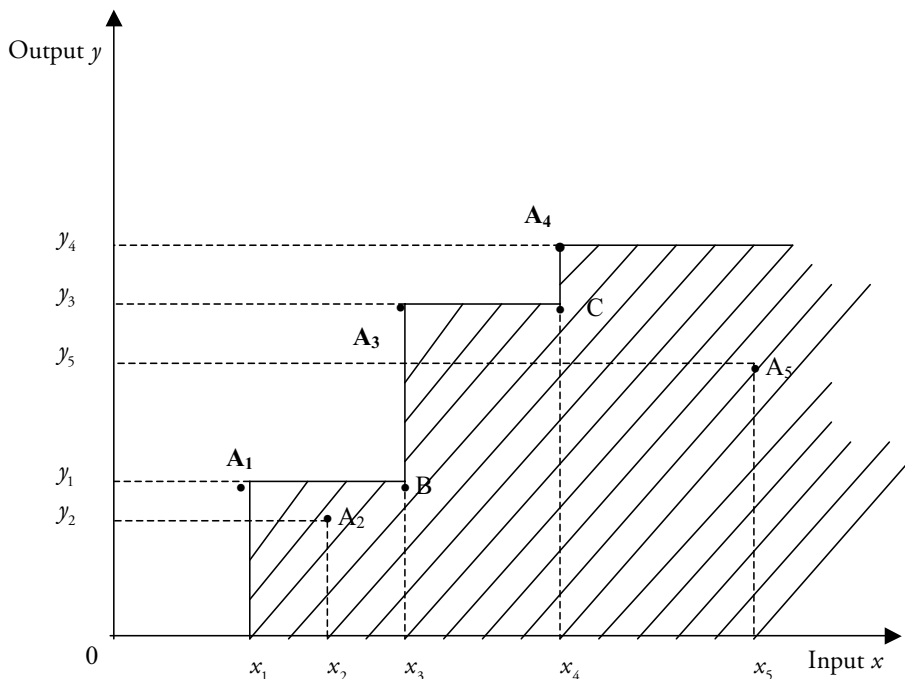


Figura 4.6 – Free Disposal Hull.

Deprins, Simar e Tulkens suggeriscono di utilizzare la misura di input efficienza calcolata in riferimento all'insieme di fabbisogno di input che, nel caso semplificato in cui il processo produttivo è rappresentato da un input e un vettore di output, si riduce a:

$$L(y) = \left\{ x \in R_+ \left| \begin{array}{l} (x, y) = (x_i, y_i) + \mu(1, 0^m) + \sum_{j=1}^m v_j(0, e_j^m), \\ (x_i, y_i) \in Z, \quad \mu \geq 0, \quad v_j \geq 0, \quad j = 1, \dots, m \end{array} \right. \right\}$$

La misura di input efficienza può essere ottenuta attraverso la seguente procedura proposta dai suddetti Autori. Per ogni punto (x_i, y_i) è necessario costruire un insieme $D(x_i, y_i)$ che include il punto stesso e tutti gli altri punti osservati (x, y) che dominano (x_i, y_i) , nel senso che la quantità di input

utilizzata è strettamente inferiore a x_i e il vettore di output prodotto y è debolmente più grande di y_i . Formalmente si ottiene l'insieme:

$$D(x_i, y_i) = \{(x_i, y_i)\} \cup \left\{ (x, y) \in Z \mid (x, y) = (x_i, y_i) + \mu(1, \mathbf{0}^m) + \sum_{j=1}^m v_j(0, \mathbf{e}_j^m), \right. \\ \left. \mu \geq 0, \quad v_j \geq 0, \quad j = 1, \dots, m \right\}$$

All'interno dell'insieme $D(x_i, y_i)$ si seleziona poi il vettore $(x_{*(i)}, y_{*(i)})$ per il quale la componente x è la più piccola e si calcola il rapporto $x_{*(i)}/x_i$. Dalla definizione di $D(x_i, y_i)$, tale rapporto assumerà valori uguali o inferiori a uno, secondo che $(x_{*(i)}, y_{*(i)})$ coincida con il punto (x_i, y_i) considerato, in tal caso $x_{*(i)}/x_i = 1$, oppure sia un punto diverso da (x_i, y_i) e, quindi, rappresenti un processo dominante, per cui $x_{*(i)}/x_i < 1$. La misura di input efficienza può essere espressa dal rapporto: $F(x_i, y_i) = x_{*(i)}/x_i$.

I suddetti Autori sottolineano che il processo $(x_{*(i)}, y_{*(i)})$, rispetto al quale viene determinata l'efficienza del vettore in esame (x_i, y_i) , è un processo realmente osservato che può quindi rappresentare un esempio reale di comportamento ottimale, nel senso che una quantità di output maggiore o uguale a quella prodotta può essere ottenuta utilizzando una quantità inferiore di input.

Si osservi che per la determinazione della misura di input efficienza si può anche risolvere il seguente sistema di programmazione lineare:

min θ

$$\text{soggetto a} \quad \sum_{b=1}^n \lambda_b x_b \leq \theta x_{ri} \quad r = 1, \dots, s$$

$$\sum_{b=1}^n y_{rb} \lambda_b \geq y_{ji} \quad j = 1, \dots, m$$

$$\sum_{b=1}^n \lambda_b = 1 \quad \lambda_b \in \{0, 1\} \quad b = 1, \dots, i, \dots, n$$

Le condizioni precedenti comportano che in qualsiasi soluzione, anche nella soluzione ottima, solamente uno dei λ_b sarà uguale all'unità mentre tutti gli altri saranno uguale a zero. Si avranno al massimo n soluzioni, delle quali, però, non tutte rappresentano processi possibili. Per individuare le soluzioni possibili

è necessario applicare una procedura che si basa sul confronto tra il livello di output prodotto dall'unità da valutare e gli output delle altre unità (Ray, 2004).

Occorre osservare che il *Free Disposal Hull* è stato concepito nell'ambito delle valutazioni di *performance* della pubblica amministrazione, per cui esso ha trovato applicazione in particolare nella valutazione di organizzazioni operanti nel settore pubblico. Solo per fare qualche esempio, si può ricordare il lavoro di Ferrari e Maltagliati (1997) per gli uffici postali nella provincia di Firenze e lo studio di Bonsi, Manchi e Pavone (1999) per le unità periferiche della Pubblica Amministrazione.

Destefanis e Pavone (1996, 1998) utilizzano l'approccio non parametrico FDH per la misurazione della *performance* degli ospedali di zona e per la valutazione dell'efficienza tecnica delle amministrazioni comunali italiane.

Viviani (1996) estende il metodo FDH applicandolo a dati *panel* allo scopo di analizzare la *performance* produttiva dei punti vendita costituenti la rete di distribuzione commerciale dell'Unicoop toscana. L'Autore suggerisce inoltre di stimare un unico modello capace di fornire sia una misura di input efficienza che un indicatore di output efficienza.

Gazzei e Viviani (1998) si avvalgono del metodo FDH per individuare i piani di produzione, attivati dai centri di responsabilità economica della pubblica amministrazione, pienamente efficienti sulla base dei quali viene stimata una funzione di costo.

4.5 RECENTI SVILUPPI DELLA METODOLOGIA NON PARAMETRICA

4.5.1 Considerazioni preliminari

Dalla sua introduzione in letteratura, come si è già avuto modo di sottolineare, la DEA si è progressivamente sviluppata ed ampliata, divenendo un ricchissimo campo di ricerca nel contesto dell'analisi della produttività ed efficienza delle organizzazioni produttive. In una recente bibliografia sulla DEA, Taveres (2002) include 3.183 contributi provenienti da 2.152 diversi autori. Le applicazioni empiriche coinvolgono i settori più disparati, dall'agricoltura all'assistenza sanitaria.

È impensabile fornire una rassegna esaustiva dei recenti sviluppi. Appare opportuno però presentare una breve introduzione ad alcune problematiche che costituiscono interessanti filoni di ricerca.

In particolare, si farà cenno ai tentativi di pervenire alla costruzione di metodologie non parametriche che possano superare le limitazioni principali di

tale approccio, rappresentate dall'eccessiva dipendenza dai valori osservati e dall'assenza di procedure statistiche di valutazione delle stime ottenute.

Si analizzeranno, brevemente, le modalità attraverso le quali l'approccio non parametrico consente l'introduzione delle variabili esterne nello studio dell'efficienza.

Infine, l'attenzione sarà rivolta al trattamento di dati gerarchici nel contesto nella valutazione della *performance* attraverso la DEA.

4.5.2 Approcci stocastici alla DEA

La DEA e le altre metodologie non parametriche, utilizzate per la valutazione dell'efficienza rispetto alla frontiera delle possibilità produttive, si sono sviluppate all'interno del cosiddetto contesto deterministico, caratterizzato dall'assenza di ipotesi relative a errori di misurazione e disturbi casuali, relativamente alle unità produttive sulla cui base viene "stimata" la frontiera dell'insieme di produzione.

A differenza di quanto accade per i modelli che descrivono le frontiere stocastiche, nel caso dei modelli di programmazione matematica le proprietà statistiche degli stimatori non sono pienamente sviluppate.

Esistono diversi approcci per lo sviluppo di una metodologia DEA di tipo stocastico e, quindi, per la traduzione del problema *economico* di determinazione della misura di efficienza tecnica in ambito non parametrico, in un problema *statistico*.

Si può considerare, ad esempio, l'interpretazione di Banker (1993) che vede la DEA come una procedura di stima di massima verosimiglianza per una frontiera deterministica, e l'approccio basato sui modelli "*chance-constrained programming*" di Charnes e Cooper (1963) sviluppati da Land, Lovell e Thore (1993).

Per una rassegna ci si può riferire, tra gli altri, a Grosskopf (1996), Simar (1996) e Simar e Wilson (2000a).

Nel rinviare a Simar (2003) per i dovuti approfondimenti, si può sottolineare in questa sede che il modello statistico più semplice ipotizza l'esistenza, in relazione al processo generatore dei dati (*Data-Generating Process, DGP*), di una funzione di densità $f(x_i, y_i)$ che genera le osservazioni (x_i, y_i) indipendenti e identicamente distribuite, per $i=1, \dots, n$, che costituiscono l'insieme di produzione possibile, Ψ .

Diverse proprietà possono caratterizzare l'insieme Ψ , tra cui la *free disposal* e la convessità, specificate nei modi usuali.

Poiché il processo generatore dei dati deve definire come il campione è generato, il caso più semplice consiste nel considerare campioni casuali di ampiezza n , per cui le osservazioni campionarie in χ_n , che rappresenta l'insieme delle

unità produttive considerate, possono essere considerate realizzazioni di variabili casuali *i.i.d* su Ψ con funzione di densità $f(x, y)$. Il modo più semplice di descrivere la $f(x, y)$, tra i diversi possibili utilizzando le coordinate cilindriche, $(x, y) \Leftrightarrow (\omega, \eta, y)$, dove (ω, η) rappresentano le coordinate polari di x , definite dal modulo $\omega = \omega(x) \in R_+^1$ e dall'angolo $\eta = \eta(x) \in \left[0, \frac{\pi}{2}\right]^{s-1}$, consiste nello

scomporre la funzione di densità congiunta di (ω, η, y) secondo la seguente espressione: $f(\omega, \eta, y) = f(\omega|\eta, y) f(\eta, y) f(y)$, nella quale sono facilmente definite tutte le densità condizionate.

Per un dato (η, y) , il punto sulla frontiera $x^\partial(y)$, rappresentante il livello efficiente di input, in un contesto orientato nel senso degli input, espresso attraverso la misura radiale di Farrell da $x^\partial(y) = \theta(x, y)x$, ha modulo:

$$\omega(x^\partial(y)) = \inf \left\{ \omega \in R_+^1 \mid f(\omega|\eta, y) > 0 \right\}.$$

Da ciò deriva che la misura di Farrell si traduce in una semplice funzione del modulo corrispondente:

$$0 \leq \theta(x, y) = \frac{\omega(x^\partial(y))}{\omega(x)} \leq 1$$

Per provare la consistenza è necessario assumere che le unità efficienti possano essere osservate con probabilità tendente a uno al crescere della numerosità campionaria. Il tasso di convergenza dipende dalla "regolarità" della frontiera, che può essere espressa, ad esempio, utilizzando la misura di Farrell $\theta(x, y)$ ed assumendo che la stessa sia differenziabile in entrambi i suoi argomenti.

Assumendo la convessità dell'insieme di produzione possibile, Ψ , si può esplicitare lo stimatore DEA della misura di efficienza di un determinato processo produttivo (x_0, y_0) , mediante:

$$\theta_{DEA}^*(x_0, y_0) = \inf \left\{ \theta \mid (\theta x_0, y_0) \in \Psi_{DEA}^*(\mathcal{X}_n) \right\}$$

dove l'insieme di produzione possibile stimato a partire dai dati osservati, caratterizzato da rendimenti costanti di scala, assume la forma:

$$\Psi_{DEA}^*(\mathcal{X}_n) = \left\{ (x, y) \in R_+^{s+m} \mid y \leq \sum_{b=1}^n \lambda_b y_b, \quad x \geq \sum_{b=1}^n \lambda_b x_b, \right.$$

$$\left. \sum_{b=1}^n \lambda_b = 1 \quad \lambda_b \geq 0 \quad \forall b = 1, \dots, n \right\}$$

Se la definizione di un modello statistico per l'individuazione degli usuali stimatori DEA e FDH della misura di efficienza tecnica non presenta particolari difficoltà, ben più articolata appare la problematica inerente la definizione delle **proprietà statistiche** che caratterizzano gli *stimatori* ottenuti.

I primi risultati sulla *consistenza* degli stimatori DEA e FDH sono stati ottenuti da Banker (1993) solamente per il caso univariato, ossia nel caso in cui ci sia un unico input, in un problema orientato nel senso degli input, oppure un unico output, nel caso in cui l'orientamento è relativo all'espansione degli output.

Il suddetto Autore dimostra la consistenza debole dello stimatore di input efficienza DEA, nel caso in cui $s = 1$ e $m \geq 1$, ossia $\hat{\theta}_{DEA}(x, y) \xrightarrow{p} \theta(x, y)$, non fornendo però indicazioni relative al tasso di convergenza.

Per quanto riguarda la distribuzione campionaria degli stimatori, risultati analitici sono stati ottenuti solo nel caso univariato per lo stimatore DEA, mentre per lo stimatore FDH sono stati ottenuti risultati asintotici generali.

Gijbels, Mammen, Park e Simar (1999) ottengono il seguente risultato asintotico per lo stimatore DEA nel caso semplificato di un solo input e un solo output:

$$\theta^*_{DEA}(x, y) - \theta(x, y) \underset{d}{\sim} F(\cdot, \cdot)$$

dove F è una funzione di distribuzione regolare nota a meno di costanti che dipendono dal processo generatore dei dati, legate alla forma della frontiera e al valore di $f(x, y)$ in corrispondenza di un punto vero sulla frontiera.

Park, Simar, e Weiner (2000) derivarono il seguente risultato, valido per il caso generale di processi produttivi che impiegato s input ($s \geq 1$), per produrre m output ($m \geq 1$), in riferimento alla distribuzione:

$$\theta^*_{FDH}(\mathbf{x}, \mathbf{y}) - \theta(\mathbf{x}, \mathbf{y}) \underset{d}{\sim} Weibull(\cdot, \cdot)$$

dove la distribuzione asintotica Weibull dipende da alcuni parametri non noti, legati al *DGP*, alla pendenza della frontiera e al valore di $f(x, y)$ vicino la frontiera.

Va osservato che i risultati precedenti devono essere impiegati con cautela nelle applicazioni empiriche per due ordini di ragioni. In primo luogo, l'utilizzazione di approssimazioni della distribuzione degli stimatori è ragionevole solo per grandi campioni. In secondo luogo, poiché le distribuzioni asintotiche dipendono da quantità non note, che devono essere stimate, si possono introdurre ulteriori fattori di disturbo nell'approssimazione delle distribuzioni campionarie.

Nel caso generale multi input e multi output la metodologia *bootstrap* (Efron e Tibshirani, 1993) si è dimostrata estremamente utile per la derivazione delle distribuzioni campionarie degli stimatori DEA.

Simar e Wilson (1998) proposero un algoritmo per implementare il metodo *bootstrap* nel contesto della stima delle frontiere di produzione, in grado di fornire un'approssimazione della distribuzione campionaria di $\hat{\theta}_{DEA}(\mathbf{x}, \mathbf{y}) - \theta(\mathbf{x}, \mathbf{y})$ nel caso multivariato ($s, m \geq 1$). In particolare, allo scopo di analizzare la sensibilità delle misure di efficienza ottenute alle variazioni campionarie, gli Autori suggeriscono l'impiego di intervalli di confidenza e di correzioni per la distorsione connessa alla procedura DEA, assumendo però ipotesi restrittive sull'omogeneità della distribuzione dell'efficienza tra le unità produttive. Il metodo è stato applicato a dati *panel* per la stima degli indici di Malmquist da Simar e Wilson (1999).

Successivamente, Simar e Wilson (2000b) estendono la metodologia proposta in precedenza, considerando strutture produttive più flessibili, caratterizzate dalla presenza di eterogeneità nella distribuzione della misura di efficienza.

Kneip, Simar e Wilson (2003) analizzano la distribuzione asintotica dello stimatore DEA della misura di efficienza considerando una tecnologia con rendimenti di scala variabili. Gli Autori propongono, inoltre, due procedure *bootstrap* necessarie per qualunque applicazione pratica. Infatti, poiché la distribuzione asintotica dello stimatore contiene parametri non noti, risulta complicato procedere nella stima e nella simulazione contemporaneamente.

Le procedure *bootstrap* suggerite oltre ad essere direttamente implementabili, godono di buone proprietà come dimostrato da Kneip, Simar e Wilson, attraverso esperimenti Monte Carlo.

Recentemente, un nuovo approccio robusto per la stima della frontiera di produzione non parametrica, denominato **frontiere di ordine m** , molto interessante, è stato introdotto da Cazals, Florens e Simar (2002).

Questo metodo innovativo non richiede ipotesi sulla convessità della tecnologia e presenta diverse proprietà desiderabili che lo rendono particolarmente utile ai fini inferenziali. Gli stimatori di ordine m sono completamente non parametrici, ma, a differenza degli stimatori DEA e FDH, sono consistenti (il tasso di convergenza è uguale a \sqrt{n}) e non soffrono del problema del "*curse of dimensionality*", derivante dal fatto che se un dato insieme di n osservazioni viene proiettato in un numero crescente di direzioni ortogonali, la distanza euclidea tra le osservazioni deve necessariamente incrementarsi.

In altre parole, nel caso degli stimatori FDH e DEA relativi alla stima dell'efficienza di processi caratterizzati da un solo input e $m \geq 1$ output, se il numero degli output si incrementa, il numero delle osservazioni deve crescere ad un tasso esponenziale per mantenere un dato *mean square error*.

Diversamente dagli stimatori DEA e FDH, gli stimatori delle frontiere di ordine m , sono, inoltre, estremamente robusti rispetto ai valori estremi.

Il concetto principale su cui si basa il metodo introdotto da Cazals, Florens e Simar è quello di “*expected minimum input function*” (o *expected maximal output function*) che risulta strettamente legato, come i suddetti Autori dimostrano, al concetto di frontiera efficiente. Lo stimatore che ne risulta è collegato allo stimatore FDH, ma a differenza di quest'ultimo non copre tutti i dati osservati ma solo una parte degli stessi. L'ordine m dello stimatore specifica le unità produttive potenzialmente efficienti che rappresentano unità di riferimento rispetto alle quali valutare la *performance* di altre unità di produzione.

Simar (2003) suggerisce l'uso degli stimatori di ordine m per la rilevazione dei dati anomali nei casi in cui si utilizzano altri stimatori non parametrici o parametrici della frontiera di produzione efficiente.

Le frontiere di ordine m sono state applicate da Wheelock e Wilson (2003) per l'analisi della *performance* delle banche commerciali statunitensi nel periodo 1984-2002. Daraio e Simar (2005) sviluppando ulteriormente l'utilizzazione del concetto di stimatore di ordine m introducono uno stimatore della frontiera di produzione di tipo DEA (ipotizzando, quindi, una tecnologia convessa), condizionato all'agire di alcune variabili ambientali, che pur non essendo né input né output del processo esercitano una notevole influenza sullo svolgimento del processo stesso.

4.5.3 L'introduzione di variabili esterne nell'analisi dell'efficienza

L'introduzione delle variabili esterne o ambientali nella determinazione dell'efficienza delle unità produttive è spinta dalla considerazione della necessità di effettuare l'analisi di efficienza in relazione a un gruppo di unità che possano definirsi omogenee.

Banker e Morey (1986a, 1986b), rilevando l'opportunità di confrontare unità produttive appartenenti al medesimo gruppo, ossia unità che condividono le medesime variabili ambientali, suggeriscono la distinzione tra fattori controllabili e non controllabili utilizzati nello svolgimento del processo produttivo.

Nella misurazione dell'efficienza di n unità produttive, che producono il vettore di output $y = (y_1, y_2, \dots, y_m) \in R_+^m$, si può suddividere il vettore degli input utilizzati nel processo in due parti, di cui la prima è costituita da s' fattori controllabili da parte del produttore, mentre la seconda è rappresentata dagli $(s - s')$ input non controllabili (*non discretionary input*).

Nel loro primo lavoro, Banker e Morey propongono la seguente formulazione per la determinazione dell'input efficienza della i -esima unità, il cui processo produttivo coinvolge esclusivamente variabili quantitative:

$$\min \theta_i - \varepsilon \left(\sum_{j=1}^{s'} s_j^- + \sum_{r=1}^m s_r^+ \right)$$

sogetto a :

$$\sum_{b=1}^n \lambda_b x_{bj} + s_j^- = \theta_i x_{jb} \quad (j = 1, \dots, s')$$

$$\sum_{b=1}^n \lambda_b x_{bj} + s_j^- = x_{bj} \quad (j = s' + 1, \dots, s)$$

$$\sum_{b=1}^n \lambda_b y_{br} - s_r^+ = y_{br} \quad (r = 1, \dots, m) \quad [4.12]$$

$$\sum_{b=1}^n \lambda_b = 1$$

$$\lambda_b \geq 0, \quad s_j^- \geq 0, \quad s_r^+ \geq 0$$

Dalla soluzione del problema di programmazione lineare precedente si ottiene la misura di efficienza radiale di Farrell, che esprime di quanto le risorse controllabili potrebbero essere simultaneamente ridotte in modo da raggiungere la piena efficienza tecnica.

L'inclusione di variabili qualitative tra gli input non controllabili può determinare alcuni problemi metodologici, in quanto l'ipotesi di convessità perde di si-

gnificato per le mutabili statistiche, dato che la combinazione convessa $\sum_{b=1}^n \lambda_b x_{bj}$

$\sum_{b=1}^n \lambda_b = 1$ con $\lambda_b \geq 0$ può non appartenere alla scala di misurazione della variabile.

Nel lavoro successivo, Banker e Morey modificano, quindi, la tecnologia cui fanno riferimento Banker, Charnes e Cooper (1984), rimuovendo la condizione di convessità per le variabili categoriali. L'insieme di produzione assume la forma seguente:

$$Z = \left\{ \begin{array}{l} (x, y): y_r \leq \sum_{b=1}^n \lambda_b y_{rb} \quad (r = 1, \dots, m); \\ x_j \geq \sum_{b=1}^n \lambda_b x_{jb} \quad (j = 1, \dots, s-1) \quad x_s \geq x_{sl} \quad \forall l \text{ tale che } \lambda_l > 0 \\ \lambda_l \geq 0; \sum_{b=1}^n \lambda_b = 1 \end{array} \right\}$$

Per rendere operativa l'espressione precedente, i suddetti Autori osservano che nel caso di variabili categoriali non controllabili, nel problema [4.12] il vincolo lineare:

$$\sum_{b=1}^n \lambda_b x_{bj} + s_j^- = x_{ij}$$

può essere sostituito dai vincoli della forma: $x_{sl} \leq x_{si}$ se $\lambda_i > 0$. In altre parole, per le DMU coinvolte nella combinazione convessa, per le quali il valore di λ_i è strettamente positivo, il valore della variabile qualitativa ordinata deve essere uguale o inferiore a quello corrispondente per l'unità i -esima.

Allo scopo di tradurre tutto ciò in un problema di programmazione lineare, Banker e Morey definiscono k nuove variabili binarie, $d_{s,b}^{(\alpha)}$, se $k+1$ sono le modalità della variabile qualitativa da considerare, che sostituiranno la variabile

$$x_{bj} \text{ nei vincoli } \sum_{b=1}^n \lambda_b x_{bj} \leq x_{ij}.$$

In tal modo il vincolo, che può assumere, ad esempio, la seguente forma:

$$\sum_{b=1}^n \lambda_b d_{s,i}^{(k)} \leq 0, \quad (k=1,2,3)$$

se l'unità i -esima appartiene alla più bassa modalità, consente di considerare come insieme di riferimento, per la valutazione dell'efficienza dell'unità i -esima, solo l'insieme delle DMU appartenenti alla medesima categoria.

Va osservato che nell'approccio di Banker e Morey, come sottolineato da Fried, Schmidt e Yaisawarng (1999), le variabili non controllabili, incluse come input e/o output del processo, influenzano esclusivamente la posizione della frontiera, mentre sono tenute costanti nel calcolo della misura di efficienza radiale, e ciò preclude la possibilità di verificare l'influenza che le stesse possono esercitare nello svolgimento del processo.

In contrapposizione all'approccio precedente, che in un unico stadio considera gli input del processo insieme ai fattori esogeni, espressi mediante variabili qualitative o quantitative, e nel quale, necessariamente, deve essere specificata la direzione dell'impatto che le medesime variabili esercitano sulla *performance* dell'unità produttiva, si è sviluppato nella letteratura non parametrica, in linea con quanto già evidenziato nell'ambito delle metodologie parametriche, un **approccio a due stadi**.

Nella formulazione tipica, ad una prima fase in cui si stima l'efficienza attraverso la DEA su un insieme di input e output, attraverso i quali si descrive il processo produttivo, segue un'analisi di regressione, che cerca di individuare le

determinanti dei livelli di efficienza, ottenuti dal primo stadio, in relazione ad un vettore di variabili esogene osservabili.

Ray (1991), ad esempio, combina l'analisi DEA con un modello regressivo per stimare l'efficienza relativa dei distretti scolastici del Connecticut. L'approccio in due stadi, suggerito dal suddetto Autore, prevede l'inclusione degli input e degli output tradizionali, ossia controllabili, nella formulazione del problema di programmazione lineare utilizzato per calcolare la misura radiale di efficienza tecnica, che costituirà la variabile dipendente in un secondo stadio, in cui attraverso un modello di regressione multipla, sono introdotte le variabili esogene, rappresentate da caratteristiche socioeconomiche dei territori nei quali si svolge il processo di produzione. I legami tra i *non discretionary input* nel processo di produzione e le misure di efficienza ottenute mediante l'applicazione della metodologia DEA, precedentemente discussi in Ray (1988) per il caso di un solo output, vengono generalizzati al caso multi output. Il modello di regressione lineare, che, come sottolinea l'Autore, dovrebbe includere un termine di errore unilaterale, viene stimato mediante i minimi quadrati modificati.

Sebbene esistano esempi in letteratura di applicazioni che prevedono la stima delle relazioni tra le variabili esogene e le misure di efficienza utilizzando gli OLS, ad esempio lo studio di Nyman e Bricker (1989), poiché la misura di efficienza assume valori tra zero e uno, i modelli più frequentemente utilizzati nella seconda fase dell'analisi sono rappresentati da modelli tobit, come in Kirjavainen e Loikkanen (1998), dove per eliminare la presenza di eventuali valori anomali viene utilizzata la procedura *jackknife* sulle misure di efficienza DEA, e da modelli probit, come, ad esempio, in Kooreman (1994).

Esistono poi molteplici contributi in letteratura la cui caratteristica principale è quella di introdurre strumenti metodologici specifici che vanno, in parte, a modificare l'approccio in due fasi.

In particolare, McCarty e Yaisawarng (1993) suggeriscono una procedura in due fasi per la misurazione dell'efficienza dei distretti scolastici del New Jersey. In una fase iniziale le misure di efficienza vengono calcolate utilizzando solo gli input controllabili, in riferimento ad una frontiera di produzione costruita mediante l'applicazione della metodologia DEA ipotizzando rendimenti variabili di scala. Nella seconda fase, le misure di efficienza vengono spiegate attraverso un vettore di variabili esplicative esterne, non controllabili, utilizzando un modello tobit.

L'aspetto innovativo del lavoro risiede nel fatto che gli Autori considerano la differenza tra la misura dell'efficienza tecnica determinata attraverso la DEA e il valore stimato tramite il modello tobit, per ciascuna unità produttiva analizza-

ta, come un indice per misurare l'efficienza tecnica "pura" attribuibile all'agire del produttore. In tal modo è possibile ottenere una scomposizione dell'efficienza complessiva individuando l'effetto delle variabili esogene non controllabili.

Fried, Lovell e Vanden Eeckaut (1993), nel primo stadio dell'analisi, determinano la *performance* delle banche di credito cooperativo (*credit union*) negli Stati Uniti, riferendosi ad una tecnologia non convessa, ossia costruendo la frontiera di produzione FDH. Successivamente, per individuare le determinanti della *performance* delle unità produttive, i suddetti Autori si avvalgono di un modello logit, nel quale la variabile dipendente dicotomica assume valore uno se l'unità è dominata e non efficiente e valore zero in caso contrario.

Riconoscendo che la mancata considerazione degli *slack* potrebbe costituire una limitazione dell'analisi, Fried, Lovell e Vanden Eeckaut propongono di spiegare l'efficienza "totale" di ciascuna unità produttiva, che comprende sia la misura radiale che l'eventuale *slack*, attraverso otto distinte regressioni, nelle quali vengono incluse un insieme di variabili ambientali. La necessità di considerare tanti modelli di regressione quante sono le variabili considerate nella determinazione dell'efficienza deriva dal fatto che gli *slack* non possono essere aggregati, essendo riferiti a variabili quantificate attraverso diverse unità di misura.

Nel lavoro di Lovell, Walters e Wood (1994) il processo di formazione secondaria negli Stati Uniti viene descritto attraverso un processo produttivo costituito da più livelli. In particolare, i suddetti Autori specificano il processo di fornitura di servizi di istruzione da parte delle istituzioni scolastiche, il processo di trasformazione da parte degli studenti in output intermedi e, infine, il processo di trasformazione dei servizi di istruzione in risultati di lungo periodo da parte degli studenti. Per ottenere un ordinamento delle unità secondo la misura di efficienza evitando i problemi derivanti dall'asimmetria della distribuzione delle misure di efficienza, Lovell, Walters e Wood applicano la procedura suggerita da Anderson e Petersen (1993)¹⁷. L'analisi delle *performance* delle scuole secondarie viene perfezionata mediante la specificazione di diversi modelli di regressione, che in relazione a ciascun livello, introducono particolari variabili di tipo esogeno. Va osservato che la modificazione dei valori di efficienza, ottenuta attraverso la procedura di Anderson e Petersen, consente di applicare gli OLS per la stima dei parametri di ciascuno dei tre modelli.

¹⁷ Allo scopo di ottenere una graduatoria di efficienza che sia in grado di distinguere anche le unità caratterizzate da una misura uguale a 1, Andersen e Petersen (1993) suggerirono un criterio che si basa sul concetto di super efficienza (*super efficiency*), ossia sulla possibilità che l'output di una unità possa essere diminuito senza che questa divenga inefficiente.

È interessante citare, a questo punto, alcuni contributi che suggeriscono l'introduzione delle variabili esogene nell'analisi di efficienza attraverso l'applicazione sequenziale di tecniche parametriche e non parametriche.

In particolare, Bhattacharyya, Lovell e Sahay (1997), dopo aver determinato le misure di efficienza di 70 aziende di credito indiane, utilizzando un modello DEA, impostano un'analisi basata su un modello di frontiera stocastica, che contiene una componente sistematica, legata a un vettore di variabili esogene, e un termine di errore composto, costituito dalle usuali componenti unilaterale e puramente casuale.

Fried, Lovell, Schmidt e Yaisawarng (2002) suggeriscono di utilizzare una procedura in tre stadi per incorporare gli effetti di variabili esterne e di disturbi casuali nella valutazione della *performance* dei produttori. Nel primo stadio viene applicata una procedura DEA per ottenere stime iniziali delle misure di efficienza tecnica. Nel secondo stadio è utilizzato un modello di frontiera stocastica per spiegare le misure precedenti in termini di un insieme di variabili ambientali, ottenendo una scomposizione della variazione in tre componenti (effetti ambientali, inefficienza e disturbi casuali). Infine, nel terzo ed ultimo stadio, dopo aver aggiustato gli input (o gli output secondo l'orientamento prescelto) in funzione dei risultati ottenuti nel secondo stadio, viene utilizzato un modello DEA per la valutazione delle *performance* delle unità produttive. I suddetti Autori applicano la procedura in tre stadi all'analisi dell'efficienza di un campione di 990 case di cura statunitensi, caratterizzate dall'operare in condizioni socio-ambientali molto diverse. I risultati mostrano che la valutazione della *performance* muta drammaticamente quando vengono incorporate le variabili esogene che solitamente vengono ignorate in un modello DEA "convenzionale".

4.5.4 La presenza di strutture gerarchiche nei dati: la *Program Evaluation*

In una delle prime applicazioni della metodologia DEA, già citata in precedenza, Charnes, Cooper e Rhodes (1981) introducono un'importante distinzione tra due tipi di efficienza: l'efficienza manageriale (*managerial efficiency*) e l'efficienza del programma (*program efficiency*).

La separazione di questi due concetti, e quindi, la conseguente definizione di due misure corrispondenti, è necessaria se si vuole evitare di imputare i risultati di un buon *management* ad un cattivo programma (e viceversa).

La singola DMU potrebbe trasformare un insieme di input (*program input*) in una serie di output (*program output*), stabiliti a livello di programma. Le unità produttive potrebbero utilizzare diverse tecnologie, condizionate dal

tipo di programma del quale fanno parte, pur appartenendo ad una medesima "industria".

Perciò la "*program efficiency*" dovrà essere misurata in riferimento al comportamento dell'unità produttiva nello svolgimento del processo all'interno della propria tecnologia; mentre la definizione della "*managerial efficiency*" implica il confronto tra i diversi programmi.

La misura radiale di efficienza tecnica per ciascuna delle n_α unità produttive appartenenti al programma α , viene determinata utilizzando la metodologia DEA, mediante la soluzione del problema di massimizzazione espresso nel modo seguente:

$$\max P_0^\alpha = \frac{\sum_{r=1}^{m_\alpha} u_r^\alpha y_{r0}^\alpha}{\sum_{j=1}^{s_\alpha} v_j^\alpha x_{j0}^\alpha}$$

$$s.t : \frac{\sum_{r=1}^{m_\alpha} u_r^\alpha y_{ri}^\alpha}{\sum_{j=1}^{s_\alpha} v_j^\alpha x_{ji}^\alpha} \leq 1; \quad i = 1, \dots, n_\alpha$$

$$u_r^\alpha, v_j^\alpha \geq 0; \quad r = 1, \dots, m_\alpha; \quad j = 1, \dots, s_\alpha$$

In considerazione dei concetti esposti in precedenza, il valore di P_0^* esprime la misura di efficienza di Farrell per ciascuna DMU considerata, in riferimento alla frontiera del proprio insieme delle possibilità produttive α , per $\alpha=1, \dots, k$.

