Università di Pisa

Dipartimento di Informatica

Technical Report : TR-99-03

Valutazioni di produttivit e misure di efficienza con applicazioni al settore del trasporto pubblico

Fernanda Farinaccio, Giorgio Gallo, Claudio Sandi

February 19, 1999

ADDR: Corso Italia 40,56125 Pisa,Italy. TEL: +39-50-887111. FAX: +39-50-887226

VALUTAZIONI DI PRODUTTIVITÀ E MISURE DI EFFICIENZA CON APPLICAZIONI AL SETTORE DEL TRASPORTO PUBBLICO

FERNANDA FARINACCIO, GIORGIO GALLO E CLAUDIO SANDI

1. Introduzione

In questa nota vengono elencate e brevemente riassunte le tecniche disponibili per misurare l'efficienza di sistemi predisposti alla produzione di beni e servizi. Lo scopo è quello di valutare l'efficacia di tali tecniche, in particolare per la valutazione di sistemi di trasporto finalizzati, più che al profitto, a scopi di pubblica utilità.

Nel primo paragrafo sono riprese definizioni caratteristiche e proprietà del classico concetto di *funzione di produzione*, in quanto ampiamente utilizzate in tutte le esposizioni seguenti.

Il paragrafo 3 presenta una panoramica delle definizioni proposte come misure di efficienza; una caratterizzazione di tali definizioni è poi riportata sulla base di apetti tecnici e/o economici (*technical and/or allocative efficiency*).

I paragrafi 4 e 5 sono dedicati alle due tecniche di valutazione di efficienza diventate ormai operative: ambedue utilizzano frontiere di produzione entro le quali sono rinchiuse (*data envelopment*) tutte le possibili realizzazioni. Nel paragrafo 4 (*econometric approach*) vengono presentati essenzialmente modelli di regressione e quindi tecniche parametriche e stocastiche; al contrario nel paragrafo 5 (*mathematical programming approach*) vengono usati modelli e metodi di ottimizzazione prevalentemente di tipo non-parametrico e deterministico.

Infine nel paragrafo 6 vengono riportate applicazioni al settore del trasporto pubblico.

2. Le funzioni di produzione

Il concetto teorico di *funzione di produzione*, come quantità di beni ottenibili sulla base di determinate risorse mediante un dato processo produttivo, è forse tra i più consolidati nelle descrizioni quantitative della teoria dell'impresa. È sulla base di tale funzione che sono normalmente evidenziate le modalità di utilizzo delle risorse disponibili, in grado di garantire i risultati più soddisfacenti. Anche le tipiche domande "what if" sono spesso formulate in termini di proprietà analitiche della funzione di produzione: presenza ed entità di economie di scala, importanza e sostituibilità delle risorse, elasticità delle grandezza coinvolte, in definitiva tutto quanto può servire ad una maggior razionalizzazione delle attività produttive. Per una trattazione esauriente della classica teoria dell'impresa si rimanda alla letteratura specializzata (si veda, ad esempio [19], Cap 3); verranno qui considerate solo le caratteristiche essenziali delle funzioni di produzione, al fine di consentire una miglior comprensione dell'argomento trattato in questa nota e cioè le misure di efficienza.

In ogni processo produttivo vengono utilizzate risorse di vario tipo (*input*) al fine di generare determinati beni e/o servizi (*output*). Il ruolo giocato da risorse e prodotti riguarda il singolo processo, in quanto una risorsa può essere resa disponibile da altre produzioni. La trasformazione degli input in output è soggetta a regole tecnologiche riassunte nei *modelli di produzione*, che esprimono in modo quantitativo il legame fra entità di risorse impiegate e l'ammontare di beni e/o servizi prodotti. Il concetto è estremamente generale e può prevedere singole relazioni funzionali discrete (dis)continue od anche sistemi di equazioni simultanee. In questo lavoro verranno considerati i modelli di produzione comunemente utilizzati nella letteratura, consistenti cioè in una *funzione di produzione* che lega un vettore $x \in R^n$ di inputs con un vettore $y \in R^m$ di outputs secondo la notazione:

$$f(x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m) = 0.$$

Per semplicità i modelli utilizzati in pratica prevedono generalmente situazioni in cui gli input concorrono alla produzione di un solo output secondo una funzione già in forma esplicita:

$$y = f(x_1, x_2, \ldots, x_p)$$

situazioni a cui ci si può ricondurre aggregando gli output in un solo criterio di valutazione. A titolo di esempio verrà spesso fatto riferimento al caso di due input che concorrono alla produzione di un output secondo la nota funzione di *Cobb-Douglas*:

$$y = Ax_1^{\alpha}x_2^{1-\alpha}$$

Tale funzione, di cui viene riportata un'immagine grafica nella figura 1, possiede le proprietà generalmente richieste alle funzioni di produzione e cioè: concavità, continuità sia della funzione che delle derivate parziali prime e seconde.

Al variare di x_1, tenendo costante il valore di x_2, la funzione di *Cobb-Douglas* dà luogo agli andamenti mostrati dalla figura 2, in cui la curva di produzione si sposta verso l'alto al crescere di x_2.

Sempre per la funzione di *Cobb-Douglas*, la figura 3 riporta alcune curve di livello, ovvero l'insieme delle coppie di valori x_1, x_2 caratterizzate da eguale valore di produzione; tali luoghi di punti vengono generalmente indicati con il termine di *isoquanti*, come ribadito nelle definizioni date più avanti.

Ulteriore proprietà della funzione di *Cobb-Douglas* è quella di essere omogenea di grado 1; se le quantità di input vengono cioè variate mediante un fattore moltiplicativo comune λ, anche l'output rimane modificato nella stessa misura:

$$A(\lambda x_1)^{\alpha}(\lambda x_2)^{1-\alpha} = \lambda A x_1^{\alpha} x_2^{1-\alpha}$$

Infine la linearità della relazione fra le quantità di input ed output espresse in logaritmi, consente, per modelli di produzione basati sulla funzione di *Cobb-Douglas*, l'utilizzo di metodi di stima estremamente consolidati:

$$\log y = \log A + \alpha \log x_1 + (1 - \alpha) \log x_2$$

come riportato in alcuni esempi del paragrafo 6.

L'elenco che segue riporta le principali grandezze caratterizzanti le funzioni di produzione. Senza sostanziale perdita di generalità le definizioni sono qui limitate al caso di due input ed un solo output; per estensioni e generalizzazioni si veda, ad esempio [19], Cap 3.

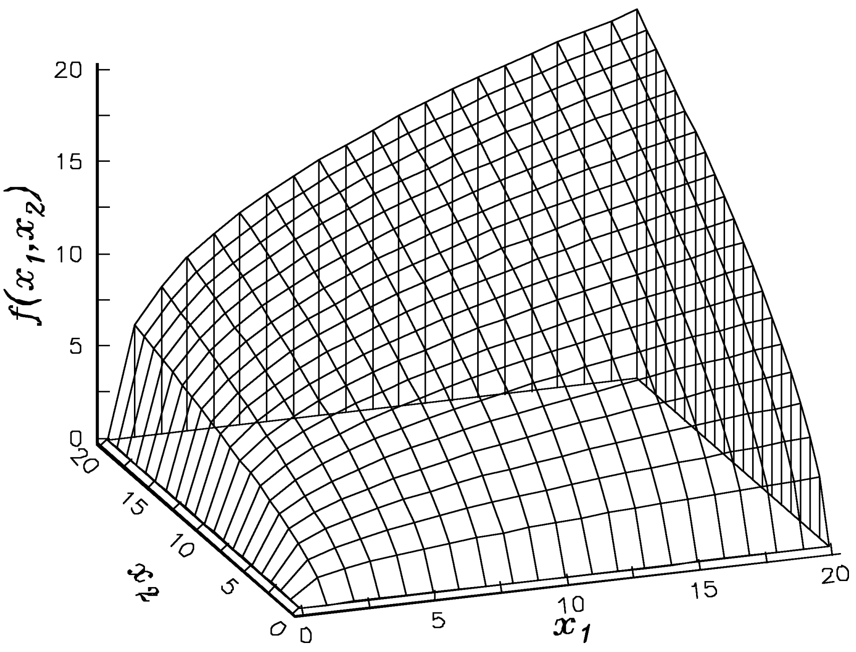

FIGURA 1. Funzione Cobb-Douglas in tre dimensioni

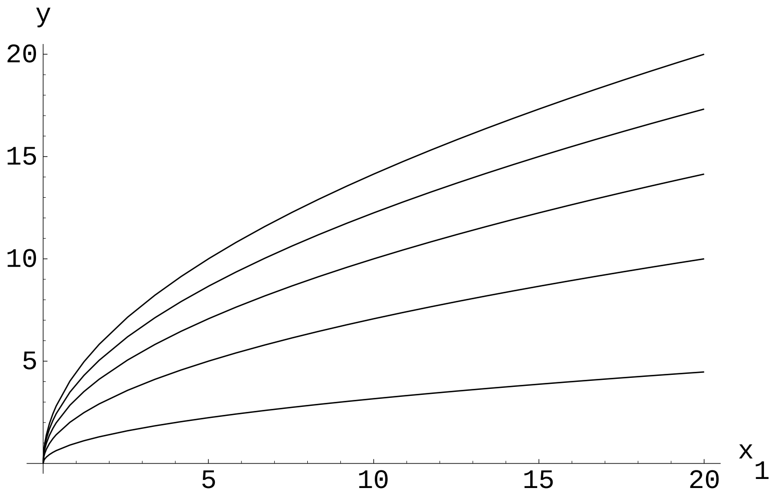

FIGURA 2. Funzione Cobb-Douglas in due dimensioni

$f(x_1, \overline{x}_2)/x_1$: *produttività media*, rispetto ad x_1, per x_2 fissato al valore $\overline{x}_2$;

$f_1 = \partial f(x_1, x_2)/\partial x_1$: *produttività marginale* rispetto ad x_1, a parità di x_2 (definizione analoga per f_2);

$f(x_1, x_2) = c$: *isoquanto*, luogo delle coppie (x_1, x_2) sufficienti a produrre il livello c di output $(y = c)$;

