

Capitolo 1

Valutazione di efficacia e di efficienza: uno sguardo d'insieme

di Giorgio Vittadini e Simona C. Minotti

1.1 Valutazione e servizi alla persona di pubblica utilità

Negli ultimi anni è cresciuta l'attenzione verso la valutazione della qualità dei servizi alla persona di pubblica utilità (SPPU), vale a dire servizi collettivi aventi lo scopo di migliorare particolari aspetti dello stato di benessere degli utenti, operanti nei settori tipici quali sanità, istruzione, formazione, assistenza, tempo libero, cultura (Gori e Vittadini, 1999).

A differenza di quanto avviene nei mercati dei beni e servizi tradizionali, studi recenti documentano come la competizione del mercato basata sulla ricerca del massimo profitto non sia sufficiente per individuare con chiarezza la qualità dei SPPU (Borgonovi, 2002).

Numerose ragioni stanno all'origine di questa affermazione: nei settori tipici del *welfare* spesso l'utente non è libero di scegliere l'agente, vi sono vincoli istituzionali, sindacali e corporativi, ed infine vengono introdotte regole etiche che limitano le modalità della concorrenza a tutela dei diritti e della sicurezza degli utenti, cosicché in tali settori si parla di "quasi mercati". Soprattutto, però, per la particolare natura dei SPPU, esiste, in questi settori, "asimmetria informativa": la qualità cioè non si manifesta fino al momento effettivo della fornitura del servizio e in molti casi, se non valutata, può rimanere occulta anche dopo tale erogazione (Moramarco, 1999). Infatti, l'esito dell'erogazione dei SPPU non sono solo gli *output* aziendali, facilmente rilevabili e quantificabili,

ma anche gli *outcome*, intesi come risultati di lungo periodo generati dagli *output*, su particolari aspetti di una condizione, stato di benessere o comportamento dell'utente, non direttamente rilevabili e quantificabili (Gori e Vittadini, 1999).

Si comprende pertanto come la costruzione di un adeguato sistema di valutazione possa aiutare: (I) gli utenti a scegliere l'agente migliore o a giudicare la bontà del servizio ricevuto, (II) l'ente pubblico finanziatore a decidere se le risorse pubbliche siano state utilizzate al meglio, (III) qualunque terzo pagante (assicurazione, investitore privato) ad allocare al meglio le risorse tra gli agenti, (IV) gli agenti a migliorare il servizio erogato.

1.2 Metodi di valutazione dei SPPU

Estendendo quanto affermato da alcuni autori (Donabedian, 1990; Opit, 1991) a riguardo dei servizi sanitari, si può affermare che la valutazione dei SPPU concerne tre dimensioni: la struttura degli agenti, il processo produttivo, i risultati conseguiti sugli utenti. A tale proposito, è importante sottolineare che i tradizionali metodi di origine aziendalistica (Epstein, 1995; Pagano e Rossi, 1999; Zangrandi, 2005) non sono sufficienti, ma è necessario costruire opportuni indicatori che permettano di misurare i tre diversi aspetti attraverso cui si manifesta la qualità dei SPPU: efficacia, efficienza e *Customer Satisfaction*.

L'efficacia si definisce come la capacità dell'agente di migliorare l'*outcome* degli utenti (Donabedian, 1988; Goldstein e Spiegelhalter, 1996); l'efficienza consiste nel confronto tra *input* utilizzati e *output* prodotti, sia in termini quantitativi che monetari (Fabbri *et al.*, 1996); la *customer satisfaction* valuta la qualità percepita dagli utenti rispetto al servizio erogato (Zanella, 2001) ed è distinta dalla qualità effettiva descritta dall'efficacia.

Un primo modo di misurare questi tre aspetti è l'utilizzo di indicatori di tipo univariato, espressi in termini di frequenze assolute o relative, tassi o rapporti (Molteni, 1997; JCAHO, 1998; Pagano, 2005). Tuttavia tali indicatori, se non utilizzati con l'ausilio di opportune metodologie statistiche, sono inadeguati allo scopo per i seguenti motivi:

1. Solitamente sono riferiti agli agenti mentre, nell'ambito delle analisi di efficacia e di *Customer Satisfaction*, l'informazione dovrebbe essere rilevata ed elaborata a livello di singolo utente.
2. Sono di tipo univariato, mentre l'efficacia si basa sul concetto di *outcome* – costruito multidimensionale sotteso a una batteria di indicatori rilevati sull'utente (Gertler, 1988) –, l'efficienza su una pluralità di *input* e *output* del processo produttivo e la *Customer Satisfaction* su una pluralità di percezioni soggettive dell'utente riferite alle diverse dimensioni del servizio.
3. Quando siano espressi in termini di frequenze relative o assolute possono al più rilevare la presenza o assenza di determinate caratteristiche; rimane tuttavia ignota l'intensità attraverso cui si manifestano variabili quantitative o l'ordine delle modalità di variabili qualitative.
4. Nell'ambito delle analisi di efficacia, essendo rilevati una sola volta dopo l'erogazione del SPPU, non sono in grado di misurare il valore aggiunto in termini di miglioramento dello stato di benessere dell'utente. Ciò comporta il rischio della *selection bias*, vale a dire di una valutazione distorta degli agenti perché non tiene conto delle caratteristiche degli utenti (soprattutto della loro condizione prima dell'erogazione del servizio) e l'ancor più grave problema della "selezione avversa", cioè del rifiuto degli utenti più complessi al fine di non peggiorare la qualità delle prestazioni (Moramarco, 1999).

Allo scopo di effettuare una corretta valutazione della qualità degli agenti, si rende quindi necessaria l'introduzione di opportune metodologie statistiche, che utilizzino dati individuali da integrare opportunamente con dati e indicatori aggregati, siano in grado di considerare simultaneamente più indicatori, consentano di trattare variabili categoriali e, nell'ambito dell'efficacia, misurarne il valore aggiunto.

Nel seguito verranno illustrate, con particolare riferimento ad efficacia ed efficienza, le particolari metodologie introdotte in letteratura al fine di rispondere ai quesiti ivi sopra sollevati.

1.3 Efficacia relativa

Le analisi di efficacia nascono nel quadro di studi epidemiologici di *clinical trials* con riferimento al concetto di efficacia come “impatto”¹, (Cislaghi *et al.*, 1995; Biggeri *et al.*, 2001).

Qualora il termine di raffronto non sia l’assenza dell’azione, né un obiettivo o *standard*, ma uno o più agenti alternativi preposti allo svolgimento dell’azione, si fa riferimento al concetto di efficacia “relativa” (Gori e Vittadini, 1999), intesa come un confronto tra diversi possibili agenti (scuole, ospedali) nell’ambito della stessa azione (istruzione, sanità) al fine di scegliere quale di essi eroghi il servizio migliore.

Per quanto detto nel precedente paragrafo a proposito di *selection bias* e “selezione avversa”, occorre effettuare valutazioni *coeteris paribus* – al netto cioè dell’influenza sugli *outcome* delle diverse caratteristiche degli utenti (*case-mix*) – adottando allo scopo metodologie statistiche denominabili nel loro complesso di *risk adjustment*. Nel seguito verranno illustrate le più comunemente utilizzate: standardizzazione, modelli lineari e logistici, modelli *multilevel*.

1.3.1 Standardizzazione diretta e indiretta

La metodologia più semplice di *risk adjustment* è la cosiddetta “standardizzazione diretta” (Blangiardo, 1997; Zaslavsky, 2001).

Dopo aver suddiviso la popolazione di utenti in q strati² ($k=1, \dots, q$), si costruisce per ciascun agente j ($j=1, \dots, p$) l’*outcome* “standardizzato” y_j come somma ponderata degli *outcome* y_{kj} rilevati sui diversi strati, con pesi pari alla proporzione di utenti π_{kj} di ogni strato:

¹ Con il termine efficacia si intende, in generale, l’analisi dei risultati conseguiti sugli utenti mediante l’erogazione del servizio. Tale definizione si declina in cinque diversi modi (Gori e Vittadini, 1999): 1. efficacia “relativa” come confronto tra agenti alternativi nell’ambito della stessa azione; 2. efficacia come rapporto tra risultati ottenuti e attesi; 3. efficacia come standard di un processo o di un prodotto; 4. efficacia come rapporto tra domanda soddisfatta e potenziale; 5. efficacia come “impatto” rispetto alla situazione in cui non sia erogato il servizio.

² Gli strati rappresentano gruppi di utenti omogenei (per esempio, in ambito sanitario, pazienti chirurgici e medici) e sono individuati dalle variabili di *case-mix*.

$$y_j = \sum_k \pi_{kj} y_{kj}, \quad (1)$$

dove $\pi_{kj} = w_{kj} / \sum_k w_{kj}$, con w_{kj} numero di utenti del k-esimo strato cui è erogato il SPPU dal j-esimo agente.

La “standardizzazione diretta” presenta tuttavia alcuni limiti: non può essere applicata ad *outcome* che non sono discreti e non può trattare più di un *outcome* simultaneamente; inoltre, per un agente che non presenti casi in un determinato strato, non è possibile calcolarne il punteggio standardizzato (Zaslavsky, 2001).

Un approccio alternativo è fornito dalla “standardizzazione indiretta”; in questo caso la ponderazione è effettuata mediante pesi $\tilde{\pi}_{kj}$, ottenuti da una teorica popolazione standard:

$$\tilde{y}_j = \sum_k \tilde{\pi}_{kj} y_{kj}. \quad (2)$$

Per valutare l’efficacia relativa \hat{u}_j del j-esimo agente si rapportano l’*outcome* osservato y_j e l’*outcome* “atteso” \tilde{y}_j :

$$\hat{u}_j = y_j / \tilde{y}_j. \quad (3)$$

Purtroppo anche la “standardizzazione indiretta” presenta alcuni dei problemi della standardizzazione diretta (Zaslavsky, 2001).

1.3.2 Modelli lineari e logistici

Un approccio più flessibile è rappresentato dai modelli lineari e logistici, suggeriti da molti studiosi (tra cui Dubois *et al.*, 1987; De Long *et al.*, 1997) e utilizzati dalle principali agenzie di *benchmark* sanitario (AHRQ, 2003; Solucient, 2003; CIHI, 2004; JCAHO, 2004; NHS, 2004).

Un modello lineare ha la seguente formulazione:

$$y_{ij} = \alpha + \beta' \mathbf{x}_{ij} + e_{ij}, \quad (4)$$

dove y_{ij} ($i=1, \dots, n_j$; $j=1, \dots, p$) è un *outcome* quantitativo riferito all’i-esimo utente cui è erogato il SPPU dal j-esimo agente; α è l’intercetta

del modello, che indica il livello medio dell'*outcome* (tra tutti gli utenti e tutti gli agenti); \mathbf{x}_{ij} è un vettore di p caratteristiche dell'utente, cui è associato il vettore $\boldsymbol{\beta}$ di parametri fissi; e_{ij} è il residuo casuale.

L'efficacia relativa del j -esimo agente è data dalla media aritmetica delle stime \hat{e}_{ij} dei residui casuali e_{ij} associati agli n_j utenti della struttura j -esima:

$$\hat{e}_j = \sum_i \hat{e}_{ij} / n_j. \quad (5)$$

Nel caso si utilizzi il modello di analisi della covarianza:

$$y_{ij} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{ij} + u_j + e_{ij}, \quad (6)$$

l'efficacia relativa del j -esimo agente è data dalla stima \hat{u}_j del parametro fisso u_j , che indica il livello medio dell'*outcome* tra tutti gli utenti sottoposti al j -esimo agente.

