

# Riforma sanitaria e produzione ospedaliera

**Daniele Fabbri**

Dipartimento di Scienze Economiche  
Università di Bologna

4-10-99

Questo lavoro si colloca nell'ambito del progetto "Matrici di mobilità sanitaria", svolto in collaborazione fra l'Agenzia Sanitaria dell'Emilia-Romagna e il Dipartimento di Scienze Economiche dell'Università di Bologna. Si ringrazia il personale dell'Agenzia Sanitaria e dell'Assessorato alla Sanità dell'Emilia-Romagna per aver messo a disposizione le banche dati necessarie. La dottoressa Lucia Nobilio ha svolto un eccellente lavoro di assistenza alla ricerca. L'autore desidera ringraziare Rosella Levaggi e un anonimo *referee* della rivista *Politica Economica* per gli utili consigli e i commenti forniti a versioni antecedenti di questo lavoro. La responsabilità di eventuali errori o omissioni restano dell'autore. Le opinioni qui espresse non sono riferibili all'Agenzia Sanitaria Regionale dell'Emilia-Romagna.

## *Sintesi*

La recente riforma del sistema di finanziamento ha reso gli ospedali responsabili dei risultati gestionali raggiunti, attribuendo loro una parte del rischio finanziario. In secondo luogo, ha reso più stringenti i vincoli di bilancio degli enti finanziatori con ripercussioni sugli incentivi al contenimento dei consumi e delle produzioni sanitarie. Infine ha ridimensionato le risorse a disposizione degli ospedali e ridisegnata la distribuzione dimensionale dell'offerta. Sono quindi all'opera almeno due dinamiche: una che tende ad incoraggiare i produttori alla riorganizzazione e all'ottimizzazione dei processi produttivi *a risorse date*, e una che tende, nello stesso tempo, a ridimensionare le risorse a loro disposizione. Nel presente lavoro valutiamo la consistenza del primo effetto scorporandolo dal secondo. Per farlo adottiamo la prospettiva di analisi economica dell'efficienza produttiva misurando l'evoluzione dell'efficienza tecnica e dell'efficienza di scala nella produzione ospedaliera dei presidi pubblici emiliano-romagnoli fra il 1994 e il 1997. I risultati dell'analisi offrono alcuni interessanti spunti per la politica economica e la regolamentazione.

La produzione di ricoveri ospedalieri sembra caratterizzata da rendimenti di scala rapidamente decrescenti. Questo risultato confermerebbe l'opportunità di perseguire politiche di accorpamento e di razionalizzazione decisamente orientate all'accrescimento del segmento più basso della distribuzione dimensionale dei presidi. L'evoluzione dell'efficienza tecnica nel periodo esaminato pare coerente con l'espressione delle potenzialità produttive dei presidi, incoraggiata nella prima fase di attuazione della riforma, cui fa seguito, nell'ultimo anno esaminato, un lieve peggioramento dovuto alla sensibile riduzione dei volumi di produzione. Questa riduzione parrebbe direttamente legata a vincoli finanziari posti dagli enti finanziatori sull'innalzamento dei volumi, piuttosto che alla riduzione degli input a disposizione. Circa il nesso fra modelli di finanziamento e efficienza degli ospedali, l'analisi dell'evoluzione delle prestazioni produttive suggerisce che, per gli ospedali in gestione, l'effetto incentivante sulla produttività sia stato inizialmente forte e efficace, mentre successivamente sia prevalso l'obiettivo del razionamento delle prestazioni. Questo ha messo in luce alcuni eccessi di dotazione nei fattori produttivi, il cui assorbimento si segnala come un obiettivo prioritario dell'azione di politica economica in questo settore. Al contrario, per le Aziende Ospedaliere l'effetto incentivante sulla produttività è inizialmente inefficace, mentre diventa rilevante, anche se non in misura statisticamente significativa, solo nell'ultimo anno. Quest'evidenza ci pare confermare la preoccupazione espressa nel Piano Sanitario Nazionale, laddove si rileva la necessità di individuare *“sistemi di valutazione della produttività delle strutture destinarie di modalità di remunerazione fondate principalmente sui costi”*, quasi che per queste gli incentivi al miglioramento delle prestazioni produttive fossero più deboli di quelli degli ospedali-azienda. A nostro parere, tuttavia, occorre estendere tale preoccupazione anche alle Aziende Ospedaliere di cui sorprende la modestia del miglioramento, soprattutto qualora si consideri anche il ruolo giocato dal progresso tecnico.

JEL : I11, D24, C61

## ***1 Introduzione***

A partire dal gennaio 1995 l'assistenza ospedaliera viene finanziata in base ad un modello articolato su due livelli [Taroni (1996)]. Al primo, la Regione, ricevuta dallo Stato la propria quota di Fondo Sanitario, trasferisce le risorse finanziarie a ciascuna Azienda Unità Sanitaria Locale (AUSL). Questa, a sua volta, finanzia le prestazioni ospedaliere erogate ai propri assistiti da parte dei fornitori accreditati: Aziende Ospedaliere, ospedali gestiti dalle AUSL, cliniche private convenzionate, operanti tanto all'interno che all'esterno del sistema sanitario regionale. Il primo livello di trasferimento avviene su base capitaria, aggiustata per le caratteristiche demografiche, epidemiologiche e sanitarie delle popolazioni assistite da ciascuna AUSL, con l'obiettivo di ripartire equamente le risorse fra gli acquirenti collettivi di prestazioni sanitarie. Il secondo livello si basa sull'applicazione di un tariffario regionale per episodio clinico, secondo la classificazione DRG. Il tariffario regionale è utilizzato direttamente per regolare le transazioni fra aziende diverse e fra aziende che non sono legate da accordi di fornitura, mentre costituisce solo un insieme di valori di riferimento per le altre transazioni. In particolare le prestazioni erogate da presidi aziendali sono di fatto finanziate in modo discrezionale. Mentre per le prestazioni oggetto di negoziazione fra aziende diverse è possibile concordare, fra AUSL e fornitore, tariffe inferiori a quella regionale in ragione del 20%<sup>1</sup>.

L'introduzione di questo sistema di finanziamento prospettico degli ospedali è stata accompagnata dal contestuale processo di aziendalizzazione dei presidi maggiori e delle USL, dall'attribuzione di responsabilità finanziaria alle Regioni, da provvedimenti di blocco nel turn-over del personale sanitario e di razionalizzazione della struttura di offerta ospedaliera<sup>2</sup>. L'effetto combinato di questi interventi ha reso gli ospedali responsabili dei risultati gestionali raggiunti, attribuendo loro una parte del rischio finanziario. In secondo luogo, ha "irrigidito" i vincoli di bilancio degli enti finanziatori con ripercussioni sugli incentivi al contenimento dei consumi e delle produzioni sanitarie. Infine ha ridimensionato le risorse a disposizione degli ospedali e ridisegnata la distribuzione dimensionale dell'offerta.

La letteratura ha ampiamente dimostrato che tutti i sistemi di finanziamento prospettico degli ospedali, attribuendo ai produttori un diritto al residuo di gestione, contengono elevati incentivi al contenimento dei costi ed al perseguimento dell'efficienza. Tuttavia la massimizzazione del residuo introduce anche forti incentivi all'adozione di comportamenti opportunistici nella erogazione delle prestazioni, in un contesto, quello sanitario, in cui l'output appropriato dipende dalle caratteristiche del paziente e quindi viene erogato condizionatamente ad una valutazione diagnostica espressa dall'erogatore. In letteratura<sup>3</sup> vengono indicati come "possibili" i seguenti

---

<sup>1</sup> L'interpretazione dei decreti 502/92 e 517/93 ha prodotto modelli regionali in parte diversi. La nostra descrizione coglie, in modo forse più calzante, i caratteri del modello emerso in Emilia-Romagna. Analisi più generali si possono trovare in Rebba e Rizzi (1998), Fabbri e Ugolini (1999) e France (1994).

<sup>2</sup> Una accurata descrizione di questi provvedimenti si trova in Mapelli (1994 e 1995).

<sup>3</sup> Coulam e Gaumer (1991) offrono una rassegna esaustiva degli effetti dell'introduzione del sistema di pagamento prospettico nell'ambito del programma Medicare statunitense.

comportamenti: la dimissione anticipata con riduzione della degenza media<sup>4</sup>; l'aumento del numero di ricoveri, in particolare, di quelli dovuti a dimissioni anticipate e successive riammissioni in ospedale; l'adozione di metodi di trattamento "standardizzati" indipendentemente dalla gravità e dalle peculiarità della casistica; la specializzazione nel trattamento dei pazienti caratterizzati da costi relativamente più bassi all'interno di ciascun DRG (*effetto lotteria* e selezione del *case-mix*) [si veda Ellis e McGuire (1996)]; l'erogazione di trattamenti di elevata qualità ai pazienti con costi effettivi più bassi della media di un certo DRG (*cream skimming*) e quella di trattamenti di qualità più bassa ai pazienti maggiormente costosi (*skimming*) al fine di attirare i primi e scoraggiare i secondi [si veda Ellis (1998)]; l'attribuzione del paziente a classi diagnostiche cui corrispondono tariffe più remunerative (*upcoding*) [Carter, Newhouse e Relles (1990)]; il trasferimento ad altre strutture, fino al rifiuto di trattare, le tipologie di pazienti ad alto costo di trattamento<sup>5</sup>.

L'opportunità di sopportare il rischio legato ai possibili effetti negativi che l'introduzione di un modello di finanziamento prospettico degli ospedali può produrre, è giustificato dalla priorità assegnata al miglioramento nella organizzazione produttiva, nell'efficienza e, in ultima istanza, alla riduzione dei costi di fornitura dei servizi. In effetti, il miglioramento dell'efficienza produttiva ha assunto, in molti passaggi delle recenti riforme, il valore di obiettivo prioritario. L'esercizio che qui proponiamo intende offrire alcune prime evidenze empiriche circa il conseguimento di tale obiettivo, e quindi dare una misura dei vantaggi ottenuti, a fronte dei rischi non noti<sup>6</sup>. In questo studio valutiamo infatti l'andamento dell'efficienza della produzione ospedaliera nel periodo che va dal 1994 al 1997<sup>7</sup>.

I concetti di efficienza definiti nella teoria economica sono numerosi. Nel nostro studio ci concentriamo sulle componenti di efficienza tecnica e di scala. L'opportunità di esaminare queste due componenti è ispirata a due ordini di considerazioni. Per quanto attiene all'efficienza tecnica, è noto che, nell'ambito delle produzioni pubbliche e di quelle altamente regolamentate, il perseguimento dell'efficienza tecnica non è conflittuale con gli altri obiettivi dell'intervento pubblico [Pestieu e Tulkens (1990)]. Esaminare l'efficienza tecnica permette, quindi, di formulare deduzioni non contaminate da considerazioni legate al contenuto redistributivo delle produzioni. Per quanto riguarda l'efficienza di scala, occorre ricordare che molti interventi di razionalizzazione della struttura di offerta ospedaliera presuppongono l'esistenza di vantaggi derivanti

---

<sup>4</sup> La riduzione della degenza media è uno dei principali comportamenti virtuosi incentivati dal sistema DRG. L'evidenza empirica al riguardo è assolutamente concorde. Esiste tuttavia la possibilità che la riduzione della degenza possa essere eccessiva e che dimissioni troppo anticipate abbiano effetti negativi sugli esiti medici per i pazienti. Cutler (1995) ad esempio ha riscontrato un significativo effetto sulla mortalità ospedaliera.

<sup>5</sup> L'evidenza empirica sul fenomeno del dumping è piuttosto limitata. I pochi studi presenti in letteratura ne ridimensionano la rilevanza. Si veda Coulam e Gaumer (1991).

<sup>6</sup> Fabbri (1999) analizza uno di questi rischi: gli opportunismi legati alla ripetizione dei ricoveri. L'evidenza, riferita agli ospedali, pubblici e privati accreditati, dell'Emilia-Romagna, non sembra testimoniare alcun significativo incremento del fenomeno. Questo appare, al contrario, del tutto sotto controllo.