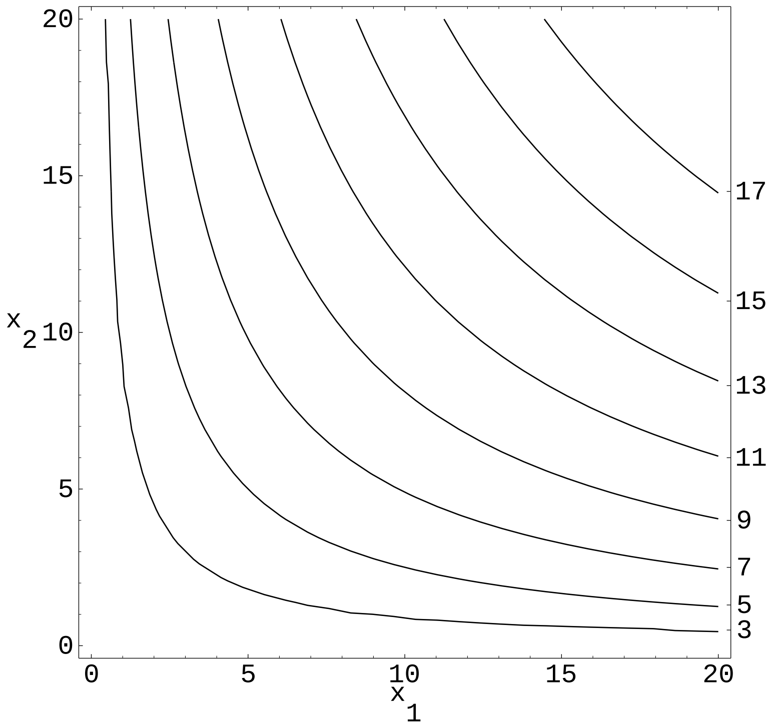

FIGURA 3. Curve di livello della funzione Cobb-Douglas

$f(\lambda x_1, \lambda x_2) = \lambda^k f(x_1, x_2)$: *omogeneità di grado k*: aumentando tutti gli input di un fattore λ, l'output aumenta di un fattore λ^k.

Si osservi che se la funzione di produzione è omogenea di grado k i rendimenti di scala sono decrescenti, costanti o crescenti a seconda che sia $k < 1$, $k = 1$, $k > 1$, rispettivamente.

In presenza di costi associati agli input e/o di ricavi associati agli output, le caratterizzazioni elencate costituiscono la base della classica teoria economica dell'impresa.

3. Frontiere di produzione e misure di efficienza

Rivolti inizialmente a contesti tipicamente industriali, gli econometristi hanno operativamente stimato le funzioni di produzione sulla base di valori medi riscontrati nei vari settori merceologici, facilitati in ciò dalla possibilità di esprimere in termini di costi /profitti tutte le componenti dei processi produttivi in esame. Non così immediati sono valutazioni e stime di attività legate ad istituzioni "no-profit" o alla produzione ed offerta di servizi. Tali situazioni, unitamente alla obiettiva difficoltà di sintetizzare in una sola funzione analitica il comportamento di sistemi complessi, hanno via via suggerito un concetto generalizzato di "Insieme delle produzioni ammissibili" , sulla base del quale ricavare, oltre alle tradizionali descrizioni riguardanti il processo produttivo, anche misure di efficienza.

Definita la produttività media come rapporto fra una misura dei beni ottenuti ed una misura delle risorse impiegate per produrli, si comprende come tale quantità possa dipendere dal processo produttivo, dal grado di efficienza e dalle situazioni ambientali in cui si opera. In questo lavoro si tenterà di isolare la componente di efficienza e misurare il suo contributo alla produttività.

Si deve a Koopmans (1951) ([20]) una delle prime definizioni formali di efficienza: *"Un'unità produttiva è tecnicamente efficiente se un aumento di un qualsiasi output richiede la riduzione di almeno un altro output o l'aumento di almeno un input e la riduzione di un qualsiasi input richiede l'aumento di almeno un altro input o la riduzione di almeno un output."*

Nei lavori pionieristici di Debreu (1951), Shephard (1953) e Farrel (1957) ([11], [26], [15]) vengono proposte le prime misure quantitative di efficienza, basate sulla *"massima diminuzione equiproporzionale di tutti gli input per ottenere gli stessi output, o il massimo aumento equiproporzionale di tutti gli output consentito dagli stessi input."* Già in queste prime fasi si può notare come l'efficienza sia estremamente legata a definizioni e criteri di valutazione. Nella figura 4 è riportato un esempio di quali input possano consentire un determinato livello di output; si tratta di una regione convessa limitata da un isoquanto lineare a tratti detta *insieme di input* e denotata con $X(y)$.

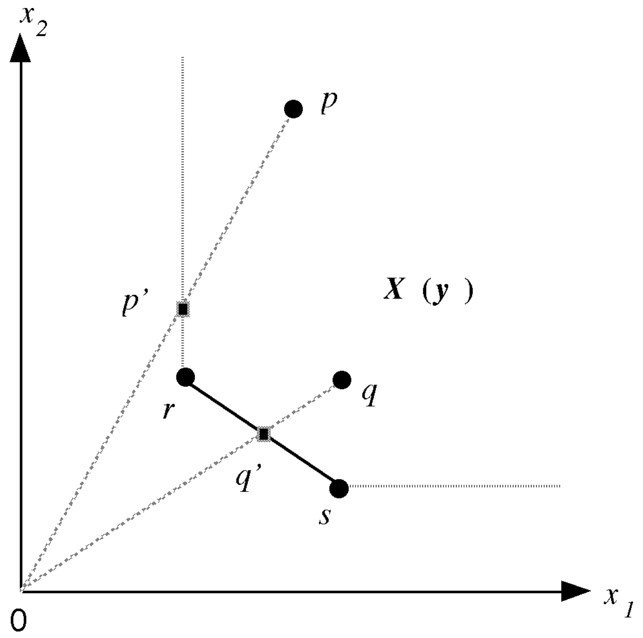

FIGURA 4. Efficienza tecnica rispetto all'insieme di input

Un impiego di risorse rappresentato dai punti p e q non può essere considerato efficiente per ottenere l'output y, in quanto quantità minori di risorse possono consentire la stessa produzione; sia p che q possono essere infatti ridotti nel modo proposto da Debreu e Farrel fino ad individuare le configurazioni efficienti p' e q'. Mentre però q' rimane efficiente anche secondo la definizione di Koopmans, non così p' che può essere ulteriormente ridotto fino alla configurazione r.

Alle definizioni di tipo tecnologico originate dalle proposte seminali sopra riportate, si sono via via aggiunte altre valutazioni di efficienza, comprensive anche di considerazioni economiche. Si supponga che le risorse siano descritte da un vettore x ed abbiano costi unitari, descritti dal vettore c; un utilizzo efficiente degli input x per la produzione degli output y può essere individuato dalla condizione:

$$Z(y, c) = \min\{cx \mid y \text{ può essere ottenuto da } x\}$$

Nella figura 5 sono illustrati, in questo caso più generale, i contributi alla mancata di efficienza dovuti a motivi tecnici (riduzione equiproporzionale pp' e slack $p'r$) ed a motivi allocativi (riduzione rs).

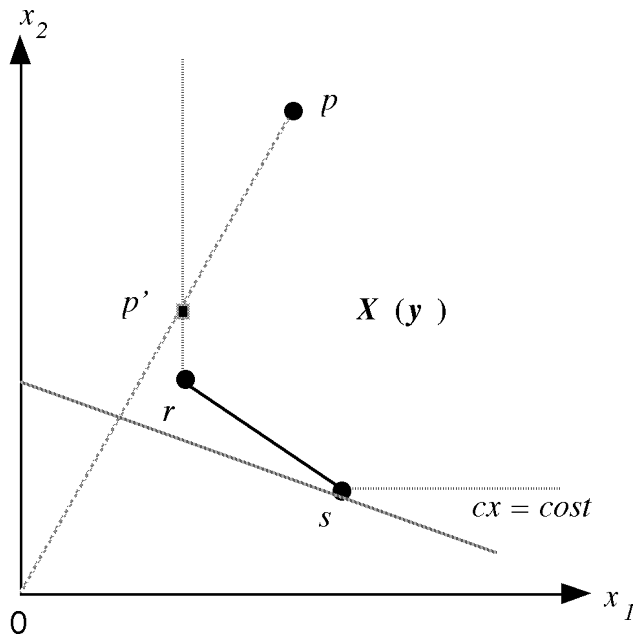

FIGURA 5. Efficienza allocativa rispetto all'insieme di input

Solamente alla fine degli anni '70 hanno avuto inizio due consistenti filoni di ricerca oggi noti con le denominazioni di *metodi parametrici* e *metodi non parametrici* per la determinazione di *frontiere efficienti.* Nei metodi parametrici, essenzialmente di tipo econometrico, viene ancora conservata l'idea di un legame analitico fra le quantità che intervengono nei processi produttivi. Tale legame però rappresenta non già una relazione deterministica fra quanto viene prodotto e quanto è stato necessario per produrlo, bensì la miglior produzione possibile a fronte di una data disponibilità di risorse. I valori reali si discostano da tale frontiera sia a causa di azioni non considerate nel modello, sia a causa di inefficienze nell'utilizzo delle risorse. I modelli sono essenzialmente di tipo stocastico, in cui i coefficienti (parametri) della funzione di frontiera vengono stimati, assieme agli scostamenti da tale comportamento ideale, dai dati sperimentali.

Nei metodi non parametrici, legati a tecniche di Programmazione Matematica ormai conosciute con la sigla *DEA*, acronimo di *Data Envelopment Analysis*, viene conservata l'idea di miglior produzione possibile a fronte di date disponibilità; ma la costruzione della frontiera e gli scostamenti dalle produzioni ideali vengono stabiliti sulla base di distanze geometriche deterministiche fra osservazioni sperimentali.