Qualora gli *outcome* siano di tipo dicotomico si utilizzano modelli logistici (Bini, 2000; Bellio e Gori, 2003; Fabbris, 2004; Chiandotto, 2005). In questo caso, il *logit* della probabilità p_{ij} ($i=1, \dots, n_j$; $j=1, \dots, p$) del verificarsi dell'evento favorevole per l' i -esimo utente della j -esima struttura viene espresso come funzione lineare del *case-mix* degli utenti:

$$\ln \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \alpha + \boldsymbol{\beta}'\mathbf{x}_{ij} = RS_{ij}, \quad (7)$$

dove RS_{ij} indica il cosiddetto *risk score* associato all' i -esimo utente della j -esima struttura (De Long *et al.*, 1997).

L'efficacia relativa del j -esimo agente è data dal rapporto:

$$\hat{u}_j = \sum_i p_{ij} / \sum_i \hat{p}_{ij} = p_j / \hat{p}_j, \quad (8)$$

dove $p_{ij} = \sum_i y_{ij} / n_j$ è la probabilità "osservata" per l' i -esimo utente della j -esima struttura e \hat{p}_{ij} è la stima della probabilità del verificarsi dell'evento favorevole dato il suo *case-mix*.

1.3.3 Modelli Multilevel

In molti casi i modelli lineari e logistici non sono adeguati a misurare l'efficacia relativa (si veda il par. 1.3.4). Per tale motivo, a partire dalla metà degli anni Ottanta, numerosi autori (tra cui Aitkin e Longford, 1986; Goldstein, 1987; Bryk e Raudenbush, 1992; Gori, Mealli e Rampichini, 1993; Goldstein e Spiegelhalter, 1996) suggeriscono di utilizzare i modelli *multilevel*.

La formulazione del modello *multilevel* (Hox, 1995) è analoga al modello di analisi della covarianza:

$$y_{ij} = \alpha + \beta' x_{ij} + u_j + e_{ij}, \quad (9)$$

con la differenza che il parametro u_j ($j=1,\dots,p$), che rappresenta l'efficacia relativa del j -esimo agente, non è fisso bensì casuale.

Le stime \hat{u}_j sono ottenute mediante l'inferenza *bayesiana* (Leyland e Boddy, 1998) e sono fornite generalmente in termini di intervalli di confidenza; in questo modo viene misurata l'incertezza associata alla posizione degli agenti nella graduatoria di efficacia (Marshall e Spiegelhalter, 2001).

Nella sua versione base il modello *multilevel* presenta le seguenti restrizioni:

- a) una sola variabile risposta (*outcome*) di tipo quantitativo; inoltre l'*outcome* è considerato come una variabile osservata rilevata ai diversi livelli gerarchici (utente, agente, territorio);
- b) assenza di variabili esplicative riferite alle caratteristiche degli agenti;
- c) modello di tipo statico, cioè riferito ad un istante temporale;
- d) ipotesi di distribuzione binomiale o normale per i residui casuali e di distribuzione multinormale per i parametri di efficacia;
- e) ipotesi di indipendenza dei parametri di efficacia;
- f) ipotesi di incorrelazione tra parametri di efficacia e variabili esplicative (non stocastiche) inerenti le caratteristiche degli utenti;
- g) ipotesi che i parametri inerenti le caratteristiche degli utenti siano fissi; in tal caso si ha un modello *multilevel* a effetti misti.

Le precedenti ipotesi sono tuttavia molto restrittive rispetto alla complessità dei casi che si incontrano nella realtà. Sono state quindi introdotte alcune generalizzazioni:

- α) Il modello *multilevel* può essere esteso alla considerazione simultanea di più *outcome* tra loro correlati (per esempio, in ambito sanitario, salute mentale e fisica). Una proposta in tale direzione è data dalla combinazione del modello SURE (Srivastava e Giles, 1987) con il modello *multilevel* (Vittadini e Sanarico, 2003; Vittadini e Minotti, 2005).

Data la sua natura di costruito multidimensionale, un *outcome* è più propriamente definibile in termini statistici come variabile latente. Per evitare i problemi di unicità connessi all'utilizzo dei modelli strutturali (Vittadini, 1989), gli *outcome* possono essere ricavati come trasformate lineari dei rispettivi indicatori, tramite tecniche quali il *Partial Least Squares* (Muthen e Speckart, 1985; Leyland e Goldstein, 2001, Dagum *et al.*, 2003).

In presenza di indicatori di tipo qualitativo o misto, occorre utilizzare metodi di *Multidimensional Scaling* quali *Alsos* o *Princals* (Young, 1981; Vittadini, 2001), che simultaneamente quantificano gli indicatori e ricavano gli *outcome* come combinazioni lineari. Qualora gli indicatori di tipo qualitativo siano influenzati dal grado di difficoltà di questionari, test o prove attitudinali da cui provengono, al fine di una loro corretta quantificazione è stata proposta la *Rasch Analysis* (Wright e Masters, 1982); tale tecnica permette di ricavare misure oggettive dell'effettiva abilità degli agenti nell'erogazione degli SPPU (Lovaglio, 2003; Gori, 2004).

- β) Introdurre le caratteristiche degli agenti (per esempio risorse, capacità organizzativa, ecc.) tra le variabili esplicative di un modello *multilevel* significa specificare un'equazione di secondo livello, che esprime i parametri casuali u_{jh} ($j=1,\dots,p$; $h=1,\dots,r$) come funzione delle caratteristiche degli agenti. In questo caso l'efficacia relativa degli agenti è valutata al netto delle caratteristiche degli utenti e degli agenti (Normand *et al.*, 1995).

Inoltre, è possibile introdurre parametri di interazione, che esprimono l'effetto congiunto di agenti e sottogruppi di utenti con particolari caratteristiche (Leyland e Boddy, 1998).

- γ) Sono stati introdotti modelli *multilevel* di tipo longitudinale per analizzare l'evoluzione temporale delle conoscenze matematiche (Rogosa, 1995; Rowan *et al.*, 2002), della domanda di assistenza ospedaliera (Verbeke e Molenberghs, 2000) e del reddito dei laureati rispetto a quello dei non laureati (Lovaglio e Vittadini, 2005); alcune

recenti proposte riguardano la considerazione di relazioni non lineari (Young, 2001; Ann Gilligan *et al.*, 2002).

- δ) In alternativa alla distribuzione binomiale e normale sono state proposte la distribuzione multinormale di ordine λ , nota anche come “*exponential power distribution*” (Subbotin, 1923; Box e Tiao, 1992), la distribuzione politomica (Daniels e Gatsonis, 1997), la distribuzione di Poisson (Langford *et al.*, 1999), la mistura di normali (Verbeke e Molenberghs, 2000), la distribuzione multivariata esponenziale (Yang, 2001). Ulteriori proposte riguardano il caso di indicatori qualitativi o misti (Marshall e Spiegelhalter, 2001) e la presenza di *outlier* (Langford e Lewis, 1998).
- ε) L’ipotesi di indipendenza dei parametri di efficacia potrebbe non essere realistica nell’ambito di modelli *multilevel* longitudinali dove uno dei livelli è rappresentato dall’utente (Heagerty e Kurland, 2001). Verbeke e Molenberghs (2000) propongono, nel caso di relazioni lineari, di utilizzare modelli ad effetti fissi, al fine di poter correggere l’impatto dovuto alla presenza di un parametro per ogni individuo mediante condizionamento.
- ζ) Quando l’ipotesi di incorrelazione tra parametri di efficacia e caratteristiche degli utenti non è soddisfatta, l’efficacia relativa non dipende esclusivamente dai parametri di efficacia, ma dipende anche dalle caratteristiche degli individui (Snijders e Bosker, 1999). Ciò può determinare il problema della “*selection bias*”, cui si è accennato precedentemente. Per ovviare a tale problema Bellio e Sartori (2003) suggeriscono di ricorrere a modelli *multilevel* con parametri di efficacia fissi e non casuali.

L’esistenza di correlazione tra parametri di efficacia e caratteristiche degli utenti può essere verificata mediante il test di Hausman (1978). Bellio e Sartori (2003) propongono invece di utilizzare il cosiddetto *decomposition approach* (Berlin *et al.*, 1999), che consiste nel sostituire il termine $\beta'x_{ij}$ nella (9) con:

$$\beta'(x_{ij} - \bar{x}_j) + \beta_1' \bar{x}_j, \quad (10)$$

dove \bar{x}_j è la media del vettore di caratteristiche degli utenti sottoposti all’agente j -esimo ($j=1, \dots, p$) e i vettori di parametri β e β_1 corrispondono alla decomposizione dell’effetto dovuto alle caratteristiche degli utenti x_{ij} ($i=1, \dots, n_j; j=1, \dots, p$) in “effetto negli agenti” (*within-cluster covariate effect*) e “effetto tra gli agenti” (*between-*

cluster covariate effect) rispettivamente (si veda a tale proposito anche Carpita, 2004 e 2005).

Nel caso vi sia “*selection bias*” si verificherà l’ipotesi di una differenza significativa tra il vettore β_1 nella (10) e il vettore β nella (9).

- η) Affinché l’effetto dovuto alle caratteristiche degli utenti possa variare tra gli agenti, si introducono i parametri casuali; in tal caso si ha un modello *multilevel* a effetti casuali.

Per la stima di modelli a effetti casuali sono stati proposti metodi di simulazione (McCulloch e Searle, 2000), tecniche numeriche (Skaug, 2002) e metodi ad hoc quali: MQL, PQL-IGLIS; RIGLS; IGLS (Langford *et al.*, 1999); PQL, PQL2 (Barbosa e Goldstein, 2000); REML (Turner *et al.*, 2001).

Data la particolare natura dell’applicazione, si considerino inoltre le seguenti possibili estensioni del modello *multilevel*:

- θ) Qualora i dati disponibili siano di natura osservazionale (ad esempio dati provenienti da archivi amministrativi), i metodi sviluppati per dati sperimentali potrebbero non essere adeguati. A tale proposito, alcuni autori hanno proposto confronti tra studi basati su dati osservazionali e sperimentali, sia in ambito parametrico (Longford e Lewis, 1988) che non parametrico (Goldstein, 1995; Leyland, 1995; Goldstein e Spiegelhalter, 1996; Carlin *et al.*, 2001).

Se i dati provengono da differenti archivi amministrativi, per la costruzione del data-set è necessario utilizzare opportune procedure di *record linkage* (Perchinunno, 2005). Inoltre, si suggerisce l’adozione di tecniche di campionamento flessibili, che permettano di tener conto dell’eventuale presenza di caratteristiche fuori controllo, di tipo ambientale o riferite ai pazienti.

Infine, qualora i dati provengano da campioni o popolazioni di numerosità elevata, i test di significatività dei parametri tipici dei modelli lineari portano quasi sempre al rifiuto dell’ipotesi nulla, con la conseguenza che tutti i parametri risultano significativi (Vroman Battle e Rakow, 1993). Per ovviare a tale inconveniente occorre introdurre metodi in grado di verificare l’eterogeneità del campione o della popolazione e di effettuare test di significatività dei parametri nell’ambito di sottocampioni della popolazione opportunamente scelti (Vittadini *et al.*, 2005).

- ι) Alcuni autori (Yang, 2001; Ann Gilinger *et al.*, 2002) consigliano l’utilizzo di modelli *multilevel* di tipo non lineare in presenza di va-

riabili di tipo misto tra loro fortemente correlate, parametri di interazione, valori mancanti (Rubin, 1984) e *outliers* (Friedman, 1991; Durand e Lombardo, 2002; Gallo *et al.*, 2001).