Dopo aver calcolato la misura di efficienza per le unità produttive in ciascuno dei k programmi, al fine di impostare un confronto tra unità appartenenti a diversi programmi, e quindi effettuare "*across-envelope efficiency comparisons*", Charnes, Cooper e Rhodes sottolineano la necessità di *proiettare*¹⁸, per ciascun programma o gruppo, le unità non efficienti sulla frontiera efficiente, in modo tale che "*all DMUs are as efficient as the most efficient among them*".

Successivamente, tutte le unità produttive, con i livelli di input aggiustati, sono incluse in un unico gruppo al fine di determinare la "*inter-envelope*", ossia

¹⁸ La proiezione di un'unità produttiva sulla frontiera efficiente, costruita secondo un orientamento nel senso degli input si ottiene mediante l'espressione: $(\theta^* x_i, y_i)$. Nel caso di un problema orientato nel senso degli output, la proiezione radiale è espressa da $(x_i, \phi^* y_i)$.

la frontiera complessiva nella cui costruzione sono coinvolti tutti i programmi, rispetto alla quale calcolare una misura di efficienza complessiva (*overall*) per ciascuna unità. La valutazione dei programmi¹⁹, o *program evaluation*, richiede, quindi, un confronto tra le distribuzioni delle misure di efficienza che risultano dalle α -frontiere e dalla “*inter-envelope*”.

Dopo aver rilevato che per misurare la distanza tra le distribuzioni possono essere utilizzate diverse statistiche, i suddetti Autori si concentrano nell’applicazione della “*divergence statistic*” di Kullback, utilizzata in precedenza per scopi analoghi anche in Charnes e Cooper (1980).

La procedura precedente è applicata da Ahn, Charnes e Cooper (1988) per l’analisi dell’efficienza delle università con dottorati di ricerca, pubbliche e private, negli Stati Uniti.

Successivamente, Byrnes (1985) riformula il concetto di *program evaluation* in un diverso contesto, per cui l’efficienza complessiva è scomposta in due componenti: *interna (within)*, esprime l’efficienza di ciascuna unità produttiva all’interno del proprio gruppo, ed *esterna (between)*, che esprime l’efficienza del programma.

Uno dei principali vantaggi dell’approccio di Byrnes consiste nel fatto che la componente di programma è derivata come rapporto tra l’efficienza complessiva e l’efficienza interna. Non risulta, quindi, necessario procedere ad un secondo calcolo dell’efficienza dopo aver proiettato tutte le unità produttive sulla rispettiva frontiera²⁰, come suggerito da Charnes, Cooper e Rhodes.

Per illustrare graficamente la scomposizione della misura di efficienza complessiva, ipotizzando una tecnologia con rendimenti di scala variabili e disposizione forte degli input e degli output, si può considerare un ipotetico insieme di unità produttive, che utilizzano un solo input per produrre un solo output, raggruppate in due distinti gruppi, denominati *A* e *B* (Figura 4.7).

La frontiera efficiente delle unità appartenenti al gruppo *B*, costruita ipotizzando un modello DEA a rendimenti di scala variabili, consiste nei segmenti che congiungono i punti B_1 , B_2 e B_3 ed è denominata “*within frontier*”. In modo analogo si può facilmente identificare la frontiera del gruppo *A*, identificata dai segmenti che congiungono i punti A_1 , A_2 e A_3 .

¹⁹Charnes, Cooper e Rhodes (1981), come si vedrà nel quinto capitolo, applicano la metodologia proposta per la valutazione del cosiddetto *Program Follow Through*, che rappresentava un importante esperimento di assistenza agli studenti svantaggiati delle scuole pubbliche negli Stati Uniti.

²⁰A quanto sembra, Byrnes (1985) fu il primo a suggerire l’uso di test non parametrici basati sui ranghi allo scopo di verificare l’esistenza di significative differenze tra le distribuzioni.

Infine, la frontiera che ricopre tutte le unità produttive, indipendentemente dal gruppo di appartenenza, ossia la frontiera complessiva (*overall frontier*) è identificata dai segmenti che congiungono i punti A_1 , A_2 e B_3 .

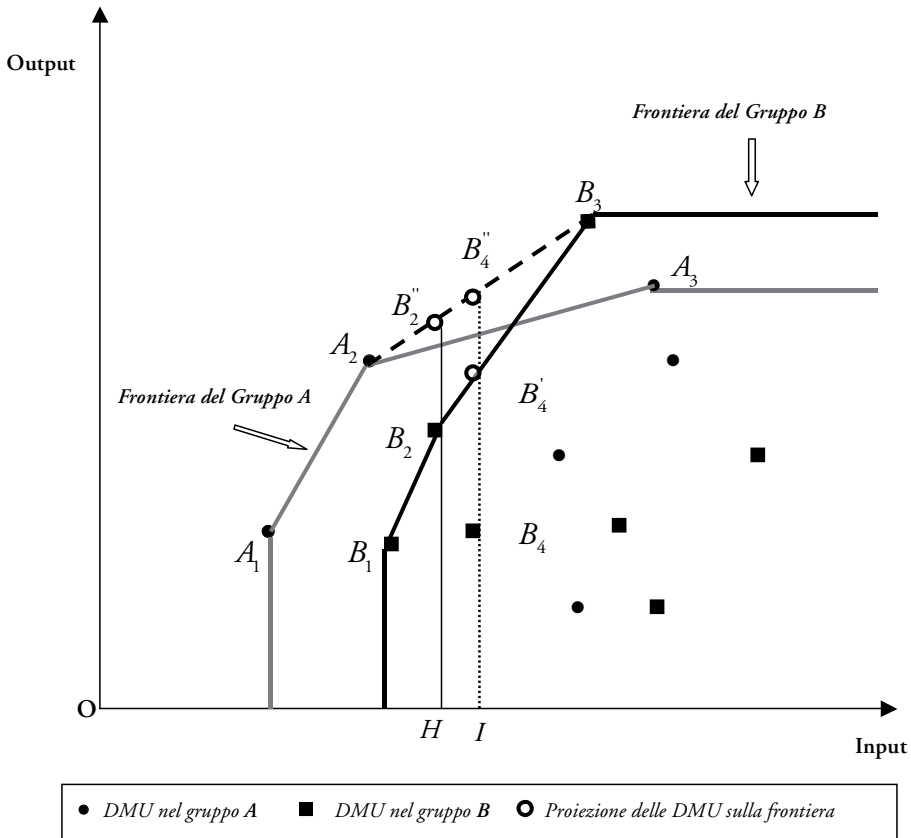


Figura 4.7 – Program Evaluation.

Seguendo un'impostazione di espansione dell'output, si può considerare, ad esempio, l'unità produttiva B_4 appartenente al gruppo B, che presenta una misura di efficienza complessiva espressa dal rapporto IB_4 / IB_4'' . Questa misura può essere scomposta nella componente interna, che, essendo definita dal rapporto IB_4 / IB_4' indica la frazione di efficienza ottenuta dall'unità produttiva B_4 in relazione ai migliori risultati produttivi nel gruppo B, e nella componente esterna, che, specificando la distanza tra la frontiera interna e quella complessiva, risulta espressa dal rapporto IB_4' / IB_4'' .

D'altra parte l'unità produttiva B_2 appartenente allo stesso gruppo, raggiunge un livello di efficienza interna uguale a 1, ossia è pienamente efficiente in relazione alla frontiera del gruppo B , ma presenta una misura di efficienza complessiva inferiore all'unità, espressa dal rapporto HB_2/HB_2^* , a causa dell'inefficienza del proprio programma rispetto agli altri considerati. Infine l'unità A_7 , posizionandosi sulla frontiera complessiva, presenta entrambe le misure di efficienza, interna ed esterna, uguali all'unità.

La procedura suddetta è stata oggetto di numerose applicazioni empiriche, tra cui si può citare, ad esempio, il lavoro di Byrnes, Färe, Grosskopf e Lovell (1988) in cui per sottoporre a verifica statistica le differenze riscontrate nell'efficienza tra i gruppi, viene utilizzata un'analisi della varianza.

Banker (1993) perseguendo l'obiettivo più ampio, consistente nell'individuare le proprietà statistiche degli stimatori DEA, sviluppa una struttura generale di test per la verifica degli effetti del raggruppamento delle unità produttive sulla misura di efficienza.

Brockett e Golany (1996), nell'ambito di un approccio analitico che ricalca quello introdotto da Charnes, Cooper e Rhodes, suggeriscono, invece, l'impiego di test non parametrici, ed in particolare del test di Mann-Whitney. Secondo i suddetti Autori, il passaggio ai ranghi garantirebbe una minore sensibilità del test ai cambiamenti nella distribuzione delle misure di efficienza derivanti dall'adozione di diversi modelli DEA. In tale ottica si pone il lavoro di Ferrari e Laureti (2005), dove accanto alla *program evaluation* si suggerisce l'applicazione di test non parametrici per la valutazione dell'efficienza nella formazione del capitale umano.

Sueyoshi e Aoki (2001) propongono l'impiego del test di Kruskal - Wallis per la valutazione di eventuali spostamenti della frontiera in diversi periodi di tempo.

Thanassoulis e Portela (2002) sviluppano un approccio basato sulla DEA per l'analisi della *performance* di un campione di studenti di 122 scuole secondarie nel Regno Unito. La metodologia suggerita dai suddetti Autori è ispirata a Charnes, Cooper e Rhodes (1981), e conduce all'individuazione di due distinte misure di efficienza, determinate a livello di singolo studente, il cui prodotto esprime l'efficienza complessiva.

Nella terminologia utilizzata da Thanassoulis e Portela l'efficienza interna, ossia calcolata rispetto alla frontiera delle possibilità produttive determinata in riferimento a ciascuna scuola, viene denominata "*pupil within school DEA efficiency*". L'efficienza esterna, esprime la distanza tra la singola frontiera e la frontiera complessiva, viene definita "*school within all schools DEA efficiency*" e viene considerata come una misura di efficienza delle unità produttive scuole e

non del singolo studente. Infine, la frontiera complessiva, che involupa tutte le osservazioni, consente la definizione della “*pupil within all schools DEA efficiency*”, ossia dell'efficienza complessiva.

Castelli, Pesenti e Ukovich (2001) suggeriscono un'interessante metodologia per il trattamento di situazioni caratterizzate dalla presenza di sub-unità non omogenee e non indipendenti raggruppate in macro-unità produttive, mostrando la diversità di soluzioni ottenibili quando ci si discosta da alcuni degli assiomi classici dell'analisi DEA, ossia omogeneità ed indipendenza.

Capitolo Quinto

LA MISURA DELL'EFFICIENZA NELL'ISTRUZIONE UNIVERSITARIA

5.1 PREMESSA

Nel ripercorrere lo sviluppo delle teorie e delle metodologie che si sono susseguite nell'analisi applicata della produzione dagli approcci tradizionali, marginalista e del sovrappiù, alle versioni più moderne, basate sulla teoria degli insiemi, nei capitoli precedenti più volte sono stati citati due tipi di fattori produttivi utilizzati per produrre beni e servizi, ossia il capitale e il lavoro.

Sebbene la trattazione sia stata riferita ad una data situazione tecnologica, e abbia, quindi, analizzato l'efficienza dei processi produttivi in un dato momento, è evidente che l'investimento in beni materiali che compongono il capitale, grazie al progresso tecnologico, determina un incremento della produttività e dell'efficienza nel tempo.

A parità di forza lavoro, capitale materiale e progresso tecnologico, si osserva, tuttavia, che la produzione e la produttività aumentano anche all'aumentare delle capacità, innate e acquisite, di imprenditori e lavoratori (Vittadini e Lovaglio, 2004) che vanno a definire il cosiddetto "capitale umano". In una tale prospettiva, quindi, l'investimento in capitale umano, che rappresenta una parte integrante del capitale stesso, come sottolineato da Kendrick (1972), assume un ruolo fondamentale nello sviluppo economico di un paese.

Nel rinviare i dovuti approfondimenti ai molteplici studi che si sono occupati della definizione e della misurazione del capitale umano, a partire dall'introduzione del concetto stesso di capitale umano, riconducibile ad Adam Smith (1948), passando per le teorie sviluppatesi nella seconda metà del ventesimo secolo, finalizzate a valutarne l'impatto sullo sviluppo economico a livello

macro¹, sino alla formulazione teorica di un approccio statistico ad opera di Dagum, che considera il capitale umano come elemento cardine nei meccanismi di spiegazione del reddito e della ricchezza delle unità economiche, individui o famiglie², si concentrerà l'attenzione, nel prosieguo, sull'istruzione e la formazione, attività attraverso le quali si può investire in capitale umano.

Si indagheranno, in particolare, gli aspetti connessi al processo attraverso il quale si "produce" capitale umano in ambito universitario, sia in termini di una sua rappresentazione nell'ambito della teoria neoclassica della produzione, sia in termini di una conseguente valutazione del grado di efficienza raggiunto nello svolgimento del processo stesso.

In relazione a quest'ultima problematica, i contributi che nei vari paesi hanno condotto l'analisi ricorrendo al concetto di frontiere di produzione sono molteplici ed interessano sistemi di formazione anche molto diversi tra loro. A partire dal lavoro di Ahn, Charnes e Cooper (1988) che costituisce una tra le prime utilizzazioni della frontiera delle possibilità produttive per la valutazione delle istituzioni universitarie, diversi studiosi si sono cimentati nella problematica di misurare l'efficienza tecnica e di costo rispetto alla frontiera delle istituzioni che operano nel campo dell'istruzione universitaria, attraverso l'utilizzazione sia dell'approccio parametrico sia di quello non parametrico, determinando in tal modo lo sviluppo di modelli sempre più flessibili per la descrizione di un processo di produzione così complesso.

La valutazione istituzionale dell'efficienza dei servizi di istruzione universitaria, che rappresentano servizi di pubblica utilità, costituisce, tuttavia, un argomento di ricerca caratterizzato da molteplici aspetti problematici (Gori e Vittadini, 1999, Ferrari, Gori, Maltagliati, Tesi, 1998) che coinvolgono entrambi gli approcci suddetti (Smith e Street, 2005).

In tal senso va nondimeno osservato che l'approccio non parametrico per la stima dell'efficienza ha privilegiato, sin dalle sue prime applicazioni empiriche, i settori non concorrenziali, ed in particolare, il settore della pubblica amministrazione, caratterizzato dalla presenza di unità produttive multi-input e multi-output, dall'assenza di una struttura competitiva per il mercato dei prodotti, dove l'obiettivo dei decisori non si riduce alla sola massimizzazione del profitto (Ganley e Cubbin, 1992, Fabbri, Fazioli e Filippini, 1996, Testi, 1996).

¹ Si possono considerare alcuni contributi degli esponenti della Scuola di Chicago, come, ad esempio, Schulz (1961), Mincer (1958) e Becker (1962).

² Tra gli innumerevoli contributi, si veda ad esempio, Dagum (1994).

Tuttavia, non sono affatto rari gli studi che utilizzano le frontiere parametriche per la valutazione dell'efficienza di organizzazioni operanti nel settore pubblico (Petretto, 1986, Gori, Mealli e Rampichini, 1993).

A prescindere dalla tecnica utilizzata per la costruzione della frontiera, la valutazione dell'efficienza e della produttività del processo di istruzione universitaria, effettuata ricorrendo ai concetti propri della microeconomia della produzione, coinvolge comunque l'analisi delle relazioni fra risorse impiegate (input) e risultati del processo (output)³.

L'attività universitaria è caratterizzata però da una molteplicità di prodotti, risorse e processi. Una scelta accurata che tenga conto di tutti gli aspetti possibili si scontra con i problemi connessi all'identificazione del processo di produzione, alla specificazione dell'unità produttiva e degli obiettivi che la caratterizzano⁴.

È necessario, quindi, riconoscere che la scelta di seguire un approccio quantitativo in un'area così complessa come quella dell'istruzione universitaria, nella quale i processi sottostanti non sono pienamente compresi nella totalità dei diversi aspetti, conduce ad analisi che risultano, talvolta, incomplete ed imperfette (Lindsay, 1982). L'individuazione dei limiti derivanti da ciascun metodo diviene perciò essenziale per la corretta interpretazione dei risultati, che possono, in casi specifici, essere affiancati da analisi complementari.

Per tali ragioni, il presente capitolo analizza le diverse fasi in cui si estrinseca la valutazione dell'efficienza delle istituzioni universitarie, calcolata rispetto alla

³ Secondo Lindsay (1982), gli studi che si sono susseguiti in letteratura sull'analisi della relazione input-output nell'istruzione universitaria potrebbero essere classificati in tre categorie. La prima raccoglie le analisi, soprattutto di costo, basate inizialmente sulla costruzione di semplici rapporti per la determinazione del costo per studente e lo studio di economie di scala e di scopo. Una seconda categoria raggruppa studi diretti all'individuazione e alla descrizione, spesso attraverso una funzione di produzione, della relazione esistente tra le risorse impiegate e i risultati ottenuti, che hanno interessato, in particolare, la formazione scolastica primaria e secondaria (Hanushek, 1979, 1986, Cooper e Cohn, 1997, Krueger, 1999, Lazeur, 2001). Infine, l'ultimo gruppo di studi esplora l'insieme di produzione caratterizzato da unità produttive multi-input e multi-output utilizzando tecniche di costruzione della frontiera delle possibilità produttive.

⁴ A tale proposito, si può osservare che per valutare un qualunque processo produttivo standard della pubblica amministrazione, in cui ricade la stragrande maggioranza delle istituzioni universitarie, almeno nel nostro Paese, e per utilizzarlo a fini comparativi, occorre identificare tutti i compiti svolti da ogni centro di responsabilità economica (Gazzei e Viviani, 1998), individuando inoltre la molteplicità di obiettivi, talvolta difficilmente quantificabili e chiaramente definibili, che caratterizza in generale il settore pubblico (Petretto e Pisauro, 1995; Venanzoni, 1998).

frontiera delle possibilità produttive, cercando allo stesso tempo di approfondire anche le problematiche esistenti. Dapprima, la fase iniziale e più delicata di tutta la “procedura”, ossia la rappresentazione formale del processo di produzione, con la specificazione delle caratteristiche del processo stesso e dell’unità produttiva che lo pone in essere. La discussione sull’individuazione della metodologia per la costruzione della frontiera, rispetto ad una tecnologia di produzione caratterizzata da specifiche proprietà, è invece accompagnata da una rassegna di alcuni lavori empirici riguardanti l’analisi dell’efficienza del settore universitario. Ciò con l’obiettivo di mostrare le possibili applicazioni dei modelli teorici per l’analisi dell’efficienza, in particolare rispetto alla frontiera non parametrica, discussi nei capitoli precedenti, in un settore così fondamentale per lo sviluppo di un paese, ossia quello della formazione del capitale umano a livello universitario.

5.2 LA RAPPRESENTAZIONE FORMALE DEL PROCESSO DI PRODUZIONE

5.2.1 La specificazione degli input e degli output.

In termini generali non si può non condividere l’affermazione di Hare e Wyatt (1992) secondo la quale “*the principal output of the higher education system is knowledge...produced by...research and teaching*”.

Gli *obiettivi principali* di una istituzione universitaria⁵ sono quindi identificabili nelle attività connesse alla fornitura di **servizi di istruzione**, alla **diffusione della conoscenza** e all’incremento della stessa attraverso l’**attività di ricerca**.

Considerando il processo di fornitura di servizi di istruzione, l’attività didattica assume un ruolo centrale, ma la connotazione della stessa dipende dal ruolo che agli studenti viene riconosciuto nello svolgimento del processo. Questi ultimi possono, infatti, essere considerati *materia prima* del processo oppure *clienti*

⁵ Thomas (1974) identifica tre approcci per la descrizione del processo di istruzione, rappresentati dalla cosiddetta funzione di produzione dello psicologo (*psychologist’s production function*), dove gli output del processo sono costituiti dai “*behavioral changes in students, including additions to knowledge, the acquisition of values, or the increased ability to relate to others*”, dalla funzione di produzione dell’economista (*economist’s production function*), che considera gli output come “*additional earnings which result from an increment of schooling, while inputs comprise the cost of that increment*”, e, infine, dalla funzione di produzione dell’amministratore (*administrator’s production function*), connessa all’attività che l’amministratore svolge nell’organizzare risorse per fornire servizi e produrre beni.

o utilizzatori di un servizio, ovvero beneficiari dei servizi offerti, assumendo in tal caso la configurazione di *input sui generis*.

L'identificazione dei diversi fattori che intervengono nel processo di istruzione, sinteticamente classificati in capitale, lavoro, energia e materiali, è oggetto di accese controversie, sia in relazione alla selezione della misura più idonea allo scopo sia in riferimento alla disponibilità di dati.

Per quanto concerne la quantificazione, va innanzitutto osservato che la misurazione dei diversi input può avvenire sia in senso fisico sia mediante l'utilizzazione di costi, sebbene questi ultimi dovrebbero riflettere il valore del contributo di ogni input alla fornitura dei servizi piuttosto che il valore intrinseco degli stessi, determinato applicando necessariamente una prospettiva esterna di mercato.

È chiaro che la metodologia adottata per la traduzione dei diversi fattori in termini quantitativi assume un'importanza determinante nella definizione di indicatori di *performance* che siano confrontabili. Ciò è testimoniato dalla vasta letteratura che ha avuto origine, in particolare in Italia, a partire da lavoro di Biggeri e Zuliani (1983), che considera la misura del tempo come criterio di omogeneizzazione. Per la quantificazione dell'input lavoro, ad esempio in relazione al personale tecnico amministrativo, che coadiuva il personale docente assolvendo compiti amministrativi e tecnici, una tale impostazione si traduce nel passaggio dal numero delle ore effettivamente lavorate alle unità di personale a tempo pieno equivalente.

La quantificazione dell'attività didattica effettivamente svolta, che non si limiti alla determinazione delle ore di lezione per corso di insegnamento, pone problemi di ripartizione del tempo di lavoro dei docenti, i quali svolgono anche attività di ricerca e funzioni di governo e di organizzazione di parti o dell'intera università.

Altri fattori produttivi, tra i quali vanno citati aule, biblioteche e laboratori informatici, contribuiscono in diversa misura allo svolgimento del processo, andando a costituire la dotazione strutturale che gli Atenei mettono a disposizione degli studenti, come utenti di diversi servizi. In relazione ad essi, la misurazione può essere effettuata attraverso la rilevazione della dimensione degli spazi, o del numero di posti effettivamente disponibili in aule e in biblioteche, del numero di postazioni in laboratori informatici, con la specificazione di ulteriori elementi che definiscano le modalità di fruizione dei servizi come, ad esempio, del numero di ore di apertura settimanali.

La *qualità* dei fattori e dei prodotti rappresenta una dimensione di estrema importanza nell'ambito della formazione universitaria.

Un ruolo particolare è rivestito dalla qualità del personale, il cui effetto dovrebbe essere adeguatamente considerato nell'analisi del processo. A tale propo-

sito, Johnes e Johnes (1993) suggeriscono di utilizzare come indicatore qualitativo del lavoro l'esperienza e l'anzianità di servizio, spesso quantificate attraverso il ricorso all'ammontare del salario percepito.

In relazione alle strutture disponibili, la qualità si traduce in un indicatore della funzionalità che potrebbe essere rappresentato dall'anno di costruzione. A tale riguardo, l'inclusione del personal computer (*pc*) tra le attrezzature è spesso complicata dalla mancanza di adeguate misure operative che esprimano la produttività potenziale del *pc*, e non il valore di acquisto o di sostituzione, che spesso non corrisponde esattamente, in generale, al ruolo svolto dalle attrezzature nello svolgimento della ricerca.

Concentrando l'attenzione sulla fase finale del processo, le difficoltà legate all'identificazione e alla misurazione degli output emergono chiaramente dalla definizione di Henderson (1971) che descrive l'istruzione come un *processo di conversione di risorse tangibili in risorse intangibili*.

Misure *ideali* degli output dovrebbero essere rappresentate, per quanto riguarda la didattica o, in generale, l'attività di formazione, da grandezze che esprimano l'incremento del capitale umano per gli studenti che ottengono la laurea.

Nelle applicazioni empiriche la quantificazione dell'incremento di capitale umano apportato è usualmente tradotta attraverso l'inclusione di misure di risultato dell'attività didattica complessiva, rappresentate, spesso, dal numero di studenti ai quali sono "forniti" servizi di istruzione (Avkiran, 2001), dal numero di esami sostenuti (Rizzi, 1999), dal numero di corsi impartiti (Pesenti e Ukovich, 1996) e dal numero di studenti laureati (Athanasopoulos e Shale, 1997).

Sebbene, come osservano Nelson e Hevert (1992), il numero di studenti dovrebbe essere accompagnato da un qualche indicatore che ne esprima la qualità, in termini di "abilità", talvolta le difficoltà connesse all'individuazione e all'ottenimento dei dati necessari a tal fine sono tali da indurre alla sola considerazione della quantità, frequentemente convertita, per scopi comparativi, nel numero di iscritti a tempo pieno equivalenti (*full time equivalent enrollments*).

Una soluzione potrebbe consistere nell'inserire una misura del successo dei laureati sul mercato del lavoro, attraverso il livello di reddito o il tempo di ingresso. A tale proposito, Johnes, Taylor e Ferguson (1987), sottolineando l'importanza di valutare la *performance* di una istituzione universitaria anche attraverso un indicatore del grado di successo dei laureati nel trovare un lavoro successivamente alla laurea, suggeriscono di utilizzare una misura del tasso di occupazione, di cui sono analizzate altresì le possibili determinanti esplicative della diversa situazione registrata tra le università del Regno Unito. Un ulteriore esem-

pio in tal senso è fornito da Avkiran (2001) che, nella valutazione dell'efficienza delle università australiane in relazione alla fornitura di servizi di istruzione, introduce come output la percentuale di studenti occupati a tempo pieno. Tali misure, che potrebbero essere ottenute attraverso indagini specifiche, dipendono, però, fortemente dalla situazione contingente del mercato del lavoro.

Per quanto concerne l'**attività di ricerca** gli indicatori di output ideali dovrebbero poter riflettere l'incremento apportato dalla ricerca alla conoscenza scientifica e il contributo fornito nel determinare cambiamenti sociali e tecnologici.

Sotto questo profilo, va osservato che il confronto di intere università in termini di efficienza nello svolgimento dell'attività di ricerca può essere inficiato dagli effetti derivanti dal differente peso e dal *mix* dei settori scientifico disciplinari presenti che possono contribuire con forme diverse e in diversa misura allo sviluppo delle conoscenze.

Sul tema Beasley (1990), in relazione all'attività svolta dai dipartimenti universitari, suggerisce di classificare l'incremento nella conoscenza (*increased knowledge*) in due categorie: l'apprendimento specifico del soggetto (*person-specific knowledge*) e la conoscenza generale (*general knowledge*).

La prima forma di incremento culturale, fortemente legata anche alle modalità dell'insegnamento, è particolarmente difficile da misurare, mentre l'incremento generale della conoscenza può, invece, essere descritto attraverso misure più "oggettive" rappresentate, ad esempio, dal numero di libri e articoli su riviste, che scaturiscono dall'attività di ricerca.

Nelle applicazioni empiriche questa funzione fondamentale che caratterizza le istituzioni universitarie viene frequentemente modellata come un processo di produzione i cui "prodotti" sono rappresentati dal numero di pubblicazioni, classificate in diverse tipologie, e dall'entità dei fondi ottenuti per lo svolgimento dell'attività stessa.

In relazione alla prima misura, sorge il problema di definizione di un sistema di pesi che consenta di aggregare pubblicazioni di diversa tipologia, che variano da articoli su riviste internazionali a rapporti tecnici interni, per giungere ad una misura complessiva di output (Johnes, 1990).

Per quanto riguarda la seconda grandezza, va osservato che numerosi studi impiegano l'ammontare dei fondi ottenuti da finanziatori esterni facendo proprie le argomentazioni di Cohn, Rhine e Santos (1989), secondo i quali "*the ability of an IHE to generate such funds is closely correlated with its research output, at least insofar as it is perceived by sponsors*".

Sull'argomento, Tomkins e Green (1988) affermano che i fondi ottenuti per ricerca riflettono il valore di mercato delle ricerche condotte e che, per tale

ragione, possono considerarsi come una valida *proxy* dell'output. Tale posizione è consolidata da Ahn, Charnes e Cooper (1988) per i quali il ricorso ai finanziamenti è motivato dalla sostanziale mancanza di misure alternative degli output che scaturiscono dall'attività di ricerca.

D'altra parte, alcuni studiosi rilevano che i fondi o le spese per ricerca costituiscono in realtà un input per lo svolgimento dell'attività stessa. In particolare, Johnes e Johnes (1993) sottolineano che i finanziamenti ricevuti vengono spesi anche per l'acquisto di attrezzature, che rappresentano veri e propri input necessari per dar luogo al processo di produzione.

La tematica relativa alla definizione di indicatori che esprimano la qualità della ricerca è forse la più dibattuta. Una prima strada possibile, come ricordano i già citati Johnes, Taylor e Ferguson, è quella di procedere ad una classifica tra le diverse pubblicazioni, articolata, ad esempio, in libri, contributi in libri, contributi in *journal*, a cui poi vengono associati dei punteggi arbitrariamente definiti. Nel caso di studi pubblicati su riviste, diventa poi essenziale valutare anche il prestigio della rivista stessa, attraverso vari strumenti, non sempre pienamente condivisi. A tale proposito, Johnes e Johnes (1993, 1995) suggeriscono una classificazione dei prodotti scientifici basata sulle caratteristiche del lavoro di ricerca e sul tipo di supporto utilizzato per la pubblicazione⁶.

Un ulteriore esempio in tale area è il ricorso al numero di citazioni. Come evidenzia però Higgins (1989), una tale strada potrebbe essere praticabile esclusivamente per effettuare confronti tra le università all'interno di una medesima area disciplinare (*subject area*), ma certamente non per comparazioni tra prodotti di diversi dipartimenti appartenenti ad una medesima università. Il suddetto Autore sottolinea che "*citations are useful in pointing to which particular staff members are perceived to have contributed to the development of a particular subject area*"

Lo sviluppo di relazioni, nazionali e internazionali, potrebbe rappresentare, in qualche modo, un indicatore della qualità della ricerca, così come la partecipazione dei membri delle università a conferenze internazionali. Infine, il grado di utilizzazione della ricerca prodotta, valutato quantitativamente attraverso il numero delle realizzazioni in azienda, potrebbe costituire, nei

⁶ La classificazione proposta da Johnes e Johnes (1993) è, in particolare, la seguente: *Papers in academic journals; Letters in academic journals; Articles in professional journals, Articles in popular journals; Authored books; Edited books; Published official reports; Contributions to edited works.*

casi in cui ciò abbia senso, un ulteriore aspetto qualitativo da includere nell'analisi.

A completare il quadro delle molteplici attività che caratterizzano le istituzioni universitarie la constatazione che le stesse, entrando in relazione con la società nel complesso, danno vita ad una molteplicità di interazioni, la cui misurazione può essere effettuata, idealmente, attraverso il progressivo accrescimento del livello culturale della società (**diffusione della conoscenza**).

Dal punto di vista operativo, le attività di interazione con la società potrebbero essere misurate dal numero di volte in cui l'istituzione universitaria appare su giornali, in programmi televisivi o radiofonici, dalle partecipazioni di docenti in commissioni pubbliche, o anche dall'eventuale attività di consulenza svolta per il settore pubblico e privato.

Concludendo si può riprendere quanto asserito da Higgins, secondo il quale: *“The university mission or key purposes or major objectives would be regarded as teaching, research and scholarship and providing various other social benefits, in particular contributions to ‘national culture’ and perhaps making useful contributions to the lives of their local communities”*.

Secondo il suddetto Autore, le università di un dato paese producendo manodopera qualificata contribuiscono alla crescita economica del paese stesso.

5.2.2 L'identificazione dell'unità produttiva

La “traduzione” in termini operativi del concetto di efficienza tecnica di una organizzazione produttiva implica il ricorso ad una teoria della produzione allo scopo di individuare i criteri attraverso i quali specificare, dapprima, il risultato dell'attività produttiva, quindi, i fattori impiegati e, infine, la tecnologia nel contesto della quale il processo ha avuto luogo.

Da quanto detto in precedenza, l'identificazione dell'unità produttiva, che, utilizzando un dato insieme di input, ottiene un dato insieme di output, costituisce un elemento necessario nell'impianto teorico e rappresenta una decisione primaria da cui non si può prescindere per una corretta e coerente applicazione delle metodologie disponibili per la stima della frontiera delle possibilità produttive.

È evidente, inoltre, che la definizione stessa del processo, e quindi la specificazione degli input e degli output da includere nell'analisi dell'efficienza dell'istruzione universitaria, è strettamente connessa alla rappresentazione dell'unità produttiva.

L'importanza di definire l'unità produttiva a cui fare riferimento, come primo passo dell'analisi, si evince chiaramente quando Ahn, Charnes e Cooper

(1988) affermano: “...in our use of DEA, these schools (doctoral-granting institutions in the U.S.) serve as Decision Making Units which are defined as units responsible for converting inputs into outputs...”

La definizione dell'unità responsabile del processo è strettamente legata anche agli obiettivi dello studio e al tipo di attività che si intende valutare. Solitamente l'attenzione è rivolta all'attività didattica o a quella di ricerca, tralasciando i servizi forniti alla società in generale.

In tale ottica si può affermare che, nel caso in cui l'interesse verta sulla misurazione dell'efficienza delle istituzioni universitarie nel complesso, l'unità produttiva potrebbe essere rappresentata dall'università stessa.

Tuttavia, la correttezza di un tale modo di procedere può essere messa in discussione dalla presenza di una forte eterogeneità tra le *unità produttive università*. La presenza all'interno di uno stesso Ateneo di diverse “anime” rappresentate dalle facoltà, caratterizzate dalla presenza di specifici settori disciplinari, può invalidare i confronti basati sulle misure di efficienza ottenute conducendo un'analisi impostata a livello di istituzione universitaria.

Nella valutazione dell'attività di ricerca è possibile “risolvere” in parte il problema della comparabilità, sviluppando la valutazione dell'efficienza per dipartimenti “omogenei”, caratterizzati dalla presenza di docenti che svolgono la propria attività in settori scientifico-disciplinari simili. Esempi in tal senso si ritrovano per il Regno Unito con i già citati contributi di Tomkins e Green (1988), Beasley (1990,1995) e Johnes e Johnes (1993).

D'altra parte, se l'obiettivo della ricerca consiste nella valutazione dell'efficienza nella formazione del capitale umano, l'analisi non può prescindere dalla considerazione di tutti quei fattori che agiscono a livello individuale e che caratterizzano il detentore stesso del capitale umano, ossia il singolo studente.

Il ruolo degli studenti è, infatti, centrale nello svolgimento del processo di formazione universitaria (Gori e Romano, 1990). Gli studenti sono i “portatori” degli output, rappresentato dai servizi di istruzione, ma sono anche degli input. Le caratteristiche individuali degli studenti determinano quanto capitale umano è realmente assorbito durante il processo. Lo studente è allo stesso tempo “materia prima” e beneficiario di un servizio di istruzione alla cui “somministrazione” prende parte attivamente.

In un tale contesto, appare ragionevole modellare il processo di formazione di capitale umano nell'università a livello individuale, considerando come unità produttiva il singolo studente che produce se stesso come laureato (Ferrari, Laureti e Maltagliati, 2001, Laureti, 2002, Ferrari e Laureti, 2005).