Per una esauriente panoramica su quanto proposto nella letteratura si rimanda ai saggi critici di Lewin e Lovell, [21], Fried, Lovell e Smith, [16] e più recentemente al volume dell'European Journal of Operational Research [12] interamente dedicato all'argomento.

4. Metodi econometrici per l'analisi di efficienza

I metodi econometrici, detti anche parametrici per la tipica presenza di forme analitiche a coefficienti incogniti da stimare, possono prevedere sia formulazioni statiche, descriventi cioè situazioni ad un determinato istante di tempo, che di tipo evolutivo (*panel data*); entrambe possono poi includere considerazioni tecnologiche e/o economiche. Nella *situazione* più semplice un output y viene prodotto usando un insieme di input $x = (x_1, x_2, \dots, x_n)$ disponibili ai prezzi $p = (p_1, p_2, \dots, p_n)$ e può consentire un ricavo unitario q, in un ambiente operativo descritto dalle variabili esogene $z = (z_1, z_2, \dots, z_k)$.

In tali situazioni l'efficienza del processo produttivo viene generalmente descritta da *frontiere di produzione*, cioè funzioni analitiche mirate a descrivere comportamenti ottimali sia sotto l'aspetto puramente tecnologico sia includendo valutazioni di costo, ricavo o consuntivi di profitto. La lista che segue riporta i principali modelli usati, ciascuno ipotizzando un ambiente operativo descritto dalle variabili z e dai parametri β caratteristici del processo produttivo in esame:

- $y = y(x, z; \beta)$**:** massimo output ottenibile dagli input x;
- $c = c(y, p, z; \beta)$**:** costo minimo per produrre l'output y acquistando gli input al prezzo p;
- $r = r(x, q, z; \beta)$**:** ricavo massimo consentito dagli input x, se il ricavo unitario di vendita per l'output è q (tale modello differisce da quello di pura produzione solo in presenza di più output);
- $p = p(q, p, z; \beta)$**:** profitto massimo acquistando gli input al prezzo p ed ottenendo un ricavo unitario q per l'output.

Mentre i modelli basati su costi, ricavi o profitti possono facilmente trattare più output, non così le frontiere di pura produzione, dove la variabile dipendente coincide con la quantità prodotta. Nel caso di output multipli, si possono comunque utilizzare frontiere di produzione relativamente ad indici, ottenuti mediante opportune aggregazioni degli output. Nel seguito si farà riferimento alla forma più semplice, in cui un legame analitico ipotizza la massima produzione ottenibile (descritta dalla variabile dipendente y) da determinate risorse (descritte dalle variabili indipendenti x), in un fissato ambiente operativo (descritto dalle variabili esogene z), utilizzando un dato processo produttivo (descritto dai parametri β):

$$y_{max} = f(x, z; \beta)$$

Quanto illustrato può comunque essere esteso anche ad altri modelli e situazioni; per una trattazione generale sui vari tipi e proprietà delle misure e delle frontiere di efficienza si veda, ad esempio, Färe, Grosskopf e Lovell [13].

4.1. **Frontiere deterministiche.** Supponendo che un eventuale termine di inefficienza entri nel modello in modo moltiplicativo, ogni unità produttiva sarà in grado di ottenere un output del tipo:

$$y_i = f(x_i, z_i; \beta)\varepsilon_i; \qquad 0 \leq \varepsilon_i \leq 1,\ (i = 1, \dots, N), \tag{1}$$

dove l'indice i identifica l'unità produttiva e la quantità stocastica ε_i denota di quanto tale unità si discosta da un comportamento ottimale. Nel modello (1) la quantità $f(x, z; \beta)$ è detta frontiera deterministica, poiché la componente stocastica è interamente contenuta nel termine di (in)efficienza ε. Nella maggioranza delle applicazioni le forme ipotizzate sono linearizzabili mediante trasformazioni opportune, come ad esempio nel caso di una funzione di *Cobb-Douglas*:

$$y = A(\prod_k x_k^{\beta_k})\varepsilon$$

che dopo trasformazione logaritmica diventa:

$$\ln y = \alpha + \sum_k \beta_k \ln x_k - u$$

dove $\alpha = \ln A$ e la quantità non negativa $u = -\ln \varepsilon$ può essere intesa come "scostamento" dalla massima efficienza possibile. Applicando a tali forme il metodo dei minimi quadrati, con vincoli di non negatività sulla u, si ottengono delle stime per i coefficienti β le cui proprietà statistiche sono intimamente legate al tipo di distribuzione delle componenti stocastiche. Così pure il metodo di massima verosimiglianza fornisce risultati strettamente dipendenti da tali distribuzioni. Se si aggiunge la soggettività sulla forma analitica delle frontiere di produzione, appare evidente la necessità di accurate valutazioni sulle misure di efficienza ottenute dai valori stimati delle u.

4.2. **Frontiere stocastiche.** Nel modello deterministico di frontiera, ogni scostamento dal comportamento ottimale viene attribuito ad inefficienza. In realtà molti fattori di perturbazione possono non essere controllabili dall'entità di cui si vuole misurare l'efficienza, né possono essere valutati quali variabili esogene. Queste considerazioni hanno suggerito la proposta di modelli in grado di individuare quanto attribuibile al caso e quanto invece ad effettiva inefficienza. Supponendo ancora le componenti stocastiche di tipo moltiplicativo, ogni unità produttiva sarà in grado di ottenere un output del tipo:

$$y_i = f(x_i, z_i; \beta)\pi_i\varepsilon_i; \qquad 0 \leq \varepsilon_i \leq 1,\ (i = 1, \dots, N),$$

dove la quantità $f(x, z; \beta)\pi$ diventa la frontiera stocastica di produzione, mentre ε rappresenta l'(in)efficienza. Anche in questo caso le forme ipotizzate sono linearizzabili mediante opportune trasformazioni; ad esempio una funzione di *Cobb-Douglas*:

$$y = A(\prod_k x_k^{\beta_k})\pi\varepsilon$$

dopo trasformazione logaritmica diventa:

$$\ln y = \alpha + \sum_k \beta_k \ln x_k + v - u$$

dove $\alpha = \ln A$, $v = \ln \pi$ rappresenta globalmente le perturbazioni casuali, mentre la quantità non negativa $u = -\ln \varepsilon$ è ancora intesa come "scostamento" dalla massima efficienza possibile, questa volta rappresentata dalla frontiera stocastica. Una prima proposta di tale approccio può essere trovata in Aigner, Lovell e Schmidt

[1], in cui sotto le ipotesi: *i)* di normalità per la componente v, *ii)* di distribuzione semi-normale (positiva) per u e *iii)* di indipendenza tra tali variabili ed i regressori x_k, si è potuto quantizzare separatamente, anche se globalmente, le due componenti stocastiche. L'utilizzo poi di serie storiche (*panel data*), invece di dati singoli di produzione, ha consentito di valutare sia la componente erratica che quella di (in)efficienza non solo globalmente, ma anche nei riguardi di ogni singola unità produttiva. Si veda ad esempio Battese e Coelli [5], come nell'ipotesi di invarianza temporale della componente di (in)efficienza, si possano corredare le suddette valutazioni anche delle principali proprietà statistiche normalmente richieste agli stimatori.

5. Metodi di programmazione matematica per l'analisi di efficienza

L'approccio matematico alla costruzione dell'insieme ammissibile di produzione, spesso indicato con la denominazione di *Data Envelopment Analysis* (*DEA*), si è sviluppato in seguito al lavoro di Charnes, Cooper e Rhodes [8], nel quale gli Autori hanno riformulato in termini di problemi di Programmazione Lineare (PL) le definizioni di efficienza di Koopmans e Farrell.

La *DEA* parte dalla conoscenza di un campione rappresentativo di un dato settore produttivo, $S_0=\{s_1, s_2, \dots, s_N\}$, dove le s_i sono unità produttive o decisionali (dette "Decision-Making Units") che si assume siano *omogenee*, cioè caratterizzate dal fatto che producono uno stesso insieme di beni o servizi, utilizzando lo stesso tipo di risorse e uguali tecnologie (o comunque molto simili). Indicando con n il numero di risorse che le unità produttive usano e con m il numero di beni o servizi che esse forniscono, possiamo rappresentare l'unità operativa s_i per mezzo della coppia (x_i, y_i), con $x_i \in \Re^n$e $y_i \in \Re^m$. I vettori x_i e y_i verranno indicati come l'*input* e l'*output* della unità produttiva e le loro componenti rappresentano le quantità delle risorse utilizzate e dei beni forniti.

L'*insieme ammissibile di produzione*, S, è l'insieme delle coppie (x, y) (*coppie ammissibili*) che rappresentano unità produttive possibili. Tale insieme rappresenta la popolazione di cui S_0 è un campione rappresentativo ed è ottenuto *inviluppando* l'insieme dei dati osservati (da cui la denominazione "data envelopment analysis"). In tale situazione i modelli *DEA* calcolano un punteggio di efficienza per ciascuna unità produttiva sulla base di un valore di "distanza" dalla frontiera di S.

Uno dei principali vantaggi dell'approccio della programmazione matematica rispetto all'approccio econometrico è che l'insieme S costruito è *non parametrico*, cioè la sua definizione non richiede la specificazione di una forma funzionale parametrica per la descrizione della tecnologia. Un altro vantaggio è che il metodo permette di trattare in modo semplice, sia dal punto di vista teorico che computazionale, il caso di tecnologie multi-input/multi-ouput. Uno dei principali svantaggi deriva invece dal fatto che la frontiera costruita è *deterministica* e, di conseguenza, l'intera deviazione di un'osservazione da essa viene attribuita all'inefficienza, senza possibilità di tener conto di errori di misurazione, omissioni di variabili, fattori di disturbo e così via.