- κ) Recentemente, l'uso di tecniche di *risk adjustment* per la costruzione di graduatorie delle strutture sanitarie è stata fortemente criticata (Lilford *et al.*, 2004), in quanto si ritiene che i confronti fra *outcome*, per quanto nell'intenzione *coeteris paribus*, non riescano a fornire una reale graduatoria tra agenti basata sull'efficacia³. Come alternativa, gli autori citati suggeriscono l'utilizzo di dati amministrativi⁴ per l'analisi delle *best practices* e a fini di miglioramenti organizzativi e gestionali. Così, per esempio, supponendo di considerare modelli lineari in cui i valori attesi dei parametri di efficacia siano le variabili dipendenti e le variabili esplicative descrivano diverse tipologie di ospedali, si possono effettuare delle post-analisi, quali la verifica dell'ipotesi di uguaglianza dei valori attesi dei parametri di efficacia tra sottogruppi di ospedali (per esempio privato vs pubblico, generale vs specializzato).

1.3.4 Ragioni per la scelta dei modelli multilevel

Quali sono le ragioni per cui è opportuno utilizzare modelli *multilevel* a effetti misti rispetto a modelli lineari o di analisi della covarianza a effetti fissi?

- 1) Negli studi di valutazione i dati hanno tipicamente una struttura gerarchica (utente, reparto, agente, territorio), che non viene colta dai modelli lineari, in quanto basati sull'ipotesi di indipendenza tra le singole osservazioni, cioè di campionamento casuale semplice. Nell'ambito dei modelli *multilevel* si ipotizza, invece, un campio-

³ In particolare, Lilford *et al.* (2004) sostengono che: “*The sensitivity of an institution's position in league tables to the method of risk adjustment used suggests that comparisons of outcomes are unlikely to tell us about the quality of care. [...] Outcome is neither a sensitive nor a specific marker for quality of care.*” (pp.1148-1149).

⁴ “*Administrative data are typically computerized, which facilitates their use in analysis because large quantities of data are available. As a result, policymakers, employers, and health plans have used administrative data for a variety of purposes: [...] to monitor the use of health services over time, to study geographic variation in the utilization of surgical and medical procedures, [...] to assess the effects of a policy change on health expenditures, [...] to evaluate differences in hospital death rates as a function of differences in the quality of care.*” (Damberg *et al.*, 1998, pp.45-46).

namento a stadi, che implica dipendenza tra le osservazioni appartenenti alla stessa sottopopolazione (Snijders e Bosker, 1999).

- 2) Nei modelli *multilevel* è possibile studiare l'effetto delle variabili di contesto e di background individuale ai diversi livelli della gerarchia (Goldstein, 1995; Snijders e Bosker, 1999).
- 3) Gli *outcome* attesi possono assumere valori sistematicamente diversi per sottogruppi di utenti cui è erogato il SPPU dal medesimo agente (Normand *et al.*, 1995). L'introduzione di effetti di interazione che rispettano la struttura gerarchica dei dati ovvia a questi problemi.
- 4) Ci possono essere problemi causati dalla diversa numerosità di sottocampioni di pazienti in diverse strutture sanitarie. Mediante l'effetto *shrinkage* (Goldstein e Spiegelhalter, 1996) le informazioni provenienti dai diversi sottocampioni pesano sulla stima dei parametri in proporzione alla numerosità dei sottocampioni medesimi.
- 5) Le graduatorie di agenti costruite sulla base dei precedenti modelli non sono sempre "robuste". In particolare, quando gli agenti presentano performance "estreme" (fortemente negative o positive) risulta rischioso misurare l'efficacia mediante un parametro fisso (Morris e Christiansen, 1996). Nei modelli *multilevel* questo problema è ovviato dal fatto che l'efficacia di ogni agente è individuata da un parametro casuale e la graduatoria, essendo stabilita sulla base di relazioni d'ordine tra gli intervalli di confidenza dei parametri, risulta "probabilizzata".
- 6) In molti casi non risultano soddisfatte le ipotesi distributive alla base del modello lineare. Come già ricordato in precedenza, l'inferenza *bayesiana*, consigliabile nell'ambito di modelli *multilevel*, permette di trattare distribuzioni non normali e *outcome* ricavati da indicatori qualitativi o misti.
- 7) In presenza di dati longitudinali, i modelli *multilevel* consentono di rappresentare la struttura delle osservazioni rispetto al tempo, in un contesto di forte flessibilità.

1.4 Efficienza relativa

Le principali metodologie di analisi dell'efficienza sono basate su alcuni concetti basilari che verranno brevemente introdotti nel seguito.

Si definisce insieme \mathbf{T} delle possibilità di produzione (o insieme di produzione) l'insieme di tutti i piani (\mathbf{x}, \mathbf{y}) realizzabili, dove il vettore \mathbf{x} r -dimensionale rappresenta gli r *input* del processo produttivo e il vettore \mathbf{y} s -dimensionale gli s *output*:

$$\mathbf{T} = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{R}_+^{r+s} : \mathbf{x} \text{ permette di produrre } \mathbf{y}\} \quad (11)$$

La frontiera di produzione, vale a dire la massima quantità di *output* ottenibile con una data quantità di *input*, è descritta analiticamente dalla funzione di produzione, che, in un'ottica *output-oriented*, associa ad ogni combinazione di *input* il massimo volume di *output*:

$$f(\mathbf{x}) = \max \{\mathbf{y} : \mathbf{y} \in \mathbf{P}(\mathbf{x})\}, \quad (12)$$

dove $\mathbf{P}(\mathbf{x})$ è l'insieme degli *output* \mathbf{y} che possono essere prodotti con un dato livello di *input* \mathbf{x} , o, in un'ottica *input-oriented*, associa ad ogni combinazione di *output* il minimo volume di *input*:

$$f(\mathbf{y}) = \min \{\mathbf{x} : \mathbf{x} \in \mathbf{L}(\mathbf{y})\}, \quad (13)$$

dove $\mathbf{L}(\mathbf{y})$ è l'insieme degli *input* \mathbf{x} che consentono di produrre un dato livello di *output* \mathbf{y} .

In una logica di massimizzazione del profitto, l'azienda cerca sempre di operare lungo la frontiera di produzione, il cui andamento dipende dalla tecnologia disponibile (si fa qui riferimento al concetto di efficienza "tecnica" introdotto da Farrel, 1957). Un'azienda è quindi efficiente se, dato un certo ammontare di *input*, produce la massima quantità di *output* o, dato un ammontare di *output*, utilizza la minima quantità di *input*. Farrel (1957) definisce inefficienza tecnica la distanza dalla frontiera di produzione, che nel caso *output-oriented*, cioè fissata una certa quantità di *output*, si può misurare attraverso il rapporto:

$$\lambda_{\mathbf{x}} = \frac{\|\mathbf{x}'\|}{\|\mathbf{x}\|}, \quad (14)$$

dove \mathbf{x}' rappresenta la quantità di *input* corrispondente alla situazione di massima efficienza e \mathbf{x} la quantità di *input* osservata.

Quanto più λ è elevato, tanto maggiore sarà la distanza tra la situazione osservata e la frontiera di produzione.

La quantità di *input* ottimale corrisponde a quella che rende λ minimo (cioè uguale ad 1):

$$E_{\mathbf{x}}(\mathbf{x}|\mathbf{y}) = \min \left\{ \lambda : \lambda_{\mathbf{x}} = \frac{\|\mathbf{x}'\|}{\|\mathbf{x}\|} \right\}. \quad (15)$$

Un concetto alternativo a quello di funzione di produzione è quello di funzione di costo, che associa ad ogni combinazione di prodotto il costo sostenuto per la sua produzione. L'andamento della funzione di costo è legato ai prezzi dei fattori oltre che alla tecnologia adottata (si fa qui riferimento alla cosiddetta efficienza "allocativa" introdotta da Farrell, 1957).

In termini pratici, la definizione o la stima di uno standard produttivo ideale risulta estremamente ardua, vale a dire la frontiera di produzione di un determinato settore è spesso per lo più un concetto ipotetico. Per tale ragione, i principali metodi matematico-statistici utilizzati nelle analisi empiriche si basano sul concetto di efficienza "relativa". Ciò significa che, non essendo possibile ottenere una stima assoluta della funzione di produzione, è necessario ripiegare su analisi di tipo comparativo tra aziende con caratteristiche simili.

I concetti ivi sopra riportati possono essere opportunamente estesi anche all'ambito dei SPPU. Misurare l'efficienza relativa significa in questo caso osservare i differenziali di risultato tra agenti che erogano SPPU simili.

1.4.1 Approccio basato su indici univariati

Il primo approccio all'analisi dell'efficienza si basa su indici di produttività univariati di tipo economico-aziendale, dati dal rapporto tra un

output e un *input*. A tale proposito, valgono le considerazioni critiche riportate nel par. 1.2. In particolare tali indici, data la loro unidimensionalità, non possono cogliere il grado di inefficienza che, come si è detto, è misurabile mediante distanze di *input* o *output*.

Per ovviare a tali problemi sono stati proposti in letteratura due approcci: uno non parametrico e l'altro parametrico.

1.4.2 Approccio non parametrico: Data Envelopment Analysis

Un approccio molto flessibile dal punto di vista delle ipotesi di partenza è l'approccio non parametrico (Charnes, Cooper e Rhodes, 1978; Bankers, Charnes e Cooper, 1984), definito tale in quanto non ipotizza alcuna forma funzionale per la frontiera di produzione. Il calcolo delle efficienze individuali si basa sulla costruzione di una frontiera di produzione mediante interpolazione matematica delle osservazioni migliori (da un punto di vista produttivo). I *peers* (così vengono chiamate le unità di frontiera) vanno perciò a formare i "nodi" del limite superiore delle possibilità di produzione e rappresentano dei punti di paragone per le strutture inefficienti.

L'esempio più classico di approccio non parametrico è la *Data Envelopment Analysis* (DEA), che affronta il problema della valutazione dell'efficienza mediante gli strumenti della ricerca operativa, ed in particolare della programmazione lineare, ed è stata sviluppata nel 1978 da Charnes, Cooper e Rhodes estendendo il lavoro pionieristico di Farrell (1957)⁵.

Date n DMU (*Decision Making Unit*), vale a dire n unità da valutare), la DEA ipotizza che ad ognuna di esse, DMU _{i} , $i=1, \dots, k$, sia associato un vettore di *input* \mathbf{x}_i , con elementi $x_{ij} \geq 0$, $j=1, \dots, r$, e un vettore di *output* \mathbf{y}_i , con elementi $y_{ki} \geq 0$, $k=1, \dots, s$. Inoltre, si suppone che per ogni DMU vi siano almeno un *input* e un *output* non nulli e che ogni DMU lavori con rendimenti di scala costanti o variabili⁶.

Per misurare l'efficienza la DEA prevede due passaggi:

⁵ Una rassegna recente sull'argomento è data da Seiford e Thrall (1990). Tra gli approcci non parametrici è degno di nota anche il Free Disposal Hull (FDH), proposto anch'esso da Charnes, Cooper e Rhodes (1978).

⁶ Rispettivamente come nel modello DEA CRS (*Constant Returns to Scale*) o VRS (*Variable Return of Scale*) (Charnes, Cooper e Rhodes, 1978).

1. Ridurre la dimensione del problema passando da r *input* ed s *output* ad un unico *input* ed un unico *output* virtuali, ottenuti come combinazioni lineari dei vari *input* ed *output* reali.

Per ciascuna DMU_i si costruiscono così i rapporti:

$$h_i(\mathbf{u}, \mathbf{v}) = \frac{\sum_k u_k Y_{ki}}{\sum_j v_j X_{ji}}, \quad (16)$$

dove $u_k, v_j \geq 0$, $k=1, \dots, r$ e $j=1, \dots, s$.