Una tale rappresentazione consente, come si vedrà in seguito, di tenere in debita considerazione, nell'analisi dell'efficienza tecnica, i diversi fattori che in-

tervengono nel processo soddisfacendo, contemporaneamente, tutte le condizioni richieste per una corretta applicazione della teoria della produzione.

5.2.3 Metodologie per l'analisi della produttività e dell'efficienza dell'istruzione universitaria

Una prima soluzione adottata nell'analisi della produttività e dell'efficienza delle istituzioni universitarie consiste nella costruzione di indicatori semplici di produttività, rappresentati, ad esempio, dal rapporto tra numero di studenti e numero di docenti, e nello sviluppo di indicatori di *performance* (*performance indicators, PI*) che includono, per quanto riguarda la didattica, misure dei costi unitari, tassi di abbandono, tassi di laurea, e misure della *performance* dell'attività di ricerca (Johnes, 1996, Latini e Sterlacchini, 1995).

Il confronto tra diverse istituzioni basato su una singola misura di output, come, ad esempio, il numero di pubblicazioni, sarebbe infatti distorto, a causa del diverso livello di input utilizzati nel processo. È necessario, quindi, introdurre degli elementi correttivi per effettuare validamente i confronti.

Sebbene gli indicatori di produttività, ottenuti rapportando ciascun output ad una misura del fattore impiegato, possono risultare interessanti e utili in alcuni contesti, soffrono, d'altra parte, di molteplici limitazioni, se presi singolarmente, in relazione alla capacità di descrivere la produttività aggregata di un'istituzione complessa come l'università⁷.

Con particolare riferimento alla valutazione della *performance* delle istituzioni nello svolgimento dell'attività di ricerca, l'esigenza di costruire indicatori che includano le diverse tipologie senza attribuire un peso definito a priori in modo arbitrario conduce all'utilizzazione della DEA (Johnes e Johnes, 1993, 1995).

Tale metodologia consente, infatti, di valutare l'attività di ricerca di un dipartimento universitario superando il problema connesso alla scelta del sistema di pesi da applicare alle diverse tipologie di pubblicazioni per giungere ad un unico output. Non esistendo, infatti, un sistema di pesi su cui ci sia un consen-

⁷ Il calcolo di indicatori quantitativi per la valutazione delle attività degli Atenei italiani è stato proposto inizialmente dalla CRUI (Conferenza dei Rettori delle Università Italiane) e successivamente dall'Osservatorio per la valutazione del sistema universitario (Comitato Nazionale per la Valutazione del Sistema Universitario con legge 370/99). Per l'aggregazione dei molteplici indicatori, raggruppati in indicatori di risultato, di risorse, di processo e di contesto, che riflettono le diverse attività svolte dalle istituzioni universitarie lo stesso Osservatorio suggeriva l'adozione di procedure matematiche tipo DEA (Doc 11/98).

so unanime, l'aggregazione delle pubblicazioni si traduce in una procedura spesso arbitraria.

La DEA si presenta però anche come uno strumento metodologico molto flessibile, attraverso il quale si può descrivere l'intero processo posto in essere dalle istituzioni universitarie, che coinvolge non solo l'attività di ricerca ma anche quella di formazione. È per tale ragione che le analisi empiriche dell'efficienza produttiva, condotte applicando diversi modelli DEA, sono moltissime e possono aiutare anche a capire quali indicatori siano più efficaci di altri nel soddisfare determinati schemi teorici.

Il ricorso all'approccio stocastico ha riguardato in particolare la stima di funzioni di costo, con particolare attenzione all'analisi delle economie di scala e di scopo delle istituzioni universitarie. Rinviando a Ehrenberg (2004) per una recente rassegna sugli studi di tipo econometrico nel campo dell'istruzione universitaria, si può in questa sede affermare che la preferenza accordata alla funzione di costo, rispetto a quella di produzione, è spesso motivata dalle difficoltà connesse al reperimento di dati sulle quantità di fattori impiegati.

Benché una trattazione della struttura dei costi che caratterizzano le istituzioni universitarie esuli dal tema principale sviluppato in questo lavoro, si ritiene importante ripercorrere molto brevemente le tappe fondamentali attraverso le quali sono state esaminate le connessioni tra università, come organizzazioni multi-prodotto, costi ed efficienza.

Sull'argomento James (1978) sottolinea che la consuetudine di considerare, nell'analisi della struttura dei costi, il trasferimento della conoscenza, come unico "prodotto" delle università, espresso, in particolare, attraverso il numero di studenti iscritti, determina una sovrastima dei costi e crea distorsioni nei risultati tra i diversi tipi di istituzioni.

Il primo studio in cui le istituzioni universitarie sono considerate come organizzazioni multi-prodotto è rappresentato dal lavoro di Cohn, Rhine e Santos (1989), che apre la strada verso una nuova concezione della struttura dei costi che caratterizza la formazione universitaria. Seguendo l'impostazione suggerita dai suddetti Autori, numerose ricerche sono state condotte, in diversi paesi, tra cui si possono citare, solo per fare qualche esempio, le analisi sviluppate da De Groot, McMahan e Volkwein (1991), Dundar e Lewis (1995), Glass, McKillop e Hyndman (1995), e, infine, da Johnes (1997).

Un approccio interessante per l'interpretazione dei risultati ottenuti dalla stima della funzione di costo, effettuata salvaguardando la comparabilità delle diverse istituzioni, è adottato dal citato lavoro di Johnes (1997), che analizza un campione di 99 università in Gran Bretagna e Irlanda del Nord, nelle quali la spesa totale è collegata a quattro output (studenti iscritti a corsi di laurea in arte,

studenti iscritti a corsi di laurea in scienze, totale studenti iscritti a corsi post lauream, e, infine, valore totale delle sovvenzioni di ricerca e contratti ricevuti). Tra i risultati, il costo addizionale associato con ogni tipo di output viene analizzato in riferimento a tre distinte “tipologie” di università, definite sulla base delle caratteristiche dei laureati “prodotti”. All’università “tipica” (*typical*) si aggiungono le università nelle quali una percentuale molto elevata (75%) di laureati appartiene a corsi di laurea in arte (*arts biased university*) e quelle in cui la percentuale maggiore di studenti (75%) consegue la laurea in corsi di tipo scientifico (*science biased university*).

Per quanto riguarda i metodi di stima utilizzati, poiché nella letteratura economica prevalente sulle funzioni di costo si assume, implicitamente, che tutti i produttori siano tecnicamente efficienti, o nella terminologia di Leibenstein (1966) siano x -efficienti, la stima della funzione di costo è prevalentemente effettuata utilizzando gli OLS e il metodo dei minimi quadrati non lineari.

Sebbene l’obiettivo principale degli studi che hanno coinvolto l’applicazione delle metodologie parametriche consista, prevalentemente, nell’analisi delle eventuali economie di costo e di scopo, nel paragrafo 5.3.2 si considereranno alcune applicazioni nelle quali l’interesse di ricerca verte anche sulla stima dell’efficienza delle istituzioni universitarie.

5.3 L’UNIVERSITÀ COME UNITÀ PRODUTTIVA

5.3.1 L’applicazione delle metodologie non parametriche

Nelle applicazioni empiriche condotte attraverso l’utilizzazione di metodologie non parametriche l’unità produttiva è spesso rappresentata dalle istituzioni universitarie considerate nel loro complesso.

Il già citato lavoro di Ahn, Charnes e Cooper (1988) è diretto alla valutazione dell’efficienza delle università degli Stati Uniti, in particolare delle università con dottorati di ricerca (*doctoral-granting*).

L’obiettivo è quello di utilizzare la DEA per lo studio dell’efficienza delle istituzioni universitarie, in alternativa ai metodi più tradizionali di regressione, evidenziando le maggiori capacità analitiche della metodologia non parametrica. L’analisi si riferisce ai dati del *National Center for Education Statistics* relativi a 161 università *doctoral-granting* degli Stati Uniti, per gli anni 1984-1985, distinguendo tra istituzioni pubbliche (108 unità) e private (53 unità) allo scopo di impostare un confronto tra le due tipologie in termini di efficienza (Tabella 5.1).

I suddetti Autori conducono le elaborazioni per due sottogruppi di università, definiti considerando la presenza (81 unità tra pubbliche e private) o l'assenza (80 unità tra pubbliche e private) di corsi di studio in medicina, allo scopo di superare l'eterogeneità presente nell'ammontare di risorse utilizzate per lo svolgimento del processo. L'efficienza tecnica ottenuta dall'applicazione del modello DEA a rendimenti di scala costanti, è scomposta in efficienza di programma ed efficienza manageriale secondo la procedura di *program evaluation*, suggerita in Charnes, Cooper e Rhodes (1981).

Per quanto concerne la rappresentazione formale del processo produttivo, è interessante rilevare che Ahn, Charnes e Cooper non effettuano una separazione tra l'attività di ricerca e quella di insegnamento e specificano gli input e gli output riportati nella Tabella 5.1.

| Unità produttive | | Input | Output |
|-----------------------------|--------------------------------------|--|---|
| | <i>Univ. con facoltà di Medicina</i> | <i>Univ. senza facoltà di Medicina</i> | |
| <i>Università pubbliche</i> | 56 | 52 | 1. Numero di studenti a tempo pieno equivalenti 2. Numero di studenti in corsi post-lauream a tempo pieno equivalenti 3. Sovvenzioni federali e contratti |
| <i>Università private</i> | 24 | 29 | 1. Spese connesse all'istruzione (stipendi per il personale docente) 2. Investimenti in strutture 3. Altre spese derivanti da attività di supporto all'attività didattica e di ricerca (biblioteca, servizi agli studenti, personale non docente, ecc.) |

Tabella 5.1 – La rappresentazione formale del processo di produzione in Ahn, Charnes e Cooper (1988).

Il riferimento è al numero di studenti, distinti per tipologia di corso (*under graduate* e *graduate*) e all'ammontare di sovvenzioni, come *proxy* dell'output ottenuto dall'attività di ricerca. Gli input impiegati nella formazione universitaria, raggruppati in lavoro e capitale, sono specificati in stipendi per docenti, investimenti in strutture e altre spese necessarie per il corretto svolgimento del processo.

I risultati ottenuti, per le università senza facoltà di medicina, riportati nella Tabella 5.2, mostrano che le università pubbliche risultano più efficienti di quelle private per quanto riguarda la misura dell'efficienza interna, determinata per

ogni gruppo separatamente. L'analisi dell'efficienza esterna, ottenuta dopo l'eliminazione della componente manageriale evidenzia invece una superiorità, in termini di efficienza media, delle unità private.

Ahn, Charnes e Cooper ripetono la medesima analisi per il gruppo rappresentato dalle università con facoltà di medicina ottenendo risultati contrastanti rispetto a quanto evidenziato in precedenza, con le università private meno efficienti delle pubbliche dal punto di vista della *program efficiency*.

La tesi dei suddetti Autori, secondo la quale le università pubbliche risultano più efficienti delle private quando entrambe le componenti dell'efficienza sono presenti nei dati (efficienza totale), sembra quindi essere solo in parte confermata dall'evidenza empirica.

| Gruppo | Efficienza interna | Efficienza esterna |
|-----------------------------|--------------------|--------------------|
| <i>Univ. senza Medicina</i> | | |
| Pubbliche | 0,853 | 0,850 |
| Private | 0,678 | 0,985 |
| <i>Univ. con Medicina</i> | | |
| Pubbliche | 0,839 | 0,997 |
| Private | 0,872 | 0,903 |

Tabella 5.2 – Risultati empirici in Ahn, Charnes e Cooper (1988): efficienza media.

La separazione tra attività didattica e di ricerca, accompagnata dall'individuazione di idonee misure di output che le specifichino, è alla base del lavoro di Abbott e Doucouliagos (2003) sull'efficienza tecnica delle università australiane. Utilizzando i dati prodotti dall'*Australian Department of Employment, Education, Training and Youth Affairs (DEETYA)*, in riferimento al 1995, i suddetti Autori specificano diverse misure di output sia per la ricerca che per la didattica, al fine di confrontare i risultati ottenuti attraverso separate analisi DEA.

In particolare, come si evince dall'esame della Tabella 5.3, tra gli indicatori disponibili per quantificare l'output dell'attività di ricerca, la scelta dei suddetti Autori ricade sul cosiddetto *Research Quantum*, che rappresenta la parte dei fondi pubblici attribuita alle università sulla base del peso che l'università stessa ricopre nell'attività di ricerca totale.

| Unità produttive | Input | Output |
|--------------------------|--|---|
| Università nel complesso | 1. Personale accademico 2. Personale non accademico 3. Spese per altri fattori (energia, biblioteche, servizi agli studenti, ecc.) 4. Stock di capitale | <i>Didattica</i> 1. Numero di studenti a tempo pieno equivalenti 2. Numero di studenti iscritti a corsi di laurea e a corsi post-lauream 3. Numero di lauree conferite di primo livello (<i>under-graduate</i>) e di secondo livello (<i>post graduate</i>) <i>Ricerca</i> 3. Sovvenzioni federali e contratti |

Tabella 5.3 – La rappresentazione formale del processo di produzione in Abbott e Doucouliagos (2003).

I fattori produttivi utilizzati nel processo sono quantificati attraverso il ricorso al personale a tempo pieno equivalente (accademico e non accademico), alle spese relative a tutti gli altri fattori impiegati nel processo, ossia spese per energia, pulizie, servizi agli studenti, biblioteche, ecc., e ad una variabile *proxy* per la misura del capitale immobilizzato⁸.

Nella Tabella 5.4 sono riportati i risultati, in termini di efficienza tecnica e di scala per alcune tra le università considerate, ottenuti dall'analisi DEA condotta considerando alternativamente, come misura dell'output per ricerca, il *Research Quantum* e il *Research Income*, ossia i fondi totali per ricerca medica e non medica ottenuti dalle università. Come si nota, a differenza di quanto accade considerando combinazioni alternative degli altri input e output, le misure di efficienza tecnica e di scala ottenute risultano sensibili alla specificazione dell'indicatore che esprime l'output dell'attività di ricerca. Sebbene alcune università, come, ad esempio, Victoria e Monash, risultino pienamente efficienti in entrambi i casi, si può osservare che l'utilizzazione del primo indicatore conduce, a livello medio, a misure di efficienza inferiori.

È opportuno qui osservare che nelle applicazioni in precedenza esaminate si considera un unico modello per la descrizione del processo che caratterizza l'unità produttiva università. Una tale impostazione, come già anticipato, può

⁸ Abbott e Doucouliagos conducono l'analisi utilizzando diverse combinazioni di fattori e prodotti. In particolare, l'output è specificato anche attraverso i fondi per ricerca ottenuti da soggetti privati.

condurre a risultati fuorvianti a causa delle diverse caratteristiche e del diverso “peso” ricoperto dalle due attività principali svolte dalle istituzioni universitarie.

| | <i>Research Quantum</i> | | <i>Research Income</i> | |
|-------------------------|-------------------------------|--------------------------------|-------------------------------|--------------------------------|
| Università | Efficienza tecnica | Efficienza di scala | Efficienza tecnica | Efficienza di scala |
| Ballarat | 1,000 | 0,916 | 1,000 | 0,944 |
| Latrobe | 1,000 | 0,981 | 1,000 | 0,931 |
| Melbourne | 1,000 | 0,974 | 1,000 | 1,000 |
| Victoria | 1,000 | 1,000 | 1,000 | 1,000 |
| Monash | 1,000 | 1,000 | 1,000 | 1,000 |
| James Cook | 0,739 | 0,991 | 0,743 | 0,983 |
| Curtin | 0,912 | 0,977 | 0,929 | 0,966 |
| Tasmania | 1,000 | 1,000 | 1,000 | 1,000 |
| New England | 0,946 | 0,994 | 0,870 | 0,974 |
| Sydney | 1,000 | 0,963 | 1,000 | 0,932 |
| Efficienza Media | 0,946 | 0,947 | 0,967 | 0,966 |
| Unità efficienti | 24 | 13 | 27 | 16 |

Tabella 5.4 – Risultati empirici in Abbott e Doucouliagos (2003): Misure di efficienza.

In riferimento a tale problematica, i lavori di Avkiran (2001) e Coelli (1996), che hanno per oggetto l'analisi dell'efficienza delle università australiane, sulla base dei medesimi dati utilizzati da Abbott e Doucouliagos, si caratterizzano per la formulazione di diverse ipotesi alternative per la rappresentazione dei processi produttivi, in relazione alla necessità di valutare, talvolta separatamente, le molteplici attività svolte dalle istituzioni universitarie.

In particolare, Avkiran (2001) sviluppa un'analisi DEA che coinvolge le 36 università australiane, per l'anno 1995, descrivendo il processo di produzione in tre modi diversi, che possono far luce sui meccanismi che guidano l'azione manageriale.

Come si nota dall'esame della Tabella 5.5 gli output assumono una configurazione specifica in relazione al tipo di funzione svolta dall'università mentre gli input sono rappresentati dal personale totale (*staff numbers*), suddi-

| Descrizione del processo (Modello) | Unità produttive | Input | Output |
|--|------------------|--|--|
| <i>Performance complessiva</i> | 36 | 1. Personale docente 2. Personale non docente | 1. Studenti iscritti a corsi di laurea (<i>undergraduate</i>) 2. Studenti iscritti a corsi post laurea (<i>postgraduate</i>) 3. Finanziamenti pubblici 4. <i>Research Quantum</i> |
| <i>Servizi di istruzione</i> | 36 | 1. Personale docente 2. Personale non docente | 1. Studenti che si iscrivono a corsi successivi nella stessa università 2. Tasso di avanzamento negli studi degli studenti 3. Percentuale di laureati occupati a tempo pieno |
| <i>Capacità di attrarre studenti non fruitori di borse di studio</i> | 36 | 1. Personale docente 2. Personale non docente | 1. Studenti 2. Studenti stranieri |

Tabella 5.5 – Risultati empirici in Avkiran (2001): unità produttive, input e output.

viso in personale docente e non docente ed espresso in termini di addetti a tempo pieno equivalente.

Secondo il primo modello, definito *performance* complessiva (*overall performance*), le università impiegano risorse umane per produrre iscrizioni, ossia “attrarre” studenti, e generare ricerca. L’inclusione del numero di iscritti come output del processo deriva dal riconoscimento dell’importanza attribuita all’allocazione delle risorse pubbliche (*Federal funding*). Al pari di Abbott e Doucouliagos (2003), Avkiran quantifica l’attività di ricerca considerando come output il *Research Quantum*.

Nel secondo modello, che si focalizza sulla somministrazione dei servizi di istruzione, gli output sono rappresentati dal numero di studenti che proseguono gli studi all’interno della medesima università, iscrivendosi a corsi successivi (*Student retention rate*); dal successo degli studenti nel percorso di studi (*Student progress rate*) e, infine, dalla proporzione di laureati in posizione lavorativa a tempo pieno sul totale dei laureati disponibili a lavorare a tempo pieno (*graduate full type employment rate*).

Nel terzo modello, diretto ad esplorare il successo delle università nell’attrarre studenti non fruitori di borse di studio e, quindi, studenti che contribuiscono al finanziamento dell’università pagando le tasse di iscrizione (*fee-paying students*), gli output sono rappresentati dal numero totale di iscritti, distinti in studenti residenti nel territorio nazionale e studenti stranieri.

Avkiran conduce un'analisi sui rendimenti di scala,⁹ accertando la presenza di rendimenti variabili. La non adeguatezza dell'ipotesi di rendimenti di scala costanti era anche suggerita dalle analisi preliminari, da cui emergevano valori negativi elevati dei coefficienti di correlazione lineare tra il personale docente, utilizzato come *proxy* della dimensione, e le misure di efficienza. A fini illustrativi e comparativi, per le medesime dieci università già considerate nella Tabella 5.4, nella Tabella 5.6 si riportano le misure di efficienza tecnica ottenute utilizzando un modello DEA-VRS, diversamente specificato in funzione delle tre rappresentazioni del processo di produzione.

| Università | Modello 1 Performance complessiva Efficienza-VRS | Modello 2 Servizi di istruzione Efficienza-VRS | Modello 3 Capacità di attrazione Efficienza-VRS |
|-------------------------|--|--|---|
| Ballarat | 1,000 | 1,000 | 1,000 |
| La Trobe | 0,891 | 0,928 | 0,153 |
| Melbourne | 1,000 | 1,000 | 0,750 |
| Victoria | 1,000 | 0,878 | 0,746 |
| Monash | 1,000 | 0,962 | 1,000 |
| James Cook | 0,815 | 0,963 | 0,182 |
| Curtin | 0,932 | 1,000 | 0,967 |
| Tasmania | 1,000 | 1,000 | 0,692 |
| New England | 1,000 | 0,909 | 0,609 |
| Sydney | 1,000 | 0,978 | 0,519 |
| Efficienza Media | 0,955 | 0,967 | 0,634 |
| Unità efficienti | 23 | 12 | 10 |

Tabella 5.6 – Risultati empirici in Avkiran (2001): Misure di efficienza.

I risultati ottenuti evidenziano un livello medio di efficienza piuttosto alto in relazione alle attività descritte dai primi due modelli, mentre per quanto

⁹ Avkiran sviluppa anche un'analisi degli *slacks* negli input, da cui trae utili indicazioni di policy per il miglioramento dei bassi livelli di efficienza che caratterizzano alcune delle università considerate.

riguarda la capacità di attrarre studenti le università non sembrano mostrare *performance* elevate. Rispetto ai risultati ottenuti da Abbott e Doucouliagos (2003), non si evidenziano particolari differenze in relazione all'adozione del modello per la *performance* complessiva. Le università identificate come efficienti da Abbott e Doucouliagos risultano pienamente efficienti anche nel lavoro di Avkiran, sebbene vi siano alcune eccezioni (ad esempio, La Trobe). È evidente che la distanza tra i risultati ottenuti nei due lavori suddetti aumenta se si considera il confronto con le misure derivanti dalla rappresentazione del processo di produzione secondo il modello per i servizi di istruzione.

Si può a questo punto osservare che l'analisi dell'efficienza delle università australiane era stata già l'oggetto del lavoro di Coelli (1996), il quale rappresentava il processo di produzione secondo tre modelli distinti. La Tabella 5.7 riassume le misure di input e di output specificate dal suddetto Autore per le diverse attività considerate.

| Descrizione del processo (Modello) | Unità produttive | Input | Output |
|---|------------------|--|--|
| <i>Modello per l'università nel complesso</i> | 36 | 1. Personale totale 2. Spese diverse dal personale | 1. Numero di studenti 2. Indice delle pubblicazioni |
| <i>Attività prettamente accademica</i> | 36 | 1. Personale docente 2. Altre spese | 1. Numero di studenti 2. Indice delle pubblicazioni |
| <i>Attività amministrativa</i> | 36 | 1. Personale amministrativo 2. Altre spese amministrative | 1. Numero di studenti 2. Personale totale |

Tabella 5.7 – La rappresentazione formale del processo di produzione in Coelli (1996).

Degna di attenzione è l'inclusione del numero di iscritti, espresso in termini di studenti a tempo pieno equivalenti (*equivalent full time student units*), in tutti e tre i modelli. Ciò evidenzia l'importanza che il suddetto Autore attribuisce all'istruzione anche in relazione all'attività amministrativa dell'università.

La valutazione dell'efficienza tecnica è effettuata applicando la metodologia DEA con rendimenti di scala variabili. L'evidenza empirica che emerge dai risultati ottenuti avvalorata la presenza di un livello di efficienza piuttosto elevato, espresso da una media uguale a 0,952 per il modello generale. Coelli conduce anche un'analisi dei rendimenti di scala, dalla quale emerge che la maggioranza delle istituzioni non efficienti presenta rendimenti di scala decrescenti.

Seguendo un'impostazione analitica che si fonda sulla separazione dei fattori principali che caratterizzano la *performance* delle istituzioni universitarie, Athanassopoulos e Shale (1997), misurano l'efficienza economica, secondo la dimensione di minimizzazione dei costi, e l'efficienza di risultato, in riferimento alla massimizzazione dell'output, di 45 università del Regno Unito nel periodo 1992-93. Di particolare interesse è l'utilizzazione di due modelli DEA distinti, nei quali gli output sono rappresentati, in entrambi i casi, dal numero di laureati di primo livello (*numbers of successful leavers*), dal numero di laureati che hanno terminato corsi post lauream, e da un indicatore di output per la ricerca, che ne riassume sia la quantità che la qualità (*weighted research index*).

Nel modello per la determinazione dell'efficienza di costo, l'obiettivo è quello di verificare se i costi di esercizio e i fondi di ricerca di ciascuna università sono adeguati in relazione agli output generati. Gli input del processo sono specificati nelle spese generali di tipo accademico e nei finanziamenti per ricerca, che rappresentano, secondo i suddetti Autori, i principali fattori che sostengono l'attività didattica e di ricerca delle università.

Nella specificazione del modello di efficienza di risultato (*outcome efficiency*), riportato nella Tabella 5.8 a titolo esemplificativo, l'attenzione è concentrata sull'analisi dei risultati che le università ottengono, date le risorse e le abilità degli studenti in entrata, espresse mediante la valutazione conseguita al termine della formazione pre -universitaria (*A-level entry score*).

| Descrizione del processo (Modello) | Unità produttive | Input | Output |
|------------------------------------|---------------------------------------|--|---|
| <i>Efficienza di risultato</i> | Università del Regno Unito (45 unità) | 1. Numero Studenti equivalenti (<i>undergraduate</i>) 2. Numero Studenti equivalenti (<i>postgraduate</i>) 3. Personale docente equivalente 4. <i>A-level</i> medio 5. Fondi per ricerca 6. Spese per biblioteche ed altre attrezzature | 1. Laureati di primo livello (<i>successful leavers</i>) 2. Laureati di livello più elevato (<i>higher degree awarded</i>) 3. Indicatore complessivo per la ricerca |

Tabella 5.8 – La rappresentazione formale del processo di produzione in Athanassopoulos e Shale (1997).

Un elemento interessante del lavoro di Athanassopoulos e Shale è rappresentato dall'introduzione, nel modello di risultato, di specifiche restrizioni imposte sul valore assunto dai pesi dei fattori. In tal modo è possibile introdurre anche giudizi esterni in relazione all'importanza attribuita dagli amministratori ai diversi input e output, allo scopo di analizzarne l'influenza nello svolgimento efficiente del processo.

Dall'esame della Tabella 5.9, nella quale sono riportati i risultati ottenuti per alcune tra le università analizzate attraverso un modello DEA-CRS per la valutazione dell'efficienza di risultato, si nota che l'introduzione del sistema di pesi denominato *value judgement 1*, che ricalca lo schema di ponderazione ufficiale seguito nella distribuzione delle risorse pubbliche alle università, produce una diminuzione dell'efficienza complessiva rispetto alla situazione senza vincoli.

| Università | Efficienza di risultato (<i>Outcome Efficiency</i>) | |
|-------------------------|---|---|
| | Nessun vincolo sui pesi Efficienza -CRS | Vincoli sui pesi - <i>Value judgement 1</i> Efficienza-CRS |
| Bath | 0,916 | 0,883 |
| Cambridge | 1,000 | 1,000 |
| East Anglia | 0,999 | 0,976 |
| Essex | 1,000 | 1,000 |
| Leicester | 0,830 | 0,802 |
| Liverpool | 0,935 | 0,899 |
| Newcastle | 0,945 | 0,901 |
| Oxford | 1,000 | 1,000 |
| Reading | 1,000 | 0,982 |
| Warwick | 1,000 | 1,000 |
| Efficienza Media | 0,972 | 0,938 |
| Unità efficienti | 27 | 14 |

Tabella 5.9 – Risultati empirici in Athanassopoulos e Shale (1997): Misure di efficienza.

Riconoscendo che le istituzioni universitarie, al pari di ogni altra organizzazione, prestano servizi a diversi *stakeholders*, definiti come individui, o gruppi di individui, che hanno un interesse nella continuazione e nella buona gestione dell'organizzazione, Sarrico, Hogan, Dyson e Athanassopoulos (1997) suggeri-

scono di valutare le *performance* delle università del Regno Unito, secondo la prospettiva degli studenti potenziali, interessati ad ottenere un titolo di studio in grado di garantire loro buone opportunità di lavoro e di successo personale.

Rispetto a tale problematica gli studenti potrebbe essere aiutati nella scelta dell'università dove studiare dalle informazioni fornite dalla guida "*The Time Good University Guide*", da cui emerge una graduatoria, denominata spesso *league table*¹⁰, tra le università britanniche sulla base di dieci indicatori di *performance*¹¹.

Allo scopo di considerare le difformi esigenze che caratterizzano le varie tipologie di studenti potenziali, in funzione degli obiettivi che possono distinguerli, i suddetti Autori introducono in un modello DEA specifiche restrizioni nei pesi che intervengono nella ponderazione degli input e degli output, specificati ricorrendo agli indicatori di *performance* utilizzati nella *Times League Table*. In tal modo, attraverso la DEA è possibile ottenere *league table* personalizzate, mediante la costruzione di una graduatoria che tenga conto sia delle caratteristiche delle singole università sia delle esigenze specifiche degli interessati.

Superando le difficoltà e le limitazioni derivanti dall'utilizzazione di un'unica classifica, costruita attribuendo lo stesso peso ai diversi indicatori, i suddetti Autori suggeriscono diverse graduatorie tra le università che differiscono dalla *Times League Table*. La Tabella 5.10 riporta il *rank* definito per gli studenti migliori (*strong student model*), ossia per quegli studenti con elevate abilità accademiche, tipicamente di 18 anni, per i quali è molto importante disporre di un alloggio all'interno del campus. Considerando solo le prime dieci università classificate attraverso la DEA si può notare che per questa tipologia di studenti le due graduatorie risultano molto simili. Ciò è confermato dal valore del coefficiente di correlazione calcolato sul complesso delle università che risulta uguale a 0,879. Per altri gruppi di studenti individuati dai suddetti Autori, come ad esempio, per gli studenti meno abili, i risultati ottenuti con la DEA possono differire anche notevolmente dalla graduatoria del *Times*.

Sull'argomento, risulta interessante il lavoro di Breu and Raab (1994), i quali utilizzano gli indicatori di *performance* sui quali si basa la graduatoria delle migliori venticinque università statunitensi, pubblicata dal *U.S. News and World Report*, per valutare l'efficienza delle medesime attraverso un modello DEA orien-

¹⁰ La costruzione di tali graduatorie, basate su molteplici indicatori di performance delle università, è diffusa in diversi paesi.

¹¹ Tra gli indicatori utilizzati si possono ricordare, ad esempio, il rapporto studenti/docenti, la percentuale di studenti a tempo pieno che alloggiano in residenze universitarie e la situazione occupazionale dei laureati dopo sei mesi dal conseguimento del titolo.

| Università | <i>Modello studenti migliori (Strong student model)</i> | | |
|-------------|---|-----------------|-------------------|
| | Nessun vincolo sui pesi Efficienza-CRS | <i>DEA rank</i> | <i>Times rank</i> |
| Cambridge | 1,000 | 1 | 1 |
| Oxford | 0,986 | 2 | 2 |
| St. Andrews | 0,875 | 3 | 4 |
| Imperial | 0,866 | 4 | 3 |
| York | 0,842 | 5 | 7 |
| Edinburgh | 0,811 | 6 | 4 |
| LSE | 0,799 | 7 | 4 |
| Warwick | 0,790 | 8 | 8 |
| Nottham | 0,789 | 9 | 8 |
| UCL | 0,787 | 10 | 8 |

Tabella 5.10 – Risultati empirici in Sarrico, Hogan, Dyson e Athanassopoulos (1997): Misure di efficienza e graduatorie.

| Descrizione del processo (Modello) | Unità produttive | Input | Output |
|------------------------------------|--|--|--|
| <i>Servizi di istruzione</i> | Migliori università negli Stati Uniti (25 unità) | <ol style="list-style-type: none"> 1. Indicatore di abilità in entrata 2. Percentuale di facoltà con dottorati 3. Rapporto docenti-studenti 4. Spese per studente in servizi di istruzioni e spese generali 5. Tasse di iscrizione per studente | <ol style="list-style-type: none"> 1. Tasso di laurea 2. Indicatore del grado di avanzamento delle matricole |

Tabella 5.11— La rappresentazione formale del processo di produzione in Breu and Raab (1994).

tato nel senso degli output con rendimenti di scala costanti. Come si evince dall'esame della Tabella 5.11, gli input e gli output che caratterizzano il processo delle università considerate sono specificati con particolare riferimento all'attività di fornitura di servizi di istruzione.

La graduatoria che deriva dall'applicazione del modello, riprodotta per le prime dieci università nella Tabella 5.12, riflette, secondo i suddetti Autori, la soddisfazione degli studenti, grazie all'impiego di alcuni indicatori specifici, e può utilmente essere confrontata con la classifica del *U.S. News* basata sulla qualità delle università.

Si verifica facilmente che le misure di efficienza ottenute per le 25 migliori università degli Stati Uniti conducono ad una graduatoria negativamente correlata con quella prodotta dal *U.S. News*. I suddetti Autori suggeriscono quindi di dedicare maggiore attenzione alla valutazione dell'efficienza delle istituzioni universitarie attraverso la DEA.

| Università | Efficienza | <i>U.S. News rank</i> |
|------------------|------------|-----------------------|
| Harvard | 0,946 | 1 |
| Yale | 0,958 | 2 |
| Stanford | 0,910 | 3 |
| Princeton | 0,958 | 4 |
| Cal.Inst.Tech | 0,899 | 5 |
| Mass .Inst.Tech. | 0,989 | 6 |
| Duke | 0,973 | 7 |
| Dartmouth | 1,000 | 8 |
| Columbia | 0,974 | 9 |
| Univ.Chicago | 0,877 | 10 |

Tabella 5.12 – Risultati empirici in Bren e Raab (1994): misure di efficienza e graduatoria.

5.3.2 L'applicazione delle metodologie parametriche

Come già anticipato, le applicazioni empiriche delle metodologie parametriche, nel contesto della valutazione dell'efficienza delle istituzioni universitarie, si sono prevalentemente focalizzate sulla stima della frontiera di costo.

Sebbene nel presente lavoro l'attenzione sia stata concentrata sulle problematiche connesse alla stima dell'efficienza tecnica, si ritiene opportuno

considerare due studi in particolare, allo scopo di illustrarne gli elementi di originalità nel contesto della valutazione della *performance* dell'istruzione universitaria.

Secondo una tale prospettiva, il lavoro di Izadi, Johnes Okrochi e Crouchley (2002) è interessante in quanto rappresenta un tentativo di pervenire ad una valutazione complessiva delle istituzioni universitarie attraverso la determinazione della misura di efficienza tecnica, che si aggiunge alla stima delle economie di scala e di scopo. Sviluppando il contributo di Johnes (1997), i suddetti Autori ottengono una stima ML della frontiera di costo C.E.S., rispetto alla quale vengono successivamente ottenute le stime dell'efficienza tecnica, attraverso la procedura suggerita da Jondrow, Lovell, Materov e Schimdt (1982).

Nel lavoro di Stevens (2001), in riferimento ad un modello di frontiera di costo stocastica, si analizza invece l'impatto esercitato sull'efficienza delle università dalle caratteristiche individuali dei soggetti coinvolti nel processo di formazione universitaria, rappresentati, in primo luogo, dagli studenti e, in secondo luogo, dal personale docente e non docente.