<sup>7</sup> I contributi in letteratura che cercano di valutare gli effetti di riforme del meccanismo di finanziamento sull'efficienza produttiva degli ospedali sono pochi. Si veda in particolare Soderlund, Csaba, Gray, Milne e Raftery (1997) e Linna (1999).

dalla concentrazione della produzione in pochi luoghi di cura. Valutare l'esistenza e misurare l'ampiezza di questi vantaggi ci sembra, a questo riguardo, presupposto di non poco rilievo. Per quanto riguarda la necessità di esaminare congiuntamente efficienza tecnica e efficienza di scala, essa emerge a causa dalla azione concomitante di due dinamiche descritte sopra: una che vuole incoraggiare i produttori alla riorganizzazione e all'ottimizzazione dei processi produttivi *a risorse date*, ovvero al perseguimento dell'efficienza tecnica, e una che tende a ridimensionare le risorse a loro disposizione, modificando la scala di produzione. Prendendo le mosse da queste premesse nel presente lavoro valutiamo la consistenza del primo effetto cercando di scorporarlo dal secondo.

Sotto il profilo dell'interesse di politica sanitaria la nostra applicazione offre evidenza circa le risposte dei produttori agli incentivi introdotti dalla riforma, in particolare le risposte differenziali date da tipologie di strutture differenti: Aziende Ospedaliere e presidi delle Aziende Sanitarie. Questa distinzione è di un certo rilievo, anche alla luce del recente Piano Sanitario Nazionale. E' infatti noto che l'introduzione del sistema di pagamento prospettico a tariffa DRG abbia interessato, in misura prevalente, le prime e solo in via marginale le seconde. Per tutte le prestazioni erogate dai presidi aziendali ai pazienti assistiti dalla AUSL la tariffa DRG ha costituito solo un prezzo di riferimento, cui ciascuna azienda ha derogato in risposta alle proprie esigenze di programmazione. Nel PSN, del resto, si indica come prioritaria, "allo scopo di integrare e completare i sistemi di remunerazione delle strutture produttive pubbliche e private, .... l'individuazione di modalità di finanziamento, commisurate ai costi standard di produzione e agli obiettivi da perseguire, delle attività escluse dal pagamento a prestazione e *contestuale definizione di sistemi di valutazione della produttività delle strutture destinatarie di tale modalità di remunerazione*" [Ministero della Sanità, PSN 1998-2000, pp. 73, il corsivo è nostro]. Il nostro studio offre l'applicazione di una metodologia consolidata di analisi dell'efficienza che si colloca a pieno titolo nell'ambito di ricerca delineato dal PSN. Del resto, la metodologia qui impiegata è già stata ripetutamente indicata come un utile strumento per le politiche di razionalizzazione delle amministrazioni pubbliche [si veda Petretto e Pisauro (1995) e Fabbri, Fazioli e Filippini (1996)].

L'analisi è condotta impiegando la Data Envelopment Analysis (DEA), su un campione rilevante di ospedali pubblici, quelli della regione Emilia-Romagna, nel quadriennio che va dal 1994 al 1997. La tecnica DEA è assai consolidata nella letteratura di Economia Sanitaria. Si tratta di una metodologia non parametrica di misurazione dell'efficienza che fa ricorso alle tecniche di programmazione matematica piuttosto che alla stima econometrica di funzioni di produzione o di costo<sup>8</sup>. La tecnica DEA opera una valutazione relativa dell'efficienza di una unità produttiva (UP) rispetto alle prestazioni di altre UP che producono il medesimo insieme di output utilizzando lo stesso insieme di input. Le applicazioni DEA nel settore sanitario sono ormai piuttosto numerose. A partire dal lavoro di Sherman (1984), questa tecnica è stata ripetutamente impiegata nel settore della produzione ospedaliera<sup>9</sup> per confrontare l'efficienza di produttori con differenti strutture di controllo [Grosskopf e Valdmanis (1987), Nyman e

---

<sup>8</sup> Per un confronto fra i due approcci in ambito sanitario si veda Banker, Conrad e Strauss (1986).

<sup>9</sup> Per una rassegna aggiornata delle applicazioni al settore sanitario si veda Hollingsworth, Dawson e Maniadakis (1999).

Bricker (1989) e Bruning e Register (1989)] e di obiettivi [Grosskopf, Margaritis e Valdmanis (1997)]; per valutare l'andamento nel tempo dell'efficienza [Burgess e Wilson (1993)] e della produttività [Burgess e Wilson (1995), Färe, Grosskopf, Lindgren e Roos (1994)]; per misurare l'efficienza tecnica e quella allocativa [Byrnes e Valdmanis (1994)].

Il lavoro è organizzato come segue. Nel prossimo paragrafo descriviamo la tecnica DEA, non prima di aver definito i concetti di input-efficienza e output-efficienza alla Debreu-Farrell. Nella sezione successiva presentiamo alcuni elementi di analisi economica della produzione ospedaliera e definiamo i concetti di efficienza che utilizzeremo nel seguito. Il quarto paragrafo presenta i dati e il modello stimato. Nel paragrafo quinto presentiamo i risultati e nella sezione conclusiva ne commentiamo e discutiamo le implicazioni.

## 2 *Efficienza tecnica e Data Envelopment Analysis*

L'efficienza della prestazione di una unità produttiva (UP), sia che essa operi nell'ambito del settore pubblico che in quello del settore privato, si definisce e si misura in termini relativi, ovvero rispetto alla prestazione di una unità produttiva presa come "pietra di paragone". Pertanto per valutare l'efficienza o l'inefficienza della prestazione di una UP reale occorre confrontarla, nei modi definiti secondo una tecnica prestabilita, con la prestazione di una UP virtuale avente la particolare proprietà di essere *la migliore prestazione ritenuta possibile alle condizioni in cui opera la UP reale*. L'elaborazione di una metodologia di misurazione dell'efficienza si sviluppa quindi in tre fasi:

1. definizione di regole e criteri teorici generali utili ad individuare le UP le cui prestazioni assolvono il ruolo di pietre di paragone;
2. scelta di una tecnica per costruire le UP pietre di paragone;
3. definizione di una regola per confrontare le prestazioni delle UP reali con quelle delle unità pietre di paragone.

Nell'approccio generale sposato in questo lavoro, la prima fase si svolge sulla base delle indicazioni fornite dalla teoria economica della produzione. Per questo è utile definire questo approccio come *approccio economico di analisi dell'efficienza*. La seconda fase si realizza attraverso l'elaborazione di appropriate tecniche econometriche e di programmazione matematica, utili ad individuare e a costruire, in modo coerente con quanto previsto nella prima fase, le UP di confronto. Infine la terza si sostanzia nella definizione delle regole e delle tecniche per misurare lo scarto fra le prestazioni osservate delle UP e quelle virtuali.

### 2.1 *Teoria della produzione e indicatori di Farrell*

L'efficienza tecnica fa riferimento alla capacità, da parte di una unità produttiva (UP), di usare le risorse in modo da produrre la maggior quantità possibile di output con i fattori produttivi a disposizione ovvero di utilizzare la minor quantità di input per produrre una quantità di output prefissata. In questo senso l'analisi dell'efficienza tecnica può essere orientata all'aumento dell'output (output-efficienza) o alla riduzione dell'input (input-efficienza). Debreu (1951) e Farrell (1957) hanno definito una utile misura di input efficienza come il complemento a uno della massima riduzione equiproporzionale di tutti gli input che permette di produrre un livello prefissato di output. In modo analogo si può definire anche la misura di output efficienza. Un po' di

notazione è a questo proposito utile per agevolare la discussione nel caso multidimensionale.

Sia  $X=(x_1, \dots, x_n) \in R^n_+$  il vettore di input utilizzato dai produttori per produrre il vettore di output  $Y=(y_1, \dots, y_m) \in R^m_+$ . Una tecnologia produttiva può essere rappresentata con un input set:

$$L(Y)=\{X: (X,Y) \in \text{insieme delle possibilità produttive}\},$$

che per ogni  $Y \in R^m_+$  possiede un isoquanto:

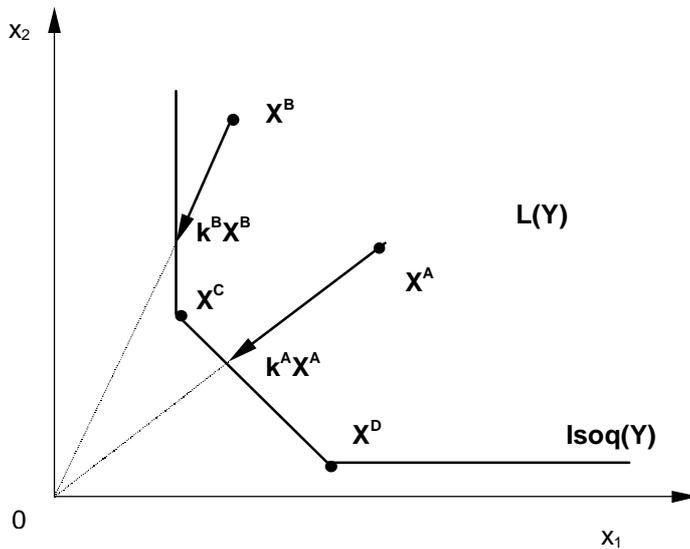
$$\text{Isoq}(Y)=\{X: X \in L(Y), kX \notin L(Y), k \in [0,1)\}$$

La misura di Debreu-Farrell orientata agli input può essere definita formalmente come segue:

$$DF_I(Y,X)= \min\{k: kX \in L(Y)\} \tag{1}$$

Dalla definizione segue che  $DF_I \leq 1$ , dove un valore di 1 indica che siamo in presenza di efficienza tecnica poiché non è possibile alcuna riduzione equiproportionale degli input, mentre un valore minore di 1 misura la presenza dell'inefficienza tecnica. La misura di output-efficienza alla Debreu-Farrell può essere agevolmente definita in modo analogo.

Figura 1: La misura di input-efficienza alla Debreu-Farrell



La misurazione empirica dell'efficienza tecnica alla Debreu-Farrell orientata agli input, richiede quindi la costruzione, a partire di un insieme di osservazioni, della mappa di isoquanti. Questa misura viene illustrata nella **Figura 1**. Nella rappresentazione con isoquanti, i vettori di input  $X^A$  e  $X^B$ , con i quali viene prodotto il vettore di output  $Y$ , possono essere ridotti in modo proporzionale, o in senso radiale, e rimanere comunque in grado di produrre il vettore di output  $Y$ . Al contrario i due vettori  $X^C$  e  $X^D$  non possono essere ridotti in senso radiale e continuare a produrre il vettore di

output  $Y$ . Di conseguenza  $DF_I(Y, X^C) = DF_I(Y, X^D) = 1 > \max\{DF_I(Y, X^A) = DF_I(Y, X^B)\}$ . Si noti che il vettore di input  $k^B X^B$  non può essere ulteriormente ridotto in senso radiale e continuare a produrre l'output  $Y$ , anche se contiene un eccesso di input <sup>10</sup>.

La tecnica DEA, che utilizzeremo in questo studio, permette di misurare la massima contrazione (espansione) radiale degli input (output) compatibile con l'insieme delle possibilità produttive esistente, o spazio di inviluppo. Pertanto l'analisi DEA consiste di due elementi: una procedura di programmazione matematica che costruisca, a partire dalle UP rilevate, lo spazio di inviluppo; in secondo luogo una procedura che ottimizzi la proiezione radiale di ciascuna UP all'interno dell'inviluppo.