2. Calcolare per ciascuna DMU_i i pesi \mathbf{u} e \mathbf{v} che rendono massimo il rapporto (16):

$$\max h_i(\mathbf{u}, \mathbf{v}) \quad (17)$$

sotto il vincolo che $0 \leq h_i(\mathbf{u}, \mathbf{v}) \leq 1$ per ogni $i=1, \dots, n$, vale a dire che per ciascuna DMU_i l'*output* virtuale non superi l'*input* virtuale (il rapporto è pari ad 1 solo per le unità efficienti). In pratica, la DEA associa ad ogni DMU un piano produttivo virtuale $(\mathbf{x}^*, \mathbf{y}^*)$ ottenuto come combinazione lineare di quelli osservati in corrispondenza alle n DMU del campione. Se tale piano virtuale coincide con il piano effettivo (\mathbf{x}, \mathbf{y}) (e quindi $h_i = 1$), allora l'unità è efficiente: non esiste alcuna combinazione delle unità del campione che, a parità di volume e di mix di *output*, consumi una quantità inferiore di *input*. Quanto più invece tale piano virtuale è distante dal piano effettivo tanto più elevata è l'inefficienza. In tempi recenti, la tecnica DEA⁷ è stata estesa in molteplici direzioni⁸.

⁷ Va detto, per la precisione, che quanto esposto in precedenza è una versione semplificata del metodo. Infatti, la condizione che h_i risulti pari ad 1 è necessaria ma non sufficiente per qualificare le relative DMU come efficienti. Sarebbe possibile, per esempio, che un'unità si posizionasse sulla funzione di produzione lineare a tratti, ma senza tuttavia essere efficiente.

⁸ Un'estensione importante del metodo è la versione VRS (*Variable Returns to Scale*) che ipotizza rendimenti di scala variabili. Tale metodo scompone il giudizio tecnico in due fattori, l'efficienza tecnica pura (capacità di trasformare le risorse in prodotti senza disturbi dovuti agli effetti negativi di una scala produttiva inadeguata), e l'effi-

1.4.3 L'approccio parametrico: i modelli di frontiera stocastica

L'approccio parametrico è rappresentato dai *modelli di frontiera stocastica* (Aigner, Lovell e Schmidt, 1977; Fried, Lovell e Schmidt, 1993; Khumbakar e Knox Lovell, 2000) che ipotizzano una distribuzione parametrica per la funzione di produzione (o di costo), che viene stimata sulla base delle osservazioni relative ad unità coinvolte nel processo produttivo. Calcolate le deviazioni individuali di ciascuna unità produttiva dalla funzione di produzione stimata, una parte della deviazione è interpretabile come disturbo stocastico, mentre l'altra parte rappresenta l'inefficienza.

Le forme funzionali maggiormente utilizzate in letteratura sono la *Cobb Douglas* (la cui trasformazione logaritmica è una funzione lineare nei parametri), preferibile per l'interpretabilità (Meeusen e Van Den Broek, 1997) e la *translogaritmica*, preferibile per la flessibilità, brevemente illustrate nel seguito.

Considerando il caso elementare in cui siano presenti una sola variabile di *output* (y) ed uno di *input* (x), la formula che identifica un modello *Cobb Douglas* (Gori *et al.*, 2002; Grassetti *et al.*, 2003) per l' i -esimo agente è la seguente:

$$y_i = \alpha x_i^\beta + e^{\varepsilon_i - u_i} \quad (i=1, \dots, n) \quad (18)$$

che in forma logaritmica risulta essere:

$$\log y_i = \alpha + \beta \log(x_j) + \varepsilon_i - u_i \quad (i=1, \dots, n) \quad (19)$$

dove ε_i è un disturbo stocastico e u_i rappresenta l'inefficienza dell'unità i -esima.

Si assume che ε_i sia una variabile casuale distribuita normalmente con media nulla e varianza σ_ε^2 e che u_i abbia una distribuzione di tipo non simmetrico con valori soltanto positivi (per esempio una distribuzione seminormale, normale troncata o esponenziale) e varianza σ_u^2 . I parametri del modello (β , σ_ε^2 e σ_u^2) vengono stimati minimizzando una

ienza dimensionale o di scala (che valuta se la DMU opera su di una scala dimensionale ottima, data quella osservata presso le altre unità del campione).

funzione obiettivo che varia a seconda del metodo di stima adottato; in un secondo tempo si ricavano le stime delle inefficienze tecniche \hat{u}_i .

Considerando invece il caso in cui siano presenti una sola variabile di *output* (y) ed due di *input* (x_1, x_2), un modello *translogaritmico* è rappresentato dalla seguente espressione:

$$\ln(y_i) = \alpha_0 + \alpha_1 \ln(x_{1i}) + \alpha_2 \ln(x_{2i}) + \alpha_3 \ln(x_{1i}) \ln(x_{2i}). \quad (i, \dots, n) \quad (20)$$

Per qualsiasi modello di tipo parametrico, la definizione di efficienza per l' i -esimo agente è sintetizzabile mediante il rapporto:

$$\text{Eff}(i) = \frac{y_i}{\hat{y}_i} \quad (21)$$

che nel caso particolare della *Cobb Douglas* implica che il termine u_i debba essere trasformato esponenzialmente al fine di ottenere le stime dell'efficienza.

Un'estensione dei modelli di frontiera stocastica sono i modelli di frontiera di costo (Battese e Coelli, 1995), utilizzabili qualora siano disponibili dati sui prezzi, e che permettono di scomporre l'inefficienza nelle sue componenti tecnica e allocativa.

Quanto finora detto fa riferimento a dati *cross-sectional*, vale a dire ottenuti in un dato istante temporale. Se invece sono disponibili dati *panel*, vale a dire misure ripetute nel tempo delle stesse grandezze, è possibile ottenere modelli maggiormente informativi. Un modello di frontiera stocastica per dati *panel*, nella versione basata sulla funzione di produzione, si può scrivere nella forma:

$$y_{it} = f(x_{it}, \beta) - u_i + v_{it}. \quad i=1, \dots, n; t=1, \dots, T \quad (22)$$

Analogamente a quanto visto in precedenza, u_i indica l'inefficienza (che si assume in versione semplificata invariante nel tempo) e v_{it} è un termine di disturbo a media nulla.

1.4.3.1 Modelli di frontiera stocastica multilevel

Come illustrato a proposito di efficacia, negli studi di valutazione la struttura gerarchica dei dati può essere utilizzata per specificare in ma-

niera più adeguata il modello. In tal modo si possono evitare errori di specificazione e problemi di *selection bias* che si hanno qualora, come spesso succede, si proceda ad aggregare poche informazioni disaggregate a livello di unità superiore.

A tale scopo si adotta anche nell'ambito della valutazione dell'efficienza il modello *multilevel* ad effetti misti (Clarke, 1996; 2002; Gori *et al.*, 2002; Grassetti *et al.*, 2003; Grassetti *et al.*, 2005), già ampiamente descritto nel par. 1.3.3. Attraverso questo approccio è possibile costruire modelli di frontiera che tengono conto della struttura gerarchica dei dati e della presenza di *input* in parte suddivisi tra i reparti ed in parte comuni a tutti i reparti di uno stesso agente.

La formulazione del modello, nella sua specificazione più semplice, si ispira ad un modello Cobb Douglas, modificato per tenere conto di eventuali differenze strutturali nella tecnica produttiva tra reparti:

$$y_{ij} = \exp \left[\left(\alpha + \sum_{k=2}^K I_{\{k\}}(j) \alpha_k \right) x_{ijk}^{\beta + \sum_{k=2}^K I_{\{k\}}(j) \beta_k} z_i^{\gamma + \sum_{k=2}^K I_{\{k\}}(j) \gamma_k} \exp(u_i + e_{ij}) \right]$$

dove gli indici i e j indicano rispettivamente l'agente (unità di secondo livello) e il reparto (unità di primo livello), y_{ij} e x_{ij} rappresentano rispettivamente l'*output* e l'*input* a livello di reparto, z_i è un *input* a livello di agente, $I_{\{k\}}$ è una variabile che identifica la tipologia di reparto. Il modello prevede la possibilità di coefficienti differenti per tipologia di reparto, sia per l'intercetta, sia per gli *input* di 1° livello, nonché per gli *input* di 2° livello, ovvero gli *input* indivisi (Gori *et al.*, 2002).

In particolare, nell'ambito della valutazione dell'efficienza degli ospedali, è stata proposta da Grassetti *et al.* (2005) una versione *translogaritmica* della funzione di produzione, al fine di catturare la non linearità della relazione esistente tra il numero di posti letto e il numero di ricoveri⁹.

⁹ Si ha in particolare:

$$\ln(\text{NHosp}_{ij}) = \alpha(\text{CMix}_{ij}) + \beta(\text{CMix}_{ij}, \text{Beds}_{ij}) + \gamma(\text{CMix}_{ij}, \text{Staff}_i) + \delta(\text{CMix}_{ij}, \text{Beds}_{ij}, \text{Staff}_i) + f(\text{Weights}) + u_i + \varepsilon_{ij},$$

dove

La peculiarità della forma funzionale *translogaritmica* è la sua flessibilità, in quanto permette di ottenere delle stime dell'elasticità rispetto ai singoli *input* dipendenti dai livelli delle covariate considerate nel modello. Le stime dell'elasticità permettono di identificare le varie condizioni produttive in essere nel sistema; inoltre, una loro analisi consente di congetturare a proposito di eventuali politiche di riallocazione degli *input* che possono portare ad un punto di equilibrio ottimale tra le strutture produttive. Infine, la forma *translogaritmica* consente di ottenere analoghi risultati per l'elasticità di scala, vale a dire calcolata rispetto a una variazione di tutti gli *input* simultanea¹⁰.

1.4.4 Confronto fra metodi non parametrici e parametrici

Presentiamo ora un breve confronto fra i diversi metodi presentati.

I metodi non parametrici, quali la DEA, presentano alcuni pregi:

- a) Flessibilità: la funzione di produzione è costruita direttamente a partire dai dati osservati, senza alcun assunto sulla sua forma fun-

$$\begin{aligned} \alpha(\text{CMix}_{ij}) &= \alpha_0 + \alpha_1 \text{std}(\ln(\text{CMix}_{ij})) \\ \beta(\text{CMix}_{ij}, \text{Beds}_{ij}) &= [\beta_0 + \beta_1 \text{std}(\ln(\text{CMix}_{ij}))][\ln(\text{Beds}_{ij}) + \beta_2 \ln(\text{Beds}_{ij})^2] \\ \gamma(\text{CMix}_{ij}, \text{Staff}_i) &= [\gamma_0 + \gamma_1 \text{std}(\ln(\text{CMix}_{ij}))][\ln(\text{Staff}_i) + \gamma_2 \ln(\text{Staff}_i)^2] \\ \delta(\text{CMix}_{ij}, \text{Beds}_{ij}, \text{Staff}_i) &= [\delta_0 + \delta_1 \text{std}(\ln(\text{CMix}_{ij}))][\ln(\text{Beds}_{ij}) + \ln(\text{Staff}_i)^2] \\ f(\text{Weights}) &= \lambda_1 u_{\text{weig},ij} + \lambda_2 u_{\text{weig},ij} + \lambda_3 u_{\text{weig},i} + \lambda_4 u_{\text{weig},i} \end{aligned}$$

con Beds_{ij} variabile di input numero di posti letto riferita al reparto j -esimo dell'ospedale i -esimo; CMix_{ij} indice di *case-mix* che rappresenta il livello di complessità dei casi trattati in un singolo reparto j di un ospedale i rispetto alla complessità media regionale; Staff_i variabile di input riferita al personale ospedaliero disponibile nell'ospedale i . La componente lineare additiva, indicata con $f(\text{Weights})$, è funzione di quattro variabili, ottenute come trasformazione della variabile peso DRG (*Disease Related Groups*), che rappresenta un indice della complessità del caso. Due di esse sono state calcolate come media dei pesi DRG nei casi trattati (a livello di reparto $\mu_{\text{Weig},ij}$ e a livello di ospedale $\mu_{\text{Weig},i}$) e mirano a depurare il fenomeno considerato dalla complessità media dei ricoveri. Le altre due variabili corrispondono alle deviazioni standard dei pesi DRG (a livello di reparto $\sigma_{\text{Weig},ij}$ e a livello di ospedale $\sigma_{\text{Weig},i}$). La considerazione di queste variabili nel modello è dettata dalla volontà di includere nella specificazione funzionale anche una qualche misura della variabilità dei casi entro i *cluster* considerati. Tuttavia si tenga presente che l'introduzione di tali variabili nella specificazione funzionale influenza solamente il livello della produzione e non il processo produttivo stesso (Grassetti *et al.*, 2005).