A tal fine, utilizzando i dati raccolti dalla *Higher Education Statistics Agency*, relativi a 80 istituzioni universitarie in Inghilterra e nel Galles, il suddetto Autore, considera un vettore di variabili esplicative attraverso l'adozione del modello di Battese e Coelli (1995).

Tra le variabili considerate come possibili determinanti dell'efficienza si ritrovano l'età, il sesso, il luogo di residenza e la classe sociale di provenienza, specificate sia per gli studenti sia per il personale docente e tecnico amministrativo.

I risultati empirici confermano l'elevata influenza esercitata dalle caratteristiche degli studenti sulla performance delle università, sebbene le stesse siano state introdotte necessariamente in termini di misure aggregate.

5.4 IL DIPARTIMENTO COME UNITÀ PRODUTTIVA

L'analisi dell'efficienza dei dipartimenti universitari, condotta attraverso l'applicazione di metodologie non parametriche, ha costituito l'oggetto di numerosi studi in diversi paesi.

Tomkins e Green (1988) analizzando l'efficienza dei dipartimenti di economia aziendale (*accounting*) nel Regno Unito sottolineano i vantaggi derivanti dall'uso della DEA, che non definisce una struttura di ponderazione degli output a priori, ma determina l'efficienza di ciascun dipartimento sulla base di un sistema di pesi che è quello più confacente all'unità da valutare. L'applicazione empirica,

che ha come scopo anche quello di verificare le effettive potenzialità dello strumento, si basa su una parte dei dati utilizzati nello studio di Groves e Perks (1984) forniti dal *British Accounting Research Register*, per gli anni 1984 e 1985.

| Descrizione del processo | Unità produttive | Input | Output |
|--|---|--|--|
| <i>Attività didattica e di ricerca</i> | <i>Dipartimenti di contabilità (20 unità)</i> | 1. Personale totale a tempo pieno (<i>full time staff numbers, FTE</i>) 2. Spese diverse (<i>non staff expenditure</i>) | 1. Studenti iscritti a corsi di laurea 2. Laureati iscritti a corsi post-lauream (<i>taught postgraduates</i>) 3 Laureati impegnati in attività di ricerca (<i>research postgraduates</i>), 4. Finanziamenti totali 5. Numero di pubblicazioni |

Tabella 5.13 – La rappresentazione formale del processo di produzione in Tomkins e Green (1988).

In particolare, i suddetti Autori stimano sei diversi modelli DEA allo scopo di valutare la sensibilità dei risultati ottenuti alla specificazione dei fattori e dei prodotti che caratterizzano il processo di produzione. I modelli si differenziano sia per il numero di input e output considerati sia per il metodo utilizzato per la quantificazione. Complessivamente i fattori e i prodotti che descrivono il processo di produzione sono riportati nella Tabella 5.13.

Dapprima Tomkins e Green confrontano la graduatoria dei 20 dipartimenti ottenuta in riferimento all'efficienza tecnica calcolata attraverso il modello DEA1, nel quale il processo è definito dall'input 1 e dai primi quattro output elencati nella Tabella 5.11, e il *rank* risultante dalla considerazione del semplice rapporto docenti/studenti (Tabella 5.12). Il confronto rivela che la considerazione di ulteriori elementi nella misurazione della performance dei dipartimenti, rappresentati anche da una semplice distinzione della tipologia degli studenti, determina un evidente miglioramento dell'analisi in quanto consente l'emergere di particolari situazioni e incoraggia le università ad esaminare gli aspetti che la rendono competitiva, e sui quali focalizzare quindi i propri sforzi. Tra i quattro dipartimenti che costituiscono la frontiera efficiente, solo il dipartimento 19, la cui attività si concentra quasi esclusivamente sulla didattica, è efficiente anche secondo il *rank* derivante dal rapporto docenti/studenti. L'esame dei risultati derivanti dall'utilizzazione dei diversi modelli DEA¹², riportati nella Tabella 5.14, mostra un

certo grado di stabilità nelle misure di efficienza derivanti dai diversi modelli. È necessario osservare che l'incremento del numero di input e output considerati determina un "naturale" aumento del numero dei dipartimenti efficienti.

Sebbene Tomkins e Green siano consapevoli di non poter giungere ad un giudizio definitivo sull'efficienza dei dipartimenti, a causa anche della mancata specificazione della qualità degli output, ritengono, tuttavia, che la nuova metodologia possa gettare maggiore luce su una tematica spesso oggetto di forti contrasti.

| Di p. | rapporto docenti/studenti Graduatoria | DEA1 Grad. | DEA1 Eff. % | DEA2 Eff. % | DEA3 Eff. % | DEA4 Eff. % | DEA5 Eff. % | DEA6 Eff. % |
|-------|---|---------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1 | 6 | 10 | 73 | 78 | 100 | 100 | 100 | 91 |
| 2 | 12 | 17 | 53 | 64 | 85 | 84 | 84 | 65 |
| 3 | 20 | 1 | 100 | 100 | 100 | 100 | 100 | 100 |
| 4 | 10 | 9 | 78 | 79 | 79 | 74 | 74 | 69 |
| 5 | 10 | 12 | 65 | 75 | 89 | 87 | 87 | 69 |
| 6 | 18 | 19 | 43 | 67 | 67 | 61 | 61 | 60 |
| 7 | 7 | 6 | 86 | 100 | 100 | 100 | 100 | 100 |
| 8 | 12 | 13 | 57 | 63 | 86 | 86 | 87 | 59 |
| 9 | 4 | 7 | 80 | 100 | 100 | 100 | 100 | 100 |
| 10 | 16 | 16 | 54 | 66 | 72 | 69 | 71 | 60 |
| 11 | 19 | 20 | 41 | 62 | 85 | 85 | 85 | 68 |
| 12 | 8 | 15 | 55 | 57 | 100 | 100 | 100 | 63 |
| 13 | 3 | 5 | 95 | 95 | 100 | 100 | 100 | 88 |
| 14 | 2 | 11 | 71 | 82 | 82 | 95 | 95 | 91 |
| 15 | 5 | 1 | 100 | 100 | 100 | 100 | 100 | 100 |
| 16 | 14 | 18 | 46 | 48 | 48 | 56 | 61 | 54 |
| 17 | 9 | 8 | 79 | 79 | 97 | 89 | 89 | 53 |
| 18 | 14 | 13 | 57 | 72 | 100 | 100 | 100 | 72 |
| 19 | 1 | 1 | 100 | 100 | 100 | 100 | 100 | 100 |
| 20 | 17 | 1 | 100 | 100 | 100 | 87 | 87 | 51 |

Tabella 5.14 – Risultati empirici in Tomkins e Green (1988): Misure di Efficienza tecnica.

¹² I modelli DEA2 e DEA3 includono rispetto al modello DEA1 rispettivamente anche il numero di pubblicazioni e altre spese. Nella DEA3 il numero di docenti viene sostituito dall'ammontare del salario corrispondente, mentre nel modello DEA4 le entrate finanziarie sono

I dipartimenti universitari di chimica e fisica nel Regno Unito rappresentano l'oggetto dell'analisi di efficienza tecnica rispetto alla frontiera delle possibilità produttive condotta da Beasley (1990, 1995).

Dopo aver descritto il processo di produzione di un dipartimento specificando misure di input e output teorici, rappresentati ad esempio dalla *person-specific increased knowledge*, l'attenzione è posta sull'identificazione dei fattori e dei prodotti per i quali i dati sono effettivamente disponibili nell'anno accademico 1986/1987. In relazione a tale problematica, il suddetto Autore rileva che un miglioramento della procedura di valutazione sarebbe sicuramente assicurato da una maggiore disponibilità di dati, in particolare in relazione alle pubblicazioni.

La traduzione degli input e degli output "ideali" in misure quantitative, riportata nella Tabella 5.15, comporta necessariamente anche l'introduzione di specifiche ipotesi, come, ad esempio, l'assunzione che l'ammontare di *increased knowledge* sia il medesimo per tutti gli studenti, indipendentemente dalle caratteristiche individuali degli stessi. I fondi raccolti per ricerca, che compaiono sia tra gli input che tra gli output, rappresentano in realtà una *proxy* che esprime il risultato dell'attività in termini di prodotti pubblicati.

Sebbene la modellizzazione finale del processo di produzione dei dipartimenti non si discosti in modo sostanziale da quella suggerita da Tomkins e Green, si può affermare che il contributo principale di Beasley consiste nel tentativo di pervenire ad una rappresentazione della tecnologia di produzione che rispecchi quanto più fedelmente possibile le caratteristiche peculiari della produzione universitaria. In una tale ottica si possono considerare i miglioramenti apportati al modello base, rappresentati dall'introduzione di vincoli sui pesi che traducono giudizi esterni sull'importanza relativa degli input e degli output considerati, e l'inclusione di misure della qualità dell'attività di ricerca.

In relazione al primo aspetto, un esempio può essere rappresentato, considerando gli studenti come output del processo di produzione (Tabella 5.15), dal vincolo per cui il peso di uno studente laureato impegnato in attività di ricerca (output 1) debba essere maggiore (o uguale) al peso di un laureato iscritto ad un corso post-lauream (output 2) e maggiore del peso di uno studente iscritto ad un corso di laurea (output 3). Tale vincolo si traduce nell'equazione: $y_3 \geq 1,25y_2 \geq 1,25^2y_1$. In riferimento alla qualità dell'attività di ricerca svolta dal dipartimento, Beasley considera la *University Grants Committee reasearch*

suddivise secondo diverse fonti. Infine, attraverso la DEA6 Tomkins e Green analizzano l'efficienza di costo aggregando le diverse spese in un'unica voce.

| Descrizione del processo | Unità produttive | Input | Output |
|--|---|--|---|
| <i>Attività didattica e di ricerca</i> | Dipartimenti di Chimica (52 unità) Dipartimenti di Fisica (50 unità) | 1. Spese generali (principalmente salari) 2. Spese per attrezzature 3. Fondi per ricerca | 1. Studenti iscritti a corsi di laurea 2. Laureati iscritti a corsi post-lauream 3. Laureati impegnati in attività di ricerca 4. Fondi per ricerca 5. Se il dipartimento ottiene il massimo punteggio (<i>star</i>) 6. Se il dipartimento ottiene un punteggio sopra la media (A ⁺) 7. Se il dipartimento ottiene un punteggio medio (A) 8. Se il dipartimento ottiene un punteggio sotto la media (A ⁻) |

Tabella 5.15 – La rappresentazione formale del processo di produzione in Beasley (1990).

ratings secondo la quale i dipartimenti vengono classificati in quattro gruppi in funzione del punteggio ottenuto (output 5, 6, 7 e 8 nella Tabella 5.15).

I risultati ottenuti, riportati solo per alcuni dipartimenti considerati nell'analisi nella Tabella 5.16, fanno apparire a livello medio i dipartimenti di Fisica (71,0%) lievemente più efficienti di quelli di Chimica (68,8%). Il suddetto Autore ritiene tali risultati non solo attendibili, ma anche in grado di fornire ai responsabili dei dipartimenti utili indicazioni su possibili miglioramenti in relazione sia all'attività di ricerca che a quella formativa.

Successivamente, in relazione alla problematica connessa alla contemporanea valutazione, per ciascun dipartimento, dell'attività di ricerca e di didattica, Beasley (1995) suggerisce un approccio non lineare grazie al quale si possono adeguatamente ripartire le risorse comuni tra le due attività.

A differenza degli Autori precedenti, Johnes e Johnes (1993) sviluppano la valutazione dell'efficienza rispetto alla frontiera non parametrica unicamente in relazione all'attività di ricerca dei dipartimenti universitari.

L'applicazione empirica è relativa ai dipartimenti di Economia del Regno Unito. Riconoscendo che uno degli svantaggi connessi all'impiego della

| Università | Dipartimenti di Chimica | Dipartimenti di Fisica |
|--------------------------------|-------------------------|------------------------|
| | Efficienza % | Efficienza % |
| Birmingham | 64 | 58 |
| Bristol | 100 | 68 |
| Cambridge | 96 | 89 |
| Essex | 59 | 100 |
| Leicester | 78 | 70 |
| London- Imperial | 89 | 89 |
| Newcastle | 58 | 81 |
| Oxford | 98 | 71 |
| Southampton | 100 | 90 |
| Warwick | 67 | 86 |
| Efficienza Media | 68,8 | 71,0 |
| Numero unità efficienti | 3 | 1 |

Tabella 5.16 – Risultati empirici in Beasley (1990): Misure di Efficienza tecnica.

metodologia DEA è rappresentato dalla sensibilità dei risultati al numero di input e output individuati, i suddetti Autori procedono alla stima di molteplici modelli DEA, ottenuti specificando diverse combinazioni dei fattori e dei prodotti considerati per la descrizione del processo di produzione¹³. Una tale strategia è perseguibile anche grazie al ricco *data set* a disposizione degli Autori che accanto a dati relativi al personale docente, in relazione a molteplici aspetti, comprende un elenco dettagliato dei lavori scientifici per ciascun membro del dipartimento, attraverso il quale è possibile ottenere una classificazione delle pubblicazioni per tipologia¹⁴. Le analisi si basano infatti su informazioni raccolte dalla *Royal Economic Society* durante i primi mesi

¹³ Successivamente, i 192 vettori esprimono le misure di efficienza dei dipartimenti, ottenute dalle diverse prove, sono analizzati implementando ulteriori tecniche statistiche, come l'analisi di classificazione e l'analisi di correlazione, allo scopo di valutare la robustezza dei risultati ottenuti.

¹⁴ Come già anticipato, le pubblicazioni sono classificate da Johnes e Johnes (1993) secondo le otto categorie: 1. *Papers in academic journals*; 2. *Letters in academic journals*; 3. *Articles in professional journals*; 4. *Articles in popular journals*; 5. *Authored books*; 6. *Edited books*; 7. *Published official reports*; 8. *Contributions to edited works*.

del 1989 nel contesto dell'indagine periodica condotta dall'istituto *Universities Funding Council*.

La Tabella 5.17 riporta, a fini illustrativi, la rappresentazione del processo di produzione adottata in tre diversi modelli DEA considerati dai suddetti Autori. È interessante evidenziare che nel modello B i finanziamenti per ricerca sono inclusi tra gli input mentre nel modello C sono considerati come output del processo. Si esplicita in tal modo il già citato dibattito esistente in letteratura sul ruolo ricoperto dai fondi da destinare all'attività di ricerca.

Le misure di efficienza tecnica ottenute attraverso la DEA sulla base dei tre modelli suddetti sono riprodotte per alcuni dipartimenti nella Tabella 5.18. La sensibilità della tecnica utilizzata alla specificazione degli input e degli output sembra essere limitata al Modello B, dove i fondi per ricerca pro capite sono inclusi tra i fattori produttivi. I risultati ottenuti dagli altri modelli evidenziano simili livelli di efficienza per i dipartimenti analizzati.

| Unità produttive | Descrizione del processo | Input | Output |
|-------------------------------------|---|--|--|
| Dipartimenti di economia (36 unità) | <i>Attività di ricerca</i> Modello A | 1. Personale impiegato in didattica e ricerca 2. Personale impiegato esclusivamente in attività di ricerca | 1. Articoli in riviste accademiche 2. Contributi in volumi 3. Articoli o comunicazioni in riviste di particolare rilevanza. |
| | <i>Attività di ricerca</i> Modello B | 1. Personale impiegato in didattica e ricerca 2. Personale impiegato esclusivamente in attività di ricerca 3. Finanziamenti per ricerca pro capite | 1. Libri 2. Articoli o comunicazioni in riviste di particolare rilevanza. |
| | <i>Attività di ricerca</i> Modello C | 1. Personale impiegato in didattica e ricerca 2. Personale impiegato esclusivamente in attività di ricerca | 1. Articoli in riviste accademiche 2. <i>Letters</i> in riviste accademiche 3. Contributi in volumi 4. Articoli o comunicazioni in riviste di particolare rilevanza. 5. Finanziamenti totali per ricerca |

Tabella 5.17 – La rappresentazione formale del processo di produzione in Johnes e Johnes (1993).

| Università | Efficienza tecnica | | |
|--------------------------------|--------------------|--------------|--------------|
| | Modello A | Modello B | Modello C |
| Aberdeen | 0,88 | 0,77 | 0,88 |
| Birmingham | 0,40 | 1,00 | 0,44 |
| Cambridge | 0,64 | 0,68 | 0,73 |
| Edinburgh | 0,40 | 0,25 | 0,40 |
| Lancaster | 0,63 | 0,46 | 0,77 |
| London (Birkbeck) | 1,00 | 1,00 | 1,00 |
| Nottingham | 0,61 | 1,00 | 0,66 |
| Sheffield | 0,37 | 0,62 | 0,43 |
| Warwick | 0,79 | 0,59 | 1,00 |
| York | 1,00 | 0,47 | 1,00 |
| Efficienza Media | 0,582 | 0,596 | 0,665 |
| Numero unità efficienti | 2 | 6 | 4 |

Tabella 5.18 – Risultati empirici in Johnes e Johnes (1993): Misure di efficienza tecnica.

Con l'obiettivo di esaminare più approfonditamente il ruolo esercitato dai finanziamenti per ricerca nella valutazione della *performance*, Johnes e Johnes (1995) ripetono l'analisi di efficienza per i medesimi dipartimenti di Economia già esaminati. In particolare, la rappresentazione del processo di produzione si caratterizza rispetto alla precedente analisi per l'inclusione, tra gli output, della totalità delle otto categorie di pubblicazioni a cui si aggiunge l'ulteriore tipologia costituita da articoli o comunicazioni pubblicati in riviste di particolare rilevanza.

Analizzando tre distinte specificazioni del processo di produzione derivanti dalle molteplici combinazioni utilizzate dai suddetti Autori si possono trarre interessanti considerazioni. Sebbene i risultati ottenuti siano piuttosto robusti rispetto alla scelta degli input e degli output da includere nell'analisi, va osservato che, in seguito all'introduzione dei fondi per ricerca pro capite come secondo input in aggiunta al personale totale, molti dipartimenti inefficienti nella prima analisi DEA1 vengono considerati pienamente efficienti (Tabella 5.19). Esaminando i risultati ottenuti dalla DEA2 per un numero limitato ma significativo di dipartimenti considerati si nota infatti un incremento generalizzato delle misure di efficienza tecnica. Sebbene i fondi per ricerca rappresentino uno degli elementi fondamentali su cui si basa la graduatoria effettuata dall'istituto *Universities Funding Council*, l'analisi evidenzia notevoli difficoltà nel considerare tali finanziamenti

come una misura dell'efficacia della ricerca. I suddetti Autori concludono che la considerazione degli stanziamenti per ricerca tra gli output conduce a risultati distorti in termini di efficienza in quanto gli stessi rappresentano chiaramente un input nel processo di produzione dei dipartimenti.

| Università | Efficienza tecnica | |
|--------------------------------|--------------------|--------------|
| | DEA1 | DEA2 |
| Aberdeen | 0,83 | 1,00 |
| Birmingham | 0,37 | 0,82 |
| Cambridge | 0,40 | 0,60 |
| Edinburgh | 0,33 | 0,42 |
| Lancaster | 0,72 | 0,95 |
| London (Birkbeck) | 1,00 | 1,00 |
| Nottingham | 0,58 | 0,88 |
| Sheffield | 0,43 | 0,76 |
| Warwick | 0,82 | 0,82 |
| York | 0,88 | 1,00 |
| Efficienza Media | 0,557 | 0,728 |
| Numero unità efficienti | 2 | 9 |

Tabella 5.19 – Risultati empirici in Johnes e Johnes (1995): Misure di efficienza tecnica.

I lavori in precedenza considerati per il Regno Unito confrontano dipartimenti di una *medesima area disciplinare* ma appartenenti a diversi Atenei. Pesenti e Ukovich (1996) e Rizzi (1999) sviluppano invece una valutazione dell'efficienza dei dipartimenti universitari a livello *inter disciplinare* considerando due università italiane, rispettivamente l'Università di Trieste e l'Università Ca' Foscari. L'analisi dell'efficienza dei dipartimenti all'interno di una medesima istituzione universitaria è stata anche oggetto di studio da parte di Sinuany-Stern, Mehrez e Barboy (1994) per la *Ben-Gurion University* in Israele.

Pesenti e Ukovich (1996) sulla base di un insieme di principi generali che dovrebbero essere seguiti per la selezione degli input e degli output, tra i quali, ad esempio, l'eliminazione di input o output altamente correlati, e la considerazione di un numero di fattori non eccessivamente elevato, specificano due input e cinque output per la rappresentazione del processo mediante il quale i dipartimenti dell'Università di Trieste svolgono attività didattica, ricerca e di *fund raising*.

In riferimento a tale modellizzazione, riprodotta nella Tabella 5.20, si può sottolineare la duplice quantificazione del fattore produttivo rappresentato dal personale docente e non docente, espresso in termini di numero di soggetti presenti ed in termini di ammontare dei salari percepiti. In tal modo i suddetti Autori intendono riflettere anche l'esperienza ed il livello di responsabilità del personale. Interessante appare inoltre la specificazione come output dell'attività didattica del numero di corsi tenuti e del numero di esami registrati in un anno.

| Descrizione del processo | Unità produttive | Input | Output |
|----------------------------|--|---|--|
| <i>Attività di ricerca</i> | Dipartimenti dell'Università di Trieste (37 unità) | 1. Fondi ordinari trasferiti dall'amministrazione centrale (spese ordinarie, acquisto libri, riviste, laboratori) 2. Personale docente (numero) 3. Personale non docente (numero) 4. Personale docente (salari) 5. Personale non docente (salari) | <i>Didattica</i> 1. numero esami effettuati 2. numero corsi attivati <i>Ricerca</i> 3. Articoli su riviste internazionali 4. Articoli presentati in conferenze internazionali 5. Articoli su riviste nazionali 6. Articoli presentati in conferenze nazionali 7. Libri 8. Altre pubblicazioni <i>Fund raising</i> 9. Fondi di ricerca |

Tabella 5.20 – La rappresentazione formale del processo di produzione in Pesenti e Ukovich (1996).

D'altra parte, la specificazione di misure che esprimano l'attività di ricerca dei dipartimenti, cui fanno riferimento diverse aree disciplinali, è invece legata al superamento di due principali difficoltà.

La prima consiste nell'individuazione di metodologie idonee ad eliminare l'eterogeneità legata alle diverse tipologie di pubblicazioni e al numero delle stesse, che, inevitabilmente, caratterizzano aree scientifico-disciplinare diverse. La seconda si concretizza nella mancanza di un indicatore oggettivo della qualità della ricerca.

Pesenti e Ukovich tentano di risolvere il primo problema considerando un indicatore specifico, costruito rapportando il numero dei prodotti (*papers*), pubblicati da ciascun ricercatore afferente al dipartimento analizzato, al numero

medio nazionale di prodotti per ricercatore in ciascuna area scientifica, sulla base dei dati forniti dalla CRUI.

Per tener conto della diversa qualità della ricerca, i suddetti Autori suggeriscono di introdurre una graduatoria tra le pubblicazioni che conduce all'aggregazione dei lavori in due gruppi, ossia prodotti internazionali e altri prodotti.

Poiché l'importanza relativa dei prodotti internazionali può essere diversa tra le diverse aree scientifico-disciplinari, viene suggerita l'introduzione di un input binario per distinguere i dipartimenti legati a discipline umanistiche, dove l'importanza può essere minore.

L'ultimo output considerato esprime l'ammontare dei fondi per ricerca raccolti da ciascun dipartimento, normalizzato per tener conto della specifica area scientifica. Tale indicatore, in mancanza di una misura oggettiva, può essere un valido strumento di valutazione della qualità della ricerca.

La frontiera delle possibilità produttive viene costruita sia rispetto ad una tecnologia convessa, utilizzando un modello DEA a rendimenti costanti ed uno a rendimenti variabili, caratterizzati entrambi dall'introduzione di vincoli sul peso delle pubblicazioni, sia rispetto alla tecnologia monotona attraverso un modello FDH.

Nel caso del modello CRS i suddetti Autori ricorrono al concetto di super efficienza (Andersen e Petersen, 1993) che, eliminando dal calcolo la DMU considerata, conduce all'individuazione di misure che possono assumere valori anche superiori all'unità. In tal modo è possibile quantificare, per un'unità efficiente, di quanto l'indice di efficienza supera quello del dipartimento che si posiziona al secondo posto.

Come si evince dall'esame della Tabella 5.21, in cui si riportano le misure di efficienza ottenute per alcuni tra i 37 dipartimenti analizzati, lo strumento migliore dal punto di vista metodologico per l'analisi della *performance* risulta il modello DEA-CRS. Ciò in conseguenza sia del basso potere discriminante manifestato dal modello DEA-VRS e dal modello FDH sia per la scarsa plausibilità dell'ipotesi di rendimenti di scala variabili in relazione all'attività didattica. Infatti secondo Pesenti e Ukovich quest'ultima attività non può dipendere dalla dimensione del dipartimento.

Interessanti risultano infine le indicazioni relative ai livelli di input da impiegare e di output da produrre per raggiungere una situazione di piena efficienza e i suggerimenti per pervenire ad una allocazione "ottimale" delle risorse.

Lo studio di Rizzi (1999) sull'efficienza dei 17 dipartimenti dell'università Ca' Foscari di Venezia affronta il problema dell'identificazione dei prodotti delle facoltà e dei dipartimenti in modo separato, a causa della duplice attività, di didattica e di ricerca, svolte dai docenti, i quali risultano riconducibili, anche se in modo diverso, ad entrambe le strutture.

| Dipartimenti | Efficienza tecnica | | |
|--------------------------------|--------------------|-----------------|--------------|
| | Modello DEA-CRS | Modello DEA-VRS | Modello FDH |
| D01 | 69,24 | 100,00 | 100,00 |
| D03 | 47,74 | 47,74 | 93,00 |
| D10 | 88,57 | 88,58 | 100,00 |
| D13 | 104,36 | 100,00 | 100,00 |
| D17 | 200,45 | 100,00 | 100,00 |
| D22 | 61,25 | 61,68 | 100,00 |
| D25 | 121,33 | 100,00 | 100,00 |
| D29 | 37,28 | 52,96 | 100,00 |
| D33 | 68,73 | 69,62 | 100,00 |
| D37 | 46,84 | 100,00 | 100,00 |
| Efficienza Media | | 87,12 | 99,81 |
| Numero unità efficienti | 12 | 18 | 36 |

Tabella 5.21 – Risultati empirici in Pesenti e Ukovich (1996): Misure di Efficienza tecnica.

In tal senso, nello studio di Rizzi (1999) vengono suggeriti due modelli, denominati *giustificazionista* e *responsabilizzante*, nei quali il modo di rappresentare la didattica assume forme diverse (Tabella 5.22). In particolare, nel primo modello si assume che il dipartimento produca servizi didattici per le facoltà, prendendo come dati le risorse assegnate dall'Ateneo e i livelli di attività definiti dalle scelte delle facoltà e degli studenti. Nel secondo modello, invece, si tende a fondere i due livelli di decisione, facoltà e dipartimento, per evidenziare che i soggetti detentori del potere decisionale sono sempre gli stessi.

Rispetto allo studio di Pesenti e Ukovich, un aspetto di particolare interesse è rappresentato dall'introduzione, tra gli input, di un indicatore della qualità della didattica, derivante dalle valutazioni espresse dagli studenti frequentanti, attraverso la compilazione degli appositi questionari¹⁵. L'attività di ri-

¹⁵ Il Comitato Nazionale per la Valutazione del Sistema Universitario (CNVSU, 2003) ha ritenuto opportuno definire e promuovere modalità comuni per lo svolgimento della rilevazione sui frequentanti e sui laureandi mediante i due documenti Doc 9/02 "Proposta di un insieme minimo di domande per la valutazione della didattica da parte degli studenti frequentanti" e Doc 4/03 "Proposta di un insieme minimo di domande per la valutazione dell'esperienza universitaria da parte degli studenti che concludono gli studi".

| Descrizione del processo | Unità produttive | Input | Output |
|-----------------------------------|--|---|---|
| <i>Modello giustificazionista</i> | Dipartimenti dell'Università Ca' Foscari di Venezia (17 unità) | 1. disponibilità di spazi per studi, uffici e laboratori; 2. disponibilità di personale docente e tecnico-amministrativo, misurata dall'indicatore "personale equivalente"; 3. ammontare dei finanziamenti per il funzionamento ricevuti dall'Ateneo; 4. numero di studenti iscritti per facoltà | <i>didattica</i> 1. numero di corsi di insegnamento attivati; 2. indicatore del volume dell'attività didattica (comprendente il numero di esami sostenuti positivamente dagli studenti e il numero di tesi di laurea); 3. indicatore di qualità della didattica (ottenuto dall'indagine sulla valutazione della didattica da parte degli studenti); <i>ricerca</i> 4. indicatore della capacità di <i>fund raising</i> (costruito rapportando i finanziamenti ottenuti per la ricerca al valore medio per docente ottenuto a livello nazionale dai fondi assegnati dal MURST in co-finanziamento). |
| <i>Modello responsabilizzante</i> | Dipartimenti dell'Università Ca' Foscari di Venezia (17 unità) | 1. disponibilità di spazi per studi, uffici e laboratori; 2. disponibilità di personale docente e tecnico-amministrativo, misurata dall'indicatore "personale equivalente"; 3. ammontare dei finanziamenti per il funzionamento ricevuti dall'Ateneo; | <i>didattica</i> 1. indicatore del volume dell'attività didattica (comprendente il numero di esami sostenuti positivamente dagli studenti e il numero di tesi di laurea); 2. indicatore di qualità della didattica (ottenuto dall'indagine sulla valutazione della didattica da parte degli studenti); <i>ricerca</i> 3. indicatore della capacità di <i>fund raising</i> (costruito rapportando i finanziamenti ottenuti per la ricerca al valore medio per docente ottenuto a livello nazionale dai fondi assegnati dal MIUR in co-finanziamento). |

Tabella 5.22 – La rappresentazione formale del processo di produzione in Rizzi (1999).

cerca viene valutata solo attraverso il *fund raising* senza introdurre il numero di pubblicazioni.

Un ulteriore aspetto distintivo rispetto al lavoro per l'Università di Trieste è rappresentato dall'applicazione sia dell'approccio non parametrico, mediante un modello DEA a rendimenti di scala costanti, sia di quello parametrico, attraverso la stima di una frontiera di produzione deterministica, identificata grazie all'introduzione di una funzione di utilità per i dipartimenti, che consente l'aggregazione dei diversi prodotti in un unico output.

Dall'esame della Tabella 5.23 emerge che utilizzando il concetto di super efficienza nel modello *giustificazionista* due soli dipartimenti sono inefficienti. Il dipartimento di Matematica mostra il livello più elevato di super efficienza pari al 163,4%. Nel modello *responsabilizzante*, ottenuto eliminando dal precedente il numero di studenti (tra gli input) e il numero di corsi attivati (tra gli output), solo nove dipartimenti risultano pienamente efficienti. Rispetto alla frontiera di produzione deterministica, ottenuta utilizzando il metodo COLS, nessun dipartimento risulta efficiente, sebbene livelli elevati di efficienza siano raggiunti dal dipartimento di Scienze economiche (88,9%) e da Economia Aziendale (76,2%).

Si può osservare che le misure di efficienza dei dipartimenti dell'Università Ca' Foscari ottenute attraverso i due modelli DEA conducono a risultati piuttosto diversi, mentre i risultati derivanti dall'utilizzazione della frontiera parametrica sono altamente correlati con le misure derivanti dal modello responsabilizzante. La concordanza di risultati ottenuti con la metodologia parametrica e con la DEA aumenta con l'introduzione di vincoli aggiuntivi sui pesi dei fattori nel modello responsabilizzante.

L'analisi dell'efficienza dei dipartimenti all'interno di una medesima istituzione universitaria è stata come già anticipato oggetto del lavoro di Sinuany-Stern, Mehrez e Barboy (1994) per la *Ben-Gurion University* in Israele. Allo scopo di verificare eventuali diversità di comportamento all'interno delle quattro facoltà¹⁶ in cui si articola l'Ateneo considerato, i suddetti Autori procedono al calcolo dell'efficienza tecnica considerando come insieme delle possibilità produttive i dipartimenti appartenenti ai diversi gruppi (*schools*).

Un modello DEA a rendimenti costanti di scala, orientato nel senso degli input, viene utilizzato per valutare l'efficienza di processi di produzione rappresentati attraverso la specificazione di due input e quattro output, descritti nella Tabella 5.24.

¹⁶ I suddetti Autori escludono dall'analisi la facoltà di Medicina a causa delle forti diversità rispetto alle altre facoltà, in particolare in relazione alle risorse disponibili.

| Dipartimenti | Efficienza tecnica | | |
|--------------------|---|---|-------------------------------------|
| | Modello DEA giustificazionista Efficienza | Modello DEA responsabilizzante Efficienza | Frontiera Deterministica Efficienza |
| Statistica | 93,8 | 93,8 | 64,7 |
| Storia e cr.Arte | 96,1 | 96,1 | 47,4 |
| Italianistica f.r. | 106,9 | 101,0 | 73,6 |
| Sc.giuridiche | 107,3 | 107,3 | 60,1 |
| Chimica | 109,8 | 63,2 | 38,1 |
| Studi storici | 111,3 | 80,0 | 56,9 |
| Chimica fisica | 114,6 | 59,3 | 33,3 |
| St.ling.lett.eur. | 135,7 | 60,3 | 48,7 |
| St.Asia orientale | 139,3 | 132,7 | 70,4 |
| Filosofia e t.s. | 139,8 | 139,8 | 76,1 |
| St.anglo.iber.am. | 142,9 | 135,0 | 66,3 |
| Econ.dir.aziend. | 149,6 | 149,4 | 76,2 |
| St.eurasiatici | 151,4 | 91,2 | 49,3 |
| Sc.ambientali | 152,4 | 58,4 | 39,0 |
| Sc.economiche | 156,1 | 156,1 | 88,9 |
| Sc.antichità v.o. | 156,2 | 114,5 | 60,5 |
| Matematica a.i. | 163,4 | 157,5 | 62,5 |

Tabella 5.23 – Risultati empirici in Rizzi (1999): Misure di Efficienza tecnica.

| Descrizione del processo | Unità produttive | Input | Output |
|--|--|--|---|
| <i>Attività di ricerca e didattica</i> | Dipartimenti dell'Università Ben-Gurion (21 unità) | 1.Spese operative (salari per personale tecnico e amministrativo, altre spese di partimento) 2.Salari per personale della facoltà (docenti, altri soggetti impegnati nella didattica) | 1.Finanziamenti totali 2. Numero di laureati 3. Numero di crediti (ore) forniti dal dipartimento. 4. Numero di pubblicazioni |

Tabella 5.24 – La rappresentazione formale del processo di produzione in Sinuany-Stern, Mehrez e Barboy (1994).

Sinuany-Stern, Mehrez e Barbooy conducono diverse analisi di sensibilità dei risultati ottenuti¹⁷, sia attraverso l'inclusione o l'eliminazione di alcuni fattori, rispetto al modello generale, sia mediante la rimozione di alcune unità produttive dall'insieme di riferimento. Rispetto al primo aspetto va osservato che attraverso l'eliminazione di alcune variabili o l'aggregazione di diversi fattori gli Autori ottengono misure di efficienza diverse.

| Dipartimenti | Efficienza tecnica | | |
|--------------------------------|--------------------|-----------------------------|--------------------------|
| | Modello generale | Modello senza pubblicazioni | Modello con spese totali |
| Geografia | 1,000 | 0,715 | 1,000 |
| Storia | 0,867 | 0,509 | 0,591 |
| Economia | 1,000 | 1,000 | 1,000 |
| Letteratura ebraica | 0,892 | 0,762 | 0,498 |
| Letteratura straniera | 1,000 | 0,805 | 0,839 |
| Filosofia | 0,754 | 0,611 | 0,515 |
| Biologia | 0,703 | 0,586 | 0,548 |
| Chimica | 0,882 | 0,615 | 0,705 |
| Matematica | 1,000 | 1,000 | 1,000 |
| Ingegneria nucleare | 0,887 | 0,544 | 0,762 |
| Efficienza Media | 0,834 | 0,687 | 0,728 |
| Numero unità efficienti | 7 | 5 | 5 |

Tabella 5.25 – Risultati empirici in Sinuany-Stern, Mehrez e Barbooy (1994): Misure di Efficienza tecnica.