## 2.2 La Data Envelopment Analysis

Vi siano  $l$  UP che impiegano  $n$  input e producono  $m$  output distinti. La  $j$ -esima UP usa  $x_{ij}$  unità di input  $i$ , con  $i = 1, \dots, n$ , per la produzione di  $y_{rj}$  unità di output  $r$ , con  $r = 1, 2, \dots, m$ . Supponiamo di disporre, per ciascuna UP, di un sistema di pesi,  $U = (u_1, \dots, u_m)$  e  $V = (v_1, \dots, v_n)$ , con cui misurare l'importanza relativa dei fattori e dei prodotti. L'efficienza di una UP può essere misurata dal rapporto tra la somma degli output prodotti, pesata con i pesi  $u$ , e la somma degli input impiegati, pesata con i pesi  $v$ <sup>11</sup>. Si tratta in sostanza di ipotizzare che esista un sistema di prezzi sulla cui base misurare il rapporto fra valore della produzione e valore delle risorse impegnate. Un simile sistema di prezzi, per ciascuna UP, non è dato. Scopo della DEA è quello di ricostruirlo, permettendo alla UP di "scegliere" i vettori di pesi in modo da massimizzare il proprio indice di efficienza sotto la condizione che nessun'altra UP ottenga, con gli stessi pesi, un indice superiore ad un valore prestabilito. Il problema, per la generica UP  $j$ -esima, può essere formulato come segue:

$$\text{Max}_{V_j, U_j} Z_j = \frac{\sum_{r=1}^m u_{rj} y_{rj}}{\sum_{i=1}^n v_{ij} x_{ij}} \quad (4)$$

$$0 \leq \frac{\sum_{r=1}^m u_{rc} y_{rc}}{\sum_{i=1}^n v_{ic} x_{ic}} \leq 1 \quad \text{per } c = 1, \dots, j, \dots, l \quad (5)$$

$$v_{ij}, u_{rj} > 0 \quad (6)$$

---

<sup>10</sup> Questa circostanza è coerente con la definizione di efficienza alla Debreu-Farrell, tuttavia non lo è con una definizione più appropriata come quella di Koopmans (1951), secondo la quale un produttore è input-efficiente qualora per ridurre un input debba necessariamente diminuire la produzione di almeno un output o accrescere l'utilizzo di almeno un altro input. Tale circostanza non si verifica nel caso del vettore  $k^A X^A$ .

<sup>11</sup> Si tratta in sostanza di ipotizzare che esista un sistema di prezzi sulla cui base misurare il rapporto fra valore della produzione e valore delle risorse impegnate.

Il programma "frazionale" definito dalla massimizzazione di (4) sotto i vincoli (5) e (6) viene condotto per ciascuna UP nel campione, generando  $l$  vettori di pesi  $U$  e  $V$ . Le misure di efficienza prodotte dal programma sono consistenti con l'interpretazione della frontiera di produzione. In particolare, valori pari a 1 implicano che la UP opera in modo pienamente efficiente, mentre valori inferiori suggeriscono che la UP opera in modo relativamente meno soddisfacente di quanto faccia una opportuna combinazione di altre UP. Charnes, Cooper e Rhodes (1978) mostrano che questo problema di programmazione frazionale e non convesso può essere sostituito da due procedure di programmazione lineare alternative: una orientata agli output e una orientata agli input. Nella prima (seconda) si impone che il denominatore (numeratore) della (4) sia pari a 1 e si massimizza il valore della produzione (il valore delle risorse) sotto opportuni vincoli. Tipicamente la soluzione di questi problemi viene ottenuta ricorrendo alla loro formulazione duale<sup>12</sup>. La definizione dei vincoli determina il tipo di rendimenti di scala che si impone all'insieme di produzione. Si possono stimare involucri a rendimenti di scala costanti (*CRS*), non crescenti (*NIRS*) e variabili (*VRS*).

### 2.3 Efficienza tecnica e efficienza di scala

La misura di Farrell di massima contrazione radiale rispetto ad una frontiera a rendimenti di scala costanti (*CRS*) riflette tanto la componente puramente tecnica dell'efficienza che quella di scala. Questo aspetto può essere colto agevolmente se si considera l'esempio della funzione di produzione monoprodotta a un solo input. Si consideri al riguardo la **Figura 2**. La tecnologia *CRS* vi è rappresentata dal raggio continuo mentre la tecnologia *VRS* dalla linea spezzata continua. Esaminiamo la UP A. Essa si colloca sulla frontiera di produzione a rendimenti variabili mentre giace all'interno di quella a rendimenti costanti. Ciò si verifica perché a quel livello di attività sono all'opera rendimenti di scala crescenti. In questo caso la lontananza dalla frontiera *CRS* (A-A') riflette la misura dell'input-inefficienza dovuta alla scarsa capacità di sfruttare i rendimenti di scala, crescenti in quel tratto della funzione di produzione, piuttosto che l'incapacità di utilizzare al meglio gli input a disposizione. Si dice quindi che la UP A è input-efficiente dal punto di vista tecnico ma inefficiente rispetto alla scala di produzione attivata. Seguendo Färe, Grosskopf e Lovell (1985), l'indicatore di input-efficienza costruito rispetto ad una frontiera *CRS*,  $DF_I(CRS)$ , può quindi essere scomposto nella componente di efficienza tecnica, data dall'indicatore di input-efficienza rispetto alla frontiera a rendimenti di scala variabili,  $DF_I(VRS)$ , e in una componente di efficienza di scala,  $ES$ , secondo la seguente regola:  $DF_I(CRS) = DF_I(VRS) * ES$ . Se  $ES < 1$  allora il mix produttivo non viene attivato alla scala efficiente. La **Figura 2** illustra la logica di tale scomposizione rispetto alla UP A". Come si vede la misura di input-efficienza alla Farrell rispetto alla frontiera *CRS*, (A"-A'), si scompone nella inefficienza di scala, colta dalla distanza (A'-A) e dall'inefficienza tecnica, colta dalla distanza (A"-A).

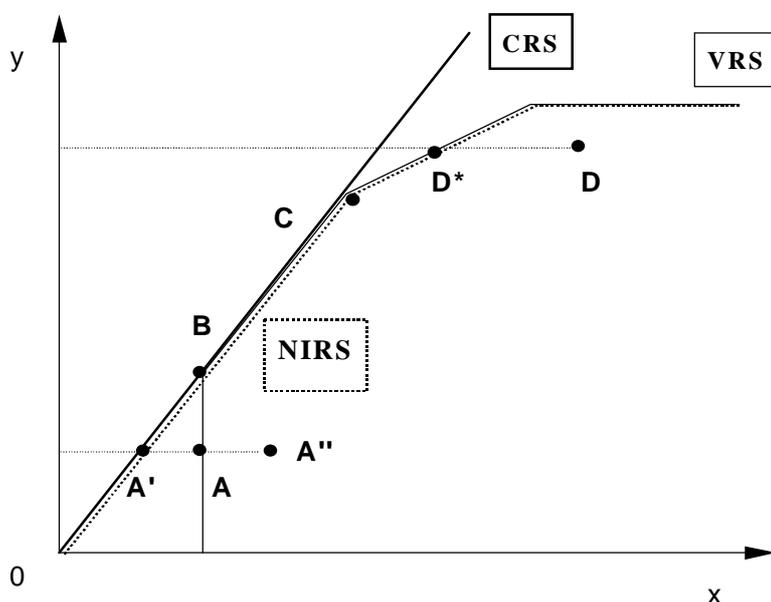
Ciò detto, per stabilire se l'inefficienza di scala è da attribuire a un sottodimensionamento o a un sovradimensionamento della UP occorre procedere, come suggeriscono Färe, Grosskopf e Lovell (1985), a confrontare gli indicatori di Farrell calcolati rispetto alle frontiere a rendimenti di scala costanti e a rendimenti di scala non crescenti (*NIRS*). Quest'ultima è rappresentata dalla linea spezzata tratteggiata. Se

---

<sup>12</sup> Si veda Ali e Seiford (1993).

questi sono indistinguibili allora si è in presenza di rendimenti di scala crescenti, ovvero la UP considerata è sottodimensionata. Viceversa, se la distanza dalla frontiera *CRS* è maggiore di quella dalla frontiera *NIRS* siamo in presenza di rendimenti di scala decrescenti, ovvero di un sovradimensionamento della UP. Nel caso della UP *A''* vediamo che l'indicatore *ES* ci segnala la presenza di inefficienza di scala. Il confronto fra distanza dalla frontiera *CRS* e frontiera *NIRS* ci conferma la presenza di rendimenti di scala crescenti. Viceversa per la UP *D* si può vedere che essa opera ad una scala non efficiente e che ciò è da attribuirsi al sovradimensionamento. Queste considerazioni portano a definire un indicatore di rendimenti di scala, *RS*, definito come segue:  $RS = DF_I(CRS) / DF_I(NIRS)$ . Siamo in presenza di rendimenti di scala crescenti (decrescenti) laddove  $RS > 1$  ( $< 1$ ).

Figura 2: La scomposizione della misura di input-efficienza alla Debreu-Farrell



Per cogliere l'esistenza di inefficienza di scala occorre quindi costruire una frontiera a rendimenti di scala costanti (*CRS*), calcolare rispetto a questa gli indici di Farrell e confrontarli con quelli calcolati rispetto ad una frontiera a rendimenti variabili (*VRS*); per valutare la direzione dell'inefficienza di scala occorre costruire una terza frontiera, a rendimenti di scala non crescenti (*NIRS*), e confrontare gli indici di Farrell calcolati rispetto ad essa con quelli *CRS*.

### 3 L'ospedale come unità produttiva

Possiamo definire l'ospedale come l'istituzione in cui individui malati o feriti ricevono trattamenti medici o chirurgici. In termini economici gli output tipici di tale istituzione sono almeno quattro: il trattamento di pazienti ricoverati, quello di pazienti visitati, l'insegnamento e la ricerca. Laddove gli ultimi tre output vengono prodotti anche da altre istituzioni, è il primo a contraddistinguere in modo peculiare l'ospedale [Butler (1995)]. Ciò premesso un'importante distinzione concettuale riguarda la corretta interpretazione di tale output.

Due sono le posizioni in letteratura. Da un lato, abbiamo la tesi secondo cui l'output proprio dell'ospedale sarebbe la fornitura del trattamento in sé. Dall'altro, si attribuirebbe la valenza di output solo al miglioramento delle condizioni di salute del paziente conseguente al trattamento ricevuto. Quest'ultima posizione è coerente con il punto di vista dei pazienti, i quali richiederebbero il trattamento ospedaliero in vista del conseguente miglioramento nelle condizioni di salute. Tuttavia, anche se concettualmente corretta, tale definizione di output sarebbe, secondo molti fautori della interpretazione alternativa [si veda ad esempio Zweifel e Breyer (1997)], empiricamente impraticabile. Se l'output dell'ospedale è dato dalla variazione nello stato di salute del paziente, allora sarebbe necessario isolare tutti i fattori di rischio per la salute del singolo paziente al fine di pervenire ad una corretta quantificazione del contributo marginale fornito dal trattamento ospedaliero. Vista l'impraticabilità di tale analisi, una soluzione soddisfacente e percorribile è quella di adottare una definizione convenzionale di output ospedaliero fondata su indicatori quali [si veda Breyer (1987)] il numero di servizi medici e assistenziali prestati (esami, operazioni chirurgiche, medicazioni, cicli di terapia, iniezioni, ...); il numero di giornate di degenza (tipicamente differenziate per il livello di intensità del trattamento); il numero di pazienti o di casi trattati (tipicamente differenziato per tipologia di trattamento).

La correttezza di questa impostazione può essere giustificata considerando che, in molti casi, il trattamento ospedaliero non sortisce alcun miglioramento nelle condizioni di salute del paziente. La relazione fra trattamento ed esito medico è, di fatto, per il paziente, aleatoria ed esposta, quindi, alla categoria del rischio. Il fatto stesso che nella generalità dei sistemi sanitari il pagamento dei trattamenti ospedalieri non sia reso contingente all'esito medico raggiunto fa ritenere che non sia questa la categoria di output economicamente rilevante quanto piuttosto il trattamento erogato. Volendo ulteriormente generalizzare, si potrebbe sostenere che, anche qualora venisse previsto un pagamento del trattamento contingente all'esito medico, la categoria corretta di output sarebbe quella del trattamento. La considerazione dell'effetto del trattamento entrerebbe come output aggiunto offerto dall'ospedale al paziente sotto forma di servizio assicurativo<sup>13</sup>.