5.1. **Insiemi ammissibili di produzione.**

L'approccio *DEA* parte dalla costruzione dell'insieme ammissibile di produzione S a partire dai dati disponibili, cioè dal campione S_0. Forniamo nel seguito alcuni dei postulati più usati per la definizione di S:

P0: (*ammissibilità delle osservazioni*)
$S_0 \subseteq S$;

P1: (*ammissibilità dei punti dominati*)
$(x, y) \in S \Rightarrow [(x', y) \in S, \forall x' \geq x] \& [(x, y') \in S, \forall y' \leq y]$;

P2.*i*: (*linearità*)
$[(x_i, y_i), (x_j, y_j) \in S] \& [\lambda_i, \lambda_j \geq 0] \Rightarrow \lambda_i(x_i, y_i) + \lambda_j(x_j, y_j) \in S$;

P2.*ii*: (*convessità*)
$[(x_i, y_i), (x_j, y_j) \in S] \& [\lambda_i + \lambda_j = 1, \lambda_i, \lambda_j \geq 0] \Rightarrow \lambda_i(x_i, y_i) + \lambda_j(x_j, y_j) \in S$;

P3: (*minimalità*)
S è il minimo insieme tra quelli che soddisfano i postulati.

Analizziamo ora più in dettaglio i postulati elencati ed il loro significato. Il primo (P0) indica semplicemente che i punti osservati devono appartenere alla regione ammissibile; in effetti sono gli unici punti di cui l'ammissibilità è sperimentalmente provata. Il postulato P1 corrisponde ad una osservazione che possiamo assumere in molti contesti ragionevole: se con un dato insieme di risorse si riesce a produrre una data quantità di output, è sempre possibile aumentare l'uso delle risorse senza diminuire la quantità di output prodotto, oppure diminuire l'output mantenendo lo stesso livello di risorse consumate; cioè è possibile realizzare dei livelli di efficienza più bassi. Con questo postulato si trascurano gli effetti di congestione che fanno ad esempio sì che aumentando oltre un certo livello la quantità di fertilizzanti per unità di terra coltivata si finisce per ridurne la produzione. Nel seguito comunque, salvo se esplicitamente affermato il contrario, noi assumeremo i postulati P0 e P1 validi.

I due postulati P2 sono invece più critici dal punto di vista del loro significato economico e comunque vanno considerati in alternativa tra di loro. Essi hanno in comune il fatto che implicano la convessità della regione ammissibile: il primo indica che se due coppie input-output sono ammissibili lo sono anche tutte quelle che appartengono al cono da esse individuate, mentre il secondo indica che tutto il segmento che unisce due coppie ammissibili è ammissibile. I coefficienti λ possono essere interpretati come *livelli di intensità* ai quale i corrispondenti processi produttivi sono (oppure potrebbero essere) attivati.

Infine il postulato P3 esprime il fatto che non abbiamo motivo di considerare ammissibili punti che non siano strettamente implicati dai postulati che abbiamo assunto come validi.

Il significato economico dei postulati sarà discusso in dettaglio nel seguito, quando esamineremo diversi tipi di insiemi ammissibili ottenibili con diverse scelte di postulati.

Nel seguito, seguendo la trattazione di Tulkens e Vanden Eeckaut [27], presenteremo tre diversi insiemi ammissibili di produzione caratterizzati dal fatto che ciascuno può essere ottenuto dal precedente attraverso un rilassamento della sua descrizione matematica.

5.1.1. *Caso 1 (dominanza).* Consideriamo innanzitutto il caso in cui si assumano validi i soli postulati P0, P1 e P3. È questo in un certo senso l'insieme minimo di postulati e corrisponde ad un atteggiamento conservativo nei riguardi dell'ammissibilità: vengono accettati come ammissibili solo i punti osservati e tutti i punti da essi dominati. Il tipo di insieme ammissibile risultante, detto Free Disposal Hull (FDH), è indicato nell'esempio della figura 6, in cui si hanno quattro unità produttive (a, b, c, d), un input e un output.

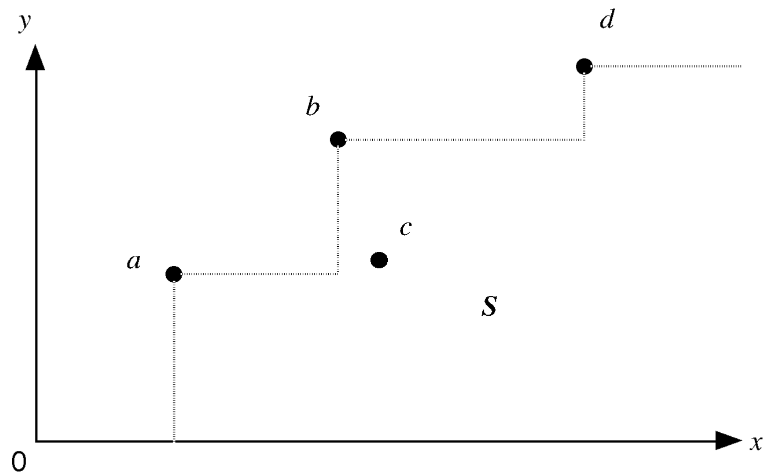

Figura 6. Insieme ammissibile Free Disposal Hull

Dal punto di vista matematico, l'insieme S è in questo caso definito come segue:

$$S = \{(x,y) \in \Re^n_+ \times \Re^m_+ : \\ x \geq \sum_{i=1}^{N} \lambda_i x^i, y \leq \sum_{i=1}^{N} \lambda_i y^i, \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \in \{0,1\}, i = 1, ..., N\}.$$

L'insieme S così definito è in un certo senso il minimo insieme che può essere derivato dal campione: gli unici punti inseriti in S, oltre a quelli già contenuti in S_0, sono quelli dominati da questi ultimi.

5.1.2. *Caso 2 (rendimenti di scala variabili).* Consideriamo ora il caso in cui si aggiunga il postulato di *convessità*, P2.*ii*, ai postulati P0, P1, P3. In questa nuova situazione l'insieme ammissibile per l'esempio considerato è rappresentato nella figura 7.

La descrizione matematica di S è:

$$S = \{(x,y) \in \Re^n_+ \times \Re^m_+ : \\ x \geq \sum_{i=1}^{N} \lambda_i x^i, y \leq \sum_{i=1}^{N} \lambda_i y^i, \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \in [0,1], i = 1, ..., N\}.$$

Si osservi che l'insieme S costruito è un poliedro con $k \leq N$ vertici corrispondenti ad un sottoinsieme delle unità decisionali osservate. Dal punto di vista economico, la frontiera di S, lineare a tratti, è detta esibire rendimenti di scala variabili (Variable Return to Scale o VRS); più precisamente, nella prima parte dell'insieme un aumento proporzionale degli input di una unità ammissibile comporta un aumento

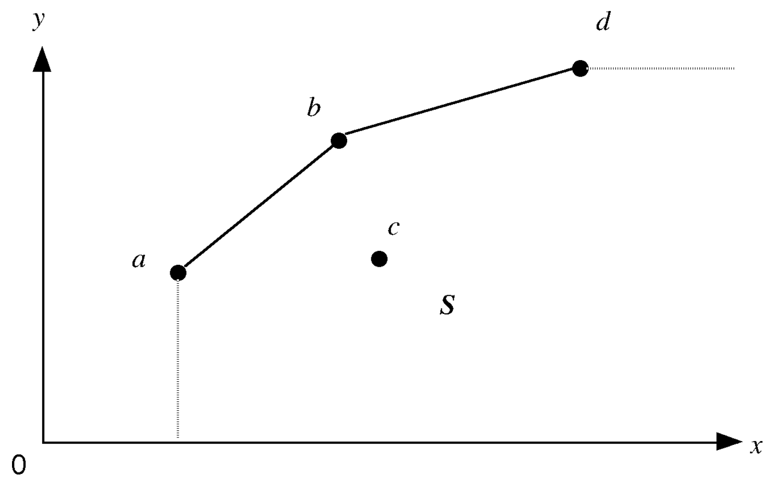

FIGURA 7. Insieme ammissibile a rendimenti variabili (VRS)

più che proporzionale degli output, mentre nella seconda parte dell'insieme un aumento proporzionale degli input è seguito da un aumento meno che proporzionale degli output. Rimandiamo a [3], [4] per la definizione e la stima dei rendimenti di scala.

5.1.3. *Caso 3 (rendimenti di scala costanti).* Se si sostituisce il postulato di *convessità*, P2.*ii*, con il postulato di *linearità*, P2.*i*, nel modello precedente, si ottiene l'insieme ammissibile della figura 8.

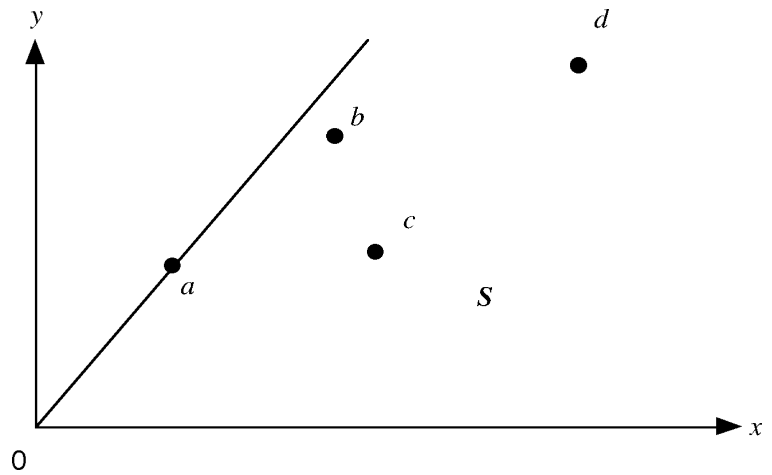

FIGURA 8. Insieme ammissibile a rendimenti costanti (CRS)

La descrizione matematica di tale insieme è:

$$S = \{(x,y) \in \Re^n_+ \times \Re^m_+ : x \geq \sum_{i=1}^{N} \lambda_i x^i, y \leq \sum_{i=1}^{N} \lambda_i y^i, \lambda_i \geq 0, i = 1, ..., N\}.$$

La frontiera di S in questo caso presenta rendimenti di scala costanti (Constant Return to Scale o CRS), poiché si assume che ad un aumento proporzionale degli input di una unità decisionale, segua un aumento proporzionale degli output, indipendentemente dalla dimensione sotto la quale si trova ad operare l'unità stessa. La proporzionalità ammessa per i coefficienti di intensità λ implica la conicità dell'insieme S.