In particolare, nella Tabella 5.25 sono riportati i risultati ottenuti costruendo la frontiera efficiente secondo tre diverse rappresentazioni del processo di produzione. Rispetto al modello generale, l'eliminazione del numero di pubblicazioni tra gli output individua un peggioramento della performance dei dipartimenti considerati. Infatti, a livello medio l'efficienza tecnica passa da 0,834 a 0,687 mentre solo cinque unità risultano efficienti rispetto alle sette dell'analisi

¹⁷ Sinuany-Stern, Mehrez e Barbooy conducono anche un esame comparativo tra i risultati ottenibili dalla DEA e da due tecniche di analisi multivariata, ossia l'analisi dei gruppi e l'analisi discriminante.

precedente. L'aggregazione dei due input in un'unica voce di spesa determina una modificazione nell'efficienza dei dipartimenti, che a livello medio sale a 0,728. Sebbene le graduatorie dei dipartimenti in termini di efficienza siano abbastanza simili, i risultati precedenti dimostrano l'importanza rivestita da una corretta rappresentazione del processo di produzione nella valutazione dell'efficienza di unità produttive complesse, quali le istituzioni universitarie.

5.5 IL SINGOLO LAUREATO COME UNITÀ PRODUTTIVA

Il processo formativo al centro del quale si trova lo studente, al tempo stesso “materia prima” e utente di un servizio a cui partecipa più o meno attivamente, è di tipo “integrato”, dipendendo anche da fattori esterni quali l'attività di ricerca ed il contesto socio-economico del territorio nel quale è situata l'università (Ferrari, Laureti e Maltagliati, 2001). Sebbene tali peculiarità rendano assai complicata la rappresentazione formale del processo, in un contesto di teoria applicata della produzione, è possibile trovare una strada che giunge ad una coerente specificazione dell'unità produttiva intorno alla quale costruire l'intero impianto per la valutazione dell'efficienza rispetto alla frontiera delle possibilità produttive.

Tale “percorso” consiste nel considerare il singolo studente come unità produttiva prendendo ispirazione da un'interessante rappresentazione dell'istruzione universitaria come processo in cui le istituzioni forniscono servizi utilizzati dagli studenti per produrre, alla stregua di veri e propri imprenditori, istruzione (Catalano e Silvestri, 1992).

Si può, infatti, innanzitutto affermare che, nello specifico, la formazione di un laureato è un processo di produzione, ed in particolare, di produzione di capitale umano, nel quale l'università, attraverso la formazione che impartisce, trasforma un materiale culturale “rozzo” in un materiale culturale “raffinato”, mediante l'utilizzazione di quei fattori produttivi culturali, intellettuali, materiali e immateriali.

È nondimeno evidente che il processo educativo è fortemente influenzato dalle abilità e dal grado di “coinvolgimento” dello studente.

La produzione di capitale umano nelle università è perciò rappresentabile formalmente attraverso un processo nel quale l'*unità produttiva produce se stessa*. In altre parole, lo studente, impiegando *input* didattici ed utilizzando risorse offerte dalla facoltà che frequenta, produce se stesso come laureato. Tale rappresentazione, peraltro non senza precedenti nel settore dei servizi alla persona (Johnston e Graham, 2001), consente di soddisfare tutte le condizioni necessarie per una corretta applicazione della teoria della produzione.

D'altra parte l'identificazione, la quantificazione e la misura dei fattori produttivi e dei prodotti associati al suddetto processo produttivo continua ad essere alquanto problematica, in riferimento ai diversi aspetti discussi in precedenza, anche nel contesto di una rappresentazione a livello di singolo laureato.

Ferrari e Laureti (2005) conducono un'applicazione empirica della modellizzazione precedente nella valutazione dell'efficienza dei laureati dell'Università degli Studi di Firenze nell'anno 1998.

Nel rinviare ai suddetti Autori i dovuti approfondimenti sulla metodologia di calcolo dei fattori e dei prodotti, si può in questa sede evidenziare (Tabella 5.26) che i beni materiali e immateriali forniti da ciascuna facoltà dell'Università di Firenze sono rappresentati dai professori, ricercatori, posti in aula, libri e dalle diverse attrezzature a disposizione degli studenti, misurati in termini fisici (numero medio di unità per laureato).

I beni materiali e immateriali impiegati nel processo dagli studenti sono rappresentati, invece, dal grado di "conoscenza accumulata" prima di accedere all'università e dalle caratteristiche individuali.

| Descrizione del processo | Unità produttive | Input | Output |
|--|---|--|---|
| <i>Formazione universitaria del capitale umano</i> | Laureati 1998 Università degli Studi di Firenze (2236 unità) | <ol style="list-style-type: none"> 1. Numero medio di professori per laureato 2. Numero medio di ricercatori per laureato 3. Numero medio di posti in aula per laureato 4. Numero medio di aule per laureato 5. Numero medio di libri per laureato 6. Numero medio di riviste per laureato 7. Numero medio di attrezzature per laureato 8. Numero medio di mobili e altro per laureato 9. Voto di diploma | <ol style="list-style-type: none"> 1. Voto medio agli esami 2. Durata degli studi (<i>bad output</i>) |

Tabella 5.26 – La rappresentazione formale del processo di produzione in Laureti e Ferrari (2005).

In relazione a tale aspetto si può affermare che sebbene la misurazione dell'abilità iniziale dello studente sulla base della performance pre-universitaria conduca spesso a risultati insoddisfacenti (Gori, Viviani e Compagnino, 1991), il voto di diploma non solo rappresenta forse l'unica misura disponibile, ma costituisce anche un aspetto estremamente importante per la quantificazione della preparazione dello studente e della sua "propensione" a ricevere servizi di istruzione¹⁸.

Va osservato che l'inclusione nel modello della durata effettiva degli studi come *bad output*, ossia come un output da minimizzare, cerca di coniugare le opposte argomentazione che vedono il tempo del conseguimento del titolo sia come un input del processo sia come un risultato dello stesso. La disponibilità di informazioni sul tempo che gli studenti realmente dedicano allo potrebbe consentire una migliore definizione del ruolo giocato dal fattore durata nel processo di formazione, che come noto costituisce uno degli aspetti più critici del sistema universitario italiano¹⁹.

Con queste basi, i suddetti Autori procedono all'applicazione della metodologia DEA per la determinazione della misura dell'efficienza tecnica rispetto alla frontiera delle possibilità di produzione di ogni processo produttivo che ha come risultato un laureato.

Tale scelta è motivata dalla necessità di considerare molteplici output e input nella descrizione del processo di formazione e di assicurare maggiore flessibilità nell'individuazione della frontiera delle possibilità produttive.

La struttura gerarchica dei dati, specificata nella circostanza per cui lo studente, come unità produttiva, svolge il suo processo all'interno di una specifica facoltà che raggruppa individui caratterizzati da una medesima tecnologia di produzione, suggerisce il ricorso nel contesto della metodologia DEA, alla procedura di *program evaluation*.

Si ottiene quindi una scomposizione della misura di efficienza tecnica calcolata per ciascuno dei circa 2,300 laureati nell'Ateneo fiorentino nell'anno 1998 presi in considerazione, che consente di riflettere il contributo attribuibile esclusivamente al singolo studente (*within efficiency*) e quello derivante dal fatto di

¹⁸ Per superare le difficoltà connesse alla variabilità dei voti nelle diverse aree di studio, si può ricorrere alla variabile espressa in termini di scarto dalla media dei voti per ciascun tipo di diploma.

¹⁹ Tra gli obiettivi della riforma universitaria introdotta con D.M. 509/1999 la riduzione della durata degli studi assume un ruolo centrale esplicitato attraverso la definizione di un percorso di studio della durata legale di tre anni che conduce al conseguimento della laurea di primo livello.

svolgere il processo in una determinata facoltà (*between efficiency*), nel raggiungimento di un determinato livello di efficienza complessiva (*overall efficiency*).

Nella Tabella 5.27 a fini puramente illustrativi sono riportati i risultati ottenuti per un numero limitato di laureati, uno per ciascuna facoltà dell'Ateneo fiorentino, in termini di efficienza complessiva, interna ed esterna.

| Laureto | Facoltà | Efficienza tecnica | | |
|-------------|--------------------|--------------------|---------|---------|
| | | Complessiva | Interna | Esterna |
| Laureato 1 | Agraria | 0,830 | 0,889 | 0,934 |
| Laureato 2 | Architettura | 0,963 | 0,963 | 1,000 |
| Laureato 3 | Economia | 0,805 | 0,911 | 0,883 |
| Laureato 4 | Farmacia | 1,000 | 1,000 | 1,000 |
| Laureato 5 | Giurisprudenza | 0,856 | 0,856 | 1,000 |
| Laureato 6 | Ingegneria | 0,837 | 1,000 | 0,837 |
| Laureato 7 | Lettere | 0,754 | 0,975 | 0,773 |
| Laureato 8 | Medicina | 0,621 | 1,000 | 0,621 |
| Laureato 9 | SS.MM.FF.NN. | 0,562 | 0,943 | 0,595 |
| Laureato 10 | Scienze Politiche | 0,902 | 0,912 | 0,990 |
| Laureato 11 | Scienze Formazione | 0,950 | 0,980 | 0,969 |

Tabella 5.27 – Risultati empirici in Laureti e Ferrari (2005): misure di efficienza per laureato.

Come si nota, alcuni laureati possono raggiungere la piena efficienza all'interno della propria facoltà, come il laureato 6 in Ingegneria, ma risultare inefficienti in riferimento alla frontiera complessiva determinata considerando anche le altre facoltà dell'Ateneo.

L'evidenza empirica che emerge dai risultati a livello aggregato per facoltà, riportati nella Tabella 5.28, conferma l'influenza esercitata dalla facoltà frequentata e dall'interazione facoltà-studente nello svolgimento del processo di formazione universitaria da parte del singolo studente.

Se si passa ora a considerare un aspetto del processo di formazione del capitale umano a livello universitario, più volte richiamato nel corso del presente capitolo, ossia quello connesso alla necessità di considerare adeguatamente il ruolo svolto dalle caratteristiche individuali degli studenti, si può osservare che le possibilità offerte dall'approccio non parametrico per l'introduzione di tali variabili, per lo più di tipo categoriale, non sono molto allettanti, traducendosi, sostanzialmente, in procedure che restringono il confronto

| Facoltà | Unità totali | Efficienza Complessiva | | | Efficienza Interna | | | Efficienza Esterna | | |
|--------------------|--------------|------------------------|------------------|--------|--------------------|------------------|--------|--------------------|------------------|--------|
| | | Media | Unità Efficienti | % fac. | Media | Unità Efficienti | % fac. | Media | Unità Efficienti | % fac. |
| Agraria | 70 | 0,733 | 0 | 0,0 | 0,941 | 8 | 11,4 | 0,779 | 0 | 0,0 |
| Architettura | 396 | 0,948 | 16 | 4,0 | 0,949 | 16 | 4,0 | 1,000 | 217 | 54,8 |
| Economia | 388 | 0,832 | 1 | 0,3 | 0,934 | 17 | 4,4 | 0,890 | 1 | 0,3 |
| Farmacia | 53 | 0,945 | 12 | 22,6 | 0,947 | 12 | 22,6 | 0,998 | 32 | 60,4 |
| Giurisprudenza | 275 | 0,935 | 19 | 6,9 | 0,935 | 19 | 6,9 | 1,000 | 268 | 97,4 |
| Ingegneria | 214 | 0,766 | 1 | 0,5 | 0,930 | 13 | 6,1 | 0,824 | 1 | 0,5 |
| Lettere | 230 | 0,791 | 2 | 0,9 | 0,967 | 22 | 9,6 | 0,818 | 2 | 0,9 |
| Medicina | 80 | 0,711 | 0 | 0,0 | 0,965 | 17 | 21,3 | 0,737 | 0 | 0,0 |
| SS.MM.FF.NN. | 170 | 0,685 | 0 | 0,0 | 0,928 | 13 | 7,6 | 0,739 | 0 | 0,0 |
| Scienze Politiche | 164 | 0,936 | 10 | 6,1 | 0,944 | 15 | 9,1 | 0,992 | 32 | 19,5 |
| Scienze Formazione | 196 | 0,918 | 7 | 3,6 | 0,966 | 13 | 6,6 | 0,950 | 8 | 4,1 |
| Totale | 2236 | 0,849 | 68 | | 0,945 | | | 0,899 | 561 | |

Tabella 5.28 – Risultati empirici in Laureti e Ferrari (2005): Misure di Efficienza per facoltà.

fra unità appartenenti a gruppi con le medesime caratteristiche, specificate dalle categorie introdotte.

L'importanza di considerare adeguatamente tali variabili, definibili come “esogene” e rappresentate, ad esempio, da genere, età, tipo di diploma, voto alla maturità, è riconducibile al fatto che esse determinano una forte eterogeneità negli studenti e condizionano gli esiti del processo formativo, in termini di completamento degli studi e di raggiungimento di una *performance* specificata dal voto conseguito alla laurea e dalla durata degli studi.

Una possibile soluzione a tale problema, nel contesto dell'approccio non parametrico all'analisi dell'efficienza tecnica, consiste nell'adottare una procedura, articolata in due stadi e suggerita in Ferrari e Laureti (2004a). In una prima fase, l'adozione di una metodologia multivariata, in particolare l'analisi delle corrispondenze multiple, consente di individuare le componenti principali che descrivono il processo di formazione, attraverso la “sintesi” di una molteplicità di variabili tra le quali sono incluse anche variabili categoriali esprimenti le caratteristiche degli studenti. Successivamente, si procede all'analisi DEA per

la determinazione dell'efficienza tecnica utilizzando le prime componenti principali, precedentemente estratte, dopo idonea trasformazione.

Un aspetto interessante dei risultati ottenuti in Ferrari e Laureti (2004a), che può essere brevemente richiamato, consiste nella "interpretazione" dei primi due fattori principali. La rappresentazione del piano principale, infatti, unitamente agli ulteriori elementi di valutazione che la procedura fornisce, consente di interpretare il primo asse fattoriale come una scala crescente di dotazione di risorse umane e di capitale. Nell'interpretazione del secondo asse giocano, invece, un ruolo importante, il voto di diploma e la residenza durante gli studi, che potrebbe in un qualche modo riflettere gli effetti negativi del pendolarismo. Lungo il semiasse positivo si ritrovano, infatti, coloro che hanno conseguito il diploma con votazioni elevate e con residenza a Firenze. Invece, nel semiasse negativo si posizionano studenti caratterizzati da voti di diploma molto bassi e residenti lontano dalla sede universitaria.

Il ricorso a metodologie parametriche può rappresentare una valida soluzione al problema dell'inclusione delle caratteristiche individuali degli studenti nella valutazione dell'efficienza tecnica, sia attraverso l'adozione di modelli stocastici, nei quali la media della variabile casuale normale troncata, che descrive il termine di inefficienza u , viene espressa mediante un vettore di variabili esplicative sia attraverso la rimozione dell'ipotesi di omoschedasticità del termine u .

Se l'utilizzazione di un modello stocastico per la stima simultanea della misura di efficienza tecnica e delle determinanti delle *performance* dei laureati, rappresenta la soluzione suggerita in Ferrari e Laureti (2004b), motivata anche dal ricorso al concetto di meta-frontiera di produzione al fine di impostare un confronto tra laureati in facoltà diverse, in Laureti (2006) la rappresentazione formale del processo di istruzione universitaria è arricchita grazie all'introduzione di una funzione esplicativa della varianza dei termini di disturbo dovuti ad inefficienza tecnica.

L'ipotesi di omoschedasticità del termine di inefficienza diviene insostenibile quando le variabili esogene, che assumono diverse modalità da individuo a individuo, specificando "l'ambiente" in cui il processo ha luogo, raggiungono dimensioni elevate.

In altre parole, appare inverosimile assumere che la variabilità del comportamento del singolo studente nello svolgimento del processo sia la stessa per ciascuna unità considerata.

In riferimento al processo di formazione universitaria descritto dalle variabili riportate nella Tabella 5.29, la suddetta Autrice stima la misura di efficienza tecnica utilizzando un modello di frontiera stocastica in cui il termine unilaterale u , descritto da una v.c. metà normale, è eteroscedastico. I parametri σ_{ui}^2

variano sistematicamente tra le unità in funzione del vettore \mathbf{z} di variabili che esprimono le caratteristiche individuali degli studenti, ossia in termini formali $\sigma_{ui}^2 = g(\mathbf{z}; \boldsymbol{\delta})$.

| Descrizione del processo | Unità produttive | Input | Output |
|--|--|--|--------------------------|
| <i>Formazione universitaria del capitale umano</i> | Laureati 1998 Università degli Studi di Firenze (2236 unità) | <ol style="list-style-type: none"> 1. Numero medio di professori per laureato 2. Numero medio di ricercatori per laureato 3. Numero medio di posti in aula per laureato 4. Numero medio di attrezzature per laureato 5. Numero medio di libri per laureato 6. Voto di diploma (scarto dalla media) | 1. Voto medio agli esami |
| Caratteristiche individuali | | <ol style="list-style-type: none"> 1. Sesso 2. Provincia di residenza (categorie) 3. Tipo di diploma (dummy) 4. Regolarità pre-universitaria (dummy) 5. Studente lavoratore (dummy) 6. Facoltà 7. Durata effettiva degli studi | |

Tabella 5.29 – La rappresentazione formale del processo di produzione in Laureti (2006).

Le stime dei parametri della funzione di produzione Cobb-Douglas e dei parametri nel modello di eteroschedasticità moltiplicativa di Harvey (1976), $\sigma_{ui}^2 = \exp(\mathbf{z}_i \boldsymbol{\delta})$, considerata nell'analisi, sono ottenute massimizzando la funzione di log verosimiglianza, che rappresenta una generalizzazione della formulazione suggerita da Aigner, Lovell e Schimdt (1977).

L'evidenza empirica che emerge dai risultati in Laureti (2006), a cui si rinvia per un approfondimento anche degli aspetti statistici, avvalorata il nesso esistente tra le caratteristiche individuali e l'inefficienza tecnica dei laureati.

Dall'esame della Tabella 5.30 si nota, in particolare, che, a parità di tutti gli altri fattori, le femmine presentano una minore inefficienza tecnica, e quindi una maggiore efficienza, rispetto ai maschi. Il tipo di diploma, raggruppato nelle due categorie, formazione liceale e non liceale, produce effetti statisticamente significativi sull'inefficienza del processo. L'aver conseguito un diploma presso un istituto tecnico o professionale, *ceteris paribus*, diminuisce il livello di efficienza raggiungibile nella formazione universitaria.

| Variabile | Coefficiente | Std. Err. | p-value | Effetto marginale su $E(u_i)$ |
|--|--------------|-----------|---------|-------------------------------|
| Sesso (Categoria di riferimento = Maschi) | -0,359 | 0,087 | 0,000 | -0,0088 |
| Residenza Categoria di riferimento = residenti a Firenze | | | | |
| Residenti nella provincia di Firenze | 0,215 | 0,085 | 0,012 | 0,0052 |
| Residenti fuori provincia di Firenze | 0,748 | 0,140 | 0,000 | 0,0183 |
| Diploma (Categoria di riferimento = Liceo) | 0,368 | 0,083 | 0,000 | 0,0090 |
| Regolarità scolastica (Categoria di riferimento = carriera regolare) | 0,359 | 0,116 | 0,002 | 0,0087 |
| Studente lavoratore (Categoria di riferimento = No) | 0,085 | 0,084 | 0,311 | 0,0021 |
| Durata degli studi | 0,165 | 0,020 | 0,000 | 0,0040 |
| Facoltà Categoria di riferimento = Giurisprudenza | | | | |
| Agraria | -0,993 | 0,277 | 0,000 | -0,0242 |
| Architettura | -1,922 | 0,209 | 0,000 | -0,0469 |
| Economia | -0,402 | 0,148 | 0,006 | -0,0098 |
| Farmacia | -0,041 | 0,278 | 0,883 | -0,0010 |
| Ingegneria | -0,777 | 0,184 | 0,000 | -0,0190 |
| Lettere | -2,189 | 0,287 | 0,000 | -0,0534 |
| Medicina | -1,267 | 0,309 | 0,000 | -0,0309 |
| SS.MM.FF.NN. | -0,177 | 0,181 | 0,329 | -0,0043 |
| Scienze Politiche | -0,757 | 0,197 | 0,000 | -0,0242 |
| Costante | -5,066 | 0,151 | 0,000 | |

Tabella 5.30 – Risultati empirici in Laureti (2006): stima dei parametri della varianza ed effetti marginali sull'inefficienza.

La quantificazione dell'effetto delle variabili esogene sull'inefficienza tecnica (Wang, 2002) può essere ottenuta attraverso la stima degli effetti marginali (ultima colonna della Tabella 5.30).

L'analisi condotta non solo produce risultati interessanti in relazione all'effetto delle variabili esogene nello svolgimento del processo di istruzione universitaria, ma consente di ottenere anche misure di efficienza per ciascun laureato comparabili tra le diverse facoltà. Gli strumenti messi a disposizione dalla teoria neoclassica della produzione per descrivere ed analizzare i processi produttivi "classici" possono quindi utilmente e coerentemente essere utilizzati per valutare "come" il capitale umano venga prodotto in ambito universitario.

Capitolo sesto

L'IDENTIFICAZIONE DELLA FRONTIERA DELLE POSSIBILITÀ PRODUTTIVE: DAI DATI ALLA STIMA

6.1 PREMESSA

La valutazione della performance di un insieme di unità produttive, attraverso l'utilizzazione di uno specifico strumento metodologico per la stima della frontiera delle possibilità produttive nel contesto della teoria neoclassica della produzione, si estrinseca in una serie di decisioni che il ricercatore deve prendere in relazione a diversi aspetti.

In primo luogo, è necessario “modellare”, ovvero rappresentare formalmente, il processo di produzione di cui si vuole valutare l'efficienza, attraverso l'identificazione delle unità produttive, degli input utilizzati e degli output prodotti. Una tale fase risulta, spesso, tutt'altro che semplice, soprattutto quando si tratta di valutare unità operanti in settori della pubblica amministrazione, dove la definizione e la successiva misurazione delle variabili che intervengono nel processo è particolarmente complessa, come emerso nei capitoli precedenti.

Lo studio attento del processo produttivo e delle caratteristiche delle unità che lo pongono in essere, nonché del contesto ambientale, guidano il ricercatore nella seconda fase, che consiste nella selezione della “migliore” rappresentazione della tecnologia di produzione, condivisa dalle unità produttive costituenti il gruppo “omogeneo” di riferimento.

Il concetto teorico di efficienza, in particolare nell'accezione di efficienza tecnica, è reso operativo attraverso l'individuazione di uno strumento metodologico che consenta quindi di individuare lo “standard ottimale” di riferimento, rispetto al quale misurare la “distanza” dell'unità in questione.

La stima delle frontiere di produzione effettuata sia sulla base di modelli parametrici, legati ad una specifica forma funzionale, sia attraverso l'utilizzo di

tecniche di programmazione matematica, per la costruzione di frontiere di tipo *best-practice*, necessita, evidentemente, di un supporto informatico per l'elaborazione dei dati.

La scelta del programma che consente di eseguire le procedure mediante le quali si riproduce il "modello" individuato è, quindi, un aspetto estremamente importante, strettamente legato alla natura del problema oggetto di analisi.

In questo capitolo si procederà ad una breve descrizione di alcuni tra i principali software disponibili, evidenziando le potenzialità di ciascuno in relazione ai diversi "modelli" che possono essere adottati e stimati nelle analisi empiriche.

L'obiettivo di questa parte finale del lavoro è quello di chiarire le scelte disponibili e quindi le procedure da seguire per la costruzione della frontiera delle possibilità produttive sia secondo un approccio non parametrico, utilizzando il programma AMPL (*A Modeling Language for Mathematical Programming*), sia secondo un approccio stocastico, applicando il software STATA (*Statistics/Data Analysis*).

A tal fine, l'esposizione sarà articolata attraverso il ricorso a semplici esempi "pratici" utilizzando dati alla base di alcuni lavori richiamati nei capitoli precedenti.

6.2 LA COSTRUZIONE DELLE FRONTIERE DI PRODUZIONE NON PARAMETRICHE

6.2.1 I programmi disponibili

La diffusione della DEA come la metodologia più flessibile a cui ricorrere per la valutazione della performance di aziende pubbliche, ma anche di imprese private, è stata accompagnata dal contemporaneo sviluppo di molteplici *software*, commerciali e gratuiti, con caratteristiche e potenzialità notevolmente eterogenee.

Nella Tabella 6.1 sono riportati gli aspetti principali che caratterizzano alcuni tra i più diffusi *packages*, in particolare, in relazione alle diverse tipologie di modelli che possono essere sviluppati attraverso il loro impiego.

Nel rinviare a Hollingsworth (2004) e, in particolare, a Barr (2004) per una dettagliata rassegna, si può osservare, in questa sede, che una versione *student* del software commerciale *DEA-Solver Professional* accompagna il testo di Cooper, Seiford e Tone (2000). È perciò possibile usufruire di un ottimo aiuto per la comprensione e l'implementazione delle metodologie non parametriche. Un

| | Frontier Analyst | PIM DEASoft | DEA Solver-Pro | DEAP | EMS |
|-----------------------|--|--|--|--------|--------|
| Prodotto da: | BANXIA Software http://www.banxia.com/ | Thanassoulis e Emrouznejad http://www.deasoftware.co.uk | SAITECH http://www.saitech-inc.com/ | Coelli | Scheel |
| Modelli | | | | | |
| CCR | SI | SI | SI | SI | SI |
| BCC | SI | SI | SI | SI | SI |
| Additivo | NO | SI | SI | NO | SI |
| FDH | NO | NO | SI | NO | SI |
| Efficienza di costo | NO | NO | | SI | NO |
| Disposizione debole | NO | NO | SI | NO | NO |
| Variabili qualitative | NO | SI | NO | NO | NO |
| Super efficienza | NO | SI | SI | NO | SI |

Tabella 6.1 – Caratteristiche di alcuni software per l'applicazione di modelli non parametrici.

discorso analogo vale per il programma *WARWICK DEA*¹ che accompagna il volume di Thanassoulis (2001).

Per quanto riguarda la possibilità di utilizzare programmi ottenibili gratuitamente, si può segnalare che il *Software EMS 1.3 (Efficiency Measurement System)* di Scheel (2000) dell'Università di Dortmund, può essere scaricato dal sito internet dell'Autore, all'indirizzo <<http://www.wiso.uni-dortmund.de/lsg/or/Scheel/EMS>>.

Gratuito è anche il programma sviluppato da Coelli (1996), denominato *DEAP 4.1*, disponibile sul sito internet del CEPA (*Centre for Efficiency and Productivity Analysis*) all'indirizzo: <<http://www.uq.edu.au/economics/cepa/deap.htm>>.

Entrambi i programmi sono corredati da manuali, predisposti dagli stessi Autori, che guidano il ricercatore nella costruzione della frontiera delle possibilità produttive caratterizzata dalla forma e quindi dalle proprietà sulle quali verte l'interesse della ricerca.

¹ Recentemente, i medesimi Autori di *WARWICK DEA* hanno sviluppato un nuovo software, che assicura notevoli performance, denominato PIM-DEAssoft V1 (*Performance Improvement Management Software*).

Si è ritenuto perciò di maggiore interesse mostrare il passaggio dalla “teoria alla pratica” non ricorrendo ai software più comunemente utilizzati, rispetto ai quali è possibile ottenere agevolmente informazioni relative all’uso, ma adottando un approccio più generale e flessibile.

Lo strumento scelto per illustrare questo percorso è rappresentato dal software AMPL, attraverso il quale è possibile “tradurre” i problemi di ottimizzazione lineare che identificano i diversi modelli non parametrici per la stima della frontiera delle possibilità produttive.

AMPL è infatti un linguaggio di modellazione per la programmazione matematica che consente di esprimere un problema di ottimizzazione in una forma che sia comprensibile da un qualsiasi risolutore. La flessibilità di AMPL è molto elevata, per cui si possono riprodurre, seguendo semplici procedure, i diversi modelli che caratterizzano l’approccio non parametrico per la stima della frontiera delle possibilità produttive presentati nel Capitolo 4.

Va osservato che in aggiunta ad AMPL sono disponibili altri software di modellazione che svolgono funzioni analoghe, tra i quali si possono ricordare GAMS (*General Algebraic Modeling System*) e Lingo, dei quali sono disponibili in rete versioni per studenti rispettivamente ai siti: <<http://www.gams.com>> e <<http://www.lingo.com>>.

La versione per studenti di AMPL, in grado di gestire problemi con un numero limitato di variabili e di vincoli, si può scaricare dal sito <<http://www.ampl.com>>, nella versione per Windows o per altri sistemi operativi, tra i quali Linux.

Per gli approfondimenti sulla sintassi e sui comandi di AMPL si rimanda al testo di Fourer, Gay e Kernighn (1993).

6.2.2 L'utilizzazione del linguaggio AMPL

Le procedure da seguire per utilizzare AMPL saranno illustrate in riferimento all’analisi dell’efficienza nella formazione universitaria.

La Tabella 6.2 riporta alcune tra le variabili (input e output) utilizzate da Beasley (1990) per la valutazione dell’efficienza dei dipartimenti di fisica e chimica delle università del Regno Unito. In particolare si considerano i soli dipartimenti di fisica, per i quali il processo è descritto da tre input, ossia spese generali (input1); spese per attrezzature (input2); fondi per ricerca (input3) e da tre output, ovvero numero di studenti iscritti a corsi di laurea (output1); numero di studenti iscritti a corsi post-lauream (output2) e numero di laureati impegnati in attività di ricerca (output3).

| Università | Input1 | Input2 | Input3 | Output1 | Output2 | Output3 |
|--------------|--------|--------|--------|---------|---------|---------|
| University1 | 528 | 64 | 254 | 145 | 0 | 26 |
| University2 | 2605 | 301 | 1485 | 381 | 16 | 54 |
| University3 | 304 | 23 | 45 | 44 | 3 | 3 |
| University4 | 1620 | 485 | 940 | 287 | 0 | 48 |
| University5 | 490 | 90 | 106 | 91 | 8 | 22 |
| University6 | 2675 | 767 | 2967 | 352 | 4 | 166 |
| University7 | 422 | 0 | 298 | 70 | 12 | 19 |
| University8 | 986 | 126 | 776 | 203 | 0 | 32 |
| University9 | 523 | 32 | 39 | 60 | 0 | 17 |
| University10 | 585 | 87 | 353 | 80 | 17 | 27 |
| University11 | 931 | 161 | 293 | 191 | 0 | 20 |
| University12 | 1060 | 91 | 781 | 139 | 0 | 37 |
| University13 | 500 | 109 | 215 | 104 | 0 | 19 |
| University14 | 714 | 77 | 269 | 132 | 0 | 24 |
| University15 | 923 | 121 | 392 | 135 | 10 | 31 |
| University16 | 1267 | 128 | 546 | 169 | 0 | 31 |
| University17 | 891 | 116 | 925 | 125 | 0 | 24 |
| University18 | 1395 | 571 | 764 | 176 | 14 | 27 |
| University19 | 990 | 83 | 615 | 28 | 36 | 57 |
| University20 | 3512 | 267 | 3182 | 511 | 23 | 153 |
| University21 | 1451 | 226 | 791 | 198 | 0 | 53 |
| University22 | 1018 | 81 | 741 | 161 | 5 | 29 |
| University23 | 1115 | 450 | 347 | 148 | 4 | 32 |
| University24 | 2055 | 112 | 2945 | 207 | 1 | 47 |
| University25 | 440 | 74 | 453 | 115 | 0 | 9 |
| University26 | 3897 | 841 | 2331 | 353 | 28 | 65 |
| University27 | 836 | 81 | 695 | 129 | 0 | 37 |
| University28 | 1007 | 50 | 98 | 174 | 7 | 23 |
| University29 | 1188 | 170 | 879 | 253 | 0 | 38 |
| University30 | 4630 | 628 | 4838 | 544 | 0 | 217 |
| University31 | 977 | 77 | 490 | 94 | 26 | 26 |
| University32 | 829 | 61 | 291 | 128 | 17 | 25 |
| University33 | 898 | 39 | 327 | 190 | 1 | 18 |
| University34 | 901 | 131 | 956 | 168 | 9 | 50 |
| University35 | 924 | 119 | 512 | 119 | 37 | 48 |
| University36 | 1251 | 62 | 563 | 193 | 13 | 43 |
| University37 | 1011 | 235 | 714 | 217 | 0 | 36 |
| University38 | 732 | 94 | 297 | 151 | 3 | 23 |
| University39 | 444 | 46 | 277 | 49 | 2 | 19 |
| University40 | 308 | 28 | 154 | 57 | 0 | 7 |
| University41 | 483 | 40 | 531 | 117 | 0 | 23 |
| University42 | 515 | 68 | 305 | 79 | 7 | 23 |
| University43 | 593 | 82 | 85 | 101 | 1 | 9 |
| University44 | 570 | 26 | 130 | 71 | 20 | 11 |
| University45 | 1317 | 123 | 1043 | 293 | 1 | 39 |
| University46 | 2013 | 149 | 1523 | 403 | 2 | 51 |
| University47 | 992 | 89 | 743 | 161 | 1 | 30 |
| University48 | 1038 | 82 | 513 | 151 | 13 | 47 |

Tabella 6.2 – Input e output per il processo di produzione dei dipartimenti di fisica (Beasley , 1990).

A partire da tali dati si presenteranno alcuni esempi in relazione alle fasi da seguire per la costruzione di diverse frontiere *best-practice*, caratterizzate da alcune proprietà che ne specificano la rappresentazione nel contesto della teoria della produzione.

Infatti, come emerso dai capitoli precedenti, la specificazione della tecnologia di produzione condiziona le misure di efficienza calcolate come distanza del processo osservato dalla frontiera delle possibilità produttive.

Tra le diverse frontiere di produzione illustrate nel Capitolo 4 si focalizzerà l'attenzione sulla costruzione del modello DEA di base, in riferimento ad una tecnologia convessa, e della frontiera di produzione riferita ad una tecnologia non monotona.

In sintesi il procedimento da seguire si articola in tre fasi principali. La prima fase consiste nello scrivere il modello in un file di testo che ha obbligatoriamente l'estensione *.mod*. Questo file, denominato file di modello, descrive la struttura logica del modello, in termini di indici, variabili, funzione obiettivo e vincoli. In altre parole, esso rappresenta una traduzione del modello teorico specificato per l'analisi della performance delle unità produttive considerate. La seconda fase si estrinseca nella specificazione dei valori numerici del problema in un file di dati, obbligatoriamente di estensione *.dat*. Sebbene sia possibile scrivere in un unico file di testo sia il modello che i dati, si consiglia di mantenere distinti questi due elementi. In tal modo è possibile sostituire le osservazioni, senza dover necessariamente riscrivere il modello, nel caso in cui l'analisi debba essere ripetuta su diversi gruppi di unità produttive utilizzando i medesimi criteri metodologici. L'ultima fase consiste nel risolvere il modello, visualizzando la soluzione ottenuta.