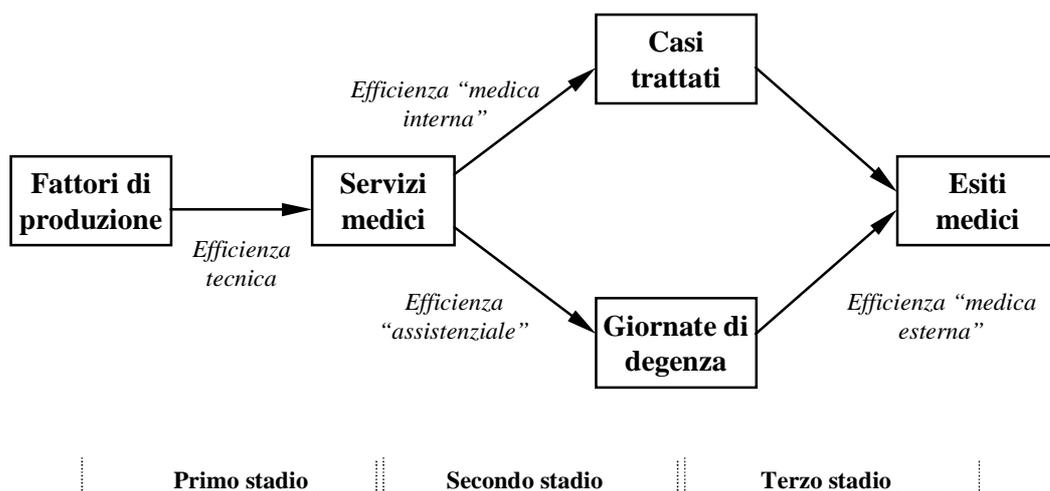
Una volta adottata l'interpretazione dell'output come trattamento occorre stabilire quale sia la misura dei trattamenti erogati più idonea a rappresentare la produzione ospedaliera. A questo riguardo, due sono le alternative: i casi trattati, inteso come numero di episodi di ricovero, e le giornate di degenza. In letteratura c'è un sostanziale consenso circa il fatto che siano gli episodi di ricovero a rappresentare, in modo più appropriato, l'output dell'attività ospedaliera. Al contrario, il numero di giornate di degenza rifletterebbe le scelte di sostituzione fra intensità del trattamento e spese alberghiere che ogni presidio può adottare. Da questo punto di vista, considerare le giornate di degenza come un output sarebbe equivalente ad assimilare un input ad un output finale. Questa valutazione è pienamente condivisibile qualora le prestazioni ospedaliere si limitino ai trattamenti per acuti. Tuttavia qualora l'ospedale operi anche come luogo di ricovero per post-acuti, pazienti stabilizzati e lungodegenti, l'aspetto

---

<sup>13</sup> Qualcosa di analogo accade nel mercato dei servizi di rappresentanza legale. Il fatto che il legale in alcuni casi venga remunerato in modo contingente all'esito della causa non incide sul fatto che oggetto del contratto è la tutela legale e non l'esito processuale di tale tutela. Rendendo l'onorario contingente all'esito processuale il legale si fa carico di parte del rischio offrendo quindi all'assistito un servizio aggiuntivo di tipo assicurativo.

della prestazione assistenziale pare meno trascurabile. In queste circostanze anche le giornate di degenza contribuiscono a descrivere l'output prodotto. Possiamo dire quindi che il numero di casi trattati riflette la componente medico-terapeutica della produzione ospedaliera, mentre il numero di giornate di degenza quella più propriamente assistenziale.

*Figura 3: Il processo produttivo ospedaliero*



Fonte: nostro adattamento da Zweifel e Breyer (1997)

La **Figura 3** fornisce una descrizione del processo produttivo ospedaliero che coinvolge le grandezze descritte. I fattori produttivi vengono utilizzati per produrre, nel primo stadio del processo produttivo, i servizi medici (esami, operazioni chirurgiche, medicazioni, cicli di terapia, iniezioni...). Questi, a loro volta concorrono, nel secondo stadio, alla produzione di episodi di ricovero e giornate di degenza. Entrambi questi prodotti intermedi si collocano immediatamente prima della realizzazione dell'esito atteso della produzione ospedaliera, ovvero il miglioramento delle condizioni di salute del paziente, gli esiti medici. In questo ambito, una definizione propria di input-efficienza tecnica nella produzione ospedaliera dovrebbe riferirsi alla capacità di utilizzare la minor quantità di input per produrre i servizi medici, ovvero riferirsi al primo stadio della nostra rappresentazione. Qualora si riesca a ricondurre agli episodi di ricovero e alle giornate di degenza, in qualità di input diretti, differenti servizi medici prodotti nel primo stadio del processo produttivo diventa possibile due ulteriori concetti di input-efficienza. L'efficienza "assistenziale" fa riferimento alla capacità di utilizzare la minor quantità di servizi assistenziali per giornata di degenza. L'efficienza "medica interna" invece si riferisce alla capacità di utilizzare la minor quantità di servizi medici per caso trattato<sup>14</sup>. Infine si può prefigurare una nozione di efficienza "medica esterna"

<sup>14</sup> Tutti gli stadi considerati coinvolgono produzioni le cui entità vengono in parte determinate dalla pressione e dalle caratteristiche della domanda. Questo aspetto non è privo di rilievo nella produzione di servizi, ancorché pubblici, a domanda individuale poiché porta a confondere la capacità di trasformare le risorse produttive in output con la capacità di attrarre clienti. In sostanza si può giudicare inefficienti produttori che seppur efficienti operano in bacini di domanda debole. Nessuna delle applicazioni DEA in letteratura affronta questo problema. Del resto, se si sposa la logica di analisi della produzione, si è usualmente disposti a tralasciare le interazioni con la domanda e a considerare quindi la

che, in una accezione orientata all'output, possiamo caratterizzare come la capacità di produrre gli esiti medici più favorevoli a parità di casi trattati e di giornate di degenza.

Un problema ulteriore nell'analisi della produzione ospedaliera è rappresentato dalla estrema eterogeneità degli output, intermedi e finali, prodotti. Il numero complessivo di casi trattati, ad esempio, offre una misura molto grossolana della produzione ospedaliera se non si qualifica il tipo e la gravità delle patologie trattate, lo stadio di avanzamento delle patologie, la presenza di patologie concomitanti, le caratteristiche complessive del paziente. Alla luce di queste considerazioni ogni paziente trattato andrebbe considerato come un output distinto, tuttavia al costo di rendere impossibile qualunque tipo di valutazione comparativa fra la produzione di ospedali diversi. Il compromesso usualmente adottato è quello di condurre le analisi classificando i pazienti trattati secondo un numero ridotto di categorie che presentino il maggior grado di omogeneità.

Nell'analisi che segue abbiamo totalmente ignorato l'efficienza medica esterna. Una sua analisi compiuta richiede strumenti di analisi differenti da quelli qui impiegati [si veda Cutler (1995)]. Per ciò che concerne i tre restanti concetti di efficienza delineati, la disponibilità dei dati ci ha indotto ad elaborarne una sintesi. La mancata rilevazione dei servizi medici e assistenziali prodotti non ci permette di cogliere quella che abbiamo definito come efficienza tecnica in senso proprio. In sostanza nell'analisi che segue abbiamo incorporato il primo stadio della produzione nel secondo immaginando che i fattori di produzione (medici, personale infermieristico, letti, ...) vengano utilizzati per "produrre" ricoveri e giornate di degenza. Pertanto l'efficienza che misureremo nel seguito incorpora tanto la capacità di utilizzare gli input per produrre in modo efficiente gli output medici e assistenziali, che l'attitudine ad impiegare questi ultimi in modo parsimonioso rispetto agli episodi di ricovero e di degenza.

#### **4 I dati e il modello stimato**

I dati<sup>15</sup> su cui si fonda la nostra analisi sono relativi all'attività degli ospedali pubblici dell'Emilia-Romagna negli anni che vanno dal 1994 al 1997. Si tratta di un periodo che ricomprende l'introduzione dei più importanti elementi di riforma del finanziamento degli ospedali, fra il 1995 e il 1996. I dati disponibili permettono di descrivere in modo esaustivo la struttura produttiva degli ospedali mentre non

---

decisione circa il livello di output da produrre come esogeno rispetto alla evoluzione della domanda. In un settore come quello sanitario, in cui la domanda è fortemente regolamentata, questa assunzione non pare fuori luogo. Solo in alcuni studi recenti, in cui si stima una funzione parametrica di costo, si affronta il problema secondo una impostazione empiricamente praticabile. Si veda Keeler e Ying (1996).

L'effetto della pressione della domanda sulla produttività dei fattori tuttavia può essere a buon titolo interpretato come un incentivo introdotta dalla riforma. l'analisi degli effetti della congestione e quindi la stima in presenza di disponibilità debole dei fattori non riflette quindi lo spirito dell'osservazione

<sup>15</sup> I dati utilizzati provengono da due banche dati dell'Assessorato Regionale alla Sanità dell'Emilia-Romagna. La prima è relativa alle schede nosologiche individuali dei ricoveri ospedalieri; la seconda ai flussi rilevati per conto del Ministero della Sanità. L'accesso a questi dati è stato reso possibile dall'Agenzia Sanitaria Regionale dell'Emilia-Romagna nell'ambito del progetto "Matrici di mobilità per motivi sanitari", svolto in collaborazione con il Dipartimento di Scienze Economiche dell'Università di Bologna.

permettono di valutarne la struttura dei costi. L'introduzione di un attendibile sistema di contabilità analitica dei costi per singolo presidio è infatti successiva al periodo qui considerato. Questo aspetto rende pressochè obbligata l'adozione di una tecnica non-parametrica di analisi dell'efficienza. Le tecniche non-parametriche infatti permettono di stimare frontiere di produzione multiprodotto molto più facilmente di quanto non consentano le tecniche parametriche. Per "ospitare" una specificazione multiprodotto in ambito parametrico occorrerebbe stimare una funzione di distanza. Questa specificazione, come del resto la stessa funzione di produzione mono-prodotto, soffre, dal punto di vista econometrico, di un problema di endogenità fra variabile dipendente (l'output) e le variabili esplicative (gli input). L'assunzione di esogenità di questi ultimi è infatti economicamente difficile da sostenere essendo input e output codeterminati nell'ambito del piano di produzione prescelto dalla UP. L'analisi non-parametrica non supera queste obiezioni di natura economica. Tuttavia la tecnica di stima non-parametrica non richiede l'esogenità degli input come ipotesi di identificazione del modello. Vista l'importanza della specificazione multiprodotto nella produzione ospedaliera, queste considerazioni ci sembrano indicare in modo piuttosto preciso che la stima dell'involuppo debba essere condotta con tecniche di programmazione matematica piuttosto che econometriche.

Lavorare su questo campione connota la nostra analisi in modo molto forte. E' evidente infatti che i risultati cui perveniamo non possono intendersi come rappresentativi dell'intera realtà nazionale: in sistemi sanitari regionali diversi da quello qui esaminato gli effetti della riforma potrebbero essere stati ben diversi. A sostegno della nostra analisi va riconosciuto peraltro la rilevanza della regione qui esaminata nel panorama nazionale, sia per la qualità della sua struttura sanitaria, che per la relativa rappresentatività del suo modello sanitario. In secondo luogo l'orizzonte temporale considerato è tale da rendere non del tutto trascurabile il ruolo del progresso tecnologico. Questo elemento è molto importante soprattutto in merito alla interpretazione dei risultati, visto che agli effetti della riforma si sommano anche gli effetti dovuti alla evoluzione delle tecnologie. La nostra analisi prescinde dalla presenza di progresso tecnologico assumendo che l'insieme di produzione a disposizione dei produttori nel nostro campione non si modifichi nel tempo. Questo ci permette di individuare una sorta di limite superiore dell'effetto della riforma, cui andrebbe sottratto l'effetto attribuibile al progresso. Questo modo di procedere non è certo privo di limiti. Tuttavia rappresenta la scelta obbligata se l'intento è quello di mettere in luce non solo gli effetti della riforma del meccanismo di finanziamento sull'efficienza tecnica ma anche i caratteri dell'insieme di produzione legati alla scala di attivazione. Infatti sarebbe possibile procedere in modo rigoroso alla analisi della produttività attraverso l'applicazione dell'indice di Malmquist. Tuttavia questa direzione di analisi sacrifica la possibilità di conseguire risultati in merito alla struttura dei rendimenti di scala<sup>16</sup>.