Si verifica facilmente che i tre insiemi ammissibili ottenuti sono l'uno contenuto nell'altro ed impongono condizioni sempre più forti sulla tecnologia; ad esempio l'insieme a rendimenti di scala costanti assume l'additività, la moltiplicatività e la divisibilità dei processi produttivi. In particolare, l'ipotesi di additività presuppone che gli output di ciascuna unità siano determinati unicamente dai suoi input, escludendo la presenza di fenomeni di interazione tra le varie unità produttive. Tuttavia tale ipotesi è in pratica difficilmente verificata poiché i moderni processi di produzione sono spesso legati da intricate relazioni di complementarietà e sostituibilità e l'utilizzo congiunto di due processi produttivi può dar luogo ad economie o diseconomie di scala.

L'opportunità di utilizzare l'uno o l'altro insieme ammissibile di produzione va valutata attentamente sulla base di constatazioni empiriche.

5.2. **Valutazione di efficienza.** La definizione dell'insieme ammissibile di produzione S permette di distinguere tra unità che operano in modo *tecnicamente* efficiente e unità inefficienti, secondo la definizione di Koopmans riportata nel paragrafo 3. Sono infatti efficienti tutte le coppie ammissibili che appartengono al sottoinsieme della frontiera di S definito come segue:

$$Eff(S) \ = \{(x,y) \in S \mid (x',y') \neq (x,y),\ x' \leq x,\ y' \geq y \Rightarrow (x',y') \notin S\}$$

Per quantificare l'inefficienza di ciascuna unità produttiva è necessario introdurre una *misura* o *indice* di efficienza. Ad esempio, si consideri il caso mostrato dalla figura 9 dove è rappresentato un insieme ammissibile a rendimenti variabili costruito a partire da cinque unità osservate.

Appare evidente che l'efficienza tecnica delle unità c ed e può essere migliorata, indipendentemente da costi o prezzi di mercato delle risorse, essendo possibile produrre le stesse quantità di output impiegando meno input. Il valore del *miglioramento* di efficienza ottenibile è in qualche modo proporzionale alla "distanza" dell'unità considerata dalla frontiera $Eff(S)$. Al contrario, le unità a, b e d non possono migliorare la loro efficienza, rimanendo nello stesso tempo ammissibili.

L'esempio precedente introduce, informalmente, una misura di efficienza. Si proverà ora a definire il concetto di efficienza in modo più formale, elencando un insieme di proprietà atte a caratterizzarla. Sia $E : S \longrightarrow \Re$ una misura di efficienza. Un insieme di proprietà di E proposte nella letteratura economica ([14], [23], [10]) è, $\forall (x,y) \in S$:

E1: $0 \leq E(x,y) \leq 1$;
E2: $E(x,y) = 1 \Leftrightarrow (x,y) \in Eff(S)$;
E3: $(x',y') \neq (x,y),\ x' \leq x,\ y' \geq y \Rightarrow E(x',y') > E(x,y)$;
E4: E è indipendente dall'unità di misura scelta.

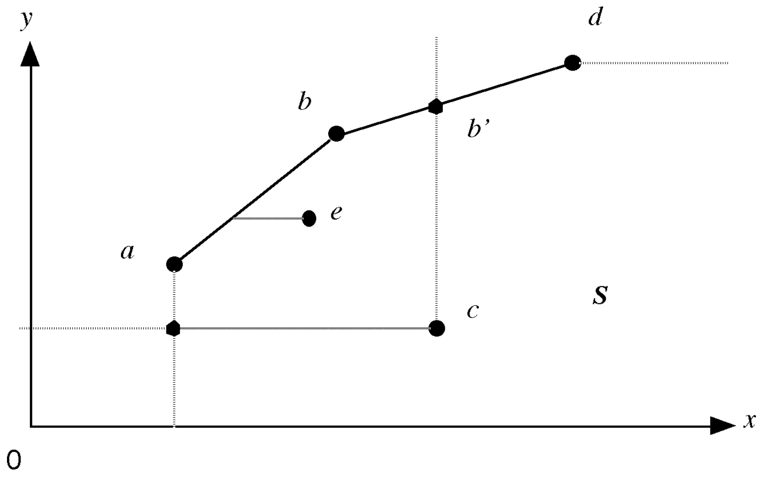

FIGURA 9. Misurazione dell'efficienza tecnica

La proprietà E1 richiede una normalizzazione della misura di efficienza, mentre E3 (*monotonicità*) comporta che l'efficienza sia decrescente all'aumentare dell'impiego degli input e crescente all'aumentare della produzione di output.

Nei modelli *DEA* si definisce una funzione D che quantifica la "distanza" tra i punti di S e la sua frontiera di efficienza, e, a partire da essa, una misura di efficienza che soddisfa le proprietà E1, E2, E3 e E4. I principali modelli *DEA* proposti in letteratura e descritti nel paragrafo 5.3 si caratterizzano per il *modo* in cui la funzione distanza è definita.

5.3. Principali modelli *DEA*.

5.3.1. *Modello additivo.* Il modello additivo definisce la distanza di una data unità (x_0, y_0) da $Eff(S)$ calcolando la massima distanza *Manhattan* o L_1 tra l'unità s_0 e le unità produttive ammissibili che la dominano. Assumendo che l'insieme di produzione S sia quello a rendimenti di scala variabili (si veda il paragrafo 5.1.2), la distanza è ottenuta risolvendo la seguente coppia duale di programmi lineari:

Primale Additivo

$$\min_{\lambda, s^+, s^-} z_0 = -\mathbf{1}^T s^+ - \mathbf{1}^T s^-$$

s.t.

$$\sum_{i=1}^{N} \lambda_i y_i - s^+ = y_0$$

$$\sum_{i=1}^{N} \lambda_i x_i + s^- = x_0$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

$$\lambda, s^+, s^- \geq 0$$

Duale Additivo

$$\max_{\mu, \nu, u} w_0 = \mu^T y_0 - \nu^T x_0 + u$$

s.t.

$$\mu^T y_i - \nu^T x_i + u \leq 0 \quad i = 1, \ldots, N$$

$$\mu^T \geq \mathbf{1}^T$$

$$\nu^T \geq \mathbf{1}^T$$

dove $\mathbf{1}$ indica il vettore di tutti uni e di dimensione opportuna. Entrambi i problemi sono non vuoti ed hanno quindi ottimo finito. È possibile ottenere soluzioni sia del primale che del duale applicando tecniche note di PL.

Il valore ottimo, $z_0^* = -\mathbf{1}^T s^{*+} - \mathbf{1}^T s^{*-}$, del primale misura la massima distanza *Manhattan* dalla frontiera efficiente e se tale valore è nullo, cioè $z_0^* = w_0^* = 0$, allora l'unità s_0 che si valuta è efficiente; al contrario, un valore ottimo diverso da zero è indice di inefficienza, essendo la distanza da $\mathit{Eff}(S)$ positiva. Nel caso in cui s_0 risulti inefficiente, le variabili di scarto ottime non nulle indicano l'ammontare e la *causa* dell'inefficienza, nel senso che se la variabile di scarto ottima relativa alla risorsa r è positiva, allora tale risorsa è una fonte potenziale di inefficienza.

Per stimare la distanza dalla frontiera di ogni unità in S_0 occorre risolvere N programmi lineari ponendo successivamente $(x_0, y_0) = (x_i, y_i)$, $i = 1, \ldots, N$.

Una misura di efficienza ottenibile dal valore ottimo del modello additivo è la seguente:

$$E(x_0, y_0) = \frac{1}{1 + \mathbf{1}^T s^{+*} + \mathbf{1}^T s^{-*}} \tag{2}$$

La misura in (2) soddisfa le proprietà postulate E1-E3, ma non la E4 in quanto dipende dall'unità di misura dei dati. Ciò è dovuto al fatto che i coefficienti associati alle variabi di scarto (fattori di scala) nella funzione obiettivo primale, $z_0 = -\mathbf{1}^T s^+ - \mathbf{1}^T s^-$, sono tutti pari a 1. Implicita in tale scelta è l'assunzione che, per l'unità produttiva che si sta valutando, il valore marginale di ciascuna unità di scarto sia lo stesso. In [2] sono proposti due modi di fissare i fattori di scala che rendono la misura di efficienza (2) invariante rispetto all'unità di misura dei dati, mentre in [10] sono definite misure di efficienza alternative alla (2) e ricavabili sempre a partire dalle soluzioni ottime del modello additivo.

Sia $(\lambda^*, s^{*+}, s^{*-})$ una soluzione ottima del problema additivo primale, allora la seguente coppia di produzione ammissibile:

$$(\tilde{x}, \tilde{y}) = (x_0 - s^{*-}, y_0 + s^{*+}) = (\sum_{i=1}^{N} \lambda_i^* x_i, \sum_{i=1}^{N} \lambda_i^* y_i)$$

individua una unità "di riferimento" efficiente, che domina l'unità in esame s_0 ed è a distanza *Manhattan* massima. Osserviamo come tale unità in generale non sarà una di quelle appartenenti al campione S_0, ma piuttosto una *unità virtuale* corrispondente ad un punto di S.

La figura 10 si riferisce ad una insieme di produzione a rendimenti variabili costruito a partire da sei unità produttive osservate, le quali impiegano un fattore di input per produrre un bene.

La frontiera efficiente è rappresentata dalla spezzata $[a, b, d]$, mentre le unità produttive corrispondenti ai punti c, e, f, sono inefficienti. Le linee tratteggiate indicano il *cammino* lungo il quale è misurata la distanza *Manhattan* nella direzione nord-ovest. Si osservi che il punto f, pur appartenendo alla frontiera di S, è inefficiente essendo dominato da d. Il punto e è relativo ad un altro modo inefficiente di produrre e il modello additivo individua un punto efficiente "di riferimento", combinazione convessa di b e d, ma non osservato.