¹⁰ Nel lavoro di Grassetti *et al.* (2005) sono presenti inoltre alcune interessanti considerazioni sull'efficienza di scala e sul problema della allocazione ottima delle risorse.

zionale; è inoltre applicabile ad un numero qualsiasi di *input* e *output*.

- b) Informatività: una singola applicazione del metodo fornisce molteplici risultati, come visto in precedenza.
- c) Semplicità di calcolo: la DEA si compone di semplici problemi di programmazione lineare, la cui soluzione si ottiene con software di tipo standard.

Accanto agli indubbi pregi, vi sono però alcuni punti critici:

- a) Eccessiva elasticità e mancanza di robustezza: essendo la frontiera costruita in maniera empirica, sulla base dei dati, vi è il rischio che i risultati dipendano fortemente dal tipo di dati a disposizione, anche in presenza di campioni di ampia numerosità.
- b) Eccessiva sensibilità alla presenza di *outlier* e difficoltà a valutare la significatività dei risultati, in quanto le misure fornite sono di tipo deterministico.

I numerosi pregi dei modelli parametrici di frontiera stocastica, sono riassumibili nei seguenti punti:

- α) Solidità strutturale: le ipotesi relative alla forma funzionale della funzione di produzione nascono dalla teoria economica e vengono sottoposte a verifica tramite le tecniche classiche dell'econometria.
- β) Realismo: i modelli di frontiera stocastica assumono che i dati siano soggetti a fluttuazioni aleatorie non imputabili all'efficienza.
- γ) Flessibilità: i modelli sono applicabili a funzioni di produzione caratterizzate sia da rendimenti di scala costanti che variabili, e più in generale da tecniche produttive differenti (Tsionas, 2002); ciò ne permette l'estensione al caso di agenti con caratteristiche differenti. Le tecniche in questione sono particolarmente appropriate al caso di dati *panel*; possibili estensioni riguardano il caso di dati *cross-sectional*. Inoltre, con l'ausilio dell'inferenza *bayesiana* è possibile utilizzare eventuali informazioni a priori sui parametri del modello e l'efficienza delle singole unità produttive (Fernández *et al.*; 2000).
- δ) L'estensione *multilevel* dei modelli di frontiera stocastica sfrutta al meglio le informazioni rilevate, fornendo risultati robusti ed evitando il già citato rischio di *selection bias*.

I limiti maggiori sono invece i seguenti:

- α) La natura parametrica dei modelli può incrementare il rischio di produrre risultati errati, qualora la funzione di produzione non sia correttamente specificata. Se la rigidità dovuta alla specificazione della funzione di produzione è superabile mediante formulazioni maggiormente flessibili, anche di tipo non lineare¹¹, più arduo è invece proporre distribuzioni degli errori differenti da quelle classiche.
- β) Il trattamento del caso multi-*output* può avvenire solo tramite l'approccio basato sulla funzione di costo (Ray; 2003) o la predisposizione di modelli multivariati misti (Snijders e Bosker, 1999). In alcune situazioni tali strade possono risultare di difficile percorribilità.

1.5 Indicatori di *outcome* e variabili di *risk adjustment*

Nei precedenti paragrafi sono stati approfonditi i principali aspetti metodologici connessi alla misura dell'efficacia e dell'efficienza dei SPPU. Tuttavia, mentre per quanto riguarda l'efficienza (associata alle tradizionali grandezze di *input* e *output* nei diversi settori) non vi sono soverchie differenze, occorre sottolineare che la validità delle analisi di efficacia dipende dalla capacità e dalla possibilità di scegliere in modo opportuno *outcome* e variabili di *risk adjustment*.

Un rapido excursus mostrerà i problemi di maggior rilievo che caratterizzano i vari settori.

1.5.1 Sanità

L'*outcome* maggiormente utilizzato, sia negli studi sperimentali basati su metodologie statistiche avanzate, che nei grandi progetti delle agenzie di *benchmark*, è il tasso di mortalità intraospedaliera (Thomas *et al.*, 1994), classificabile come *outcome* clinico. Non mancano tuttavia per-

¹¹ Soluzioni semiparametriche sono state recentemente formulate, si veda Fan, Li e Weersink (1996).

plexità circa l'adeguatezza del suo utilizzo¹², anche per la difficoltà di individuare indicatori di *case-mix* appropriati ai fini di una valutazione *coeteris paribus*¹³.

Sono state pertanto formulate le seguenti proposte alternative:

- a) In luogo del tradizionale tasso di mortalità, Iezzoni *et al.* (1996) propongono di utilizzare, in alcuni casi, *outcome* clinici alternativi in connessione con opportuni indicatori di *case-mix*; entrambi sono ottenuti a partire dai dati clinici, che rappresentano la più accurata fonte di informazione su diagnosi, tipi di trattamenti e fattori di rischio dei pazienti (Lovaglio, 2003; Vittadini e Sanarico, 2003; Auxilia *et al.*, 2004; Vittadini e Minotti, 2005). L'estensione di tale proposta alla generalità dei casi appare tuttavia piuttosto problematica. Infatti, in primo luogo, in presenza di *outcome* e indicatori di *case-mix* connessi a particolari patologie, risulta piuttosto difficoltoso riuscire a tener conto dello stato generale di salute del paziente, nonché delle diverse tipologie delle strutture sanitarie (Leyland e Boddy, 1998). Occorre inoltre considerare che le informazioni relative ai pazienti sono usualmente conservate in cartelle cliniche o registri scritti a mano, senza procedure standard di registrazione e unicità di identificatori individuali. Infine, il processo di reperimento dei dati risulta nel complesso molto costoso e richiede tempi piuttosto lunghi (Damberg *et al.*, 1998).
- b) Le principali agenzie internazionali di *benchmark* (AHRQ, 2003; SOLUCIENT, 2003; CIHI, 2004; JCAHO, 2004; NHS, 2004) propongono

¹² "The relationship of death rates to quality of care remains controversial and unproven. Unless findings are adjusted for patient characteristics, conclusions about quality based on an evaluation of [patient] outcomes may be erroneous." (Iezzoni *et al.*, 1996, p. 1379). "Mortality may depend more on the mix of patients reaching hospital in the first place, rather than the quality of care given once admitted. In hospital mortality in particular may be prone to bias and manipulation" (Goldstein e Spiegelhalter, 1996, p. 399).

¹³ "Measures of severity of illness at admission to hospital have been criticized for not fully taking into account known discrepancies in outcomes associated with social background and other factors" (Goldstein e Spiegelhalter, 1996, p. 399). "The majority of variation in annual hospital death rates for the four conditions studied (stroke, pneumonia, myocardial infarction, and congestive heart failure) is chance variability that results from the relatively small numbers of patients treated in most hospitals in a year. [...] Risk adjustment methods do not show whether the unexplained difference in mortality rates results from differences in effectiveness of care or unmeasured differences in patient risk at the time of admission." (Jencks *et al.*, 1988, p. 3611).

di utilizzare *proxy* degli *outcome* clinici che, espresse sotto forma di percentuali o tassi, descrivono la frequenza di trattamenti o risultati classificabili come *best practice*, in quanto, in molti casi, sono associabili a decorsi positivi delle patologie trattate; alcune agenzie (JCAHO, 2004) suggeriscono di considerare anche i cosiddetti “eventi sentinella”, intesi come eventi inaspettati, che provocano usualmente la morte o serie ferite fisiche e/o psicologiche, o rischi conseguenti ad esse. Valgono per questi indicatori le stesse difficoltà descritte precedentemente per gli *outcome* clinici. Inoltre, per rilevarli, è necessario costruire un sistema di rilevazione generale, quale quello adottato dalle agenzie di *benchmark*.

- c) Un ulteriore strumento alternativo è rappresentato dai questionari sulla qualità della vita (Johansson, 1995; Johannensons e Johansson, 1997), che possono riferirsi alle condizioni di salute del paziente nel suo complesso, come per esempio le scale SF-12 e SF-36 (Ware, 1993), o ad aspetti particolari quali il suo stato funzionale fisico-mentale dopo cure riabilitative, come la scala FIM (Tesio, 2003). La risposta ad ogni *item* del questionario è espressa mediante modalità di tipo ordinale ed è rilevata tramite intervista. Non mancano problemi anche in questo caso, poiché le informazioni raccolte possono essere influenzate da aspetti ambientali indipendenti dal tipo di cura e, soprattutto in caso di patologie acute, possono risultare insufficienti a descrivere lo stato clinico del paziente. Inoltre, i costi delle interviste possono essere particolarmente rilevanti, soprattutto nel caso di studi *panel* (Damberg *et al.*, 1998).
- d) In luogo di *outcome* clinici o *outcome* sulla qualità della vita, alcuni autori (Epstein, 1995; Goldstein e Spiegelhalter, 1996; Leyland e Boddy, 1998; Vittadini *et al.*, 2005) suggeriscono di utilizzare *outcome* di contesto (quali per esempio ritorni in sala operatoria, ricoveri ripetuti, ri-ospedalizzazioni), facilmente reperibili a partire dagli archivi amministrativi. In questa tipologia possono essere classificati anche gli *outcome* ottenibili attraverso metodi quali il *disease staging* o l'APR DRG, che riclassificano le informazioni desumibili dal DRG in modalità di tipo ordinale sulla base della gravità della malattia del paziente ricoverato (Iezzoni *et al.*, 1996)¹⁴. I dati amministrativi presentano il vantaggio di contenere informazioni rela-

¹⁴ Molte perplessità a riguardo di tali metodi nascono dal fatto che la morte corrisponde sempre alla modalità con frequenza più elevata, qualunque ne sia la causa.

tive a un ampio numero di individui; tuttavia possono risultare insufficienti a descrivere il reale stato clinico del paziente e l'efficacia degli agenti. Non mancano poi problemi legati all'accuratezza dei dati, in quanto per esempio potrebbe accadere che differenti codici procedurali siano utilizzati per descrivere il medesimo evento patologico.

1.5.2 Università

Lockheed e Hanushek (1994) classificano le analisi di efficacia ed efficienza dell'università secondo un'ottica interna o esterna al sistema (si veda Tabella 1).

Tabella 1.1

	INTERNA	ESTERNA
EFFICACIA	Effetto dell'ateneo o corso di laurea sulla capacità di apprendimento dello studente	Effetto dell'ateneo o corso di laurea sulla capacità lavorativa del laureato
EFFICIENZA	Analisi costi – ricavi aziendali dell'investimento universitario	Rapporto tra costi individuali e collettivi e benefici monetarizzabili dell'istruzione universitaria

Per quanto riguarda l'efficacia interna, gli *outcome* possono essere ricavati da test volti a misurare il grado di apprendimento, dai voti di laurea o dalle votazioni degli esami, una volta eliminato l'effetto dovuto alla diversa difficoltà e/o al diverso metro di giudizio attraverso opportune metodologie statistiche, quali per esempio la *Rasch Analysis* (Lei *et al.*, 2001); per l'efficienza interna si utilizzano invece le informazioni aziendali.