In particolare, in riferimento ad una **tecnologia convessa caratterizzata da rendimenti di scala costanti** (Charnes, Cooper, e Rhodes, 1978) si può determinare la misura di output efficienza radiale, definita come funzione di distanza. Il procedimento di calcolo è esprimibile dal punto di vista teorico attraverso un problema di programmazione lineare che assume la forma seguente (Modello CCR):

$$\begin{aligned}
 & \max \phi_i \\
 & \text{soggetto a :} \\
 & \sum_{h=1}^n y_{rh} \lambda_h \geq y_{ri} \phi_i \quad r = 1, \dots, m \\
 & \sum_{h=1}^n \lambda_h x_{jh} \leq x_{ji} \quad j = 1, \dots, s \\
 & \lambda_h \geq 0; \quad h = 1, \dots, i, \dots, n
 \end{aligned} \tag{6.1}$$

La soluzione conduce alla proiezione radiale sulla frontiera efficiente della i -esima unità produttiva, espressa da $(\mathbf{x}_i, \phi^* \mathbf{y}_i)$, mediante l'espansione del vettore di output che la caratterizza mantenendo fissi gli input impiegati. La misura di efficienza è ottenuta attraverso il rapporto $ET_o^{CCR}(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{\phi^*}$.

La struttura logica del problema precedente viene descritta nel file di testo, illustrato nella Figura 6.1, che può essere denominato, ad esempio, *deaCCR.mod*.

```

set dmus ordered;          #decision making units

set vars;
set inputs within vars;   #inputs utilizzati nella produzione
set outputs within vars;  #outputs ottenuti dal processo

param obs {vars,dmus};
param not {vars};
param inc {dmus};
param ji;
param jj;

var phi >= 0;
var lambda {dmus} >= 0;

maximize efficiency: phi;

s.t. ifront {i in inputs}:
    sum{j in dmus: inc[j] >= ji} obs[i,j]*lambda[j] <= not[i];

s.t. ofront {r in outputs}:
    sum{j in dmus: inc[j] >= ji} obs[r,j]*lambda[j] - not[r]*phi >= 0;

```

Figura 6.1 – Modello per la determinazione dell'output-efficienza CCR. *File: deaCCR.mod.*

Passando alla seconda fase della procedura, si osservi che il file di dati (con estensione *.dat*) può essere costruito secondo l'impostazione riprodotta nella Figura 6.2, nella quale, per esigenze di chiarezza, sono stati indicati i valori assunti dalle variabili solo per alcune unità.

Come si nota, è necessario specificare le unità produttive, indicandone la denominazione, gli input e gli output che descrivono il processo e che costituiscono l'insieme delle variabili (**vars**).

Giungendo al termine della procedura si può osservare che in quest'ultima fase è possibile specificare un particolare solutore, come ad esempio, CPLEX, per la programmazione lineare e intera, e MINOS, per la programmazione non lineare continua.

```

-----
set dmus :=
Uni1
Uni2
Uni3
.
.
.
Uni36
Uni37
.
.
.
Uni49
Uni50
;

set vars := Input1 Input2 Input3 Output1 Output2 Output3;

set inputs :=Input1 Input2 Input3;
set outputs :=Output1 Output2 Output3;

param obs (tr):
      Input1 Input2 Input3 Output1 Output2 Output3:=
Uni1  528    64    254    145    0    26
Uni2  2605   301   1485   381   16   54
Uni3  304    23    45    44    3    3
.      .      .      .      .      .
.      .      .      .      .      .
.      .      .      .      .      .
.      .      .      .      .      .
Uni36 1251    62    563    193   13   43
Uni37 1011   235   714    217    0   36
.      .      .      .      .      .
.      .      .      .      .      .
.      .      .      .      .      .
Uni49  206    1     72    16    0    6
Uni50 1193   95   485    240    0   32
;
param inc :=
Uni1  1
Uni2  1
Uni3  1
.
.
.
Uni36 1
Uni37 1
.
.
.
Uni49 1
Uni50 1
;
param ji := 1;
-----

```

Figura 6.2 – Parte del *file* di dati: *deaDATI.dat*.

Come noto, la procedura di ottimizzazione dovrebbe essere ripetuta 50 volte, ossia un numero di volte uguale a quello dei dipartimenti analizzati. In altre parole, il numero dei problemi da massimizzare è uguale al numero di unità produttive considerate. Un modo per automatizzare la sequenza di ottimizzazioni e produrre i risultati relativi in un unico file di output, consiste nello scrivere un file di comandi con estensione *.run*, che, nell'esempio in esame assume la forma rappresentata in Figura 6.3. Si osservi che le prime due righe di comando richiamano il modello e i dati su cui si sta lavorando.

```
-----
model deaCCR.mod;
data deaDATI.dat;

let {i in inputs} not[i] := obs [i,first(dmus)];
let {r in outputs} not[r] := obs [r,first(dmus)];

if inc[first(dmus)] >= ji then {
display jj;
solve;
printf "%10s%10.5f%10i\n", first(dmus),phi,inc[first(dmus)] >deaCCR.out;
display lambda >deaCCR.out
};

repeat{
let jj := jj + 1;
if jj > card(dmus) then break;

let {i in inputs} not[i] := obs [i,member(jj,dmus)];
let {r in outputs} not[r] := obs [r,member(jj,dmus)];

if inc[member(jj,dmus)] >= ji then {
display jj;
solve;
printf "%10s%10.5f%10i\n", member(jj,dmus),phi,inc[member(jj,dmus)] >deaCCR.out;
display lambda > deaCCR.out
};
};
;
-----
```

Figura 6.3 – File: *deaCCR.run*.

A questo punto si dispone di tutto ciò che è necessario per effettuare l'analisi dell'efficienza delle 50 unità produttive che costituiscono il campione di riferimento. Per eseguire i comandi contenuti in *deaCCR.run* occorre dapprima eseguire l'applicazione *Ampl.exe* e, quindi, digitare il comando: include *deaCCR.run*.

Il file di risultati specificato, che nell'esempio è denominato *deaCCR.out*, conterrà le soluzioni di ottimizzazione del problema [6.1].

Per ciascuna unità produttiva viene indicata la misura di distanza, da cui si deriva quella di efficienza, i valori dei pesi λ_i intervenuti nell'individuazione delle *peers*, o unità di riferimento che specificano la frontiera efficiente “più vicina” all'unità considerata.

Come si evince dalla Figura 6.4, in cui si riportano i risultati ottenuti solo per le prime tre unità considerate nell'esempio, il dipartimento di fisica dell'università1, che rappresenta la prima unità produttiva, è output efficiente nel senso di Farrell, in quanto il valore di ϕ_{hi} è uguale a 1 e tutti i λ_i sono uguali a zero tranne per l'unità considerata, che giace, appunto, sulla frontiera efficiente.

```

      Uni1      1.00000
lambda [*] :=
Uni1 1          Uni21 0          Uni33 0          Uni45 0
Uni10 0         Uni22 0          Uni34 0          Uni46 0
Uni11 0         Uni23 0          Uni35 0          Uni47 0
Uni12 0         Uni24 0          Uni36 0          Uni48 0
Uni13 0         Uni25 0          Uni37 0          Uni49 0
Uni14 0         Uni26 0          Uni38 0          Uni5 0
Uni15 0         Uni27 0          Uni39 0          Uni50 0
Uni16 0         Uni28 0          Uni4 0           Uni6 0
Uni17 0         Uni29 0          Uni40 0          Uni7 0
Uni18 0         Uni3 0           Uni41 0          Uni8 0
Uni19 0         Uni30 0          Uni42 0          Uni9 0
Uni2 0          Uni31 0          Uni43 0
Uni20 0         Uni32 0          Uni44 0
;

      Uni2      1.62675
lambda [*] :=
Uni1 3.6173     Uni21 0          Uni33 0          Uni45 0
Uni10 0         Uni22 0          Uni34 0          Uni46 0
Uni11 0         Uni23 0          Uni35 0.583971  Uni47 0
Uni12 0         Uni24 0          Uni36 0          Uni48 0
Uni13 0         Uni25 0          Uni37 0          Uni49 0
Uni14 0         Uni26 0          Uni38 0          Uni5 0
Uni15 0         Uni27 0          Uni39 0          Uni50 0
Uni16 0         Uni28 0          Uni4 0           Uni6 0
Uni17 0         Uni29 0          Uni40 0          Uni7 0.368423
Uni18 0         Uni3 0           Uni41 0          Uni8 0
Uni19 0         Uni30 0          Uni42 0          Uni9 0
Uni2 0          Uni31 0          Uni43 0
Uni20 0         Uni32 0          Uni44 0
;

      Uni3      1.19439
lambda [*] :=
Uni1 0.0204143 Uni21 0          Uni33 0          Uni45 0
Uni10 0         Uni22 0          Uni34 0          Uni46 0
Uni11 0         Uni23 0          Uni35 0          Uni47 0
Uni12 0         Uni24 0          Uni36 0          Uni48 0
Uni13 0         Uni25 0          Uni37 0          Uni49 0
Uni14 0         Uni26 0          Uni38 0          Uni5 0.110902
Uni15 0         Uni27 0          Uni39 0          Uni50 0
Uni16 0         Uni28 0.200674 Uni4 0           Uni6 0
Uni17 0         Uni29 0          Uni40 0          Uni7 0
Uni18 0         Uni3 0           Uni41 0          Uni8 0
Uni19 0         Uni30 0          Uni42 0          Uni9 0
Uni2 0          Uni31 0          Uni43 0
Uni20 0         Uni32 0          Uni44 0.0645625
;

```

Figura 6.4 – Parte del file di risultati: *deaCCR.out*.

Per la seconda unità considerata, l'efficienza tecnica si ottiene attraverso il rapporto $ET_0^{CCR}(\mathbf{x}_2, \mathbf{y}_2) = \frac{1}{1,62675} = 0,615$. La frontiera efficiente, rispetto alla quale la DMU uni2 è stata valutata, viene definita dalle unità 1, 35 e 7 che rappresentano, quindi, le cosiddette *peers*.

Per quanto riguarda la determinazione degli *slack*, negli input o negli output, e, quindi, il passaggio dalla misura radiale di Farrell, determinata rispetto all'isoquanto, alla misura definita rispetto al sottoinsieme efficiente, che riproduce il concetto di Pareto - Koopmans, si può utilizzare il modello illustrato nella Figura 6.5.

```

set dmus ordered;          #decision making units

set vars;
set inputs within vars;    #inputs to production
set outputs within vars;   #outputs to production

param obs {vars,dmus};
param not {vars};
param inc {dmus};
param ji;
param jj;

var lambda {dmus} >= 0;
var alfa {inputs} >= 0;
var beta {outputs}>= 0;
var phi;

maximize efficiency: phi;

maximize slack: sum{i in inputs} alfa[i] + sum{r in outputs} beta[r];

s.t. ifront {i in inputs}:
    sum{j in dmus: inc[j] >= ji} obs[i,j]*lambda[j] + alfa[i] = not[i];

s.t. ofront {r in outputs}:
    sum{j in dmus: inc[j] >= ji} obs[r,j]*lambda[j] - beta[r] = not[r]*phi;

```

Figura 6.5 – Modello per la determinazione degli *slack*. File: *slack.mod*.

I risultati, raggruppati nel file *slack.out*, mostrano, per ciascuna unità produttiva, nel caso siano presenti, sia gli *slack* negli input, raggruppati nel vettore *alfa*, sia quelli negli output, inclusi nel vettore denominato *beta*.

La Figura 6.6 riproduce i risultati ottenuto dalla soluzione del programma descritto nel modello *slack.mod* solo per i primi quattro dipartimenti considerati nell'esempio.

```

      Uni1   1.00000
alfa [*] :=
;

beta [*] :=
;

      Uni2   1.62675
alfa [*] :=
Input3  157.422
;

beta [*] :=
Output3  41.2361
;

      Uni3   1.19439
alfa [*] :=
;

beta [*] :=
Output3  4.71312
;

      Uni4   1.55013
alfa [*] :=
Input2   288.636
Input3   160.682
;

beta [*] :=
Output3   5.36665
;

```

Figura 6.6 – Parte del file di risultati: *slack.out*.

Seguendo un procedimento analogo al precedente è possibile determinare la misura di input-efficienza tecnica, ottenuta cercando la contrazione massima del vettore degli input mantenendo fissi gli output prodotti.

Va osservato che la scelta dell'orientamento può essere determinata da particolari esigenze, dettate talvolta dalle caratteristiche dei processi produttivi di cui si vuole misurare l'efficienza.

D'altra parte, nel caso in cui la tecnologia presenti rendimenti di scala costanti, la misura di output efficienza come già anticipato coincide con quella di input efficienza, per cui la scelta dell'orientamento non appare discriminante.

Risultati divergenti si ottengono invece nel caso in cui la tecnologia sia caratterizzata dalla presenza di rendimenti di scala variabili (VRS) per cui occorre prestare notevole attenzione alla scelta della "direzione" rispetto la quale si vuole impostare il problema di ottimizzazione.

A tale proposito si illustrerà l'implementazione del modello introdotto da Banker, Charnes e Cooper (1984). La Figura 6.7 riproduce il modello DEA con rendimenti di scala variabili orientato nel senso degli input, per cui il problema consiste nell'individuazione del valore di q che riduce proporzionalmente tutti gli input impiegati mantenendo la capacità di produrre le date quantità di output.

Da notare che l'ultima riga riproduce il vincolo sulla somma dei pesi, che deve, per i rendimenti di scala variabili, essere uguale ad uno.

```
-----
set dmus ordered;          #decision making units

set vars;
set inputs within vars;   #inputs to production
set outputs within vars;  #outputs to production

param obs {vars,dmus};
param not {vars};
param inc {dmus};
param ji;
param jj;

var theta >= 0;
var lambda {dmus} >= 0;

minimize efficiency: theta;

s.t. ifront {i in inputs}:
    sum{j in dmus: inc[j] >= ji} obs[i,j]*lambda[j] <= not[i]*theta;

s.t. ofront {r in outputs}:
    sum{j in dmus: inc[j] >= ji} obs[r,j]*lambda[j] >= not[r];

s.t. convex:
    sum{j in dmus: inc[j] >= ji} lambda[j] = 1;
-----
```

Figura 6.7 – Modello per la determinazione dell'input efficienza VRS. *File: deaBCC.mod.*

Anche in questo caso, il file per la risoluzione del modello con estensione *.run*, che consente la ripetizione automatica dei singoli problemi di massimizzazione per ciascuna unità produttiva, riproduce esattamente quello illustrato nella Figura 6.3. Si dovrà semplicemente cambiare nella prima riga la denominazione del modello, che ora diviene *deaBCC.mod*, e sostituire nell'ottava riga dei comandi un nuovo file in cui saranno raccolti i risultati, che, ad esempio, può essere denominato *deaBCC.out*.

```

Uni1 1.00000 1
lambda [*] :=
Uni1 1 Uni25 0 Uni40 0
Uni10 0 Uni26 0 Uni41 0
Uni11 0 Uni27 0 Uni42 0
Uni12 0 Uni28 0 Uni43 0
Uni13 0 Uni29 0 Uni44 0
Uni14 0 Uni3 0 Uni45 0
Uni15 0 Uni30 0 Uni46 0
Uni16 0 Uni31 0 Uni47 0
Uni17 0 Uni32 0 Uni48 0
Uni18 0 Uni33 0 Uni49 0
Uni19 0 Uni34 0 Uni5 0
Uni2 0 Uni35 0 Uni50 0
Uni20 0 Uni36 0 Uni6 0
Uni21 0 Uni37 0 Uni7 0
Uni22 0 Uni38 0 Uni8 0
Uni23 0 Uni39 0 Uni9 0
Uni24 0 Uni4 0
;

Uni2 1.00000 1
lambda [*] :=
Uni1 0 Uni25 0 Uni40 0
Uni10 0 Uni26 0 Uni41 0
Uni11 0 Uni27 0 Uni42 0
Uni12 0 Uni28 0 Uni43 0
Uni13 0 Uni29 0 Uni44 0
Uni14 0 Uni3 0 Uni45 0
Uni15 0 Uni30 0 Uni46 0
Uni16 0 Uni31 0 Uni47 0
Uni17 0 Uni32 0 Uni48 0
Uni18 0 Uni33 0 Uni49 0
Uni19 0 Uni34 0 Uni5 0
Uni2 1 Uni35 0 Uni50 0
Uni20 0 Uni36 0 Uni6 0
Uni21 0 Uni37 0 Uni7 0
Uni22 0 Uni38 0 Uni8 0
Uni23 0 Uni39 0 Uni9 0
Uni24 0 Uni4 0
;

Uni3 1.00000 1
lambda [*] :=
Uni1 0 Uni25 0 Uni40 0
Uni10 0 Uni26 0 Uni41 0
Uni11 0 Uni27 0 Uni42 0
Uni12 0 Uni28 0 Uni43 0
Uni13 0 Uni29 0 Uni44 0
Uni14 0 Uni3 1 Uni45 0
Uni15 0 Uni30 0 Uni46 0
Uni16 0 Uni31 0 Uni47 0
Uni17 0 Uni32 0 Uni48 0
Uni18 0 Uni33 0 Uni49 0
Uni19 0 Uni34 0 Uni5 0
Uni2 0 Uni35 0 Uni50 0
Uni20 0 Uni36 0 Uni6 0
Uni21 0 Uni37 0 Uni7 0
Uni22 0 Uni38 0 Uni8 0
Uni23 0 Uni39 0 Uni9 0
Uni24 0 Uni4 0
;

```

Figura 6.8 – Parte del file di risultati: *deaBCC.out*.

La Figura 6.8 riproduce i risultati della stima della frontiera non parametrica con rendimenti variabili per le prime tre unità produttive, al fine di effettuare un confronto con quanto emerso nel caso di rendimenti di scala costanti, riportati nella Figura 6.4.

È interessante osservare che il dipartimento di fisica dell'università 2 e quello dell'università 3 risultano in questo caso entrambi pienamente efficienti nel senso di Farrell. Infatti, la frontiera VRS, essendo "più vicina" alle unità osservate, produce misure di efficienza radiali maggiori, o al limite uguali, a quelle ottenibili dalla frontiera a rendimenti di scala costanti.

Per illustrare il procedimento che necessita dell'utilizzazione di AMPL, in riferimento all'analisi dell'efficienza rispetto a tecnologie non monotone (Färe, Grosskopf e Lovell, 1985), ossia caratterizzate dalla rimozione dell'ipotesi di disponibilità forte degli input e degli output, si può considerare il problema seguente, orientato nel senso degli output:

$$\begin{aligned}
 & \max \phi_i \\
 & \text{soggetto a :} \\
 & \sum_{h=1}^n y_{rh} \lambda_h \geq y_{ri} \phi_i \quad r = 1, \dots, m \\
 & \sum_{h=1}^n \lambda_h x_{jh} = x_{ji} \quad j = 1, \dots, s \\
 & \lambda_h \geq 0; \quad h = 1, \dots, i, \dots, n
 \end{aligned} \tag{6.2}$$

Le proprietà considerate nel problema [6.2] sono relative a rendimenti di scala costanti, disponibilità forte degli output e disponibilità debole degli input.

Il file *weakinp.mod*, riprodotto nella Figura 6.9, descrive la struttura logica del modello precedente.

Il file di dati e quello dei comandi assumono la medesima forma degli esempi precedenti. I risultati che si ottengono sono riportati per le medesime unità produttive considerate nel modello CCR nella Figura 6.10.

È possibile determinare una misura di **congestione** negli input (Färe e Grosskopf, 1983a, 1998) attraverso il rapporto tra le precedenti misure di efficienza, ottenute risolvendo il problema [6.1] e [6.2].

Ad esempio per la seconda unità produttiva considerata si determinano le misure di efficienza riferite alla tecnologia con rendimenti di scala costanti (Figura 6.4), nella quale si considera dapprima la proprietà di disposizione forte degli input e poi la versione debole. Utilizzando la simbologia adottata nel paragrafo 4.3 si ha:

```

set dmus ordered;          #decision making units

set vars;
set inputs within vars;    #inputs to production
set outputs within vars;  #outputs to production

param obs {vars,dmus};
param not {vars};
param inc {dmus};
param ji;
param jj;

var phi >= 0;
var lambda {dmus} >= 0;

maximize efficiency: phi;

s.t. ifront {i in inputs}:
    sum{j in dmus: inc[j] >= ji} obs[i,j]*lambda[j] = not[i];

s.t. ofront {r in outputs}:
    sum{j in dmus : inc[j] >= ji} obs[r,j]*lambda[j] - not[r]*phi >= 0;

```

Figura 6.9 – File: *weakinp.mod*.

$$F^o(\mathbf{x}_2, \mathbf{y}_2) = \frac{1}{1,62675} = 0,615 \quad W^o(\mathbf{x}_2, \mathbf{y}_2) = \frac{1}{1,61786} = 0,618$$

Quindi, la misura di congestione per l'unità considerata può derivarsi (Färe, Grosskopf e Lovell, 1994) attraverso il rapporto:

$$C(\mathbf{x}_4, \mathbf{y}_4) = \frac{F^o(\mathbf{x}_2, \mathbf{y}_2)}{W^o(\mathbf{x}_2, \mathbf{y}_2)} = \frac{0,615}{0,618} = 0,995$$

È interessante osservare che la terza unità (Uni3) non presenta congestione negli input (le due misure precedenti sono identiche) sebbene essa non risulti comunque pienamente efficiente.

Sulla base di quanto emerso da questa breve guida per l'implementazione delle metodologie non parametriche per la stima dell'efficienza tecnica, si può affermare che l'utilizzo di AMPL consente la massima flessibilità. E' possibile scrivere il modello che si desidera, introducendo vincoli aggiuntivi, come, ad esempio, il vincolo necessario per considerare tecnologie caratterizzate da rendimenti di scala non crescenti, in cui si impone che la somma dei pesi sia inferiore o uguale a uno, oppure specificando ulteriori variabili, come ad esempio variabili cosiddette *non-discretionary*.

Ovviamente anche la scelta di determinare la misura di efficienza radiale, contraendo il vettore degli input (*input-oriented*), o espandendo il vettore degli output (*output-oriented*) si traduce, come visto, in semplici modifiche nel file del modello.

```

-----
Uni1 1.00000
lambda [*] :=
Uni1 1 Uni25 0 Uni40 0
Uni10 0 Uni26 0 Uni41 0
Uni11 0 Uni27 0 Uni42 0
Uni12 0 Uni28 0 Uni43 0
Uni13 0 Uni29 0 Uni44 0
Uni14 0 Uni3 0 Uni45 0
Uni15 0 Uni30 0 Uni46 0
Uni16 0 Uni31 0 Uni47 0
Uni17 0 Uni32 0 Uni48 0
Uni18 0 Uni33 0 Uni49 0
Uni19 0 Uni34 0 Uni5 0
Uni2 0 Uni35 0 Uni50 0
Uni20 0 Uni36 0 Uni6 0
Uni21 0 Uni37 0 Uni7 0
Uni22 0 Uni38 0 Uni8 0
Uni23 0 Uni39 0 Uni9 0
Uni24 0 Uni4 0
;

Uni2 1.61786
lambda [*] :=
Uni1 3.08424 Uni21 0 Uni33 0 Uni45 0
Uni10 0 Uni22 0 Uni34 0 Uni46 0
Uni11 0 Uni23 0 Uni35 0.506892 Uni47 0
Uni12 0 Uni24 0 Uni36 0 Uni48 0
Uni13 0 Uni25 0.58498 Uni37 0 Uni49 0
Uni14 0 Uni26 0 Uni38 0 Uni5 0
Uni15 0 Uni27 0 Uni39 0 Uni50 0
Uni16 0 Uni28 0 Uni4 0 Uni6 0
Uni17 0 Uni29 0 Uni40 0 Uni7 0.594223
Uni18 0 Uni3 0 Uni41 0 Uni8 0
Uni19 0 Uni30 0 Uni42 0 Uni9 0
Uni2 0 Uni31 0 Uni43 0
Uni20 0 Uni32 0 Uni44 0
;

Uni3 1.19439
lambda [*] :=
Uni1 0.0204143 Uni21 0 Uni33 0 Uni45 0
Uni10 0 Uni22 0 Uni34 0 Uni46 0
Uni11 0 Uni23 0 Uni35 0 Uni47 0
Uni12 0 Uni24 0 Uni36 0 Uni48 0
Uni13 0 Uni25 0 Uni37 0 Uni49 0
Uni14 0 Uni26 0 Uni38 0 Uni5 0.110902
Uni15 0 Uni27 0 Uni39 0 Uni50 0
Uni16 0 Uni28 0.200674 Uni4 0 Uni6 0
Uni17 0 Uni29 0 Uni40 0 Uni7 0
Uni18 0 Uni3 0 Uni41 0 Uni8 0
Uni19 0 Uni30 0 Uni42 0 Uni9 0
Uni2 0 Uni31 0 Uni43 0
Uni20 0 Uni32 0 Uni44 0.0645625
;
-----

```

Figura 6.10 Parte del file di risultati: *weak.out*.

Si segnala infine l'esistenza di un sito internet estremamente interessante e utile, denominato *Neos server for Optimization (NEOS)*, il cui indirizzo è il seguente: <<http://neos.mcs.anl.gov/neos/>> . Seguendo le procedure indicate è possibile sottoporre, gratuitamente, ad un risolutore, selezionato tra i molteplici disponibili in elenco, problemi di programmazione lineare di tipo DEA.

Le opzioni presenti sono molteplici, i file necessari sono il file del modello (con estensione *.mod*), il file dei dati e quello dei comandi, dal quale è necessario eliminare solo le prime due righe nelle quali venivano specificati il modello e i dati, che vanno inviati direttamente a NEOS.

6.3 LA STIMA DELLE FRONTIERE DI PRODUZIONE STOCASTICHE

6.3.1 I programmi disponibili

Le frontiere stocastiche possono essere stimate utilizzando sia programmi econometrici generali, come LIMDEP (Green, 1995), Shazam, GAUSS, SAS, STATA, sia *software* specificatamente formulati, come FRONTIER 4.1.

La Tabella 6.3 riassume sinteticamente alcune caratteristiche, in relazione ai modelli stocastici che possono essere stimati, dei programmi più frequentemente utilizzati nelle applicazioni per la stima parametrica delle frontiere di produzione, in particolare FRONTIER 4.1 (Coelli, 1996) è scaricabile gratuitamente dal sito <<http://www.uq.edu.au/economics/cepa/frontier.htm>>).

Nel rinviare a Sena (1999) per gli approfondimenti relativi alle caratteristiche tecniche dei due software posti a confronto dall'Autore, ossia LIMPEP 7.0 e FRONTIER 4.1, si può in questa sede rilevare che gli elementi a favore del software FRONTIER 4.1 sono rappresentati dal maggior numero di modelli per dati *panel* che consente di stimare.

In particolare, utilizzando FRONTIER 4.1 è possibile stimare agevolmente il modello di Battese e Coelli (1995) dove il termine di efficienza è spiegato attraverso un vettore di variabili esogene a livello di singola unità produttiva.

Per quanto riguarda invece l'analisi dei dati *cross-section*, FRONTIER 4.1 può stimare frontiere sia di produzione che di costo al pari di LIMDEP e STATA, ma senza la possibilità di utilizzare la v.c. esponenziale per descrivere il termine di errore unilaterale.

È opportuno osservare che il programma FRONTIER 4.1 è accompagnato da un manuale (Coelli, 1996) che guida l'interessato nell'utilizzazione del programma, fornendo anche esempi applicativi molto utili per apprendere il funzionamento.

| | LIMDEP 7.0 | FRONTIER 4.1 | STATA 8.1 |
|---|---------------|-----------------|--------------|
| Funzione di produzione su dati cross-section | | | |
| Distribuzione per u metà normale | SI | SI | SI |
| Distribuzione per u Normale troncata | SI | SI | SI |
| Distribuzione per u Esponenziale | SI | NO | SI |
| <i>Introduzione effetto variabili esogene</i> | | | |
| Attraverso parametrizzazione media | SI | SI | SI |
| <i>Funzione di produzione su dati panel</i> | | | |
| invarianza nel tempo dell'efficienza | | | |
| Distribuzione per u metà normale | SI | SI | NO |
| Distribuzione per u Normale troncata | SI | SI | SI |
| efficienza variabile nel tempo | | | |
| Distribuzione per u metà normale | NO | SI | NO |
| Distribuzione per u Normale troncata | NO | SI | SI |

Tabella 6.3 – Caratteristiche di alcuni software per l'applicazione di modelli parametrici.

Va osservato che il testo di Green (2000) include una versione per studenti di LIMDEP 7.0. Le limitazioni di questa versione, denominata EA/LimDep, sono relative alla dimensione del *data set* che è possibile analizzare, al numero dei parametri che possono essere inclusi nel modello, e altre procedure utilizzabili. Nel manuale e negli esempi riportati nel CD rom allegato, sono illustrate le procedure da seguire per riprodurre l'applicazione dei modelli della frontiera stocastica trattati nel testo.

Si è ritenuto perciò di maggiore interesse mostrare come realizzare praticamente la maggior parte delle analisi illustrate nel Capitolo 3, utilizzando il programma Stata 9.0, nella versione Intercooled Stata.

6.3.2 L'utilizzazione del programma Intercooled Stata 9.0

Per riprodurre il modello della frontiera stocastica per dati *cross-section* espresso da $\ln y_i = \ln f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i - u_i$ in Stata 9.0 è necessario procedere come segue.

Per prima cosa, devono essere svolte le operazioni preliminari di apertura dell'appropriato file di dati, attraverso il comando: *use "C:nome del file.dta"*.

In tal modo si chiede a Stata di “aprire” il file di dati cancellando qualsiasi altro file di dati eventualmente già presente in memoria.

Al fine di illustrare i diversi modelli della frontiera stocastica la cui stima è ottenibile tramite Stata, si è deciso di considerare solo una parte dei dati utilizzati in Ferrari e Laureti (2004b) e, in particolare, i dati relativi ai laureati nella facoltà di Economia dell’Ateneo di Firenze nell’anno 1998 riportati nella Tabella 6.4. Il file di riferimento è denominato, quindi, *economia.dta*.

Come esposto nel capitolo precedente, l’unità produttiva di riferimento è rappresentata dal singolo studente che, utilizzando risorse offerte dalla facoltà in cui è iscritto e risorse derivanti dal proprio impegno personale, produce se stesso come laureato.

È importante sottolineare che per ottenere un file di dati leggibile da Stata 9.0 si può utilizzare il *software* Stata Transfer che consente di trasformare qualsiasi tipo di file, ad esempio un file excel, in un file con estensione *.dta*.

Per l’applicazione dell’approccio stocastico è di estrema importanza la corretta specificazione della forma funzionale per caratterizzare la tecnologia di produzione $f(x_i; \beta)$, che può essere espressa da qualsiasi funzione lineare o linearizzabile, come la Cobb-Douglas o la translog. La selezione di una particolare funzione comporta, quindi, la predisposizione nel file dei dati delle variabili trasformate da utilizzare nella stima.

Si può passare a questo punto alla scelta della v.c. che descrive il termine di inefficienza u . Questa fase è particolarmente delicata in quanto le stime dell’efficienza tecnica sono sensibili al tipo di specificazione effettuata.

L’ipotesi usuale sul termine di errore puramente stocastico v_i è che sia descritto da una v.c. normale con media nulla e varianza costante².

Se si ipotizza di scegliere una funzione di produzione Cobb-Douglas, che con tre input assume la forma $f(x_1, x_2, x_3) = Ax_1^\alpha x_2^\beta x_3^\gamma$, è indispensabile passare ai logaritmi creando specifiche variabili nel file di dati “*economia.dta*”.

Nel caso in cui il termine u_i è descritto mediante una v.c. metà normale non è necessario specificare alcun comando aggiuntivo, in quanto quest’ultima rappresenta l’ipotesi di *default*.

Per scopi puramente illustrativi, si selezionano i seguenti tre input: i) attrezzature per studente e ii) libri per studente, che esprimono le risorse offerte dalla facoltà, e infine iii) voto di maturità, che rappresenta l’abilità di ciascun studente in entrata. L’output del processo è rappresentato dalla media dei voti agli esami.

² Si noti che è possibile prevedere con Stata anche l’ipotesi di eteroschedasticità del termine simmetrico v_i .

Nella Tabella 6.4, per facilitare la lettura, sono riportati i valori assunti dalle variabili precedenti solo per alcuni dei 388 studenti dell'Ateneo fiorentino che hanno conseguito la laurea in Economia nell'anno 1998.

| Laureato | Durata anni | Voto medio | docenti | attrezzature | libri | periodici | voto diploma |
|----------|-------------|------------|----------|--------------|----------|-----------|--------------|
| 1 | 7,451140 | 26,12 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 57 |
| 2 | 4,467580 | 24,42 | 0,021955 | 0,103806 | 32,93611 | 0,124293 | 60 |
| 3 | 5,464840 | 28,47 | 0,021396 | 0,096775 | 32,26231 | 0,124168 | 60 |
| 7 | 9,268040 | 25,24 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 52 |
| 8 | 7,442920 | 24,41 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 45 |
| 9 | 8,445660 | 23,40 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 44 |
| 10 | 7,456620 | 27,32 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 49 |
| 11 | 8,265300 | 26,08 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 50 |
| 12 | 6,462100 | 27,26 | 0,020903 | 0,090971 | 31,84685 | 0,125256 | 56 |
| 67 | 6,456620 | 24,41 | 0,020903 | 0,090971 | 31,84685 | 0,125256 | 38 |
| 68 | 9,462100 | 24,88 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 42 |
| 69 | 8,456620 | 24,48 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 41 |
| 70 | 9,265300 | 29,04 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 52 |
| 71 | 8,462100 | 25,11 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 48 |
| 80 | 9,456620 | 26,84 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 46 |
| 81 | 8,270780 | 26,04 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 42 |
| 82 | 7,464840 | 26,48 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 47 |
| 83 | 8,462100 | 26,76 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 36 |
| 195 | 6,117810 | 26,04 | 0,020903 | 0,090971 | 31,84685 | 0,125256 | 50 |
| 196 | 8,115070 | 26,16 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 52 |
| 197 | 7,117810 | 26,65 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 46 |
| 198 | 9,101370 | 25,37 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 37 |
| 199 | 7,106850 | 23,52 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 40 |
| 200 | 7,886990 | 24,42 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 52 |
| 298 | 7,884250 | 26,47 | 0,02042 | 0,080293 | 31,7781 | 0,130969 | 46 |
| 299 | 7,106850 | 27,91 | 0,020727 | 0,085496 | 31,85008 | 0,128334 | 36 |
| 308 | 10,448400 | 26,07 | 0,020193 | 0,075954 | 31,80358 | 0,133597 | 37 |
| 309 | 7,109590 | 26,48 | 0,02072 | 0,085496 | 31,85008 | 0,28334 | 54 |
| 310 | 4,881510 | 27,65 | 0,021396 | 0,096775 | 32,26231 | 0,124168 | 60 |
| 311 | 5,967580 | 23,79 | 0,020903 | 0,090971 | 31,84685 | 0,125256 | 37 |

Tabella 6.4 – Variabili utilizzate per la descrizione del processo di formazione universitaria (laureati nella Facoltà di Economia).