Il nostro campione è costituito da 45 unità produttive rilevate su quattro anni, dal 1994 al 1997. Abbiamo qui considerato i soli presidi pubblici con esclusione dei presidi monospécialistici. Per presidio intendiamo l'agglomerato di stabilimenti sancito dalla

---

<sup>16</sup> L'analisi con indici di Malmquist permette di decomporre le variazioni della produttività di ciascuna delle UP in una componente attribuibile al progresso tecnico, ovvero allo spostamento della frontiera produttiva, ed in una componente ascrivibile al miglioramento dell'efficienza tecnica intesa come avvicinamento alla frontiera. Per far questo occorre assumere che la tecnologia sia a rendimenti di scala costanti.

delibera regionale di recepimento della norma contenuta nella legge finanziaria del 1995, in cui si stabiliva la necessità di procedere alla chiusura o alla razionalizzazione degli stabilimenti con meno di 120 posti letto. Questo processo di razionalizzazione ha raggiunto, in Emilia-Romagna, un assetto stabile nel 1996. La mappa delle aggregazioni degli stabilimenti nei 45 presidi contenuti nel nostro campione è appunto quella emersa in quell'anno. Le osservazioni relative ai due anni precedenti sono state ricostruite accorpando gli stabilimenti, allora autonomi, secondo la medesima mappa di aggregazioni. In questo modo restano esclusi dal campione le osservazioni relative a presidi chiusi nell'ambito dell'arco di tempo considerato<sup>17</sup>. La mancanza del dato relativo al personale per i presidi di Riccione, nell'anno 1995, e di San Secondo Parmense, nell'anno 1997, ci ha indotto ad escludere questi due presidi dall'intero campione. Pertanto il campione è un panel bilanciato composto di 43 unità osservate su 4 anni.

La costruzione delle osservazioni di presidio procede dai dati di reparto, opportunamente costruiti e aggregati. A livello di reparto sono state eliminate le osservazioni che presentavano valori mancanti sui letti, pur in presenza di una attività di ricovero, e valori mancanti sull'attività di ricovero, pur in presenza di letti. Per ognuno degli ospedali del campione sono stati ricostruiti, a partire da fonti ufficiali, i dati di attività descritti nella **Tabella 1**. Per quanto concerne l'attività di ricovero abbiamo qui considerato la sola attività in regime ordinario. Questo elemento rende in effetti specifici i risultati della nostra analisi, soprattutto vista la crescente rilevanza dei ricoveri in day-hospital. Ci ripromettiamo di estendere l'analisi anche a questa dimensione dell'attività in un successivo momento.

*Tabella 1: Descrizione del campione.*

	Minimo	Massimo	Media	Dev std
<b>INPUT</b>				
Letti	14	2356	401	415
Medici	2	1117	174	208
Infermieri	20	2468	486	513
Altro personale	9	1336	176	229
<b>OUTPUT</b>				
Dimessi*	0.5	75.4	14.6	15.1
Punti DRG*	0.4	77.7	12.8	14.3
<b>INDICATORI</b>				
Degenza media	4.5	23.9	8.4	2.3
Tasso di occupazione	60.8	95.0	77.8	6.4
Peso medio DRG	0.654	1.300	0.831	0.101
Concentraz. Casistica	0.238	0.700	0.395	0.090

\*Dati espressi in migliaia. Gli indicatori sono ottenuti nel campione come medie ponderate sui dimessi di ciascun reparto appartenente al presidio.

La **Tabella 1** presenta i dati del nostro campione e descrive la specificazione dell'insieme di produzione. Il campione si presenta piuttosto disomogeneo, tanto sotto il profilo delle dimensioni produttive che per la tipologia di attività. Assumere, come è necessario fare nell'ambito delle stime parametriche, l'esistenza di un'unica relazione di produzione fra input e output, parrebbe, in queste condizioni, piuttosto forzato. La DEA,

<sup>17</sup> Si tratta di 8 stabilimenti: Medicina, Crevalcore, Cervia, Russi, Bagnacavallo, Brisighella, Catelbolognese e Modigliana.

conducendo un'analisi locale dell'insieme di produzione, permette di superare questa difficoltà.

Il modello che stiamo stimando consiste di quattro input e due output. **Letti** è calcolato come il numero di letti ordinari e a pagamento mediamente presenti nell'arco dell'anno nella struttura. **Medici**, l'ammontare complessivo di personale medico, comprende i primari, gli aiuti e gli assistenti medici, gli odontoiatri e l'altro personale laureato. **Infermieri** contiene il personale infermieristico-assistenziale definito come il complesso del personale infermieristico in senso stretto, il personale tecnico-sanitario, il personale con funzioni di riabilitazione, gli operatori tecnici di assistenza e gli ausiliari specializzati socio-assistenziali. Infine l'**Altro personale** comprende il personale con funzioni dirigenziali e gestionali, ovvero direttori e vicedirettori sanitari, statistici e ausiliari specializzati tecnico economici. Per quel che riguarda gli input la disaggregazione prescelta ci sembra descriva in modo piuttosto soddisfacente il contributo dei fattori produttivi, in particolare di quelli rispetto ai quali è più rilevante il dibattito di politica economica. La mancanza di indicatori relativi alla disponibilità di attrezzature mediche è nel contesto della nostra applicazione di secondaria importanza, visto il livello molto aggregato dell'analisi.

*Tabella 2: Descrizione del campione: 1994-1997.*

ANNO	1994		1995		1996		1997	
	Media	std	media	std	media	std	media	std
INPUT								
Letti	433	453	422	435	405	413	344	363
Medici	190	225	169	205	169	206	167	204
Infermieri	488	512	487	517	489	526	479	516
Altro personale	180	215	184	238	175	241	166	230
OUTPUT								
Dimessi*	14.6	15.3	15.0	15.4	15.0	15.4	13.9	14.9
Punti DRG*	12.1	13.1	13.1	14.3	13.5	15.1	12.6	14.9
INDICATORI								
Degenza media	9.1	2.7	8.4	1.5	8.1	1.4	8.2	3.1
Tasso di occupazione	78.2	6.5	76.6	6.7	77.3	6.7	78.9	5.4
Peso medio DRG	0.796	0.088	0.822	0.097	0.847	0.107	0.860	0.104
Concentraz. Casistica	0.417	0.101	0.385	0.086	0.382	0.083	0.398	0.088

\*Dati espressi in migliaia. Gli indicatori sono ottenuti nel campione come medie ponderate sui dimessi di ciascun reparto appartenente al presidio.

Per quanto concerne gli output abbiamo adottato una specificazione molto semplificata che coglie tuttavia le tre principali dimensioni dell'attività in regime ordinario. **Dimessi** contiene il numero complessivo degli episodi di ricovero, come rilevato dalla apertura e chiusura di scheda nosologica, nell'arco dell'anno considerato. Alla luce della nostra precedente discussione si tratta dell'output in senso più proprio del processo di produzione ospedaliera. Poiché il dato viene qui costruito a partire dai reparti, si considerano in questo aggregato i dimessi dal reparto. **Punti DRG** è dato dall'ammontare complessivo di punti DRG prodotti dal singolo presidio. Questa produzione offre una indicazione del consumo complessivo di risorse per la cura dei pazienti ricoverati. Ospedali con case-mix più complesso tenderanno ad avere una maggiore produzione di punti DRG. Inserire questa produzione permette quindi di

neutralizzare, almeno in parte, l'effetto dovuto alla diversa composizione della casistica<sup>18</sup>.

Dalla lettura dei dati in **Tabella 2** vediamo che nel periodo esaminato la dimensione media dei presidi si è progressivamente ridotta e resa più omogenea. Media e deviazione standard di tutti gli input considerati si riducono infatti nell'intero periodo esaminato. La riduzione dei posti letti è intervenuta in particolare nell'ultimo anno considerato. Per quel che riguarda i medici la riduzione più sensibile si osserva nel 1995, mentre per il personale infermieristico e per l'altro personale si registra una sostanziale stabilità nei primi tre anni e una riduzione nel 1997. La produzione aumenta, in termini di dimessi e di punti DRG, fino al 1996 per poi ridursi bruscamente nel 1997. Nel primo biennio, la riforma parrebbe quindi aver incentivato i produttori ad aumentare le produzioni. Solo nel 1997 sembrerebbero diventare attivi i vincoli di razionamento sul consumo delle prestazioni imposti, al fine di rispettare i bilanci, dagli enti finanziatori. Per quel che riguarda la produzione di giornate di degenza, essa si riduce sistematicamente sull'intero periodo tanto che la degenza media si riduce di 1 giorno nell'arco dei quattro anni. Pur in presenza di una riduzione dei posti letto e di una concomitante crescita della produzione nei primi anni non si riscontra una significativa variazione del tasso di occupazione dei letti. In questo senso non sembra che la disponibilità dai posti letto influisca in modo significativo sui comportamenti di produzione dei presidi. Per quanto attiene alla complessità della casistica vediamo che essa è aumentata. Questo è un effetto noto in letteratura e da ascrivere in larga parte al miglioramento della pratica di codifica dei pazienti piuttosto che ad un effettivo incremento della complessità dei pazienti [Carter, Newhouse e Relles (1990)].

Volendo collocare queste variazioni rispetto al trend è interessante notare che fatti 100 ricoveri in regime ordinario, posti letto e durata media della degenza nel 1978, nel 1994 abbiamo rispettivamente 93, 60 e 69. In sostanza nel periodo considerato l'andamento della degenza media e dei posti letto è in linea con il trend, mentre quello dei ricoveri si discosta sensibilmente nella direzione indotta dagli incentivi della riforma nel primo biennio di applicazione per ritornare a calare nel 1997. Da questo punto di vista possiamo caratterizzare il 1996 come anno di massima espansione della produzione e il 1997 come anno in cui diventano effettivi gli aggiustamenti introdotti per controllare la crescita dei volumi e della spesa, non ultimo il sostanziale dimagrimento delle capacità produttive sotto tutti i profili esaminati.

Sulla scorta di questi dati abbiamo quindi stimato il modello assumendo che i rendimenti di scala fossero variabili e misurando sull'inviluppo, per ciascun presidio, la massima contrazione radiale del vettore degli input.

---

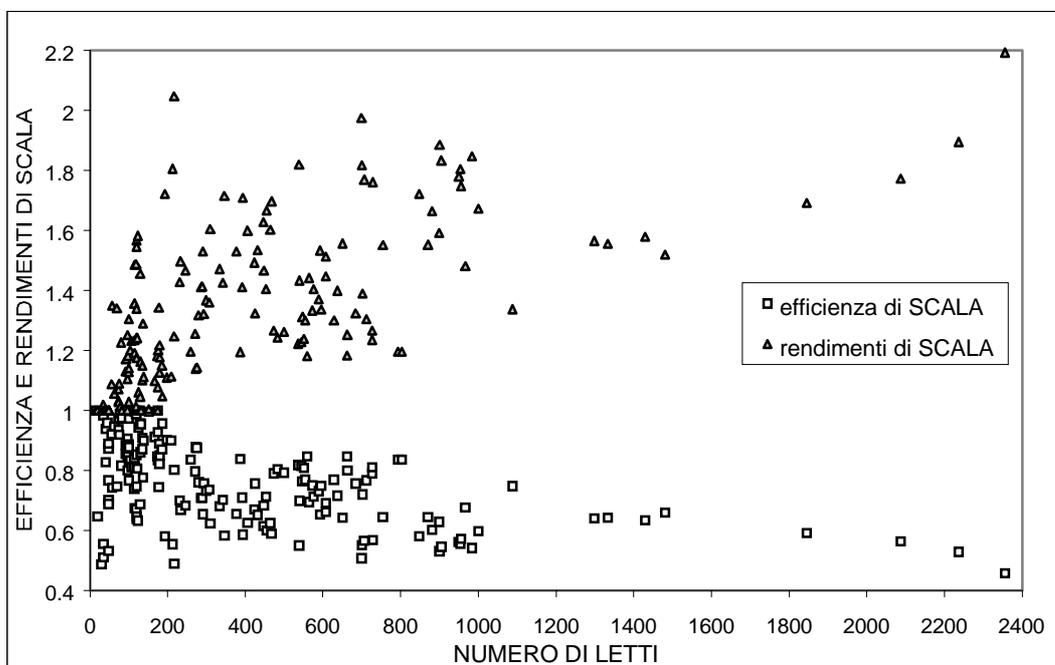
<sup>18</sup> In una specificazione preliminare del modello avevamo introdotto anche la **Degenza**, intesa come ammontare complessivo di giornate di degenza prodotte dal presidio. Si tratta di una misura dalla duplice natura. Da un lato essa riflette la capacità di produrre gli output assistenziali. Dall'altro, come dicevamo sopra, riflette scelte relative al processo di trattamento. Nel periodo della riforma la degenza ha progressivamente perso la prima connotazione per assumere la seconda. Per questa ragione l'abbiamo qui esclusa.