Per ciò che concerne l'efficacia esterna, al fine di definire la probabilità di successo individuale nel mondo del lavoro, dovuta al corso di

laurea e al netto dei fattori individuali economico-ambientali¹⁵, gli *outcome* utilizzabili sono:

- a) situazione occupazionale, rapidità e livello di inserimento nel mercato del lavoro¹⁶;
- b) grado di utilizzo delle competenze acquisite¹⁷;
- c) grado di efficacia del titolo di studio¹⁸;
- d) soddisfazione del laureato, considerata, a 5 anni dalla laurea, un *outcome* “oggettivo” della reale efficacia degli studi (Zanella, 2001).

Per ciò che concerne l’efficienza esterna, si dovrebbe invece valutare il capitale umano proveniente dall’investimento in istruzione superiore, vale a dire l’incremento di reddito percepito e di ricchezza accumulata dall’individuo nel ciclo vitale. Il reddito mensile a cinque anni dalla laurea può essere considerato una prima *proxy* di tale capitale umano definibile nel ciclo vitale (ALMALAUREA, 2003).

Quali variabili di *case-mix* si possono utilizzare:

- a) caratteristiche socio-economiche (età, sesso, stato civile, ricchezza della famiglia di origine);
- b) numero di anni di lavoro effettivo;
- c) tipo e durata di formazione professionale o post laurea, che identifica l’effetto della formazione sul lavoro o sul capitale umano;

¹⁵ Si hanno molti lavori in letteratura basati su *outcome* ed indicatori che possono considerarsi *proxy* del capitale umano potenziale (Bini e Chiandotto, 2003; Chiandotto *et al.*, 2004; Chiandotto e Variale, 2004); tipicamente vengono misurati ad una distanza temporale variabile dalla laurea (solitamente di 1, 3 o 5 anni, ma preferibilmente di 5, come in ALMALAUREA, 2003).

¹⁶ Percentuale di neo-laureati che: trovano occupazione entro un certo periodo (Bini, 1999); sono in cerca di lavoro dopo un certo periodo; trovano un lavoro corrispondente alle loro attese; cercano un nuovo lavoro dopo il conseguimento del titolo. Si considerano anche i tempi medi di ricerca del lavoro (Bini, 1999; Biggeri *et al.*, 2001). Tali percentuali sono traducibili attraverso modelli statistici, quali i modelli *logit*, in termini di probabilità di: “occupazione”, “disoccupazione”, “ricollocazione” del neo-laureato in cerca di lavoro (Bini, 1999); proporzioni di: contratti trasformati da atipici a tempo indeterminato; passaggi a posizione di carriera superiore entro un certo periodo dall’assunzione.

¹⁷ Si considerano ad esempio indicatori di coerenza tra la formazione specifica acquisita e l’attività lavorativa svolta (Fabbris *et al.*, 2002), la specificità o la necessità del titolo di studio per svolgere un determinato lavoro, anch’essi riesprimibili in termini di probabilità attraverso opportune metodologie statistiche quali i modelli *logit* (Bini, 1999).

¹⁸ Integra il criterio precedente con il criterio della necessità legale del titolo di studio per il lavoro svolto.

- d) informazioni sul curriculum scolastico rilevabili attraverso voti di maturità e test all'ingresso, che permettono di identificare la capacità individuale degli studenti, indipendentemente dalla diversa difficoltà e dai criteri di valutazione;
- e) informazioni sul curriculum universitario non attribuibili all'efficacia dei corsi di laurea;
- f) informazioni riguardanti le caratteristiche del territorio. Sono le variabili di *case-mix* più difficili da quantificare ed anche le più importanti, se l'obiettivo è confrontare atenei e/o corsi di laurea di differenti aree geografiche, senza che il risultato venga inficiato dalle diversità delle condizioni occupazionali e sociali. Qualora non sia possibile effettuare un confronto *coeteris paribus* rispetto alle caratteristiche territoriali, la valutazione può avvenire secondo un criterio longitudinale che preveda anzitutto una descrizione del miglioramento (o peggioramento) nel tempo dei gruppi disciplinari o dei corsi di studi dei singoli atenei evitando il confronto con altri.

1.5.3 Formazione

Numerosi lavori di Gori ed altri autori (Gori, Mealli e Rampichini, 1993; Bellio e Gori, 2003) hanno approfondito gli aspetti metodologici connessi alla valutazione degli interventi di formazione. Vi è da rilevare che, con l'accordo Stato-Regioni del 18 febbraio 2000, tali argomentazioni sono state prese in considerazione ai fini dell'accreditamento. Si specificano così le nozioni di efficacia, che "confronta quanto è stato realizzato con quanto era stato originariamente programmato, ovvero le realizzazioni, i risultati e gli impatti effettivi rispetto a quelli attesi o stimati", e di efficienza, che indica "in che misura le risorse si sono tradotte in realizzazioni o in risultati".

Se ne deducono tre criteri fondamentali per l'individuazione degli indicatori da utilizzare per le valutazioni di efficacia ed efficienza degli interventi di formazione:

1. capacità di attuazione dei progetti, dei loro obiettivi e delle attività previste (efficacia);
2. capacità di svolgere l'attività a costi contenuti (efficienza);
3. capacità di ottenere buoni risultati in rapporto agli obiettivi inerenti i singoli allievi (efficacia).

Mentre il primo e il secondo criterio riguardano il comportamento delle sedi operative nei confronti della realizzazione dei progetti, il terzo valuta i risultati raggiunti a livello di ogni singolo individuo coinvolto nell'attività. Ogni indicatore viene quindi messo a confronto con dei valori soglia, così come è normalmente previsto per l'accreditamento, al fine di stabilire se la sede operativa rispetti o meno gli standard prefissati.

Ciò premesso, i principali indicatori di *outcome* adottati sono:

- a) tasso di attuazione dei progetti (corsi rendicontati/corsi approvati), con il quale si vuole cogliere la capacità di portare a buon fine i progetti per i quali si è ricevuto l'incarico di attuazione;
- b) tasso di attuazione dei destinatari (numero allievi frequentanti/numero allievi preventivati), con il quale si vuole cogliere la capacità di rispettare gli impegni formativi assunti relativamente al numero di allievi coinvolti;
- c) tasso di attuazione dei costi (costo rendicontato/costo preventivato), con il quale si vuole stimolare le sedi a rispettare il preventivo concernente i costi;
- d) indice di costo, con il quale si vuole stimolare le sedi ad una riduzione dei costi medi effettivi per destinatario rispetto a quelli previsti;
- e) tasso di successo (numero allievi formati/numero allievi avviati), con il quale si vuole stimolare le sedi a considerare tra i propri obiettivi la conclusione con esito positivo del percorso formativo/orientativo da parte di ogni allievo coinvolto;
- f) indicatore di monitoraggio (numero schede destinatari trasmesse/numero destinatari), con il quale si vuole stimolare le sedi a fornire le informazioni richieste per il monitoraggio.

1.5.4 Scuola

Gli studi più recenti riguardanti l'efficacia della scuola (Aitkin e Longford, 1986; Goldstein, 1987, 1995, 1997; Goldstein e Woodhouse, 2000; Gori *et al.*, 2001) introducono il *background* individuale quale fattore di aggiustamento degli *outcome* al livello più basso della gerarchia e stimano l'effetto esercitato da variabili operanti a livelli superiori della gerarchia stessa (per esempio l'effetto dei pari a livello di classe o l'effetto del contesto a livello di scuola). Da questo punto di vista i di-

versi settori di ricerca, in particolare quello della valutazione dell'efficacia (*School Effectiveness*) e quello che studia gli aspetti del miglioramento della scuola (*School Improvement*), si trovano a dover affrontare due ordini di problemi:

- 1) scelta di un criterio per misurare le differenze tra *outcome* cognitivi delle scuole;
- 2) scelta di una metrica per valutare la variazione degli *outcome*.

Riguardo il primo problema sono state proposte diverse soluzioni. Una prima opzione, adottata in Gran Bretagna, raccomanda di pubblicare le cosiddette *status-LT* (*league table*), contenenti i valori medi dei risultati conseguiti dai ragazzi nelle scuole (Plewis e Goldstein, 1998). Tuttavia, le *status-LT* hanno avuto in passato numerosi effetti negativi ed hanno sollevato molte polemiche circa la mancata considerazione del differente *background* degli studenti. L'alternativa è quella di calcolare il valore aggiunto di conoscenze attribuibili all'efficacia della scuola, al netto di caratteristiche personali e di *background* culturali, al fine di evitare problemi di *selection bias* e selezione avversa già menzionati in precedenza. In letteratura sono stati proposti due metodi per la misura del valore aggiunto (Good *et al.*, 1975; Linn e Slinde, 1977): il "gain score", corrispondente alla differenza ($y_2 - y_1$) tra *outcome post-test* (y_2) e *outcome pre-test* (y_1) e il "residual gain score", in cui l'efficacia è misurata dai residui della regressione del *post-test* rispetto al *pre-test* ($y_2 - E[y_2|y_1]$). È stato dimostrato (Willet, 1988) che l'*outcome pre-test* è afflitto da errori di misura, la stima dei parametri della regressione dipende fortemente dalla composizione del campione e da eventuali *outlier* così come la stima del "residual gain score" dipende fortemente da una serie di fattori che non consentono di pervenire ad una misura dell'evoluzione dell'*outcome* per ogni individuo. Ne consegue che il "residual gain score" è uno stimatore inefficiente ed inconsistente della vera variazione (*true gain score*). Anche Rogosa *et al.* (1982, 1985, 1995) dimostrano che il "residual gain score" non è affidabile.

Per ciò che concerne il secondo problema, vi sono due requisiti fondamentali da rispettare per la misura della variazione degli *outcome* cognitivi:

- a) costanza del costrutto di cui si vuole misurare la variazione nel corso del tempo (la nozione di variazione perderebbe infatti di significato qualora i contenuti ed il costrutto cambiassero nel corso del tempo);

- b) costanza dello strumento e della scala utilizzati per misurare l'*outcome* (posto che il costrutto sia qualitativamente costante nel tempo, misurare la variazione nelle quantità risulterebbe infatti impossibile se lo strumento di misura o la scala cambiassero in maniera incognita).

Rispetto ad altre misure, quali il “rango percentile”, frequentemente utilizzato a fini di comparazione degli *outcome* cognitivi, o il “*normal curve equivalent*”, rappresentato dal valore corrispondente al rango percentile di una distribuzione normale, risulta preferibile lo “*scale score*”, misura riferita ai contenuti anziché ad una popolazione o ad un campione di riferimento. Lo “*scale score*”, ottenuto attraverso il modello di Rasch, permette di identificare l'abilità e il livello di conoscenza dello studente al netto del grado di difficoltà della prova e/o di severità del giudicante (Bond e Fox, 2001). Le misure di *outcome* così ottenute, corrispondenti ad un certo grado di sviluppo dell'apprendimento, sono collocabili su una scala ad intervallo, attraverso la quale possono essere comparate sia orizzontalmente che verticalmente, da un anno all'altro, attraverso procedure di “equalizzazione” (Bond e Fox, 2001)¹⁹.

1.6 Bibliografia

- Aigner D.J., Lovell C.A.K., Schmidt P. (1977), “Formulation and Estimation of Stochastic Frontier Production Function Models”, *Journal of Econometrics*, 6, 21-37.
- Aitkin M., Longford N. (1986), “Statistical modelling issues in school effectiveness studies”, *Journal of the Royal Statistical Society-Series A*, 149(1), 1-43.
- AHRQ, Agency for Healthcare Research and Quality (2003), *National Healthcare Quality Report*, Department of Health and Human Services, Rockville, www.ahrq.gov/data/hcup.