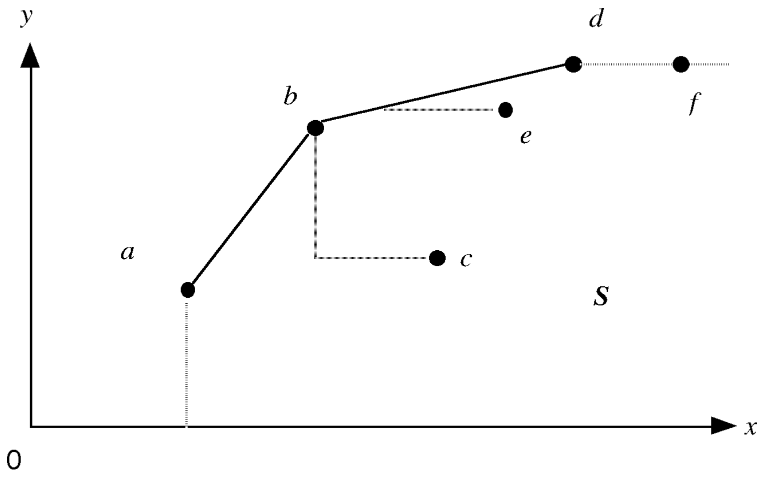

FIGURA 10. Utilizzo della distanza Manhattan

5.3.2. *Modelli orientati.* Il modello additivo suggerisce un modo per migliorare le prestazioni di una unità inefficiente, individuando una unità "di riferimento", efficiente, ottenuta dalla prima diminuendo alcuni input e/o aumentando alcuni output. Nell'ipotesi in cui gli output siano esogeni all'unità in esame, o comunque non facilmente modificabili, è opportuno definire una funzione di distanza differente, che dia maggiore enfasi alla diminuzione degli input.

I modelli *DEA* orientati alla minimizzazione degli input (input-oriented) misurano l'efficienza di una data unità s_0, calcolando la massima contrazione equiproporzionale degli input necessaria per raggiungere la frontiera efficiente. In tal modo la riduzione degli eccessi di consumo di input è ottenuta conservando lo stesso rapporto (mix tecnologico) tra i fattori impiegati dall'unità produttiva.

La funzione distanza in questo caso è definita rispetto all'insieme di input:

$$\mu^* = \max\{\mu : (1-\mu)x_0 \in X(y_0)\} \tag{3}$$

Lo scalare $\theta^* = 1 - \mu^*$ viene detto *tasso di riduzione radiale* e $\theta^* x_0$ rappresenta il minimo vettore degli input che produce lo stesso output, mantenendo invariate le proporzioni con cui i diversi fattori di produzione sono utilizzati. Il vettore ridotto $\theta^* x_0$ non necessariamente appartiene a *Eff*(S), essendo possibili ulteriori riduzioni non radiali di input e/o output. Per tener conto di tutte le possibili inefficienze, si ricorre alla seguente funzione distanza:

$$D(x_0, y_0) = \mu^* + \varepsilon(\mathbf{1}^T s^{+*} + \mathbf{1}^T s^{-*}), \tag{4}$$

dove μ^* è definito in (3), mentre le variabili di scarto sono soluzioni ottime del problema di massimizzazione della distanza *Manhattan* rispetto al punto $(\theta^* x_0, y_0)$:

$$\max\{\mathbf{1}^T(\theta^* x_0 - x) + \mathbf{1}^T(y - y_0) \mid (x, y) \in S \text{ e } x \leq \theta^* x_0,\ y \geq y_0\}.$$

La quantità $\varepsilon > 0$ che compare nella definizione di distanza (4) è detta valore "non archimedeo" ([6], [7]). "ε" simbolizza un elemento che è definito essere più piccolo di ogni numero reale positivo e permette di esprimere l'importanza prioritaria data allo spostamento radiale, rispetto alla somma degli scarti.

Una possibile misura di efficienza definibile a partire dalla distanza (4) è la seguente:

$$E(x_0, y_0) = 1 - D(x_0, y_0) = \theta^* - \varepsilon(\mathbf{1}^T s^{+*} + \mathbf{1}^T s^{-*}). \tag{5}$$

Si dimostra che la misura di efficienza definita in (5) soddisfa tutte le proprietà E1-E4, mentre la sola misura di efficienza radiale, θ^*, detta misura di efficienza di Farrell, non soddisfa la E2[1].

Il modello *DEA BCC* (Banker, Charnes, Cooper [4]), input-oriented, calcola la misura di efficienza (5) di una data unità s_0, assumendo ancora una volta un insieme ammissibile di produzione a rendimenti variabili. La sua formulazione risulta dalla seguente coppia di programmi lineari:

Primale BCC input-oriented

$$\begin{aligned}
\min_{\theta,\lambda,s^+,s^-} \; & z_0 = \theta - \varepsilon\mathbf{1}^T s^+ - \varepsilon\mathbf{1}^T s^- \\
\text{s.t.} \; & \\
& \sum_{i=1}^{N} \lambda_i y_i - s^+ = y_0 \\
& \sum_{i=1}^{N} \lambda_i x_i + s^- = \theta x_0 \\
& \sum_{i=1}^{N} \lambda_i = 1 \\
& \lambda, s^+, s^- \geq 0
\end{aligned}$$

Duale BCC input-oriented

$$\begin{aligned}
\max_{\mu,\nu,u} \; & w_0 = \mu^T y_0 + u \\
\text{s.t.} \; & \\
& \nu^T x_0 = 1 \\
& \mu^T y_i - \nu^T x_i + u \leq 0 \\
& i = 1, \ldots, N \\
& \mu^T \geq \varepsilon\mathbf{1}^T \\
& \nu^T \geq \varepsilon\mathbf{1}^T
\end{aligned}$$

Considerando lo stesso indice di efficienza (5), ma un insieme di produzione a rendimenti costanti (si veda il paragrafo 5.1.3, caso 3) si ricava la formulazione del modello *CCR* input-oriented, proposto inizialmente da Charnes, Cooper e Rhodes

[1] In realtà Farrell per distinguere tra punti di frontiera efficienti e non efficienti, aveva introdotto il concetto di "punto all'infinito", che nei modelli *DEA* ha portato all'introduzione della quantità ε.

(1978) (si veda [8])

Primale CCR input-oriented	**Duale CCR input-oriented**
$\min_{\theta,\lambda,s^+,s^-} z_0 = \theta - \varepsilon\mathbf{1}^T s^+ - \varepsilon\mathbf{1}^T s^-$	$\max_{\mu,\nu} w_0 = \mu^T y_0$
s.t.	s.t.
$\sum_{i=1}^{N} \lambda_i y_i - s^+ = y_0$	$\nu^T x_0 = 1$
$\sum_{i=1}^{N} \lambda_i x_i + s^- = \theta x_0$	$\mu^T y_i - \nu^T x_i \leq 0 \quad i = 1, \ldots, N$
$\lambda, s^+, s^- \geq 0$	$\mu^T \geq \varepsilon\mathbf{1}^T$
	$\nu^T \geq \varepsilon\mathbf{1}^T$

A differenza dil modelli a rendimenti variabili, nel primale del modello CCR non compare il vincolo di convessità ($\sum_{i=1}^{N} \lambda_i = 1$) e nel duale non è piú presente la variabile non vincolata u.

Applicando i modelli radiali BCC e CCR orientati alla minimizzazione degli input, la condizione necessaria e sufficiente affinché s_0 risulti efficiente (cioè a distanza zero da *Eff* (S)) è che:

$$\theta^* = 1 \text{ e } s^{*+} = s^{*-} = 0.$$

Se l'unità s_0 non è efficiente (cioè $\theta^* < 1$ oppure $\theta^* = 1$ e non tutte le variabili di scarto nulle), allora dalla soluzione ottima del primale si individua la seguente coppia di produzione ammissibile "di riferimento", che domina (x_0, y_0):

$$(\tilde{x}, \tilde{y}) = (\theta^* x_0 - s^{*-}, y_0 + s^{*+}) = (\sum_{i=1}^{N} \lambda_i^* x_i, \sum_{i=1}^{N} \lambda_i^* y_i).$$

Si consideri, ad esempio, la figura 11 che illustra un insieme di input $X(y_0)$.

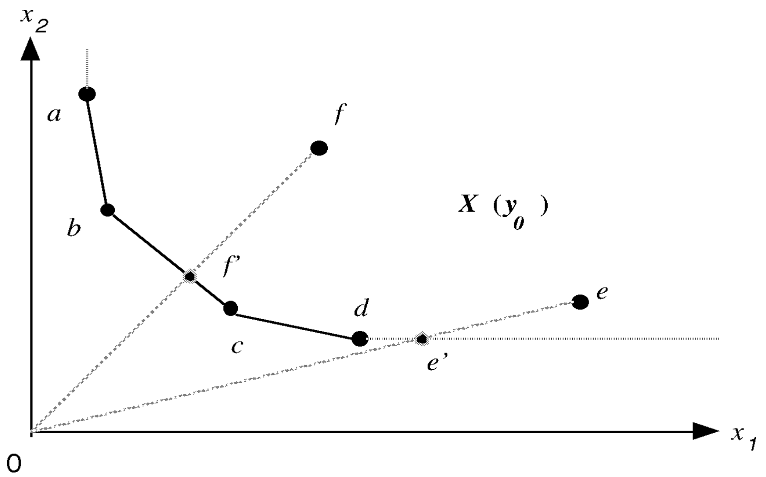

FIGURA 11. Utilizzo della distanza radiale input-oriented

La massima contrazione ottenibile per f, mantenendo l'ammissibilità della coppia di produzione, è rappresentata, geometricamente, dal rapporto tra i segmenti

$[f', f]$ e $[0, f]$. Analogamente, la massimizzazione della contrazione degli input per il punto e, porta all'individuazione del punto di frontiera e'. Si osservi che il punto e' non è efficiente, essendo possibile ottenere un ulteriore miglioramento, diminuendo il solo input 1.