Per stimare il modello prescelto si utilizza il comando *frontier* specificando in un'unica riga dapprima il prodotto del processo e poi a seguire i diversi input.

consente di ottenere utilizzando le informazioni derivanti dalla distribuzione condizionata $f(u|\varepsilon)$.

In particolare, come già evidenziato, la stima puntuale di ciascun termine di inefficienza u_i , ossia \hat{u}_i , da sostituire nell'espressione $ET_i = \exp\{-\hat{u}_i\}$, può essere ottenuta utilizzando come stimatore la moda $M(u_i|\varepsilon_i)$, secondo l'espressione [3.17], che qui si riporta per facilitare la lettura si ha:

$$M(u_i|\varepsilon_i) = -\varepsilon_i \left(\frac{\sigma_u^2}{\sigma^2} \right) \quad \text{se } \varepsilon_i \leq 0$$

$$M(u_i|\varepsilon_i) = 0 \quad \text{se } \varepsilon_i > 0$$

Più frequentemente si ricorre alla media $E(u_i|\varepsilon_i)$ della distribuzione condizionata di u dato ε , espressa da:

$$E(u_i|\varepsilon_i) = \sigma \cdot \left[\frac{\phi(\varepsilon_i \lambda / \sigma)}{1 - \Phi(\varepsilon_i \lambda / \sigma)} - \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right]$$

A tal fine si deve utilizzare il comando *. predict*, contenuto nell'insieme di procedure denominate *frontier post estimation*, attraverso il quale si crea un vettore i cui valori, salvati automaticamente nel file di dati, sono specificati dalla seconda parte del comando.

In particolare se si vuole stimare direttamente la misura di efficienza tecnica, utilizzando lo stimatore $ET_i = E(\exp\{-u_i\}|\varepsilon_i)$ suggerito da Battese e Coelli (1988), si può scrivere: **. predict Technical_efficiency, te**. In tal modo, le misure di efficienza tecnica saranno raccolte nel vettore *Technical_efficiency*.

Se, invece, si è interessati alla stima dell'efficienza tecnica utilizzando la moda come stimatore dell'inefficienza è necessario innanzitutto ottenere la stima \hat{u}_i , scrivendo il comando *. predict "nome variabile", u* nel quale può essere specificato il nome della variabile che sarà creata nel file di dati.

Nel caso in cui la scelta ricada sulla media condizionata, la stima della componente di inefficienza per ciascuna unità produttiva, è ottenibile utilizzando il comando *. predict nome variabile, m*.

Dopo aver ottenuto le stime di u per ciascuna unità produttiva si può derivare la misura di efficienza tecnica attraverso $ET_i = \exp\{-\hat{u}_i\}$ facilmente calcolabile in Stata 9.0 attraverso il comando: *generate float ET = exp(-Media)* nel quale *Media* rappresenta il nome assegnato al vettore che raccoglie le stime dell'inefficienza.

Si ritiene importante osservare che, come già anticipato, i risultati ottenuti in termini di efficienza tecnica possono differire applicando uno stimatore piuttosto che un altro. La Tabella 6.5 riporta, per il gruppo di laureati considerati nell'esempio, i risultati ottenuti dalle procedure precedenti, che come anticipato, vanno ad integrare il file di dati *economia.dta*.

| Laureato | Efficienza tecnica $ET_i = E(\exp\{-u_i\} \varepsilon_i)$ | Efficienza tecnica $ET_i = \exp\{-[E(u_i \varepsilon_i)]\}$ | Efficienza tecnica $ET_i = \exp\{-[M(u_i \varepsilon_i)]\}$ |
|----------|--|--|--|
| 1 | 0,93244 | 0,93217 | 0,93232 |
| 2 | 0,83443 | 0,83419 | 0,83419 |
| 3 | 0,96914 | 0,96896 | 0,97573 |
| 7 | 0,92760 | 0,90820 | 0,90820 |
| 8 | 0,90829 | 0,94535 | 0,94605 |
| 9 | 0,88503 | 0,93597 | 0,93621 |
| 10 | 0,97656 | 0,92732 | 0,92740 |
| 11 | 0,95367 | 0,90802 | 0,90803 |
| 12 | 0,96133 | 0,88477 | 0,88477 |
| 67 | 0,92433 | 0,92406 | 0,92411 |
| 68 | 0,94274 | 0,94247 | 0,94299 |
| 69 | 0,92908 | 0,92881 | 0,92890 |
| 70 | 0,99064 | 0,99061 | 1,00000 |
| 71 | 0,92955 | 0,92928 | 0,92938 |
| 80 | 0,97876 | 0,97864 | 0,99456 |
| 81 | 0,96975 | 0,96958 | 0,97672 |
| 82 | 0,96435 | 0,96414 | 0,96856 |
| 83 | 0,98664 | 0,98658 | 1,00000 |
| 195 | 0,94125 | 0,94099 | 0,94142 |
| 196 | 0,95144 | 0,95119 | 0,95248 |
| 197 | 0,96980 | 0,96963 | 0,97680 |
| 198 | 0,96916 | 0,96898 | 0,97575 |
| 199 | 0,89379 | 0,89353 | 0,89353 |
| 200 | 0,89766 | 0,89739 | 0,89739 |
| 298 | 0,97080 | 0,97063 | 0,97846 |
| 299 | 0,99109 | 0,99106 | 1,00000 |
| 308 | 0,98047 | 0,98036 | 0,99901 |
| 309 | 0,94959 | 0,94933 | 0,95041 |
| 310 | 0,95043 | 0,95018 | 0,95135 |
| 311 | 0,90731 | 0,90704 | 0,90705 |

Tabella 6.5 – Stime relative al termine di efficienza.

Per stimare in Stata il modello suggerito da Stevenson (1980), che generalizzando il modello di Aigner, Lovell e Schmidt (1977) descrive il termine di inefficienza mediante una v.c. normale troncata, è sufficiente indicare nel comando *frontier*, al termine dell'elenco degli input considerati, il tipo di distribuzione selezionata, attraverso l'espressione *distribuzione (tnormal)*.

La Figura 6.12 riproduce il comando e i risultati che derivano dall'applicazione del modello precedente.

universitaria, dalle caratteristiche individuali degli studenti, possono essere introdotte nell'analisi utilizzando Stata 9.0 seguendo gli approcci suggeriti in letteratura.

Una prima soluzione, come emerso dalla trattazione precedente, consiste nella riproduzione del modello di Battese e Coelli (1995), che si estrinseca nell'ipotizzare una v.c. normale troncata per il termine u_i , ossia $N(m_i, \sigma_u^2)$, la cui media è funzione di un insieme di variabili esplicative, ossia $m_i = \mathbf{z}_i \boldsymbol{\delta}$, in cui \mathbf{z}_i è il vettore di variabili associate all'efficienza tecnica e $\boldsymbol{\delta}$ è il vettore di coefficienti che devono essere stimati.

Il comando *frontier* si modifica introducendo le variabili esplicative, che a fini puramente illustrativi sono specificate dal genere, dal tipo di diploma, nelle categorie liceo classico e scientifico, liceo linguistico, istituto commerciale e istituto tecnico, e da una variabile *dummy* che indica se lo studente ha svolto attività lavorativa durante gli studi.

A tal fine si aggiunge al comando l'espressione: *cm(femmina linguistico commerciale tecnico lavoro)*

Nella Figura 6.13 è riprodotto l'output di Stata 9.0, dal quale possono trarsi informazioni sull'effetto esercitato dalle variabili introdotte come esplicative.

Il programma consente, inoltre, di introdurre le caratteristiche individuali degli studenti considerandole come variabili esplicative della varianza del termine che descrive l'inefficienza tecnica.

Nel modello di frontiera stocastica in cui il termine u è descritto da una v.c. metà normale si può ipotizzare che la varianza non sia costante tra le unità produttive (Laureti, 2006). Si è in presenza di eteroschedasticità del termine di inefficienza.

Per rimuovere l'ipotesi di omoschedasticità, in riferimento alle medesime variabili considerate nell'esempio precedente, è necessario utilizzare il comando: *uhet(femmina linguistico commerciale tecnica lavoro)*.

La Figura 6.14 riproduce i comandi e alcuni risultati relativi alla stima del modello eteroscedastico. La funzione di log-verosimiglianza, che costituisce una generalizzazione dell'espressione suggerita da Aigner, Lovell and Schmidt (1977), assume in tal caso la forma seguente:

$$\ln L(y|\boldsymbol{\beta}, \lambda, \sigma_v^2, \boldsymbol{\delta}) = \text{constant} - \frac{1}{2} \sum_i \ln [g(\mathbf{z}_i; \boldsymbol{\delta}) + \sigma_v^2] + \sum_i \ln \Phi \left(-\frac{\varepsilon_i \lambda_i}{\sigma_i} \right) - \frac{1}{2} \sum_i \frac{\varepsilon_i^2}{\sigma_i^2}$$

dove $\sigma_{ii}^2 = g(\mathbf{z}_i; \boldsymbol{\delta})$ e $\sigma_{ii}^2 = g(\mathbf{z}_i; \boldsymbol{\delta})$ può essere specificata nella forma moltiplicativa $\sigma_i^2 = \exp(\mathbf{z}_i \boldsymbol{\delta})$ nella quale il vettore \mathbf{z} include un termine costante.

```

-----
      tm
-----
  Statistics/Data Analysis 9.0 Copyright 1984-2005
                             StataCorp
                             4905 Lakeway Drive
                             College Station, Texas

> 77845 USA                               800-STATA-PC           ht
> tp://www.stata.com                       979-696-4600          st
> ata@stata.com                            979-696-4601 (fax)

Single-user Stata for Windows perpetual license:
  Serial number: 1990519724
  Licensed to: Tiziana laureti
              Tuscia

Notes:
  1. (/m# option or -set memory-) 1.00 MB allocated t
> o data

1 . use "C:\economia.dta", clear
2 . frontier media_voti attrezzature libri voto_diploma, distribution(tnormal) cm(femmina linguist
> ale tecnica lavoro limitrofe lontane)

Iteration 0: log likelihood = 638.09138 (not concave)
Iteration 1: log likelihood = 651.34995
Iteration 2: log likelihood = 656.6751
Iteration 3: log likelihood = 658.4415 (not concave)
Iteration 4: log likelihood = 658.55751
Iteration 5: log likelihood = 658.96254
Iteration 6: log likelihood = 659.23158
Iteration 7: log likelihood = 659.4882
Iteration 8: log likelihood = 659.56345
Iteration 9: log likelihood = 659.66959
Iteration 10: log likelihood = 659.69762
Iteration 11: log likelihood = 659.72588
Iteration 12: log likelihood = 659.72715
Iteration 13: log likelihood = 659.73034
Iteration 14: log likelihood = 659.73039

Stoc. frontier normal/truncated-normal model      Number of obs =      388
                                                    Wald chi2(3) =     141.55
Log likelihood = 659.73039                        Prob > chi2 =      0.0000

-----
| media_voti |      Coef.  Std. Err.  z  P>|z|  [95% Conf. Interval]
-----+-----
| media_voti |
| attrezzature | .1068267   .0386515   2.76  0.006   .0310711   .1825823
| libri        | .9364983   .5942815   1.58  0.115   -.228272   2.101269
| voto_diploma | .1511577   .0158311   9.55  0.000   .1201293   .1821861
| _cons       | -.2336338   2.128496   -0.11  0.913   -4.405408   3.938141
-----+-----
| mu         |
| femmina     | -.0222961   .0085814   -2.60  0.009   -.0391153   -.0054768
| linguistico | -.0249394   .0263899   -0.95  0.345   -.0766627   .026784
| commerciale | .0140951   .0084137   1.68  0.094   -.0023955   .0305856
| tecnica     | .0303246   .012423   2.44  0.015   .0059759   .0546733
| lavoro      | .0067039   .007728   0.87  0.386   -.0084427   .0218505
| limitrofe   | .0172927   .0080543   2.15  0.032   .0015066   .0330788
| lontane     | .0602992   .015613   3.86  0.000   .0296982   .0909002
| _cons       | .0394304   .0194122   2.03  0.042   .0013832   .0774775
-----+-----
| /lnsigma2   | -5.819454   .1653455   -35.20  0.000   -6.143525   -5.495382
| /ilgtgamma  | 1.719926    .4560866   3.77  0.000   .8260125   2.613839
-----+-----
| sigma2     | .0029692   .0004909   .0021473   .0041057
| gamma      | .8481193   .0587499   .6955111   .9317469
| sigma_u2   | .0025183   .0004839   .0015698   .0034667
-----

```

Figura 6.13 – Risultati Modello Cobb-Douglas *Normale-Normale troncata* (Battese e Coelli, 1995).

BIBLIOGRAFIA

- Abbott, M. e Doucouliagos, C. (2003) The efficiency of Australian universities: A data envelopment analysis, *Economics of Education Review*, vol. 22, pp. 89-97.
- Afriat, S.N. (1972) Efficiency Estimation of Production Functions, *International Economic Review*, vol. 13, pp. 568-598.
- Ahn, T., Charnes, A. e Cooper, W.W. (1988) Some statistical and DEA evaluation of relative efficiencies of public and private institutions of higher learning, *Socio-Economic Planning Sciences*, vol. 22, n. 6, pp. 259-269.
- Aigner, D.J. e Chu, S.F. (1968) On Estimating the industry production function, *American Economic Review*, vol. 58, pp. 826-839.
- Aigner, D.J., Amemiya, T. e Poirier, D.J. (1976) On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function, *International Economic Review*, vol. 17, n. 2, June, pp. 377-396.
- Aigner, D.J, Lovell, C.A.K. e Schimdt, P.(1977) "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics*, vol. 6, pp. 21-37.
- Ali, A.I. e Seiford, L.M. (1993) The Mathematical programming Approach to Efficiency Analysis, in *The Measurement of Productive Efficiency: Techniques and Applications*, Fried H.O., Lovell, C.A.K. e Schmidt S. (Eds.), Oxford University Press, New York, pp. 120-159.
- Ali, M. e Flinn J.C. (1989) Profit Efficiency among Basmati Rice Producers in Pakistan Punjab, *American Journal of Agricultural Economics*, vol. 71, n. 2, pp. 303-310.

- Allen, R.G.D. (1947) *Mathematical Methods of Economists*, Macmillan, London.
- Amemiya, T. (1973) Regression Analysis when the dependent variable is truncated normal, *Econometrica*, vol. 41, pp. 997-1016.
- Anderson, P. e Petersen, N.C. (1993) A procedure for ranking efficient units in Data Envelopment Analysis, *Management Science*, vol. 39, n. 10, pp. 1261-1264.
- Arrow, K.J., Chenery, H.B. Minhas, B.S. e Solow R.M. (1961) Capital-labor Substitution and Economic Efficiency, *The Review of Economics and Statistics*, vol. XLIII, n. 3, pp. 225-250.
- Athanassopoulos, A. e Shale, E. (1997) Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis, *Education Economics*, vol. 5, n. 2, pp. 117-134.
- Audibert, M. (1997) Technical Inefficiency Effects among Paddy Farmers in the Villages of the "Office du Niger", Mali, West Africa, *Journal of Productivity Analysis*, vol. 8, n. 4, pp. 379-394.
- Avkiran, N.K. (2001) Investigating technical and scale efficiencies of Australian universities through data envelopment analysis, *Socio-Economic Planning Sciences*, vol. 35, pp. 57-80.
- Azzalini, A. (1985) A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 12, pp. 171-178.
- Azzalini, A. (1986) Further results on a class of distributions which includes the normal ones, *Statistica*, vol. 46, pp. 199-208.
- Banker, R.D. (1984) Estimating the Most Productive Scale Size using Data Envelopment Analysis, *European Journal of Operational Research*, vol. 17, n. 1, pp. 35-44.
- Banker, R.D. (1993) Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundations, *Management Science*, vol. 39, pp. 1265-1273.
- Banker, R.D., Charnes, A. e Cooper, W.W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, vol. 30, n. 9, pp. 1078-1092.
- Banker, R.D. e Morey, R.C. (1986a) Efficiency analysis for exogeneously fixed inputs and outputs, *Operations Research*, 34 (4): 513-21.
- Banker, R.D. e Morey, R.C. (1986b) The use of categorical variables in data envelopment analysis, *Management Sciences*, 32 (12): 1613-27.

- Barr, R., (2004) DEA Software Tools and Technology: A State-of-the-Art Survey, in W.W. Cooper, L. M. Seiford, and J. Zhu, eds. *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston, pp. 539-566.
- Battese, G.E. e Broca, S.S. (1997) Functional Forms of Stochastic Frontier Production Functions and Models for Technical Inefficiency Effects: A Comparative Study for Wheat Farmers in Pakistan, *Journal of Productivity Analysis*, vol. 8, pp. 395-414.
- Battese, G.E. e Coelli, T.J. (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics*, vol. 38, n. 3, pp. 387-399.
- Battese, G.E. e Coelli, T.J. (1992) Frontier production functions, technical efficiency and panel data with application to paddy farmers in India, *Journal of Productivity Analysis*, vol. 3, pp. 153-169.
- Battese, G.E. e Coelli, T.J. (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics*, vol. 20, pp. 325-332.
- Battese, G.E., Coelli, T.J. e Colby T.C. (1989) Estimation of Frontier Production Functions and the Efficiencies of Indian Farms Using Panel Data from ICRI-SAT's Village Level Studies, *Journal of Quantitative Economics*, vol. 5, pp. 327-348.
- Battese, G.E., e Corra, G.S. (1977) Estimation of a Production Frontier model: with application to the pastoral zone of eastern Australia, *Australian Journal of Agricultural Economics*, vol. 21, No 3, pp. 169-179.
- Battese, G.E., e Rao Prasada, D.S (2002) Technology Gap, Efficiency and a Stochastic Metafrontier Function, *International Journal of Business and Economics*, vol. 1, n. 2, pp. 87-93.
- Battese, G.E., Rao Prasada, D.S. e O'Donnell, C.J. (2004) A Metafrontier Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies, *Journal of Productivity Analysis*, vol. 21, pp. 91-103.
- Bauer, P.W. e Hancock, D. (1993) "The Efficiency of the Federal Reserve in Providing Check Processing Services", *Journal of Banking & Finance*, vol. 17, April, No. 2/3, pp. 287-311.
- Beasley, J.E. (1990) Comparing University Departments, *OMEGA, Int.J.Mgmt.Sci*, vol. 2, pp. 171-183.
- Beasley, J.E. (1995) Determining Teaching and Research Efficiencies, *Journal of the Operational Research Society*, vol. 46, pp. 441-452.

- Becker, G. (1962) Investment in Human Capital: A Theoretical Analysis, *Journal of Political Economy*, LXX (5), Part 2, pp. 9-49.
- Berger, A.N. e Mester, L.J. (1997) "Inside the black box: What explains differences in the efficiencies of financial institutions?", *Journal of Banking & Finance*, vol. 21, pp. 895-947.
- Berndt, E., Darrough, M. e Diewert, W.E. (1977) Flexible Functional Forms and Expenditure Distributions: An Application to Canadian Consumer Demand Functions, *International Economic Review*, vol. 18, pp. 651-675.
- Berndt, E., Hall, B.H., Hall, R.E. e Hausman, J.A. (1974) Estimation and Inference in Non-linear Structural Models, *Annals of Economics and Social Measurement*, vol. 4, pp. 653-665.
- Bessent, A. e Bessent, W. (1980) Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis, *Educational Administration Quarterly*, vol. 16, n. 2, pp. 57-75.
- Bessent, A., Bessent, W., Kennington J., Reagan B. (1982) An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District, *Management Science*, vol. 28, n. 12, pp.1355-1367.
- Bhattacharyya, A., Kumbhakar, S.C., e Bhattacharyya A. (1995) Ownership Structure and Cost Efficiency: A Study of Publicly Owned Passenger-Bus Transportation Companies in India, *Journal of Productivity Analysis*, vol. 6, n. 1, April, pp. 47-62.
- Bhattacharyya, A., Lovell, C.A.K., e Sahay, P. (1997) The impact on the productive efficiency of Indian commercial banks, *European Journal of Operational Research*, vol. 98, pp. 322-345.
- Bianchi, T. (1993) Elementi economico-aziendali per un'analisi del grado di efficienza delle banche, *Bancaria*, n. 11.
- Biggeri, L., e Zuliani, A. (1983) La misura della produttività-efficienza nella pubblica amministrazione italiana: i risultati di una ricerca, *Città e regione*, IX, n. 4.
- Boles, J.N., (1967) Efficiency Squared- Efficient Computation of Efficiency Indexes, *Western Farm Economic Association, Proceeding 1966*, Pullman, Washington, pp. 137-142.
- Boles, J.N. (1971) The 1130 Farrell Efficiency System-Multiple Products, Multiple Factors, *Giannini Foundation of Agricultural Economics*, February.
- Bonsi, L., Manchi P. e Pavone A. (1999) L'Analisi dell'efficienza tecnica nelle unità periferiche della pubblica amministrazione, Istituto nazionale di Statistica.

- Borts, G.H., e Mishan E.J. (1962) Exploring the “Uneconomic Region” of the Production Function, *Review of Economic Studies: Journal of the Economic study society*, vol. 29, pp. 300-312.
- Breu, T.M. and Raab, R.L. (1994), Efficiency and Perceived Quality of the Nation’s “Top 25” National Universities and National Liberal Arts Colleges: An Application of Data Envelopment Analysis to Higher Education. *Socio-Economic Planning Sciences*, 28, pp. 33-45.
- Bronfenbrenner, M. (1944) “Production Functions: Cobb-Douglas, Interfirm, Intrafirm”, *Econometrica*, vol. 12, n. January, pp. 35-44.
- Bronfenbrenner, M. e Douglas, P.H. (1939) “Cross-section studies in the Cobb-Douglas function”, *The Journal of Political Economy*, vol. 47, n. 6, pp. 761-783.
- Brockett, P.L., Cooper, W.W., Wang, Y., e Shin, H.C. (1998), Congestion and inefficiency in chinese production before and after the 1978 economic reforms, *Socio-Economic Planning Sciences*, vol. 32, pp. 1-20.
- Brockett, P.L. e Golany, B. (1996) Using Rank Statistics for Determining programmatic Efficiency Differences in Data Envelopment Analysis, *Management Science*, vol. 42, n. 3, Marzo, pp. 466-472.
- Bruschini, G. e Viviani, A. (2002) L'efficienza delle filiali nell'impresa-banca: dall'analisi micro al bechmarking, *Economia Bancaria*, n. 12, pp. 18-26.
- Burley, H. T. (1976) Productive Efficiency Measures, *IX International Symposium on Mathematical Programming*, Budapest, Hungary, August.
- Burley, H.T. (1980) Productive Efficiency in U.S. Manufacturing: A Linear Programming Approach, *The Review of Economics and Statistics*, vol. 62, n. 4, pp. 619-622.
- Byrnes, P. (1985), The effect of ownership on efficiency: A nonparametric programming approach. PhD Dissertation, Southern Illinois University, Carbondale, Illinois.
- Byrnes, P., Färe, R. e Grosskopf, S.(1984) “Measuring Productive Efficiency: An Application to Illinios Strip Mines”, *Management Science*, vol. 30, n. 6, June, pp. 671-681.
- Byrnes, P., Färe, R., Grosskopf, S. e Lovell, C.A.K. (1988) “The Effect of Unions on Productivity: U.S. Surface Mining of Coal”, *Management Science*, vol. 34, n. 9 September, pp. 1037-1053.
- Cambini, C. e Filippini, M. (2003) Competitive Tendering and Optimal Size in the Regional Bus Transportation Industry: An Example form Italy, *Annals of Public and Cooperative Economics*, vol. 74, n. 1, pp. 163-182.

- Cardani, A.M., Castagna, M. e Galeotti, M. (1991) La misurazione dell'efficienza economica: un'applicazione al settore bancario italiano, *Ricerche Economiche XLV*, pp. 57-77.
- Castelli, L., Pesenti, R. e Ukovich, W. (2001) DEA-like models for efficiency evaluation of specialized and interdependent units", *European Journal of Operational Research*, vol. 132, n. 2, pp. 274-286.
- Caudill, S.B. e Ford, J.M. (1993) "Biases in Frontier Estimation due to Heteroscedasticity", *Economics Letters*, Vo.41, pp. 17-20.
- Caudill, S.B., Ford, J.M. e Gropper, D.M. (1995) "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity", *Journal of Business & Economic Statistics*, vol. 13, N0.1, January, pp. 105-111.
- Catalano, G. e Silvestri, P. (1992) Il finanziamento del sistema universitario italiano. *Ministero del Tesoro, Commissione tecnica della spesa pubblica*, Roma.
- Cazals, C., Florens, J.P. e Simar, L. (2002) Nonparametric frontier estimation: a robust approach, *Journal of Econometrics*, vol. 106, pp. 1-25.
- Chambers, R.G., Chung, Y. e Färe, R. (1996) Benefit and Distance Functions, *Journal of Economic Theory*, vol. 70, pp. 407-419.
- Charnes, A. e Cooper, W.W. (1962) Programming with linear fractional functionals, *Naval Research Logistic Quarterly*, vol. 9, n. 3/4, pp. 181-185.
- Charnes, A. e Cooper, W.W. (1963) Deterministic Equivalents for Optimizing and Satisficing Under Chance Constraints, *Operations Research, the Journal of the Operations Research Society of America*, vol. 11, pp. 18-39.
- Charnes, A. e Cooper, W.W. (1973) An explicit general solution in linear fractional programming, *Naval Research Logistic Quarterly*, vol. 20, n. 3.
- Charnes, A. e Cooper, W.W. (1980) Auditing and Accounting for Program Efficiency and Management Efficiency in Not-for-Profit Entities, *Accounting, Organizations and Society*, vol. 5, n. 1, pp. 87-107.
- Charnes, A. e Cooper, W.W. (1985) Preface to topics in Data Envelopment Analysis, *Annals of Operations Research*, vol. 2, pp. 59-94.
- Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M. e Steetz, J. (1985) Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions, *Journal of Econometrics*, vol. 30, n. 1/2, pp. 91-107.
- Charnes, A., Cooper, W.W., Lewin, A.Y. e Seiford, L.M. (1995) *Data Envelopment Analysis: Theory, methodology and Applications*, Kluwer Academic Publishers, Boston.

- Charnes, A., Cooper, W.W. e Rhodes, E. (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research*, vol. 2, pp. 429-444.
- Charnes, A., Cooper, W.W. e Rhodes, E. (1979) Short Communication: Measuring the efficiency of decision making units, *European Journal of Operational Research*, vol. 3, p. 339.
- Charnes, A., Cooper, W.W. e Rhodes, E. (1981) Evaluating Program and Managerial Efficiency: an application of Data Envelopment Analysis to Program Follow Through, *Management Science*, vol. 27, n. 6, June, pp. 668-697.
- Cobb, C.W. e Douglas, P.H. (1928) "A Theory of Production", *The American Economic Review Supplement*, vol.18.
- Coelli, T. (1996) Assessing the performance of Australian universities using data envelopment analysis. Internal report, Centre for Efficiency and Productivity Analysis, University of New England, Armidale, NSW, Australia.
- Coelli, T. (1996) A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program, CEPA Working Paper 96/08, Centre for Efficiency and Productivity Analysis, University of New England, Armidale, NSW, Australia.
- Coelli, T. (1998) A multi-stage methodology for the solution of oriented DEA models, *Operations Research Letters*, vol. 23, pp. 143-149.
- Coelli, T., Rao Prasada, D.S e Battese, G. (1999) An introduction to efficiency and productivity analysis, Kluwer Academic Publisher.
- Cohn, E., Rhine, S. e Santos, M.C. (1989) Institutions of higher education as multi-product firms: Economies of scale and scope, *Review of Economics and Statistics*, vol. 71, pp. 284-290.
- CNVSU (2003) Università obiettivo valutazione 2, Documenti, Edizioni Le Monnier.
- Cooper, S.T. e Cohn, E. (1997) Estimation of a Frontier Production Function for the South Carolina Educational Process, *Economics of Education Review*, vol. 16, n. 3, pp. 313-327.
- Cooper, W.W., Deng, H., Huang, Z.M. e Li, S.X. (2002) A one-model approach to congestion in data envelopment analysis, *Socio-Economic Planning Sciences*, vol. 36, pp. 231-238.
- Cooper, W.W., Seiford, L.M. e Tone, K. (2000) *Data Envelopment Analysis*, Dordrecht: Kluwer Academic Press.
- Cooper, W.W., Seiford, L.M. e Zhu, J. (2000) A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA, *Socio-Economic Planning Sciences*, vol. 34, pp. 1-25.

- Cooper, W.W., Seiford, L.M. e Zhu, J. (2001) Slacks and congestion: A response to comments by Färe and Grosskopf, *Socio-Economic Planning Sciences*, vol. 35, pp. 1-11.
- Cornwell, C., Schmidt, P. e Sickles, R.C. (1990) Production Frontiers with Cross-sectional and Time-series Variation in Efficiency Levels, *Land Economics*, vol. 54, pp. 156-186.
- Christensen, L.R., Jorgenson, D.W., Lau, L.J., (1971) Conjugate duality and the transcendental logarithmic production function, *Econometrica*, vol. 39, pp. 255-256.
- Christensen, L.R., Jorgenson, D.W. e Lau, L.J. (1973) Transcendental logarithmic production frontiers, *The Review of Economics and Statistics*, vol. 55, pp. 28-43.
- Dagum, C. (1994) Human Capital, Income and Wealth Distribution Models and their Applications, Proceedings of the American Statistical Association, Business and Economic Statistics Section, pp. 253-258.
- Daraio, C. and L. Simar (2005), Introducing environmental variables in nonparametric frontier models: a probabilistic approach, *Journal of Productivity Analysis*, vol. 24, 1, pp. 93-121.
- Debreu, G. (1951) The Coefficient of Resource utilization, *Econometrica*, vol. 19, n. 3, pp. 273-292.
- De Groot, H., McMahon, W.W. e Volkwein, J.F. (1991) The cost structure of American research universities, *Review of Economics and Statistics*, vol. 73, pp. 424-431.
- Deprins, D. e Simar, L. (1983) On Farrell Measures of Technical Efficiency, *Recherches Economiques de Louvain*, vol. 49, n. 2, pp. 123-137.
- Deprins, D. e Simar, L. (1989) Estimating Technical Inefficiencies with Corrections for Environmental Conditions with an Application to Railway Companies, *Annals of Public and Cooperative Economics*, vol. 60, n. 1, January-March, pp. 81-102.
- Deprins, D., Simar, L. e Tulkens, H. (1984) Labor-Efficiency in Post Offices, in *The Performance of Public Enterprises: Concepts and Measurement*, a cura di Marchand, M., Pestieau, P. e Tulkens, H., North Holland: Elsevier Science Publications, pp. 243-267.
- Destefanis, S. e A. Pavone (1996) L'analisi dell'efficienza nell'ambito dell'approccio FDH. Un'estensione e un'applicazione per gli ospedali di zona, *Quaderni di Ricerca* n. 3, Roma, ISTAT.

- Destefanis, S. e A. Pavone (1998) Servizi istituzionali e servizi a carattere produttivo. Un'analisi dell'efficienza delle amministrazioni comunali italiane, Atti della XXXIX Riunione Scientifica SIS, 14-17 aprile, Sorrento.
- Diewert, W.E. (1971) An Application of the Shephard duality theorem: A Generalized Leontiev production function, *Journal of Political Economy*, vol. 79, n. 3, pp. 481-507.
- Diewert, W.E. (1973) *Separability and A Generalization of the Cobb-Douglas Cost, Production and Indirect Utility Functions*, Research Branch, Department of Manpower and Immigration Canada, pp. 1-44.
- Dominguez-Molina, J.A., Gonzalez-Farias, G. and Ramos-Quiroga, R. (2004). Skew-normality in stochastic frontier analysis. In *Skew-elliptical distributions and their applications* (ed. M. G. Genton), capitolo 13, 223–242. Chapman & Hall/CRC, London.
- Douglas, P.H. (1976) The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some New Empirical Values, *Journal of Political Economy*, vol. 84, n. 5, pp. 903-915.
- Drezé, J., Kmenta, J. e Zellner, A. (1966) Specification and Estimation of Cobb-Douglas Production Function Models, *Econometrica*, vol. 34, n. 4, pp. 784-795.
- Dundar, H. e Lewis, D.R. (1995) Departmental Productivity in American universities: Economies of scale and scope, *Economics of Education Review*, vol. 14, pp.119-144.
- Efron, B. e Tibshirani, R.J. (1993) *An introduction to the Bootstrap*, Champman Hall, Inc., New York.
- Ehrenberg, R.G. (2004) Econometrics studies of higher education, *Journal of Econometrics*, vol. 121, pp. 19-37.
- Fabbri, D. (1998) La stima di frontiere di costo nel trasporto pubblico locale: una rassegna e un'applicazione", *Economia Pubblica*, anno XXVIII, n. 3, pp. 55-94.
- Fabbri, D. (2002) Misurare l'efficienza nella produzione ospedaliera, in *Data Mining, Web Mining e CRM*, a cura di Camillo, F. e G. Tassinari, Franco Angeli, Milano.
- Fabbri, D., Fazioli, R. e Filippini, M. (1996) *L'intervento pubblico e l'efficienza possibile*, Il Mulino, Bologna.
- Färe, R. (1975) Efficiency and the Production Function, *Zeitschrift für Nationalökonomie*, vol. 35, pp. 317-324.
- Färe, R. e Grosskopf, S. (1983a) Measuring Congestion in Production, *Zeitschrift für Nationalökonomie*, vol. 43, pp. 251-271.

- Färe, R. e Grosskopf, S. (1983b) Measuring Output Efficiency, *European Journal of Operational Research*, vol. 13, pp. 173-179.
- Färe, R. e Grosskopf, S. (1996) *Intertemporal Production Frontier: with dynamic DEA*, London, Kluwer.
- Färe, R. e Grosskopf, S. (1998) Congestion: a Note, *Socio-Economic Planning Sciences*, vol. 32, n. 1, pp. 21-23.
- Färe, R. e Grosskopf, S. (2000) Slacks and congestion: a comment, *Socio-Economic Planning Sciences*, vol. 34, pp. 27-33.
- Färe, R. e Grosskopf, S. (2001) When can slacks be used to identify congestion? An answer to W.W. Cooper, L. Seiford and J.Zhu, *Socio-Economic Planning Sciences*, vol. 35, pp. 217-221.
- Färe, R., Grosskopf, S. e Lovell, C.A.K. (1983) The Structure of Technical efficiency, *Scandinavian Journal of Economics*, vol. 85, pp. 181-190.
- Färe, R., Grosskopf, S. e Lovell, C.A.K. (1985) *The Measurement of Efficiency of Production*, Kluwer-Nijhoff, Boston.
- Färe, R., Grosskopf, S. e Lovell, C.A.K. (1987) Non Parametric Disposability Tests, *Zeitschrift Für Nationalökonomie*, vol. 47, n. 1, pp. 77-85.
- Färe, R., Grosskopf, S. e Lovell, C.A.K. (1994) *Production Frontiers*, Cambridge University Press, Cambridge.
- Färe, R. Grosskopf, S., Lovell C.A.K. e Pasurka, C. (1989) Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach, *The Review of Economics and Statistics*, vol. 71, pp. 90-98.
- Färe, R., Grosskopf, S. e Weber, W.L. (1989) Measuring district performance, *Public Finance Quarterly*, vol. 17, n. 4, pp. 409-428.
- Färe, R. e Lovell, C.A.K (1978) Measuring the Technical Efficiency of Production, *Journal of Economic Theory*, vol. 19, pp. 150-162.
- Färe, R., e Svensson, L. (1980) Congestion of Production Factors, *Econometrica*, vol. 48, n. 7, Novembre, pp. 1745-1753.
- Farrell, M.J. (1957) The Measurement of Productive Efficiency of Production, *Journal of the Royal statistical society. Series A, General*, vol. 120, pp. 253-281.
- Farrell, M.J. e Fieldhouse, M. (1962) Estimating efficient production functions under increasing returns to scale, *Journal of the Royal statistical society. Series A, General*, vol. 125, pp. 252-267.
- Fazioli, R., Filippini, M. e Künzle, M. (2003) Valutazione dell'efficienza delle compagnie di bus italiane e svizzere, in *L'efficienza nei servizi pubblici*, Banca d'Italia, Roma, pp. 171-210.