¹⁹ Il comitato scientifico dell'INVALSI suggerisce quindi di ricavare informazioni longitudinali (per esempio a cadenza biennale) sul valore aggiunto in termini di *outcome* cognitivi degli studenti, punto di partenza per la costruzione di banche di *item*, validate con procedure che si rifanno alla teoria della misurazione di tipo Rasch (Bond e Fox, 2001).

- ALMALAUREA (2003), Il profilo dei laureati 2002, Consorzio Interuniversitario ALMALAUREA, Bologna (vedi anche indagini 1998-2001), <http://www.almalaurea.it/universita/profilo/profilo2002/index/shtml>
- Ann Gilligan M., Kneusel R.T., Hoffmann R.G., Greer A.L., Nattinger A.B. (2002), "Persistent differences in sociodemographic determinants of breast conserving treatment despite overall increased adoption", *Medical Care*, 40(3), 181-189.
- Auxilia F., Mapelli V., Rossi, C. (2004), "Organizzazione del sistema sanitario, metodi di finanziamento, ricerca della qualità", in Pagano A., Vittadini G. (a cura di), *Qualità e valutazione delle strutture sanitarie*, Etas, Milano, 5-19.
- Bankers R.D., Charnes A., Cooper W.W. (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, 30(9), 1078-1092.
- Barbosa M.F., Goldstein H. (2000), "Discrete response multilevel models for repeated measures: an application to voting intentions data", *Quality and Quantity*, 34(3), 323-330.
- Battese G.E., Coelli T.J. (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data", *Empirical Economics*, 20(2), 325-32.
- Bellio R., Gori E. (2003), "Impact evaluation of job training programmes: selection bias in multilevel models", *Journal of Applied Statistics*, 30, 85-97.
- Bellio R., Sartori N. (2003), "Extending conditional likelihood in models for stratified binary data", *Statistical Methods & Applications*, 12, 121-132.
- Berlin J.A., Kimmel S.E., Ten Have T.R., Sammel, M.D. (1999), "An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty", *Biometrics*, 55, 470-476.
- Biggeri L., Bini M., Grilli L. (2001), "The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job", *Journal of the Royal Statistical Society-Series A*, 162(2), 293-305.
- Bini M. (1999), "Valutazione dell'efficacia dell'istruzione universitaria rispetto al mercato del lavoro", *Rapporto di Ricerca 3/99*, Osservatorio per la Valutazione del Sistema Universitario, MURST, Roma.
- Bini M. (2000), "Il modello di regressione logistica per la stima della chance occupazionali dei laureati" in ALMA LAUREA e Università

- degli Studi di Bologna (a cura di), *Condizione occupazionale dei laureati 1997 ad un anno dalla laurea*, Ed. Asterisco, Bologna.
- Bini M., Chiandotto B. (2003), “La valutazione del sistema universitario italiano alla luce della riforma dei cicli e degli ordinamenti didattici”, *Studi e Note di Economia*, 2, 29-61.
- Blangiardo G.C. (1997), *Elementi di demografia*, Il Mulino, Bologna.
- Bond T.G., Fox M.C. (2001), *Applying The Rasch Model: Fundamental Measurement in the Human Sciences*, Lawrence Erlbaum Associates, London.
- Borgonovi E. (2002), “Che cos’è il Welfare Mix?” in Vittadini G. (a cura di), *Liberi di scegliere*, ETASLIBRI, Milano, 170-188.
- Box G.E.P., Tiao G.C. (1992), *Bayesian Inference in Statistical Analysis*, Wiley, Chichester.
- Bryk A.S., Raudenbush S.W. (1992), *Hierarchical Linear Models, Applications and Data Analysis Methods*, Sage Publications, Newbury Park.
- Carlin J.B., Wolfe R., Brown C.H., Gelman A. (2001), “A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes”, *Biostatistics*, 2, 397-416.
- Carpita M. (2004), “Analisi statistica multivariata per la valutazione della job satisfaction” in Pagano A., Vittadini G. (a cura di), *Qualità e valutazione delle strutture sanitarie*, Etas, Milano, 245-263.
- Carpita M. (2005), “Sulla specificazione dei modelli multivello lineari” in Squillante M., D’Ambra L. (a cura di), *Metodi, modelli e tecnologie dell’informazione a supporto delle decisioni*, Franco Angeli Ed., Milano.
- Charnes A., Cooper W.W., Rhodes E., (1978), “Measuring the Efficiency of Decision Making Units”, *European Journal of Operation Research*, 2(6), 429-444.
- Chiandotto B., Bertaccini B. (2003), “Profili e sbocchi occupazionali dei laureati e diplomati dell’Ateneo Fiorentino 1999”, *Working Paper*, Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze.
- Chiandotto B., Bacci S., Bertaccini B. (2004), “I laureati e diplomati dell’Ateneo Fiorentino dell’anno 2000: profilo e sbocchi professionali”, *Working Paper*, Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze.

- Chiandotto B., Varial, R. (2004), “Tempi di conseguimento del titolo nell’Ateneo fiorentino”, *Working Paper*, Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze.
- CIHI, Canadian Institute for Health Information (2004), www.secure.cihi.ca/cihiweb/splash.html.
- Cislaghi C., Biggeri A., Braga M., Lagazio C., Marchi M. (1995), “Exploratory tools for disease mapping in geographical epidemiology”, *Statistics in Medicine*, 14(21/22), 2363-2382.
- Clarke A. (1996), “Why are We Trying to Reduce Length of Stay? Evaluation of the Costs and Benefits of Reducing Time in Hospital Must start from the Objectives that Govern the Change”, *Quality & Safety in Health Care*, 5, 172-179.
- Clarke A. (2002), “Length of In-Hospital Stay and its Relationship to Quality of Care”, *Quality & Safety in Health Care*, 11, 209-210.
- Dagum C., Vittadini G., Lovaglio G., Costa, M. (2003), “A microeconomic recursive model of human capital, income and wealth determination: specification and estimation”, *Proceedings of the Meeting of the American Statistical Association – Business & Economic Statistics Section*, 1119-1126.
- Damberg C., Kerr E.A., Mc Glynn E.A. (1998), “Description of Data Sources and Related Issues” in Mc Glynn E.A., Damberg C., Kerr E.A., Brook, R.A. (a cura di), *Health Information Systems. Design Issues and Analytical Application*, RAND Health Corporation, 5, 43-76.
- Daniels M.J., Gatsonis C. (1997), “Hierarchical polytomous regression models with applications to health services research”, *Statistics in Medicine*, 16, 2311-2325.
- De Long E.R., Peterson E.D., De Long, D.M. Muhlbaier L.H., Hackett S., Mark D.B. (1997), “Comparing risk-adjustment methods for provider profiling”, *Statistics in Medicine*, 16, 2645-2664.
- Donabedian A. (1988), “The quality of care. How can it be assessed?”, *JAMA*, 260(12), 1743-1748.
- Donabedian A. (1990), *La qualità dell’assistenza sanitaria. Principi e metodologie di valutazione*, La Nuova Italia Scientifica, Roma.
- Dubois R.W., Brook R.H., Rogers W.H. (1987), “Adjusted hospital death rates: a potential screen for quality of medical care”, *American Journal of Public Health*, 77, 1162-1167.
- Durand J.F., Lombardo R. (2003), “Interaction terms in Non-linear PLS via Additive Spline Transformation” in M. Schader, Gaul W., Vichi

- M. (a cura di), *Between Data Science and Applied Data Analysis*, Springer, Berlin, 22-29.
- Epstein A. (1995), "Performance Reports on Quality-Prototypes. Problems and Prospects", *New England Journal of Medicine*, 333, 57-61.
- Fabbri D., Fazioli R., Filippini M. (1996), *L'intervento pubblico e l'efficienza possibile*, Il Mulino, Bologna.
- Fabbris L., Gasparotto C., (2000), "Modelli di valutazione della qualità della didattica universitaria" in Fabbris L. (a cura di), *CAPTOR 2000, Qualità della didattica e sistemi computer-assisted*, CLEUP, Padova, 27-43.
- Fabbris L., Martini M.C., Rota G. (2003), "Il valore di mercato dei titoli di laurea e diploma di Padova a sei mesi dal conseguimento" in Carli Sardi L., Del Vecchio F. (a cura di), *Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari*, CLEUP, Padova, 171-194.
- Fabbris L. (2004), "Professionalità e occupabilità dei laureati e dei diplomati dell'Università di Padova", *Quaderno Pharos*, 7, Osservatorio sul Mercato Locale del lavoro, CLEUP, Padova.
- Fan Y., Li Q., Weersink A. (1996), "Semiparametric Estimation of Stochastic Production Frontier Models", *Journal of Business and Economic Statistics*, ASA, vol. 14(4), 460-468.
- Farrel M.J. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society-Series A*, 120, 3, 253-281.
- Fernández C., Koop G., Steel M. (2000), "A Bayesian analysis of multiple-output production frontiers", *Journal of Econometrics*, 98, 47-79.
- Fried H.O., Lovell C.A.K., Schmidt S.S. (1993), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.
- Friedman J.H. (1991), "Multivariate adaptive regression splines", *The Annals of Statistics*, 19(1), 1-67.
- Gallo M., Lombardo R., D'Ambra L. (2001), "Multilevel co-inertia analysis by PLS", *CLADAG 2001– Meeting of the Classification and Data Analysis Group of Italian Statistical Society*, Università degli Studi di Palermo.
- Gertler P.J. (1988), "A latent-variable model of quality determination", *J Bus Econ Stat*, 9, 241-252.

- Goldstein H. (1987), *Multilevel Models in Educational and Social Research*, Charles Griffin & Co Ltd, London.
- Goldstein H. (1995), *Multilevel Statistical Models*, Second Edition, Edward Arnold London, Wiley, New York.
- Goldstein H., Spiegelhalter D.J. (1996), "League tables and their limitations: statistical issues in comparisons of institutional performance", *Journal of the Royal Statistical Society-Series A*, 159, 385-443.
- Goldstein H. (1997), "Methods in school effectiveness research", *School Effectiveness and School Improvement*, 8, 369-395.
- Goldstein H., Woodhouse G. (2000), "School Effectiveness Research and Educational Policy", *Oxford Review of Education*, 26(3-4), 353-363.
- Good T.L., Biddle B.J., Brody J.E. (1975), *Teachers make a difference*, Lanham, MD: University Press America.
- Gori E., Mealli F., Rampichini C. (1993), "Indicatori di efficienza ed efficacia per la valutazione dell'attività di formazione professionale", *Statistica*, 53(3), 501-533.
- Gori E., Vittadini G. (1999), "La valutazione dell'efficienza ed efficacia dei servizi alla persona. Impostazione e metodi" in Gori E., Vittadini G. (a cura di), *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano, 121-241.
- Gori E., Pagani L., Seghieri C. (2001), "L'efficacia dell'istruzione media superiore rispetto alla carriera universitaria dei maturi" in *Atti del Convegno Intermedio Sis 2001, Processi e metodi statistici di valutazione – Sessioni Spontanee*, Roma 4-6 Giugno.
- Gori E., Grassetti L, Rossi C.(2002), "Linear Models in Efficiency Analysis: Evidence from Validation Procedures", *Statistica Applicata*, Vol. 14, n. 3, 247-275.
- Gori E. (2004), "Indicatori di efficienza ed efficacia per l'accreditamento delle agenzie di formazione professionale ed orientamento/inserimento lavoro" in Elias G. (a cura di), *Accreditamento degli enti di formazione e di orientamento al lavoro*, Armando Ed., Roma, 47-68.
- Grassetti L.,Gori E., Bellio R. (2003), "Efficiency estimation of hospital services: a survey and multilevel developments" in *Atti del Convegno Intermedio Sis 2001, Analisi Statistica Multivariata per le scienze economico-sociali, le scienze naturali e la tecnologia – Sessioni Plenarie*, Napoli 9-11 giugno.