## 5 Risultati

### 5.1 Efficienza tecnica e di scala nel periodo 1994-1997

Iniziamo la presentazione dei risultati descrivendo le principali caratteristiche della funzione di produzione esaminata. Per approfondire l'argomento abbiamo calcolato gli indicatori di rendimenti di scala,  $RS$ , e li abbiamo collocati su un grafico insieme all'indicatore di efficienza di scala,  $ES$ . Il **Grafico 1** ne illustra l'andamento, per l'intero campione di quattro anni, rispetto alla dimensione dell'ospedale, misurata dal numero dei letti. Dal grafico si evince che i presidi con più di 200 letti presentano un sistematico scostamento dalla dimensione efficiente. Viceversa, solo alcuni dei presidi con meno di 200 letti operano alla scala ottimale. La semplice analisi visiva rivela la presenza di una correlazione negativa fra inefficienza di scala e dimensione, confermata dal coefficiente di correlazione di Pearson negativo (-0.475) e altamente significativo. In sostanza al crescere della dimensione del presidio tanto maggiore è il grado di allontanamento dalla dimensione efficiente. Viceversa è altrettanto visibile la correlazione positiva fra dimensione e rendimenti di scala (0.634). Ricordiamo che, per le UP che si discostano dalla dimensione efficiente, ovvero per le quali  $ES < 1$ , valori maggiori di 1 dell'indicatore di rendimenti di scala,  $RS$ , segnalano la presenza di rendimenti decrescenti, mentre si ha evidenza di rendimenti di scala crescenti laddove l'indicatore assume valore unitario. Quindi solo i presidi con meno di 200 letti si collocano sul tratto crescente della curva di prodotto marginale, mentre i presidi con più di 200 letti si collocano nel tratto decrescente. L'inefficienza di scala dei presidi maggiori è chiaramente da ascrivere al loro sovradimensionamento.

Grafico 1: Efficienza di scala e rendimenti di scala.



La **Tabella 3** presenta i risultati della stima del modello per quanto attiene all'evoluzione dell'efficienza, tecnica e di scala, nel periodo considerato. I dati presentati sono valori di sintesi della distribuzione degli indicatori. Per quanto riguarda i primi, si

tratta, lo rammentiamo, della massima contrazione radiale del vettore di input impiegati compatibile con la produzione “efficiente” del vettore di output effettivamente prodotto. Dai risultati si evince che nell’arco dei quattro anni l’efficienza tecnica della produzione è aumentata. Infatti, se nel 1994 la produzione ospedaliera in regime ordinario, negli ospedali pubblici dell’Emilia-Romagna poteva essere ottenuta in modo tecnicamente efficiente impiegando, in media, il 77% degli input effettivamente impiegati, ovvero con una riduzione di circa il 23%, nel 1997 il risparmio realizzabile nell’utilizzo degli input si riduce al 15%. Analogamente, la percentuale di ospedali efficienti nel senso di Debreu-Farrell passa dal 7% del campione, nel 1994, al 30% nel 1997. Quindi, l’effetto della riforma sull’incremento dell’efficienza tecnica può collocarsi, fatto salvo l’effetto del progresso tecnico, su un intervallo di valori che raggiunge al massimo il 15%, in termini di risparmio realizzato nell’uso dei fattori produttivi.

Per quanto riguarda l’indicatore di efficienza di scala, vediamo che esso resta ben al di sotto del valore unitario per tutto il periodo considerato. Solo il 2% dei presidi presenti nel campione può considerarsi, rispetto alla nostra analisi, operante ad una dimensione di scala efficiente. Inoltre vediamo che l’efficienza di scala sembra aumentare lievemente. Tuttavia la percentuale dei presidi correttamente dimensionati non aumenta. Sia per quel che riguarda l’efficienza tecnica che quella di scala si riscontra un incremento netto fra il 1994 e il 1996 e una successiva riduzione nel 1997, anche se quest’ultima non sovverte l’incremento complessivo di efficienza nel raffronto 1997-1994.

*Tabella 3: Efficienza di Farrell e efficienza di scala nel periodo 1994-1997. Stima di una frontiera unica.*

ANNO	1994	1995	1996	1997	1994	1995	1996	1997
	FARRELL				SCALA			
Media	0.774	0.827	0.878	0.868	0.740	0.743	0.767	0.754
Deviaz. Standard	0.150	0.134	0.121	0.153	0.149	0.141	0.143	0.142
Minimo	0.426	0.529	0.563	0.417	0.388	0.510	0.487	0.358
Massimo	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000
% UP efficienti	7%	12%	23%	30%	0%	2%	2%	2%
Pr[rigetto $H_0(1994)$ ]*	-	0.031	0.000	0.001	-	0.534	0.225	0.310
Pr[rigetto $H_0(1995)$ ]*	-	-	0.020	0.046	-	-	0.189	0.319
Pr[rigetto $H_0(1996)$ ]*	-	-	-	0.454	-	-	-	0.651

\*Area di rigetto dell’ipotesi nulla di uguaglianza dell’efficienza delle UP rispetto alle UP dell’anno di riferimento. Il test utilizzato è quello di Mann e Whitney.

Questa evoluzione è coerente con l’espressione delle potenzialità produttive dei presidi, incoraggiata nella prima fase di attuazione della riforma, e successivamente moderata dalla successiva imposizione di un freno all’incremento dei volumi complessivi di produzione imposta nell’ultimo anno esaminato. L’imposizione pare trascendere la limitazione imposta sui fattori produttivi. Infatti la riduzione degli input non comporta di per sé un peggioramento dell’efficienza produttiva. Questa, in presenza di una concomitante riduzione dei volumi di prestazioni, dovrebbe, in condizioni generali, rimanere invariata. Al contrario, riscontriamo qui un lieve peggioramento della performance dei presidi nell’ultimo anno, dovuta ad una riduzione degli output più che proporzionale rispetto a quella degli input. Questo andamento offre un’evidenza indiretta del fatto che nel 1997 sia stato imposto un limite sui volumi di produzione attraverso canali finanziari. Nel 1997 si riscontra la percentuale più elevata di ospedali

pienamente efficienti e nel contempo l'indicatore di Farrell mostra una elevata variabilità testimoniata dal valore della deviazione standard. Parrebbe quindi esservi nel nostro campione una forte diversificazione nei modelli comportamentali in risposta ai vincoli di razionamento.

Per validare la rilevanza statistica di questi risultati abbiamo usato un test non-parametrico. Le ipotesi che vengono tipicamente testate nelle analisi di efficienza coinvolgono il confronto fra due gruppi di UP al fine di valutare se uno dei due è più efficiente dell'altro, ovvero se sono stati estratti popolazioni contraddistinte da distribuzioni differenti<sup>19</sup>. Nel nostro caso si tratta di formulare un test relativo alla significatività dell'incremento di efficienza fra le UP osservate negli anni, e in particolare di valutare se la tendenza centrale della distribuzione degli indicatori di efficienza del 1995, 1996 e 1997 sia significativamente maggiore di quella relativa al 1994. Nel nostro studio facciamo ricorso a un test non parametrico molto popolare nelle applicazioni DEA, quello di Mann-Whitney<sup>20</sup>. Le ultime tre righe della **Tabella 4** presentano il livello di probabilità con cui rigettare l'ipotesi nulla di uguale efficienza fra l'osservazione relativa all'anno in colonna e l'osservazione dell'anno di riferimento a favore dell'ipotesi che le UP dell'anno considerato siano più efficienti. La distribuzione asintotica del test è descritta in Siegel (1956). Come si vede dai risultati, il test di Mann-Whitney indica che l'incremento di efficienza tecnica rispetto al 1994 è sempre statisticamente significativo. L'incremento del 1995 è significativo al 95% di probabilità, mentre quello degli anni successivi al 99%. Considerazioni analoghe si possono formulare se si prende come anno di riferimento il 1995. Viceversa fra il 1997 e il 1996 non si riscontra un incremento di efficienza significativo.

Per quel che riguarda l'efficienza di scala osserviamo che le variazioni registrate non risultano in nessun caso significative. Nel periodo considerato l'efficienza di scala è rimasta invariata.

## 5.2 *Modelli di finanziamento e efficienza degli ospedali*

La riforma si applica in modo differenziato e quindi determina effetti differenziati sui diverse elementi del sistema di offerta: ospedali in gestione diretta e Aziende Ospedaliere, *in primis*. Si è infatti soliti affermare che la riforma del meccanismo di finanziamento interessi in via prevalente queste ultime, dal momento che i primi risultano finanziati, da parte delle AUSL, in modo sostanzialmente discrezionale. Peraltro è importante rilevare che la riforma non si risolve solamente nella trasformazione del meccanismo di pagamento ma coinvolge un più generale processo di razionalizzazione di cui è importante valutare appunto l'impatto differenziale. La **Tabella 4** presenta a questo riguardo, una disaggregazione dell'analisi dell'efficienza tecnica e di scala per tipo di ospedale. Abbiamo considerato, oltre alla distinzione fra Aziende Ospedaliere e presidi in gestione, anche la disaggregazione di questi ultimi fra

---

<sup>19</sup> Banker (1993) dimostra che la distribuzione asintotica dello stimatore DEA dell'inefficienza tecnica è la medesima di quella vera. Tuttavia, come rileva l'autore, non si può utilmente applicare il Teorema del Limite Centrale per costruire test delle ipotesi sulla distribuzione degli scarti di efficienza in quanto i parametri che rappresentano la media e la varianza degli scarti di efficienza non sono noti. Pertanto occorre utilizzare test non parametrici o formulare alcune ipotesi circa la distribuzione dell'inefficienza. Per una trattazione completa di questi temi si veda Banker (1996).

<sup>20</sup> Applicazioni del test di Mann-Whitney negli studi DEA sul settore ospedaliero si ritrovano in Valdmanis (1992) e Grosskopf e Valdmanis (1987).

1°, 2° e 3° livello stabilita dalla Regione, nel 1996, al fine di graduare le tariffe applicate al punto DRG<sup>21</sup>. Nel primo livello rientrano 24 piccoli presidi, nel secondo 7 presidi di medie dimensioni e nel terzo 7 ospedali di alta specialità. Questi ultimi possiedono, alla luce delle valutazioni formulate dalla Regione Emilia-Romagna, caratteristiche strutturali non distinguibili, sotto il profilo della capacità di produrre, dalle Aziende Ospedaliere, tanto che si vedono riconosciuti la medesima tariffa per punto DRG.