Il problema duale *CCR* input-oriented ha una formulazione equivalente in forma frazionaria. Infatti, assumendo che ogni unità produttiva osservata utilizzi quantità di fattori di input positive, il duale può essere riscritto come

Modello CCR in forma frazionaria

$$\max_{\mu,\nu} w_0 = \frac{\mu^T y_0}{\nu^T x_0}$$

$$s.t.$$

$$\frac{\mu^T y_i}{\nu^T x_i} \leq 1 \qquad i = 1, \ldots, N \tag{6a}$$

$$\mu^T \geq \varepsilon \mathbf{1}^T \tag{6b}$$

$$\nu^T \geq \varepsilon \mathbf{1}^T \tag{6c}$$

Il problema (6) ha un interessante interpretazione economica; infatti il modello frazionario *CCR* considera come indice di efficienza il rapporto $\frac{\mu^T y_0}{\nu^T x_0}$ tra una somma pesata di output e una somma pesata di input, generalizzando al caso multi-input/multi-output una tipica misura di produttività. Il problema frazionario (6) incorpora un procedimento di assegnazione di pesi alle varie risorse di input/output in modo tale da massimizzare la produttività dell'unità s_0 che si sta valutando, con i vincoli (6a) (cosiddetti vincoli tecnologici) che la produttività di tutte le unità osservate risulti inferiore ad una costante di normalizzazione. La misura di efficienza per s_0 è definita come il valore ottimo del problema frazionario ed è una misura di efficienza *relativa* poiché dipende dalle altre unità in S_0.

Un problema frazionario (6) è risolto per ogni unità da valutare, quindi la definizione dei pesi non è unica, ma ciascuna unità è lasciata libera di determinare il sistema dei pesi che meglio la valorizza. I vincoli (6b) e (6c) impongono che ad ogni risorsa venga assegnato un peso strettamente positivo, in modo tale da evitare che la misura di produttività sia ottenuta dando valore nullo a qualche input e/o output.

Procedendo in modo analogo a quanto fatto per i modelli input-oriented è possibile definire una funzione di distanza radiale rivolta alla massimizzazione degli output (output-oriented) e formulare opportunamente dei problemi di programmazione lineare che la calcolino su un insieme delle possibilità produttive a rendimenti costanti o variabili. Per un'introduzione generale alla metodologia *DEA* si veda [9], mentre si rimanda a [24] per una vasta bibliografia della letteratura *DEA*.

6. Applicazioni ed esempi

6.1. **Efficienza del management nel sistema ferroviario europeo.** Gathon e Pestieau [17] affrontano il problema della separazione della componente di inefficienza dovuta al management e di quella dovuta ai vincoli all'autonomia dei manager derivanti dal contesto in cui l'unità decisionale opera. Il caso che viene considerato è quello delle aziende ferroviarie europee. L'approccio scelto dagli autori è di tipo parametrico, a due fasi.

Si assume una funzione di produzione di tipo *Cobb-Douglas*

$$\ln y_i = \alpha + \beta \ln x_i + \varepsilon_i,$$

dove y_i è l'output e x_i il vettore degli input dell'unità decisionale i.

Una prima regressione lineare consente di stimare i valori $\widehat{\alpha}$, $\widehat{\beta}$. Indicando poi, per ogni i, con z_i una variabile che esprime il livello di autonomia del management, si effettua una nuova regressione in modo da individuare la dipendenza da z_i degli scarti ottenuti:

$$\ln y_i - \widehat{\alpha} - \widehat{\beta} \ln x_i = \delta + \gamma z_i + \eta_i.$$

Si ottengono così i valori $\widehat{\delta}$, $\widehat{\gamma}$ e $\widehat{\eta}_i$. Se indichiamo con z^+ e $\widehat{\eta}^+$ i massimi valori assunti dalle variabili z_i e $\widehat{\eta}_i$, possiamo stimare la frontiera *best practice* della funzione di produzione:

$$\ln y_i^F = \widehat{\alpha} + \widehat{\beta} \ln x_i + \widehat{\delta} + \widehat{\gamma} z^+ + \widehat{\eta}^+.$$

L'efficienza è allora data dalla:

$$e = \frac{y_i}{y_i^F} = \exp[(\widehat{\eta}_i - \widehat{\eta}^+) + \widehat{\gamma}(z_i - z^+)] = E_m E_r,$$

avendo indicato con $E_m = \exp(\widehat{\eta}_i - \widehat{\eta}^+)$ la componente manageriale dell'efficienza e con $E_r = \exp[\widehat{\gamma}(z_i - z^+)]$ il fattore di correzione dipendente dai vincoli esterni.

Il modello proposto porta pertanto ad una decomposizione moltiplicativa dell'efficienza in due componenti, una dipendente dal management e l'altra dall'ambiente in cui il management opera:

$$E = E_m E_r.$$

Nello studio sono state considerate le aziende ferroviarie di 19 nazioni europee nel periodo 1961-88. Per ciascun anno e per ciascuna azienda sono stati considerati:

- in output, il numero totale di tonnellate-kilometri lordi movimentati (merci + passeggeri);
- in input, il numero totale dei vagoni e motrici disponibili, il personale, la lunghezza totale delle linee elettrificate e quella delle linee non elettrificate, e la proporzione del traffico merci rispetto al totale.

Per quel che riguarda la misura di autonomia del management, gli autori hanno condotto un'indagine principalmente basata su un questionario con oltre 50 punti, inviato alle diverse aziende. L'indagine mirava a valutare l'autonomia tariffaria, quella commerciale ed il livello di autonomia nei riguardi delle assunzioni del personale. Il risultato è stato la costruzione di un indice con valori fra 40 e 100; il valore 100, corrispondente all'autonomia massima del campione, è stato attribuito all'azienda svizzera, *BLS*, mentre il valore minimo (40) a quella Finladese, *VR*. In questa classifica l'azienda italiana, *FS*, si colloca esattamente al centro con valore 67. L'assunzione implicita nel modello usato è che all'aumentare dell'autonomia aumenti l'efficienza, il che risulta confermato dai risultati della regressione ($\widehat{\gamma} > 0$).

Analizzando i risultati relativi al periodo 1986-88 (il periodo cui si riferisce l'inchiesta) si ottengono diversi risultati interessanti. Ad esempio, l'azienda finlandese, che si colloca in posizione intermedia per quel che riguarda l'efficienza globale, risulta al primo posto considerando l'efficienza del management; questo indicherebbe che l'inefficienza di quest'azienda è in gran parte solo apparente, e che, nei limiti fissati dai vincoli esterni, il management opera ad un elevato livello di efficienza. Molto diverso è invece il caso dell'azienda svizzera la cui inefficienza è quasi tutta dovuta al management. Intermedio è il caso di *FS*, per cui i risultati evidenziano la possibilità di migliorare in modo sostanziale l'efficienza del management.

6.2. **Effetti della deregulation nell'aviazione civile.** Il lavoro che qui descriviamo, dovuto a Good, Roller and Sickles [18] mira a valutare gli effetti di una politica di *deregulation*, attraverso un confronto fra le aereolinee europee e quelle americane nel periodo 1976-1986. In quel periodo, mentre il trasporto aereo negli Usa era ampiamente liberalizzato, quello europeo era caratterizzato da un regime sostanzialmente protezionistico.

L'analisi riguarda le 8 maggiori aereolinee europee (Air France, Alitalia, British Air, Iberia, KLM, Lufthansa, Sabena e SAS) e le 8 maggiori aereolinee americane (American, Continental, Delta, Eastern, Northwest, Pan Am, Transworld e United).

Come *input* sono stati considerati i costi annuali del personale, dei materiali di consumo (incluso il carburante) ed il valore della flotta disponibile. Come *output* sono stati considerati i ricavi dei servizi passeggeri, merci e dei servizi accessori; per i primi due si è considerato il ricavo per tonnellata-kilometro. Vengono poi anche considerate cinque indicatori delle caratteristiche operative delle aziende: il fattore di carico (che viene assunto come una misura della qualità del servizio), una misura della lunghezza delle tratte servite, la lunghezza totale della rete, la percentuale della flotta costituita da jumbo e la percentuale costituita da aerei a turboelica.

Un elemento interessante di questo lavoro è che l'analisi viene effettuata sia con un approccio di tipo parametrico che con la metodologia non-parametrica di tipo *DEA*.

6.2.1. *Modello parametrico.* Il modello parametrico è basato su una funzione di produzione di tipo *Cobb-Douglas* ed è molto simile a quello a due fasi del paragrafo precedente. Nella prima fase su utilizza il modello

$$\ln y_{it} = \beta \ln x_{it} + \gamma z_{it} + \varepsilon_{it}, \qquad i = 1, \ldots N,\ t = 1, \ldots, T,$$

dove l'indice i si riferisce alle aereolinee, l'indice t al tempo (anno), y_{it} è una misura aggregata degli output, mentre x_{it} è il vettore degli input e z_{it} quello delle 5 caratteristiche operative dell'azienda i al tempo t. La regressione fornisce le stime $\widehat{\beta}$ e $\widehat{\gamma}$. Una seconda regressione basata sulla relazione

$$\ln y_{it} - \widehat{\beta} \ln x_{it} - \widehat{\gamma} z_{it} = \alpha_i + \delta_i t + \eta_{it},$$

consente di separare la componente dello scarto dipendente dall'azienda (costante più trend) da quella casuale. Si ottengono così le stime $\widehat{\alpha}_i$ e $\widehat{\delta}_i$ e i valori $\alpha_{it} = \widehat{\alpha}_i + \widehat{\delta}_i t$.

Ponendo $\alpha_t = \max_i\{\alpha_{it}\}$, ed operando come nell'esempio precedente ricaviamo una misura dell'efficienza dell'azienda i al tempo t:

$$e_{it} = \exp(\alpha_{it} - \alpha_t).$$

6.2.2. *Modello non-parametrico.* Anche nell'analisi di tipo non-parametrico si è proceduto in più fasi, in modo da tenere conto della dimensione temporale. Nel seguito sono descritti brevemente i passi della procedura proposta.

1. In una prima fase ogni coppia aereolinea-anno è stata considerata come una unità distinta. Si utilizzato un modello *DEA* sull'insieme di questi dati ($16 \times 11 = 176$ unità decisionali), ottenendo i valori di efficienza θ_{it}, uno per ogni aereolinea ($i = 1, \ldots, 16$) e per ogni anno ($t = 1, \ldots, 11$).