- Grassetti L. Gori E. Minotti S.C. (2005), "Multilevel flexible specification of the production function in health economics", *IMA Journal of Management Mathematics (Special Healthcare Issue)*, 16(4), 383-398.
- Hausman J.A. (1978), "Specification tests in econometrics", *Econometrica*, 46, 1263-1271.
- Heagerty P.J., Kurland B.F. (2001), "Misspecified maximum likelihood estimates and generalized linear mixed models", *Biometrika*, 88, 973-985.
- Hox J.J. (1995), *Applied Multilevel Analysis*, TT-Publikaties, Amsterdam.
- Iezzoni L.I., Ash A.S., Shwartz M., Daley J., Hughes J.S., Mackiernan Y.D. (1996), "Judging Hospitals by Severity-Adjusted Mortality Rates: the Influence of the Severity-Adjustment Method", *American Journal of Public Health*, 86(10), 1379-1387.
- JCAHO, Joint Commission on Accreditation of Healthcare Organizations (1998), *National Library of Healthcare Indicators*, Health Plan and Network Edition.
- JCAHO, Joint Commission on Accreditation of Healthcare Organizations (2004), www.jcaho.org.
- Jencks S.F., Daley J., Draper D., Thomas N., Lenhart G., Walker J. (1988), "Interpreting hospital mortality data. The role of clinical risk adjustment", *JAMA*, 260, 3611-3616.
- Johannensons M., Johansson P.O. (1997), "Quality of Life and the WTP for an increased Life Expectancy at an Advanced Age", *Journal of Public Economics*, 65, 219-228.
- Johansson P.O. (1995), *Evaluating Health Risks: An Economic Approach*, Cambridge University Press.
- Khumbakar S.C., Knox Lovell C.A. (2000), *Stochastic Frontiers Analysis*, Cambridge University Press, New York.
- Langford I.H., Leyland A. Rasbasch J., Goldstein H. (1999), "Multi-level modelling of the geographic distribution of the disease", *Journal of the Royal Statistical Society-Series C*, 48, 253-268.
- Lei P.W., Bassiri D., Schulz E.M. (2001), "Alternatives to the grade point average as a measure of academic achievement in college", *ACT Research Report Series*, 4.
- Leyland A.H. (1995), "Examining the relationship between length of stay and readmission rates for selected diagnoses in Scottish hospi-

- tals”, *IMA Journal of Mathematics Applied in Medicine and Biology*, 12, 175-184.
- Leyland A.H., Boddy F.A. (1998), “League Tables and Acute Myocardial Infarction”, *The Lancet*, 351, 555-558.
- Leyland A.H., Goldstein H. (a cura di) (2001), *Multilevel Models of Health Statistics*, Wiley, London.
- Lilford R., Mohammed M.A., Spiegelhalter D.J., Thomson R. (2004), “Use and Misuse of Process and Outcome Data in Managing Performance of Acute Medical Care: Avoiding Institutional Stigma”, *The Lancet*, 364, 1147-1154.
- Linn R.L., Slinde J.A. (1977), “The Determination of the Significance of Change between pre- and posttesting Periods”, *Review of Educational Research*, 47, 121-150.
- Lockheed M.E., Hanushek E.R. (1994), “Concepts of Educational Efficiency and Effectiveness” in Husén T., Postlethwaite T. N. (eds) *The International Encyclopedia of Education – Second Edition*, Vol.3, Oxford: Pergamon, 1779-1784.
- Longford I.H., Lewis T. (1998), “Outliers in Multivariate Data”, *Journal of the Royal Statistical Society–Series A*, 161, 121-160.
- Lovaglio P.G. (2003), “The estimate of Customer Satisfaction in a Reduced Rank Regression framework”, *Total Quality Management*, 16, 33-44.
- Marshall E.C., Spiegelhalter D.J. (2001), “Institutional Performance” in Goldstein H., Leyland A.H. (a cura di), *Multilevel Modelling of Health Statistics*, Wiley, Chichester, 127-142.
- McCulloch C.E., Searle R.E. (2000), *Generalized, Linear and Mixed Models*, Wiley, New York.
- Meeusen W., Van Den Broeck J. (1997), “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error”, *International Economic Review*, 18, 435-444.
- Molteni M. (1997), *Le Misure di Performance nelle Aziende Non-Profit di Servizi alla Persona*, CEDAM, Padova.
- Moramarco V. (1999), “Qualità ed efficienza economica” in Gori E., Vittadini G. (a cura di), *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano, 1-22.
- Morris C.N., Christiansen C.L. (1996), “Hierarchical Models for Ranking and for Identifying Extremes, with application” in Bernardo J.M., Berger J.O. Dawid A.P., Smith A.F.M. (a cura di), *Bayesian Statistic*, 5, Oxford University press, Oxford, 277-298.

- Muthen B., Speckart G. (1985), "Latent variable probit ANCOVA: treatment effects in the California Civil Addict Programme", *The British Journal of Mathematical and Statistical Psychology*, 38, 161-170.
- NHS, National Health Service (2004), *A Commentary on Star Ratings 2002-2003*, www.nhs.uk./root/starratings/explained.asp.
- Normand S.L.T., Glickman M.E., Ryan T.J. (1995), "Modelling mortality rates for elderly heart attack patients: profiling hospitals" in Gatsonis C. *et al.* (a cura di), *The Co-Operative Cardiovascular Project. Case Studies in Bayesian Statistics*, Springer, New York, 435-456.
- Opit L.J. (1991), "The Measurement of Health Service Outcomes" in Holland *et al.* (a cura di) *Oxford Text Book of Public Health*, vol. 3, Oxford University Press, Oxford, 159-172.
- Pagano A., Rossi C. (1999), "La valutazione dei servizi sanitari" in Gori E., Vittadini G. (a cura di), *Qualità e valutazione nei servizi di pubblica utilità*, Etas, Milano, 263-294.
- Pagano A. (a cura di) (2005), "Indicatori di qualità, banche dati, benchmarking delle strutture sanitarie" in Pagano A., Vittadini G. (a cura di), *Qualità e valutazione delle strutture sanitarie*, ETASLIBRI, Milano, 61-119.
- Perchinunno P. (2005), "Le tecniche di *record linkage*: approcci statistici", *Annali del Dipartimento di Scienze Statistiche – Università degli Studi di Bari*, Vol. IV, 501-527.
- Plewis I., Goldstein, H. (1998), "Excellence in Schools: a Failure of Standards", *British Journal of Curriculum and Assessment*, 8, 17-20.
- Ray S.C. (2003), "Measuring Scale Efficiency from the Translog Multi-Input, Multi-Output Distance Function", *Working Paper 2002-25*, Department of Economics Working Paper Series – University of Connecticut.
- Rice N., Leyland A. (1996), "Multilevel models: applications to health data", *Journal of Health Services Research & Policy*, 3, 154-164.
- Rogosa D.R., Brand D., Zimowski M. (1982), "A Growth Curve Approach to the Measurement of Change", *Psychological Bulletin*, 90, 726-748.
- Rogosa D.R., Willett J.B. (1985), "Understanding Correlates of Change by Modeling Individual Differences in Growth", *Psychometrika*, 50, 203-228.

- Rogosa D.R. (1995), "Myths and methods: "Myths about longitudinal research" plus supplemental questions" in Gottman J.M. (a cura di), *The Analysis of Chang*, Lawrence Erlbaum Associates, New Jersey, 3-65.
- Rowan B.R., Correnti R., Miller R.J. (2002), "What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools", *Teachers College Record*, 104(8), 1525–1567.
- Rubin D.B. (1984), "Bayesianly justifiable and relevant frequency calculations for the applied statistician", *The Annals of Statistics*, 12(4), 1151-1172.
- Scott A., Shiell A. (1997), "Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model", *Journal of Health Economics*, 16, 323-342.
- Seiford L. M., Thrall R.M. (1990), "Recent Developments in DEA: the Mathematical-Programming Approach to frontier Analysis", *Journal of Econometrics*, 46(1-2), 7-38.
- Skaug H.J. (2002), "Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models", *Journal of Computational and Graphical Statistics*, 11, 458-470.
- Snijders T., Bosker R. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications, London.
- Solucient (2003), *Solucient's 100 Top Benchmark Hospitals*, www.solucient.com.
- Srivastava V.K., Giles D.E.A. (1987), *Seemingly Unrelated Regression Equations*, Marcel Dekker, New York.
- Subbotin M.T. (1923), "On the law of frequency of errors", *Matem. Sbornik*, 31, 296-301.
- Tesio L. (2003), "Measuring Behaviours and Perceptions: Rasch Analysis as a Tool for Rehabilitation Research", *Journal of Rehabilitation Medicine*, 35(3), 105-115.
- Thomas N., Longford N.T., Rolph J.E. (1994), "Empirical Bayes Methods for Estimating Hospital Specific Mortality Rates", *Statistics in Medicine*, 13, 889-903.
- Tsionas E.G. (2002), "Stochastic Frontier Models with Random Coefficients", *Journal of Applied Econometrics*, 17, 127-147.

- Turner R.M., Omar R.Z., Thompson S.G. (2001), "Bayesian methods of analysis for cluster randomised trials with binary outcome data", *Statistics in Medicine*, 20, 453-472.
- Verbeke G., Molenberghs G. (2000), *Linear Mixed Models for Longitudinal Data*, Springer, New York.
- Vittadini G. (1989), "Indeterminacy problems in the LISREL model", *Multivariate Behavioral Research*, 24, 397-414.
- Vittadini G. (2001), "On the use of multivariate regression models in the context of multilevel analysis" in Borra, S. *et al.* (a cura di), *Advances in Classification and Data Analysis*, Springer, 225-232.
- Vittadini G., Sanarico M. (2003), "Metodologia per lo studio dell'efficacia relativa di strutture sanitarie", *Statistica Applicata*, Vol. 14(3), 331-359.
- Vittadini G., Minotti S.C. (2005), "A methodology for measuring the relative effectiveness of healthcare services", *IMA Journal of Management Mathematics (Special Healthcare Issue)*, 16(3), 239-254.
- Vittadini G., Sanarico M., Berta P. (2005), "Testing procedures for multilevel models with administrative data", *CLADAG 2005 – Meeting of the Classification and Data Analysis Group of Italian Statistical Society*, Università degli Studi di Parma.
- Vroman Battle, M., Rakow E.A. (1993), "Zen and the art of reporting differences in data that are not statistical significant", *IEEE Transactions on Professional Communication*, 36(2), 75-80.
- Ware J.E. (1993), *SF-36. Health Survey. Manual and Interpretation – Guide*, The Health Institute, New England Medical Center, Boston.
- Wright B.D., Masters G. (1982), *Rating Scale Analysis*, MESA Press, Chicago.
- Yang M. (2001), "Multinomial Regression" in Leyland A. and Goldstein H. (a cura di), *Multilevel Models of Health Statistics*, Wiley, London.
- Young F. (1981), "Quantitative analysis of qualitative data", *Psychometrika*, 46, 357-388.
- Zanella A. (2001), "Valutazioni e modelli interpretativi: ricerca di un quadro complessivo" in *Atti del Convegno Intermedio SIs 2001, Processi e metodi statistici di valutazione – Sessioni Plenarie*, Roma 4-6 Giugno.
- Zangrandi A. (2005), "La gestione delle aziende sanitarie: alla ricerca di qualità ed efficienza (con particolare riferimento alle organizza-

- zioni pubbliche)” in Pagano A., Vittadini G. (a cura di), *Qualità e valutazione delle strutture sanitarie*, ETASLIBRI, Milano, 31-59.
- Zaslavsky A. (2001), “Statistical issues in reporting quality data: small samples and casemix variation”, *Int. J. Qual. Health Care*, 13(6), 481-488.