*Tabella 4: Efficienza di Farrell e efficienza di Scala. Disaggregazione per livello di ospedale. Stima di una frontiera unica.*

LIVELLO	1° LIVELLO	2° LIVELLO	3° LIVELLO	AOSP
	FARRELL			
Media	0.787	0.880	0.884	0.950
Deviaz. Standard	0.158	0.092	0.101	0.070
Minimo	0.417	0.712	0.665	0.800
Massimo	1.000	1.000	1.000	1.000
% UP efficienti	14%	11%	21%	45%
	SCALA			
Media	0.808	0.709	0.710	0.595
Deviaz. Standard	0.146	0.092	0.100	0.066
Minimo	0.358	0.550	0.531	0.456
Massimo	1.000	0.878	0.847	0.748
% UP efficienti	3%	0%	0%	0%

Dai risultati contenuti nella **Tabella 4** si evince che al crescere del livello degli ospedali aumenta il grado di efficienza tecnica e diminuisce quello di efficienza di scala. Volendo razionalizzare la graduazione delle tariffe adottate in Regione alla luce della nostra analisi possiamo quindi osservare che i differenziali tariffari vanno in sostanza a coprire i differenziali di efficienza di scala e a stimolare di converso il miglioramento dell'efficienza tecnica nell'uso delle risorse. In particolare gli ospedali in gestione di 1° livello mostrano un inefficiente grado di utilizzo dei fattori: potrebbero infatti produrre lo stesso output utilizzando un quantitativo inferiore di input pari al 22% di quanto effettivamente utilizzato. D'altro canto mostrano di essere chiaramente i meglio dimensionati nel campione. La deviazione standard ci dice che questo è anche il gruppo di ospedali contraddistinto dal più elevato grado di disomogenità tanto per quel che riguarda l'efficienza tecnica che quella di scala. Gli ospedali di 2° e 3° livello paiono essere del tutto simili sotto i profili qui esaminati, tanto che la nostra analisi getta alcuni dubbi in merito alla rilevanza della classificazione adottata per distinguerli. Volendo ritenere comunque razionale questa classificazione siamo portati a ritenere che la indistinguibilità delle prestazioni sotto il profilo produttivo possa essere ascritta alla insufficiente descrizione dei contenuti qualitativi dell'output che abbiamo adottato nella nostra specificazione. Infine, per quanto concerne le Aziende Ospedaliere, vediamo che esse sono le più efficienti in senso tecnico, ma le più lontane dal dimensionamento

<sup>21</sup> Recependo le Linee Guida 1/1995, in cui si suggerisce di individuare “le tipologie di erogatori rispetto alle quali articolare il sistema tariffario sulla base di criteri che, nello specifico ambito regionale, costituiscano attualmente significativi fattori di variabilità dei costi ospedalieri quali, ad esempio, la complessità della casistica trattata, la complessità funzionale e/o la dotazione di personale”, la Regione Emilia-Romagna ha proceduto a collocare i propri ospedali in tre classi, specificando proporzioni di abbattimento tariffario per ciascuna classe.

efficiente. Inoltre si riscontra una significativa differenza fra queste e gli altri ospedali di 3° livello, che, al pari loro, si vedono riconosciuta la fascia tariffaria più elevata<sup>22</sup>.

*Tabella 5: Efficienza di Farrell e efficienza di Scala nel periodo 1994-1997. Disaggregazione per livello di ospedale. Stima di una frontiera unica.*

ANNO	EFFICIENZA DI FARRELL				EFFICIENZA DI SCALA			
	1994	1995	1996	1997	1994	1995	1996	1997
<b>1° LIVELLO</b>								
Media	0.720	0.773	0.837	0.819	0.792	0.802	0.824	0.813
Standard deviation	0.161	0.141	0.138	0.174	0.163	0.138	0.146	0.143
% effic.	4%	8%	17%	25%	0%	4%	4%	4%
Pr[rigetto H <sub>0</sub> (1994)]*	-	0.122	0.005	0.019	-	0.484	0.242	0.394
Pr[rigetto H <sub>0</sub> (1995)]*	-	-	0.054	0.128	-	-	0.245	0.359
Pr[rigetto H <sub>0</sub> (1996)]*	-	-	-	0.566	-	-	-	0.664
<b>2° LIVELLO</b>								
Media	0.797	0.881	0.922	0.920	0.683	0.712	0.727	0.712
Standard deviation	0.075	0.087	0.059	0.099	0.110	0.102	0.096	0.072
% effic.	0%	14%	14%	14%	0%	0%	0%	0%
Pr[rigetto H <sub>0</sub> (1994)]*	-	0.024	0.006	0.013	-	0.169	0.169	0.203
Pr[rigetto H <sub>0</sub> (1995)]*	-	-	0.153	0.222	-	-	0.241	0.500
Pr[rigetto H <sub>0</sub> (1996)]*	-	-	-	0.351	-	-	-	0.673
<b>3° LIVELLO</b>								
Media	0.825	0.882	0.934	0.896	0.709	0.678	0.732	0.720
Standard deviation	0.103	0.090	0.084	0.111	0.066	0.120	0.105	0.118
% effic.	0%	0%	43%	43%	0%	0%	0%	0%
Pr[rigetto H <sub>0</sub> (1994)]*	-	0.139	0.024	0.112	-	0.625	0.241	0.283
Pr[rigetto H <sub>0</sub> (1995)]*	-	-	0.071	0.283	-	-	0.203	0.169
Pr[rigetto H <sub>0</sub> (1996)]*	-	-	-	0.695	-	-	-	0.525
<b>AOSP</b>								
Media	0.936	0.940	0.938	0.988	0.616	0.592	0.594	0.577
Standard deviation	0.086	0.085	0.075	0.016	0.116	0.050	0.039	0.048
% effic.	40%	40%	40%	60%	0%	0%	0%	0%
Pr[rigetto H <sub>0</sub> (1994)]*	-	0.542	0.542	0.232	-	0.768	0.699	0.768
Pr[rigetto H <sub>0</sub> (1995)]*	-	-	0.458	0.232	-	-	0.458	0.699
Pr[rigetto H <sub>0</sub> (1996)]*	-	-	-	0.125	-	-	-	0.542

\* Area di rigetto dell'ipotesi nulla di uguaglianza dell'efficienza delle UP rispetto alle UP dell'anno di riferimento. Il test utilizzato è quello di Mann e Whitney.

La **Tabella 5** presenta l'evoluzione delle prestazioni produttive per questi quattro gruppi di presidi. Da essa si evince che nel 1996 l'efficienza tecnica dagli ospedali in gestione diretta è significativamente superiore a quella del 1994 per tutti i tipi di presidio, in particolare per quelli di 2° livello. Le Aziende Ospedaliere invece non migliorano la propria efficienza tecnica in misura statisticamente rilevante su nessuno degli anni considerati. L'andamento dell'efficienza tecnica individuato sopra (miglioramento nel 1995 e 1996 con successivo lieve peggioramento nel 1997) pare contraddistinguere in particolare i presidi in gestione diretta di 1° e 3° livello. Per i primi si vede che il risparmio possibile aumenta dal 16% al 18%. Per i secondi passa dal 7.5% al 10.5%. E' interessante notare che il primo peggioramento si accompagna ad un aumento del numero di presidi pienamente efficienti, mentre il secondo lascia invariato

<sup>22</sup> Questa evidenza ci pare riveli una incongruenza nell'impostazione dello schema tariffario. Infatti, se l'elevata dimensione di scala è ritenuto il prerequisito per offrire prestazioni di altissima specialità non pare ragionevole accordare lo stesso trattamento tariffario agli ospedali in gestione di dimensione ampiamente inferiore.

questo numero. Ciò sta a segnalare una forte diversificazione dei profili di prestazione all'interno dei due gruppi, che diventano, nell'anno esaminato, più disomogenei al loro interno, di quanto non fossero nel 1994. In chiave di politica questo ci porta a osservare che si concentrano in questi due gruppi i maggiori spazi di miglioramento, in particolare negli ospedali di 1° livello. Questi producono infatti circa il 20% dei ricoveri, presentano rendimenti di scala elevati, un elevato spazio di miglioramento dell'efficienza tecnica e un'elevata variabilità all'interno del gruppo. Per quanto attiene alla efficienza di scala, tutti i miglioramenti registrati non risultano statisticamente significativi.

## **6 *Discussione***

La recente riforma del sistema di finanziamento degli ospedali ha sensibilmente modificato gli incentivi cui questi rispondono. Innanzi tutto li ha resi responsabili, quantomeno in misura maggiore, dei risultati gestionali raggiunti, attribuendo loro una parte del rischio finanziario. In secondo luogo ha reso più stringenti i vincoli di bilancio degli enti finanziatori con immediate ripercussioni sugli incentivi al contenimento dei consumi e delle produzioni sanitarie, compresa quella ospedaliera. Infine, ha seriamente ridimensionato le risorse a disposizione degli ospedali e ridisegnata la distribuzione dimensionale dell'offerta. Sono quindi all'opera almeno due dinamiche: una che tende ad incoraggiare i produttori alla riorganizzazione e all'ottimizzazione dei processi produttivi *a risorse date*, e una che tende, nello stesso tempo, a ridimensionare le risorse a loro disposizione. Prendendo le mosse da queste premesse nel presente lavoro abbiamo offerto una valutazione della consistenza del primo effetto cercando di scorporarlo dal secondo. Per cogliere queste trasformazioni, abbiamo adottato la prospettiva di analisi economica dell'efficienza produttiva, misurando l'evoluzione dell'efficienza tecnica e dell'efficienza di scala nella produzione degli ospedali pubblici dell'Emilia-Romagna, fra il 1994 e il 1997.

Valgono, rispetto all'interpretazione dei risultati, gli usuali richiami alla cautela riconducibili ai limiti dello strumento impiegato: omissione di output rilevanti, omissione e non misurabilità di alcuni input, difficoltà di controllo rispetto alla composizione della casistica [Newhouse (1994)]. Gran parte di questi limiti sono comuni anche alle tecniche parametriche di misurazione dell'efficienza. A sostegno della nostra preferenza per le tecniche non parametriche va detto che, laddove non siano disponibili dati affidabili in merito alla struttura dei costi, esse si prestano, più di quelle parametriche, ad ospitare specificazioni multiprodotto della funzione di produzione. Rispetto all'interpretazione dei risultati va ricordato inoltre che la nostra applicazione trascura l'effetto del progresso tecnico. Pertanto gli incrementi di efficienza misurati, poiché si riferiscono ad un intervallo di tempo su cui non è ragionevole escludere la presenza di progresso tecnico, non sono interamente dovuti all'effetto della riforma nel meccanismo di finanziamento, ma ne costituiscono un limite superiore.

I risultati dell'analisi ci offrono alcuni interessanti spunti per la politica e la regolamentazione. Innanzi tutto, sulla scorta delle nostre valutazioni, la produzione di ricoveri ospedalieri sembra caratterizzata da rendimenti di scala rapidamente decrescenti. Infatti, abbiamo riscontrato la presenza di inefficienza di scala nei presidi con più di 200 letti, mentre solo alcuni dei presidi con meno di 200 letti sembrano operare alla scala ottimale. L'analisi di correlazione conferma che al crescere della dimensione del presidio tanto maggiore è il grado di allontanamento dalla dimensione

efficiente. Questo risultato risente in modo forte dell'effetto della diversa composizione della casistica trattata dagli ospedali. La specificazione che abbiamo adottato, poiché controlla solo parzialmente per l'effetto del case-mix, potrebbe portare a sottostimare la dimensione di scala efficiente. Tuttavia, qualora venisse, anche parzialmente confermato, suggerirebbe di perseguire politiche di accorpamento e di razionalizzazione orientate all'accrescimento del segmento più basso della distribuzione dimensionale dei presidi. Sotto questo profilo, il provvedimento di chiusura dei presidi con meno di 120 posti-letto non è incoerente con la nostra evidenza. L'analisi sul quadriennio ci rivela che l'efficienza di scala non si modifica, tanto che siamo portati a ritenere che la razionalizzazione delle dimensioni produttive non abbia sortito effetti significativi, né positivi né negativi. Tuttavia ha contribuito a spostare la distribuzione della capacità produttiva per classe dimensionale nella giusta direzione, almeno a giudicare da quanto è accaduto in Emilia-Romagna (si veda la **Tabella 6**).