2. Si è quindi effettuata una regressione per misurare la dipendenza dei punteggi θ_{it} dal tempo t e dal vettore z_{it} delle 5 caratteristiche dell'input e dell'output indicate sopra, utilizzando il modello lineare:

$$\theta_{it} = \alpha_i + \delta_i t + \gamma z_{it} + \varepsilon_{it}.$$

Indicando con $\widehat{\alpha}_i, \widehat{\delta}_i$ e $\widehat{\gamma}$ le stime dei parametri, i valori $\alpha_{it} = \widehat{\alpha}_i + \widehat{\delta}_i t + \widehat{\gamma} z_{it}$ rappresentano una misura della componente dell'efficienza che dipende dal tempo e dalle caratteristiche qualitative proprie di ciascuna azienda.

3. Vengono a questo punto determinati i valori:

$$\begin{aligned} \alpha_t &= \max_i \{\alpha_{it}\}, \\ e_{it} &= \exp(\alpha_{it} - \alpha_t). \end{aligned}$$

Il primo individua l'azienda più efficiente per ciascun anno, mentre il secondo fornisce una misura di efficienza dell'azienda i nell'anno t. I valori $\exp(\alpha_t - \alpha_{t-1})$ e $\exp(e_{it} - e_{i(t-1)})$ forniscono rispettivamente la misura del cambiamento tecnologico nel settore e quella della crescita di produttività per l'azienda i, fra l'anno t-1 e l'anno t.

6.2.3. *Conclusioni.* I risultati ottenuti con entrambi gli approcci indicano che le compagnie Usa sono più efficienti di quelle europee di circa il 20%. Pertanto, per gli autori, i risultati dell'indagine suggeriscono che «i benefici della deregolamentazione e delle privatizzazioni in termini di efficienza e di produttività possono essere grandi per il trasporto aereo europeo». Va però detto che non sono stati inseriti nei modelli indicatori significativi della qualità del servizio fornito, e che non mancano aspetti nei risultati che lasciano perplessi[2].

6.3. **Un modello per l'allocazione di risorse.** In questo modello la metodologia *DEA* è usata come strumento per determinare l'allocazione dei fondi per le spese di trasporto fra i 100 distretti scolastici dello stato nordamericano del North Carolina[25].

Ciascun distretto dispone di una flotta di autobus per il trasporto degli studenti. Il costo di trasporto è finanziato dallo Stato. Il problema affrontato nello studio è quello di fornire dei criteri per la distribuzione dei fondi che incentivino un uso efficiente delle risorse. La principale difficoltà nasce dalla disomogeneità dei diversi distretti. Ad esempio, la distribuzione degli alunni sul territorio ed il tipo di strade

[2] Ad esempio, il modello parametrico colloca, nel 1986, la *Pan Am* e la *Trans World* al secondo posto a parimerito per efficienza, mentre quello non-parametrico, nello stesso anno, le colloca rispettivamente al secondo ed al terzo posto; negli anni successivi a quelli dello studio, la prima compagnia è fallita, mentre la seconda, dopo essere stata sul punto di fallire, è stata fortemente ridimensionata.

hanno un notevole effetto sulle spese, con il risultato di falsare le misure di efficienza ove la metodologia *DEA* fosse applicata senza correttivi.

Si è utilizzato un approccio articolato in quattro fasi.

1. Si è applicata la tecnica DEA ai 100 distretti, usando 2 input ed 1 output. Come *input* sono stati scelti il numero degli autobus effettivamente utilizzati per il trasporto degli alunni e la spesa aggregata per le operazioni di trasporto. Come *output* è stato scelto il numero medio giornaliero di alunni trasportati. Sono stati così ottenuti dei valori di efficienza $E_1, E_2, \ldots, E_{100}$.

2. Per mezzo di un modello di regressione lineare di tipo *stepwise* sui valori di efficienza ottenuti al passo precedente, si è espressa l'efficienza in funzione di 19 variabili caratterizzanti i distretti. Si è visto in pratica che le variabili che spiegano la quasi totalità della varianza dell'efficienza sono: il logaritmo della densità territoriale degli alunni, il numero medio di viaggi giornalieri per alunno, la percentuale di viaggi per alcune attività speciali in cui gruppi di studenti svolgono attività in scuole al di fuori di quelle di appartenenza, la percentuale di strade non asfaltate, la percentuale della superficie del territorio designata zona per usi ricreativi, il logaritmo del rapporto fra lunghezza totale delle strade e superficie del territorio. Qusto modello ha fornito i valori attesi di efficienza $\widehat{E}_1, \widehat{E}_2, \ldots, \widehat{E}_{100}$.

3. I valori degli output sono stati corretti, per ogni distretto i, utilizzando la formula

$$Y_i' = Y_i \frac{\overline{E}}{\widehat{E}_i},$$

dove Y_i è il valore effettivo dell'output del distretto i, mentre $\overline{E}$ è il valore medio delle efficienze ricavate al passo 1. L'idea che sta dietro questo tipo di correzione è che se per un distretto il valore $\widehat{E}_i$ è alto, ciò vuol dire che il distretto è favorito dalle condizioni in cui opera; l'opposto accade quando il valore è basso. È quindi naurale una correzione che riduca il valore dell'output nel primo caso e lo aumenti nel secondo.

4. I valori di output corretti sono stati utilizzati per una nuova applicazione della *DEA*. Sono stati così ottenuti i valori finali di efficienza.

I valori di efficienza sono stati utilizzati per la distribuzione dei fondi. A ciascun distretto è stata rimborsata una proporzione delle spese totali pari al minimo fra 1 (rimborso totale) e il suo valore di efficienza aumentato di 0.1. I risultati sono stati incoraggianti: dopo il primo anno di applicazione l'efficienza dei diversi distretti è andata aumentando.

Riferimenti bibliografici

[1] Aigner, D.J., C.A.K. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models", Journal of Econometrics 6/1 (1977) 21-37.

[2] Ali, A.I., C.S. Lerme, and L.M. Seiford, "Components of Efficiency Evaluation in Data Envelopment Analysis", European Journal of Operational Research 80, (1995) 462-73.

[3] Banker, R.D., "Estimating most productive scale size using Data Envelopment Analusis", European Journal of Operational Research 17 (1984) 35-44.

[4] Banker, R.D., A. Charnes, and W.W. Cooper, "Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis", Management Science 30/9 (1984), 1078-92.

[5] Battese, G.E., and T.J. Coelli, "Prediction of arm-level technical efficiency with a generalized frontier production function and panel data", Journal of Econometrics 38 (1988) 387-399.

[6] Boyd, G. and R. Färe, "Measuring the efficiency of decision making units: A comment", European Journal of Operational Research 15 (1984) 331-333.

[7] Charnes, A. and W.W. Cooper,'The non-Archimedean CCR ratio efficiency analysis: A rejoinder to Boyd and Färe", European Journal of Operational Research 15 (1984) 333-334.

[8] Charnes, A., W.W. Cooper, and E. Rhodes,"Measuring the efficiency of decision making units", European Journal of Operational Research 2 (1978) 429-444.

[9] Charnes, A., W.W. Cooper, A.Y. Lewin, and L.M. Seiford eds. "Data Envelopment Analysis: theory, methodology, and applications", Kluwer Academic Publisher, Norwell, MA (1994).

[10] Cooper, W.W., and J.T. Pastor "Generalized Efficiency Measures (GEMS) and Model Relations for Use in DEA", to appear in Journal of Productivity Analysis.

[11] Debreu, G., "The coefficient of Resource Utilization", Econometrica 19, (1951) 273-292.

[12] EJOR, "Special issue on productivity analysis: parametric and non paramentric applicatons", European Journal of Operational Research 80 (1995).

[13] Färe, R., S. Grosskopf, and C.A.K. Lovell, "Production Frontiers", Camridge University Press, (1994).

[14] Färe, R. and C.A.K. Lovell, "Measuring the Technical Efficiency of Production", Journal of Economic Theory 19, (1978), 150-62.

[15] Farrell, M. J., "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A, General 120, (1957) 253-281.

[16] Fried, H.J., Lovell C.A.K., and Schmidt S.S., "The Measurement of Productive Efficiency", New York, Oxford University Press (1993).

[17] Gathon, H.-J. and P. Pestieau, "Decomposing efficiency into its managerial and its regulatory components: The case of European Railways", European Journal of Operational Research 80 (1995) 500-507.

[18] Good, D.H., L.-H. Roller, and R.C. Sickles, "Airline efficiency differences between Europe and the US: Implications for the pace of EC integration and domestic regulation", European Journal of Operational Research 80 (1995) 508-518.

[19] Henderson, J.M. and R.E. Quandt, "Micro-Economic Theory: A Mathematical Approach", McrGraw-Hill (1971) Cap 3 , 52-102.

[20] Koopmans, T.C., "An Analysis of Production as an Efficient Combination of Activities", in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph No. 13, New York: John Wiley and Sons,Inc(1951).

[21] Lewin, A.Y. and C.A. Knox Lovell eds., "FRONTIER ANALYSIS: Parametric and non parametric approaches", Journal of Econometrics 46, Vol.1/2, (1990).

[22] Lewin, A.Y. and C.A. Knox Lovell eds., "Special issue on productivity analysis: parametric and non parametric applications", European Journal of Operational Research 80 (1995).

[23] Russel, R.R., "Continuity of Measures of Technical Efficiency", Journal of Economic Literature. Classification Numbers:022,226. Academic Press, Inc (1990).

[24] Seiford, L.M., "A Bibliography for Data Envelopment Analysis (1978-1996)", Annals of Operations Research 73 (1997) 393-438.

[25] Sexton, T.R., S. Sleeper and R.E. Taggart Jr., "Improving pupil transportation in North Carolina", Interfaces 24 (1994) 87-103.

[26] Shephard R.W., "Cost and Production Function", Princeton N.J., Princeton University Press (1953).

[27] Tulkens, H. e P. Vanden Eeckaut, "Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects", European Journal of Operational Research 80 (1995) 474-499.