- Ferrari, G. e Ganugi, P. (1993) Economie di scala ed efficienza produttiva nel comparto tessitura dell'industria tessile di Prato, *Quaderni del Dipartimento Statistico*, vol. 37, Firenze.
- Ferrari, G., Gori, E., Maltagliati, M., Tesi, G. (1998) Efficienza e gruppi isorisorse nei servizi alla persona: una trattazione congiunta attraverso funzioni di costo e tecniche A.I.D., Note di Ricerca n. 4, Dipartimento di Scienze Statistiche, Udine.
- Ferrari, G. e Laureti, T. (2004a) Evaluation of University Graduates Technical Efficiency Using Data Envelopment Analysis Combined with Multiple Correspondence Analysis: the case of the University of Florence, Proceedings of the International Conference on Quantitative Methods, Virt.
- Ferrari, G. e Laureti, T. (2004b) "Multi-Factor Efficiency Analysis of the Florence University Graduates" Atti della XLII Riunione Scientifica SIS, 9-11 giugno, Bari.
- Ferrari, G. e Laureti, T. (2005) "Evaluating the Technical Efficiency of Human Capital formation in the Italian University: Evidence from Florence", *Statistical Methods and Applications*, vol. 14, n. 2.
- Ferrari, G., Laureti, T. e Maltagliati, M. (2001) La valutazione DEA dell'efficienza del processo formativo dell'Ateneo fiorentino, *Atti del Convegno intermedio SIS*, 4-6 giugno, 2001.
- Ferrari, G. e Maltagliati, M. (1997) La misura dell'efficienza del lavoro negli uffici postali: un'analisi per la provincia di Firenze, Working Papers, n. 71, Dipartimento di Statistica 2G.Parenti", Università degli Studi di Firenze.
- Ferrier, G.D. e Lovell, C.A.K. (1990) Measuring Cost Efficiency in Banking Econometric and Linear Programming Evidence, *Journal of Econometrics*, vol. 46, pp. 229-245.
- Førsund, F.R. e Hjalmarsson, L. (1979) Generalized Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants, *Economic Journal*, vol. 89, pp. 294-315.
- Førsund, F.R., Lovell, C.A.K. e Schmidt, P. (1980) A Survey of Frontier Production Functions and of their relationship to Efficiency Measurement, *Journal of Econometrics*, vol. 13, pp. 5-25.
- Førsund, F.R. e Sarafoglou, N. (2002) On the Origins of Data Envelopment Analysis, *Journal of Productivity Analysis*, vol. 17, pp. 23-40.
- Fourer, R., Gay, D.M., and Kernighan, B.W. (1993), *AMPL: A Modelling Language for Mathematical Programming*, The Scientific Press, S. Francisco.

- Fried, H.O., Lovell, C.A.K. e Vanden Eeckaut, P. (1993) Evaluating the performance of US credit unions, *Journal of banking and Finance*, Vo.17, pp. 251-265.
- Fried, H.O., Lovell, C.A.K., Schmidt, S.S. e Yaisawarng, S. (2002) Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis, *Journal of Productivity Analysis*, vol. 17, pp. 157-174.
- Fried, H.O., Schmidt, S.S. e Yaisawarng, S. (1999) Incorporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency, *Journal of Productivity Analysis*, Vo. 12, pp. 249-267.
- Frisch, R. (1965) *Theory of Production*. Dordrecht: D. Reidel. Rand MsNally Chicago.
- Fuss, M., Mc Fadden, D. e Mundlack, Y. (1978) *A Survey of Functional Forms in the Economic Analysis of Production*, in *Production Economics: a dual approach to theory and application*, volume 1 The theory of Production, Fuss M. e Mc Fadden D. editori, Noth Holland publishing company.
- Ganley, J.A. e Cubbin, J.S. (1992) *Public Sector Efficiency Measurement. Applications of Data Envelopment Analysis*, North-Holland, Amsterdam.
- Gazzei, D.S. (1993) La misurazione dell'efficienza della produzione nelle Amministrazioni pubbliche: il caso delle Poste Italiane, *Note Economiche*, XXIII, n. 1, pp. 143-166.
- Gazzei, D.S., Lemmi, A. e Viviani, A. (1997) *Misure statistiche di performance produttiva. Un percorso di metodi e di evidenze empiriche*, CLEUP, Padova.
- Gazzei, D.S. e Viviani, A. (1998) *La performance della pubblica amministrazione: problemi di metodo e di misura*, Atti della XXXIX Riunione Scientifica SIS, 14-17 aprile, Sorrento.
- Gazzei, D.S. e Viviani, A. (2001) L'analisi dell'efficienza nel modo bancario. Dall'approccio macro all'approccio micro, *Studi e Note di Economia*, n. 2, pp. 45-70.
- Georgescu-Roegen, N.(1973) *Analisi economica e processo economico*, Sansoni, Firenze (ed. originale inglese 1966).
- Gijbels, I., Mammen, E., Park, B.U. e Simar, L. (1999) On Estimation of Monotone and Concave Frontier Functions, *Journal of the American Statistical Association*, vol. 94, pp. 220-228.
- Giusti, F. (1994) *Modelli neoclassici di produzione. Evoluzione storica e ricerca quantitativa*, Università degli Studi di Roma "La Sapienza", Dipartimento di teoria economica e metodi quantitativi per le scelte politiche, Roma.

- Glass, J.C., McKillop, D.G. e Hyndman, N. (1995) Efficiency in the Provision of University Teaching and Research: An Empirical Analysis of UK Universities, *Journal of Applied Econometrics*, vol. 10, n. 1, pp. 61-72.
- Gori, E., Grassetti L., Minotti, S.C. (2004) Multilevel Flexible Specification of the Production Function in Health Economics, n. 15 aprile 2004, Serie Verde, Metodi quantitativi e Informatica.
- Gori, E., Mealli, F. e Rampichini, C. (1993) Indicatori di efficienza ed efficacia per la valutazione dell'attività di formazione professionale, *Statistica*, anno LIII, n. 3.
- Gori, E. e Romano, F.M. (1990) I risultati dell'istruzione universitaria: il ruolo dei fattori individuali e di struttura, *Economica e Diritto del Terziario*, n. 2, Genova.
- Gori, E., Viviani, A. e Compagnino, A. (1991) L'efficienza del "processo produttivo" di istruzione universitaria, Working Papers, n. 37, Dipartimento Statistico, Università degli Studi di Firenze.
- Gori, E. e Vittadini, G. (1999) *La valutazione dell'efficienza ed efficacia dei servizi alla persona*. Impostazioni e metodi, in *Qualità e Valutazione nei servizi di pubblica utilità*, a cura di Gori E. e Vittadini G., Etas, Milano.
- Greene, W.H. (1980) Maximum Likelihood Estimation of Econometric Frontier Functions, *Journal of Econometrics*, vol. 13, n. 1, May, pp. 27-56.
- Greene (2000) *Econometric Analysis*, Prentice Hall, 4th Edition.
- Griliches, Zvi. (1963) The Sources of Measured Productivity Growth: U.S. Agriculture, 1940-1960, *Journal of Political Economy*, vol. 81, n. 4, August, pp. 331-46.
- Grosskopf, S. (1986) The Role of the Reference Technology in Measuring Productive Efficiency, *The Economic Journal*, vol. 96, n. 382, pp. 499-513.
- Grosskopf, S. (1996) Statistical Inference and Nonparametric Efficiency: A Selective Survey, *Journal of Productivity Analysis*, vol. 7, pp. 161-176.
- Groves, R.E.V. e Perks, R.W. (1984) The Teaching and Researching of Accounting in UK Universities, A Survey, *The British Accounting Review*, vol. 16, n. 2, pp. 10-20.
- Guarini, R. e Tassinari, F. (1990) *Statistica Economica. Problemi e metodi di analisi*, Il Mulino, Bologna.
- Hanoch, G. e M. Rothschild, (1972): Testing the assumptions of production theory: a nonparametric approach, *Journal of Political Economy*, vol. 80, n. 2, pp. 256-275.

- Hanushek, E.A. (1979) Conceptual and Empirical Issues in the Estimation of Educational Production Functions, *The Journal of Human Resources*, vol. 14, n. 3, pp. 351-388.
- Hanushek, E.A. (1986), The Economics of Schooling: Production and Efficiency in Public Schools, *Journal of Economic Literature*, XXIV, pp. 1141-1177.
- Harvey, A.C. (1976) Estimating regression models with multiplicative heteroscedasticity, *Econometrica*, 44, 3, pp. 461-465.
- Hayami, Y. (1969) Sources of Agricultural Productivity Gap among Selected Countries, *American Journal of Agricultural Economics*, Vo.51, pp. 564-585.
- Hayami, Y.E e Ruttan, V.W. (1970) Agricultural Productivity Differences Among Countries, *American Economic Review*, vol. 60, pp. 895-911.
- Hayami, Y.E e Ruttan, V.W. (1971) *Agricultural Development: An International Perspective*, Baltimore: John Hopkins University Press.
- Hare, P.G. e Wyatt, G. (1992) Economics of academic research and its implications for higher education, *Oxford Review of Economic Policy*, vol. 8, n. 2, pp. 48-66.
- Hartman, T.E. e Storbeck J.E. (1996) Input congestion in loan operations, *International Journal of production Economics*, vol. 46, pp. 413-421.
- Hatry, H.P. (1982) Misurazione della performance: principi e tecniche. Un'analisi generale del problema negli enti locali, *Problemi di Amministrazione Pubblica*, Fasc.2, FORMEZ.
- Henderson, A.D. (1971) Some unresolved problems. Remarks delivered on the occasion of his citation by the Higher Education Colloquium. Citato da Toombs W. *Productivity: The burden of success*, American Association for Higher Education, Washington.
- Hicks, J. (1932) *The Theory of Wages*, Macmillan, London.
- Higgins, J.C. (1989) Performance measurement in universities, *European Journal of Operational Research*, vol. 38, pp. 358-368.
- Hildenbrand, G. e Liu, T. (1965) Manufacturing Production Function in the United States 1957, Ithaca.
- Hjalmarsson, L., Kumbhakar, S.C. e Heshmati (1996) DEA, DFA and SFA: A Comparison, *Journal of Productivity Analysis*, vol. 7, n. 2/3, pp. 303-327.
- Hoch, I. (1955) Estimation of production Function Parameters Combining Time-Series and Cross-Section Data, *Econometrica*, vol. 30, January, pp. 34-53.

- Hoffman, A.J. (1957) Discussion on Mr.Farrell's Paper, *Journal of the Royal Statistical Society, Series A, General*, vol. 120, Part.3, p. 284.
- Hollingsworth, B. (2004) Non Parametric Efficiency Measurement, *The Econometric Journal*, vol. 114, pp. 307-311.
- Horrace, W.C. e Schmidt, P. (1996) Confidence Statements for Efficiency Estimates from Stochastic Frontier Models, *Journal of Productivity Analysis*, vol. 7, n. 2/3, pp. 257-282.
- Huang, C.J. e Liu, J.T. (1994) Estimation of a Non-Neutral Stochastic Production Function, *Journal of Productivity Analysis*, vol. 5, n. 2, June, pp. 171-180.
- Izadi, H., Johnes, G., Okrochi, R. e Crouchley, R. (2002) Stochastic frontier estimation of a CES cost function: the case of higher education in Britain, *Economics of Education Review*, vol. 21, pp. 63-71.
- James, E. (1978) Product mix and cost disaggregation, *Journal of Human Resources*, vol. 13, pp. 157-186.
- Johnes, G. (1990) Measures of research output: university departments of economics in the UK, 1984-88, *Economic Journal*, vol. 100, pp. 556-560.
- Johnes, G. (1996) Performance assessment in higher education in Britain, *European Journal of Operational Research*, vol. 89, pp. 18-33.
- Johnes, G. (1997) Costs and Industrial Structure in contemporary British Higher Education, *The Economic Journal*, vol. 107, pp. 727-737.
- Johnes, J. e Johnes, G. (1993) Measuring the research performance of UK Economic Departments: An Application of Data Envelopment Analysis, *Oxford Economic Papers*, vol. 45, pp. 332-347.
- Johnes, J. e Johnes, G. (1995) Research funding and performance in U.K. university departments of economics: A frontier analysis. *Economics of Education Review* vol. 14, n. 3, pp. 301-314.
- Johnes, G., Taylor, J. e Ferguson, G. (1987) The Employability of new graduates: a study of differences between UK universities, *Applied Economics*, vol. 19, pp. 695-710.
- Johnson, W.E. (1913) The Pure Theory of Utility Curves, *Economic Journal*, vol. 23, pp. 483-513.
- Johnson, N.L., Kotz, S. e Balakrishnan, N. (1994) *Continuous Univariate Distributions*, vol. 1, second edition, Wiley Series in probability and Mathematical Statistics.
- Johnston, R. e Graham, C. (2001), *Service Operations Management*, Prentice Hall, Essex, UK.

- Jondrow, J., Lovell, C.A.K., Materov, I.S. e Schimdt, P. (1982) On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, vol. 19, pp. 233-238.
- Kalirajan, K.P. (1990) On Measuring Economic Efficiency, *Journal of Applied Econometrics*, vol. 5, pp. 75-85.
- Kendrick, J.W. (1956) Productivity Trend: Capital and Labor, *The Review of Economics and Statistics*, vol. 38, n. 3, pp. 248-257.
- Kendrick, J.K. (1972) The Treatment of Intangible Resources as Capital, *The Review of Income and Wealth*, n. 1, marzo, pp. 109-125.
- Kirjavainen, T. e Loikkanen, H.A. (1998) Efficiency Differences of Finnish Senior Secondary Schools: An Application of DEA and Tobit Analysis, *Economics of Education Review*, vol. 17, n. 4, pp. 377-394,
- Kneip, A., Simar, L. e Wilson, P.W. (2003) Asymptotics for DEA Estimators in Non-parametric Frontier Models, working paper, Department of Economics, University of Texas, Austin.
- Kopp, R.J. (1981) The Measurement of Productive Efficiency: A Reconsideration, *Quarterly Journal of Economics*, vol. 96, n. 3, pp. 477-503.
- Kopp, R.J. e Mullah, J. (1990) Moment-based Estimation and Testing of Stochastic Frontier Models, *Journal of Econometrics*, vol. 46, pp. 165-183.
- Koopmans, T.C. (1951) Analysis of Production as an Efficient Combinations of Activities, trad. italiano in Koopmans, T.C. (1987).
- Koopmans, T.C. (1953) Activity Analysis and its applications, *American Economic Review*, vol. 43, n. 2, pp. 406-414.
- Koopmans, T.C. (1987) *Econometria, analisi delle attività, crescita ottimale*, Il Mulino, Bologna.
- Kooreman, P. (1994) Nursing home care in The Netherlands: a nonparametric efficiency analysis, *Journal of Health Economics*, vol. 13, pp. 301-316.
- Krueger, A.B. (1999) Experimental Estimates of Education production Functions, *The Quarterly journal of Economics*, CXIV, pp. 497-532.
- Kumbhakar, S.C. (1990) Production Frontiers, Panel Data and Time-Varying Technical Inefficiency, *Journal of Econometrics*, vol. 46, pp. 201-211.
- Kumbhakar, S.C., Ghosh, S. e McGuckin, J.T. (1991) A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms, *Journal of Business and Economic Statistics*, vol. 9, n. 3, Luglio, pp. 279-286.

- Kumbhakar, S.C. e Lovell, C.A.K. (2000) *Stochastic Frontier Analysis*, Cambridge University Press.
- Kuznets, S. (1990) *Popolazione, tecnologia, sviluppo*, Il Mulino, Bologna.
- Land, K.C., Lovell, C.A.K. e Thore, S. (1993) Chane-Constrained Data Envelopment Analysis, *Managerial and Decision Economics*, vol. 14, n. 6, pp. 541-554.
- Latini, M.E. e Sterlacchini, A. (1995) L'efficienza del sistema universitario italiano e di alcune facoltà marchigiane, *Economia Marche*, XIV, n. 1, pp. 61-83.
- Laureti, T. (2002) Misure di efficienza del processo di formazione universitaria nell'Ateneo fiorentino, Tesi di dottorato, Dipartimento di Statistica "G. Parenti", Firenze.
- Laureti, T. (2006) Modelling Exogenous Variables in Human Capital Formation through a Heteroscedastic Stochastic Frontier, Mimeo.
- Lazear, E.P. (2001) Educational Production, *The Quarterly Journal of Economics*, vol. CXVI, n. 3, pp. 777-803.
- Lee, L.F. (1983) A Test for Distributional Assumption for the Stochastic Frontier Functions, *Journal of Econometrics*, vol. 22, n. 3, August, pp. 245-267.
- Lee, Y. e Schmidt, P. (1993) A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency," In H. Fried and K. Lovell, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York, 1993.
- Leibenstein, H. (1966) Allocative Efficiency vs "X-efficiency," *The American Economic Review*, vol. 56, n. 2, June, pp. 392-415.
- Levaggi, R. (1994) Parametric and Non-Parametric Approach to Efficiency: The Case of Urban Transport in Italy, *Studi Economici*, vol. 49, n. 53, pp. 67-88.
- Lindsay, A.W. (1982) Institutional Performance in Higher Education: The Efficiency Dimension, *Review of Educational Research*, vol. 52, n. 2, pp. 175-199.
- Lovell, C.A.K. (1993) Production Frontiers and Productive Efficiency, in *The Measurement of Productive Efficiency: Techniques and Application*, Fried H., Lovell, C.A.K. e Schmidt, S. (Eds.), Oxford University Press.
- Lovell, C.A.K. e Schmidt, P. (1988) A Comparison of Alternative Approaches to the Measurement of Productive Efficiency, in *Application of Modern production Theory. Efficiency and productivity*, Dogramaci A., Färe R. (eds.) pp. 3-32, Kluwer Academic Publishers, Boston.

- Lovell, C.A.K., Walters, L.C. e Wood, L.L. (1993) Stratified Models of Education Production using Modified DEA and Regression Analysis, in *Data Envelopment Analysis: Theory, Methodology and Application*, a cura di Charnes, A., Cooper, W.W., Lewin, A.Y. e Seiford, L.M.
- Luenberger, D.G. (1992) Benefit Functions and Duality, *Journal of mathematical Economics*, vol. 21, pp. 461-481.
- Marchese, U., Ferrari, C. e Benacchio, M. (2000) L'efficienza tecnica delle imprese terminaliste dedicate alla movimentazione di container: un approccio DEA (*Data Envelopment Analysis*), *InforMare*, Università degli Studi di Genova, Dipartimento di Economia e Metodi Quantitativi, Sezione di Economia dei Trasporti e Geografia Economia.
- Marschak, J. e Andrews, W.H. (1944) Random Simultaneous Equations and the Theory of Production, *Econometrica*, vol. 12, n. 3-4, pp. 143-205.
- Materov, I.S. (1981) On full identification of the stochastic production frontier model, *Ekonomika i Matematicheskie Metody* 17, 784-788.
- McCarty, T. e Yaisawarng, S. (1993) Technical Efficiency in New Jersey School District, in *The measurement of Productive Efficiency: Techniques and Applications*, Fried H.O., Lovell C.A.K. e Schmidt S. (Eds.), Oxford University Press.
- McDonald, J. (1996) Note: A Problem with the Decomposition of Technical Inefficiency into Scale and Congestion Components, *Management Science*, vol. 42, n. 3, March, pp. 473-474.
- McFadden, D. (1966) *Cost, Revenue and Profit Functions: A cursory review*, Working paper n. 86, University of California, Berkeley. Versione riprodotta in Fuss and McFadden, 1978.
- Meeusen, W. e van den Broeck, J. (1977) Efficiency estimation from Cobb-Douglas Production Functions with composed error, *International Economic Review*, vol. 18, n. 2, June, pp. 435-444.
- Mester, L.J. (1993) Efficiency in the savings and loan industry, *Journal of Banking & Finance*, vol. 17, April, n. 2/3, pp. 267-286.
- Mincer, J. (1958) Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, vol. 55, pp. 281-302.
- Mubarik, A. e Flinn, J.C. (1989) Profit Efficiency Among Basmati Rice Producers in Pakistan Punjab, *American Agricultural Economics Association*, vol. pp. 303-310.
- Mundlak, Y. e Hoch, I. (1965) Consequences of alternative specifications in estimation of Cobb-Douglas production functions, *Econometrica*, vol. 33, n. 4, pp. 814-828.

- Mundlak, Y. e Hellinghausen, R. (1982) The Intercountry Agricultural Production Function: Another View, *American Agricultural Economics Association*, vol. 64, pp. 664-671.
- Nelson, R. e Hevert, K.T. (1992) Effects of class size on economies of scale and marginal costs in higher education, *Applied Economics*, vol. 24, pp. 473-482.
- Nesticò, A. e Prosperetti, L. (1991) *Produzione e Produttività*, in *Statistica Economica*, a cura di Marbach, G., Utet, Torino.
- Nesticò, A. e Viviani, A. (1987) Funzioni Medie e funzioni frontiera nella misurazione dell'efficienza: il caso dei trasporti pubblici locali, *Note Economiche*, vol. 1, pp. 131-154.
- Nyman, J.A. e Bricker, D.L. (1989) Profit incentives and technical efficiency in the production of nursing home care, *The Review of Economics and Statistics*, vol. 56, pp. 586-594.
- Park, B., Simar, L. e Weiner, C. (2000) The FDH Estimator for Productivity Efficiency Scores, *Econometric Theory*, vol. 16, pp. 855-877.
- Pareto, V. (1906) *Manual of Political Economy 1971*, traduzione edizione 1927, New York: Augustus M. Kelley.
- Pesenti, R. e Ukovich, W. (1996) Evaluating academic activities using DEA, Working Paper, Dipartimento di Elettrotecnica, Elettronica ed Informatica, Università di Trieste.
- Petretto, A. (1986) L'Approccio Econometrico per la misurazione dei risultati delle imprese pubbliche locali, *Politica Economica*, a.II, n. 2 pp. 203-224.
- Petretto, A. e Pisauro, G. (1995) La Riforma della Pubblica Amministrazione sotto il profilo dell'analisi economica: i controlli e la ristrutturazione degli uffici, *Politica Economica*, a.XI, n. 1, aprile.
- Petretto, A. e Viviani, A. (1984) An Econometric Model for Cross-Section Analysis of the Production of Urban Transport Service, *Economic Notes*, vol. 13, pp. 35-65.
- Pitt, M.M. e Lee, L.F. (1981) The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry, *Journal of Development Economics*, vol. 9, pp. 43-64.
- Ray, S.C. (1998) Data Envelopment Analysis, Non-discretionary Inputs and Efficiency: An Alternative Interpretation, *Socio-Economic Planning Science*, vol. 22, pp. 167-178.
- Ray, S.C. (1991) Resource-use Efficiency in Public Schools: A Study of Connecticut Data, *Management Science*, vol. 37, n. 12, pp. 1620-1628.

- Ray, S.C. (1998) Measuring Scale Efficiency from a Translog Production Function, *Journal of Productivity*, vol. 11, pp. 183-194.
- Ray, S.C. (2004) *Data Envelopment Analysis. Theory and Techniques for Economics and Operations Research*, Cambridge University Press.
- Reder, (1943) An Alternative Interpretation of the Cobb-Douglas Function, *Econometrica*, vol. 11, n. 11, pp. 259-264.
- Reifschneider, D. e Stevenson, R. (1991) Systematic Departures from the Frontier: A Framework for the Analysis of Firm inefficiency, *International Economic Review*, vol. 32, n. 3, August, pp. 715-723.
- Resti, A. (1994) Il dibattito su efficienza delle banche ed economie di scala: il contributo della Data Envelopment Analysis ed un'applicazione a dati italiani, *Politica Economica*, n. 2, Agosto.
- Richmond, J. (1974) Estimating the Efficiency of Production, *International Economic Review*, vol. 15, pp. 515-521.
- Rizzi, D. (1999) L'efficienza dei dipartimenti dell'Università Ca' Foscari di Venezia via DEA e DFA, Nota di Lavoro 99.09, Dipartimento di Scienze economiche, Università Ca' Foscari di Venezia.
- Ruggiero, J. e Bretschneider, S. (1995) The weighted Russell measure of technical efficiency, *European Journal of Operational Research*, vol. 108, pp. 438-451.
- Russell, R. (1985) Measures of Technical Efficiency, *Journal of Economic Theory*, vol. 35, pp. 109-126.
- Samuelson, P. (1947) *Foundations of Economic Analysis*, Harvard University Press
- Sarrico, C.S., Hogan, S.H., Dyson, R. and Athanassopoulos, G.A.D. (1997), Data Envelopment Analysis and University Selection. *Journal of the Operational Research Society*, 48, pp. 1163-1177.
- Scheel, H. (2000) EMS: Efficiency Measurement System Users Manual, Version 1.3, Universität Dortmund, Dortmund, Germany.
- Scheel, H. (2001) Undesirable outputs in efficiency valuations, *European Journal of Operational Research*, vol. 132, pp. 400-410.
- Schmidt, P. (1976) On the Statistical Estimation of Parametric Frontier Production Function, *Review of Economics and Statistics*, vol. 58, n. 2, pp. 238-239.
- Schmidt, P. e Sickles, R. (1984) Production Frontiers with Panel Data, *Journal of Business and Economic Statistics*, vol. 2, n. 4, pp. 367-374.
- Schmookler, J. (1952) The Changing Efficiency of the American Economy, 1869-1938," *Review of Economics and Statistics*, vol. 34, n. 3, 2 14-23.

- Schultz, T. (1961) Investment in Human Capital, *American Economic Review*, LI, vol. 1, pp. 1-17.
- Seiford, L.M. e Zhu, J. (2002) Modeling undesirable factors in efficiency evaluation, *European Journal of Operational Research*, vol. 142, pp. 16-20.
- Seitz, W.D. (1971) Productive Efficiency in the Steam-Electric Generating Industry, *Journal of Political Economy*, vol. 79, N, 4 July/August, pp. 878-886.
- Selmi, M. (1991) La valutazione dell'efficienza economica dei servizi pubblici locali tramite modelli di frontiera: il caso della azienda municipalizzata di Modena, Greta Associati, Venezia.
- Sena, V. (1999) Stochastic Frontier Estimation: A Review of the Software Options, *Journal of Applied Econometrics*, vol. 14, pp. 579-586.
- Shephard, R.W. (1953) *Cost and Production Functions*, Princeton University Press, Princeton.
- Shephard, R.W. (1970) *The Theory of Cost and production Functions*, Princeton University Press, Princeton.
- Simar, L. (1996) Aspects of Statistical Analysis in DEA-Type Frontier Models, *Journal of Productivity Analysis*, vol. 7, pp. 177-185.
- Simar, L. (2003), Detecting Outliers in Frontiers Models: a Simple Approach, *Journal of Productivity Analysis*, vol. 20, pp. 391-424.
- Simar, L., Lovell, C.A.K. e Vanden Eeckaut, P. (1994) Stochastic Frontiers Incorporating Exogenous Influences on Efficiency, Discussion Paper n. 9403, Institut de Statistique, Université Catholique de Louvain, Louvain-la Neuve, Belgium.
- Simar, L. e Wilson, P.W. (1998) Sensity Analysis of Efficiency Score: How to Bootstrap in Nonparametric Frontier Models, *Management Science*, vol. 44, n. 11, pp. 49-61.
- Simar, L. e Wilson, P.W. (1999) Estimating and Bootstrapping Malmquist Indices, *European Journal of Operations Research*, vol. 115, pp. 459-471.
- Simar, L. e Wilson, P.W. (2000a) Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis*, vol. 7, pp. 177-185.
- Simar, L. e Wilson, P.W. (2000b) A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics*, vol. 27, n. 6, pp. 779-802.
- Sinuany-Stern, Z., Meherez, A. e Barboy, A. (1994) Academic departments efficiency via DEA, *Computers and Operations Research*, vol. 5, pp. 543-554.
- Smith, A. (1948) Ricerche sopra le cause della ricchezza delle nazioni, UTET, Torino.

- Smith, P.C. e Street, A. (2005) Measuring the efficiency of public services: the limits of analysis, *Journal of the Royal Statistical Society, A*, vol. 168, 2, pp. 401-417.
- Solow, R. (1957) Technical Change and the Aggregate Production Function, *Review of Economics and Statistics*, vol. 39, August, pp. 313-320.
- Sraffa, P. (1960) *Produzione di merci a mezzo merci*, Einaudi, Torino.
- Starrett, D.A. (1977) Measuring Return to Scale in the Aggregate, and the Scale Effect of Public Goods, *Econometrica*, vol. 45, n. 6, pp. 1439-1455.
- Stevens, P.A. (2001) The determinants of economic efficiency in English and Welsh universities, National Institute of Economic and Social Research, London, Discussion Paper, n. 185.
- Stevenson, R.E. (1980) Likelihood Functions for Generalized Stochastic Frontier Estimation, *Journal of Econometrics*, Vol13, n. 1, May, pp. 57-66.
- Stigler, G.J. (1976) The Xistence of X-efficiency, *The American Economic Review*, vol. 66, N1, pp. 213-216.
- Sueyoshi, T. e Aoki, S. (2001) A use of a nonparametric statistic for DEA frontier shift: the Kruskal and Wallis rank test, *Omega*, vol. 29, pp. 1-18.
- Tani, P. (1986) *Analisi microeconomica della produzione*, NIS, Roma.
- Tani, P. (1989) *La rappresentazione della tecnologia produttiva nell'analisi microeconomica: problemi e recenti tendenze*, in *Le teorie economiche della produzione*, a cura di Zamagni S. , Il Mulino, Bologna.
- Taveres, G. (2002) *A Bibliografy of Data Envelopment Analysis (1978-2001)*, RUTCOR Research Report RRR 01-02, January, Rutgers University, NJ.
- Testi, A. (1996) La valutazione non parametrica dell'efficienza nella produzione di ricoveri ospedalieri, *Economia Pubblica*, anno XXVI, n. 3, 81-121.
- Thanassoulis, E. (2001) *Introduction to the Theory and Application of Data Envelopment Analysis*, Boston Kluwer Academic Press.
- Thanassoulis, E. e Portela, M.A.S. (2002), School Outcomes: Sharing the Responsibility Between Pupil and School. *Education Economics*, 10, 183-207.
- Thiry, B. e Tulkens, H. (1987) Allowing for technical inefficiency in parametric estimates of production functions- With an application to urban transit firms, Center for Operations Research & Econometrics, University Lowvaine, Louvain La Neuve, Belgium.
- Thomas, J.A. (1974) *The productive school: A system analysis approach to educational administration*. New York, John Wiley.

- Timmer, C. P. (1971) Using a probabilistic frontier production function to measure technical efficiency, *Journal of Political Economy*, vol. 79, pp. 776-794.
- Tomkins, C. e Green, R. (1988). An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. *Financial Accountability and Management*, vol. 4, n. 2, pp. 147-164.
- Vaggi, G. (1987) Produzione, in Dizionario di Economia Politica, a cura di Lunghini, G. e D'Antonio, R., Boringhieri, Torino.
- Van den Broeck, J., Forsund, F.R., Hjalmarsson, L. e Meeusen, W. (1980) On the estimation of deterministic and stochastic frontier production functions – A comparison, *Journal of Econometrics*, vol. 13, pp. 117-138.
- Varian, H. (1978) *Microeconomic Analysis*, Norton & Co., London.
- Varian, H. (1984) The Nonparametric Approach to production Analysis, *Econometrica*, vol. 52, n. 3, pp. 579-597.
- Varian, H. (1999) *Intermediate Microeconomics. A modern Approach*, 5th. Ed, Norton & Company, New York.
- Venanzoni, G. (1996) Forma funzionale e specificazione dell'output nella misura dell'efficienza produttiva. Il caso degli ospedali pubblici, Atti della XXXVIII Riunione Scientifica della Società Italiana di Statistica, Rimini.
- Vitali, O. (1999) *Statistica economica*, Cacucci, Bari.
- Vittadini, G. e Lovaglio, P. (2004) Fattori materiali e immateriali del capitale umano, in *Capitale Umano la ricchezza dell'Europa*, a cura di Vittadini, G., Edizioni Angelo Guerini e Associati SpA, Milano.
- Viviani, A. (1996) Tecnologie di produzione ed efficienza nella distribuzione commerciale, Atti della XXXVIII Riunione Scientifica della Società Italiana di Statistica, Rimini.
- Viviani, A., Mealli, F. e Pierotti, F. (1999) Funzioni frontiera su dati panel: il caso della distribuzione commerciale, *Rivista di Statistica Applicata*, vol. 11, n. 1, pp. 77-92.
- von Neumann, J. (1937) A Model of General Economic Equilibrium, in *Ergebnisse eines mathematischen Kolloquiums, 1935-36*, a cura di Menger K. (Tradotto e ristampato in *Review of Economic Studies*, vol. 13, n. 1, p. 1-9).
- Waldman, D.M. (1984) Properties of Technical Efficiency Estimators in the Stochastic Frontier Model, *Journal of Econometrics*, vol. 25, pp. 353-364.
- Walras (1874) *Elements of Pure Economics: Or the theory of social wealth*. 1954 translation of 1926 edition, Homewood, Ill.: Richard Irwin.

- Walters, A.A. (1963) Production and Cost Functions: An Econometric Survey, *Econometrica*, vol. 31, n. 1-2, pp. 1-66.
- Wang, H.J. (2002) Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic frontier Model, *Journal of Productivity Analysis*, vol. 18, pp. 241-253.
- Wang, H.J. e Schmidt, P. (2002) One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels, *Journal of Productivity Analysis*, vol. 18, pp. 129-144.
- Weinstein, M.A. (1964) The sum of values from a normal and a truncated normal distribution, *Technometrics*, vol. 6, pp. 104-105.
- Wheelock, D.C. e Wilson, P. (2003) Robust nonparametric estimation of efficiency and technical change in U.S. commercial banking, Working Papers 2003-037, Federal Reserve Bank of St. Louis.
- Winsten, C.B. (1957) Discussion on Mr. Farrell's Paper, *Journal of the Royal Statistical Society, Series A, General*, vol. 120, Part.3, pp. 282-284.
- Zamagni, S. (1994) *Economia politica*, Utet Torino.
- Zappa, G. (1957) *La produzione nell'economica delle imprese*, Giuffrè Ed., Milano.
- Zellner, A., Kmenta, J. e Dreze, J. (1966) Specification and Estimation of Cobb-Douglas Production Function Models, *Econometrica*, vol. 34, n. 4, pp. 784-795.
- Zieschang, K. (1985) An Extended Farrell Efficiency Measure, *Journal of Economic Theory*, vol. 33, pp. 387-396.