*Tabella 6: Distribuzione dimensionale della capacità di offerta. Ripartizione dei letti disponibili per classe dimensionale dell'ospedale.*

	1994	1995	1996	1997
Presidi con ...				
Meno di 100 letti	12.2%	11.3%	13.8%	13.7%
Fra 100 e 200 letti	18.1%	18.0%	17.0%	22.7%
Più di 500 letti	69.7%	70.7%	69.2%	63.5%
Totale	100.0%	100.0%	100.0%	100.0%

L'evoluzione dell'efficienza tecnica pare segnalare la successione di due regimi distinti. Nel primo biennio essa è coerente con l'espressione delle potenzialità produttive dei presidi, fortemente incoraggiata nella prima fase di attuazione della riforma. A questa fa seguito, nell'ultimo anno esaminato, un lieve peggioramento dovuto alla sensibile riduzione dei volumi di produzione. Tale riduzione parrebbe da ricondursi direttamente all'imposizione di vincoli finanziari sull'innalzamento dei volumi da parte degli enti finanziatori, piuttosto che alla riduzione degli input a disposizione. Questa evidenza testimonia la presenza di un'elevata capacità produttiva investita. Tale capacità, da riferirsi in particolare al personale, è all'origine di quasi il 60% del costo di produzione dei servizi ospedalieri. La difficoltà con cui si procede alla sua riduzione parrebbe all'origine del trade-off fra razionamento delle prestazioni e raggiungimento di un elevato livello di efficienza produttiva da noi riscontrato.

L'analisi del nesso fra modelli di finanziamento e efficienza degli ospedali ci offre alcune interessanti deduzioni. Innanzi tutto, al crescere del livello di complessità e di qualità degli ospedali aumenta il grado di efficienza tecnica e diminuisce quello di efficienza di scala. Alla luce della nostra analisi, la graduazione delle tariffe adottate nella regione Emilia-Romagna pare ispirata dalla volontà di coprire i differenziali di efficienza di scala e a stimolare, di converso, il miglioramento dell'efficienza tecnica nell'uso delle risorse. L'evidenza che abbiamo costruito in merito all'evoluzione delle prestazioni produttive per questi quattro gruppi di presidi ci conferma che, nel periodo considerato, questo obiettivo è stato effettivamente raggiunto. Nel primo biennio della riforma l'efficienza tecnica dagli ospedali in gestione diretta migliora in modo generalizzato, per poi peggiorare nel 1997. Le Aziende Ospedaliere, al contrario, manifestano un'efficienza costante nel primo biennio e un miglioramento, anche se non significativo, nel quarto anno. Il risultato dei primi emerge, nel primo biennio, a causa dell'incremento nella produzione di ricoveri e di punti DRG in concomitanza con la

riduzione degli input, letti e personale medico in particolare. A ciò fa seguito, nell'ultimo anno, una brusca riduzione in entrambe le produzioni pur in presenza di una sostanziale invarianza nel personale. Viceversa il risultato delle seconde emerge a causa dell'incremento dei punti prodotti, bilanciato nel primo biennio da un incremento nel personale, e mantenuto nell'ultimo anno pur in presenza di una drastica riduzione nei posti letto e una lieve riduzione nelle dotazioni di personale.

Queste evidenze ci permettono di trarre alcune considerazioni in merito all'efficacia del sistema di incentivi introdotti dalla riforma. Innanzi tutto parrebbe che per gli ospedali in gestione l'effetto incentivante sulla produttività sia stato inizialmente forte ed efficace, mentre successivamente sia prevalso l'obiettivo del razionamento delle prestazioni. Questo ha messo in luce alcuni eccessi di dotazione nei fattori produttivi, il cui assorbimento si segnala come un obiettivo prioritario dell'azione di politica economica in questo settore. Al contrario, per le Aziende Ospedaliere, l'effetto incentivante sulla produttività è inizialmente inefficace, mentre diventa rilevante, anche se non in misura statisticamente significativa, solo nell'ultimo anno. Questo risultato è piuttosto sorprendente e va opportunamente interpretato. Infatti, l'introduzione del pagamento a tariffa ha coinvolto, in particolar modo, le Aziende Ospedaliere. E' lecito quindi ritenere che questi produttori siano fortemente incentivati al perseguimento dell'efficienza produttiva. Del resto, nel Piano Sanitario Nazionale si rileva la necessità di individuare *“sistemi di valutazione della produttività delle strutture destinatarie di modalità di remunerazione fondate principalmente sui costi”*, quasi che per queste gli incentivi al miglioramento delle prestazioni produttive siano più deboli di quelli degli ospedali-azienda. La nostra evidenza parrebbe contraddire questa preoccupazione, addirittura rovesciandone la direzione. In realtà, ci pare che il richiamo all'attenzione verso le prestazioni produttive degli ospedali in gestione sia ben fondato. Nel nostro studio, infatti, rileviamo un miglioramento dell'efficienza in corrispondenza di una fase di crescita della produzione. Tale miglioramento viene meno nel momento in cui i vincoli di razionamento sulle prestazioni vengono fatti valere: se i consumi di ricoveri ospedalieri vengono portati a livelli più contenuti emerge, in questo segmento, un eccesso di capacità produttiva. Sotto questo profilo, gli incentivi del sistema di finanziamento nel caso di identità fra ente gestore e ente finanziatore, l'Azienda Unità Sanitaria Locale, parrebbero insufficienti, alla luce della nostra analisi, a stimolare congiuntamente il miglioramento dell'efficienza produttiva dei presidi esclusi dal meccanismo di finanziamento a tariffa e il contenimento delle prestazioni prodotte dagli stessi. Con ciò giustificando la preoccupazione espressa nel PSN.

Più sorprendente è invece la modestia del miglioramento indotto nelle Aziende Ospedaliere che la nostra evidenza sembra suggerire. A rafforzare questa indicazione si consideri che, se il progresso tecnico si sviluppa in particolare in questi ospedali, visto anche che svolgono funzioni di insegnamento e ricerca, allora i nostri risultati giungono addirittura a rivelare la presenza di effetti di inefficienza sulle Aziende Ospedaliere. A cosa debba ricondursi, in questo caso, l'inefficacia del meccanismo di pagamento a tariffa nell'incentivare il miglioramento delle prestazioni produttive non è agevole a dirsi. Tuttavia occorre considerare che le Aziende Ospedaliere operano, nella realtà regionale considerata, in assenza di competizione a livello provinciale, mentre la componente di domanda proveniente da fuori provincia tende ad essere piuttosto contenuta. Ora, se nell'ambito degli accordi di fornitura con l'Azienda Sanitaria del territorio di riferimento, le tariffe vengano fissate a livelli coerenti con le caratteristiche della struttura produttiva e di costo dell'Azienda Ospedaliera, ecco che il contenuto

incentivante della concorrenza a prezzo fisso viene meno. A questo riguardo, ci pare che la preoccupazione espressa nel PSN debba estendersi ai contenuti degli accordi di fornitura fra Aziende Sanitarie e Aziende Ospedaliere, incoraggiando quegli elementi di valutazione comparativa delle condizioni di fornitura che paiono sinora non sufficientemente sviluppati.

### **Bibliografia**

- Ali, A. I. e Seiford, L. M., 1993, "The Mathematical Programming Approach to Efficiency Analysis", in Fried, H.O., Knox Lovell, C.A. e Schmidt, S.S., a cura di, *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, 1993, pp. 120-159.
- Banker, R.D., 1993, "Maximum Likelihood, Consistency and Data Envelopment Analysis : a Statistical Foundation", *Management Science* 39, pp. 1265-1273.
- Banker, R.D., 1996, "Hypothesis Tests Using Data Envelopment Analysis", *Journal of Productivity Analysis* 7, pp. 139-159.
- Banker, R.D., Conrad, R.F. e Strauss, R.P., 1986, "A Comparative Application of DEA and Translog Methods : an Illustrative Study of Hospital Production", *Management Science* 32, pp. 40-44.
- Breyer, F., 1987, "The Specification of a Hospital Cost Function. A Comment on the Recent Literature", *Journal of Health Economics* 6, pp. 147-157.
- Bruning, E.R. e Register, C.A., 1989, "Technical Efficiency within Hospitals: Do Profit Incentives Matter?", *Applied Economics* 21, pp. 1217-33.
- Burgess, J.F. e Wilson, P.W., 1993, "Technical Efficiency in Veterans Administration Hospitals", in Fried, H.O., Knox Lovell, C.A. e Schmidt, S.S., a cura di, *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, 1993, pp. 335-51.
- Burgess, J.F. e Wilson, P.W., 1995, "Decomposing Hospital Productivity Changes, 1985-1988: A Nonparametric Malmquist Approach", *Journal of Productivity Analysis* 6, pp. 343-63.
- Butler, J., 1995, *Hospital Cost Analysis*, Kluwer Academic Publishers, Dordrecht.
- Byrnes, P. e Valdmanis, V., 1994, "Analyzing Technical and Allocative Efficiency of Hospitals", in Charnes, A., Cooper, W., Lewin, A.Y. e Seidorf, L.M., a cura di, *Data Envelopment Analysis*, Kluwer Academic Publishers.
- Carter, G., Newhouse, J. e Relles, D., 1990, "How Much Change in the Case Mix Index Is DRG Creep?", *Journal of Health Economics* 9, pp. 411-435.
- Charnes, A., Cooper W. W., e Rhodes, E., 1978, "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research* 2, pp. 429-444.
- Coulam, R.F. e G.L. Gaumer, 1991, "Medicare's Prospective Payment System: a Critical Appraisal", *Health Care Financing Review*, supplement, pp. 45-77.
- Cutler, D.M., 1995, "The Incidence of Adverse Medical Outcomes under Prospective Payment", *Econometrica* 63, pp. 29-50.
- Debreu, G., 1951, "The Coefficient of Resource Utilization", *Econometrica* 19, pp. 273-292.
- Ellis, R.P. e T.G. McGuire, 1996, "Hospital Response to Prospective Payment: Moral Hazard, Selection, and Practice-Style Effects", *Journal of Health Economics* 15, pp. 257-77.
- Ellis, R.P., 1998, "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins", *Journal of Health Economics* 17, pp. 537-556.

- Fabbri, D. e C. Ugolini, 1999, "Pagamento prospettico, modelli sanitari regionali e funzioni di controllo delle prestazioni ospedaliere", in Fabbri, D. e G. Fiorentini, a cura di, *Regolamentazione e Finanziamento dei Servizi Pubblici*, Carocci, Roma.
- Fabbri, D., 1999, "Comportamenti opportunistici e riforma sanitaria: ricoveri ripetuti e trasferimenti negli ospedali dell'Emilia-Romagna", in Fiorentini, G., a cura di, *Rapporto sul Servizio Sanitario Nazionale*, 1999, Il Mulino, Bologna (in corso di pubblicazione).
- Fabbri, D., Fazioli, R. e Filippini, M., 1996, *L'Intervento Pubblico e l'Efficienza Possibile*, Il Mulino, Bologna.
- Färe, R., Grosskopf, S., e Lovell, C.A.K., 1985, *The Measurement of Productive Efficiency*, Kluwer, Boston.
- Färe, R., Grosskopf, S., Lindgren, B. e Roos, P., 1994, "Productivity Developments in Swedish Hospitals: a Malmquist Output Index Approach", in Charnes, A., Cooper, W., Lewin, A.Y. e Seidorf, L.M., a cura di, *Data Envelopment Analysis*, Kluwer Academic Publishers.
- Farrell, M. J., 1957, "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society A* 120, pp. 253-281.
- France G., 1994, "Modelli relazionali tra i servizi sanitari nazionali e i fornitori di prestazioni", in G. France, a cura di, "Concorrenza e Servizi Sanitari", *Quaderni per la Ricerca*, n. 33, Istituto di Studi sulle Regioni, CNR, Roma.
- Grosskopf, S. e Valdmanis, V., 1987, "Measuring Hospital Performance: A Non-parametric Approach", *Journal of Health Economics* 6, pp. 89-107.
- Grosskopf, S., Margaritis, D. e Valdmanis, V., 1997, "Comparing Teaching and Non-teaching Hospitals: a Frontier Approach", Discussion Paper No. 97-29, Department of Economics, Southern Illinois University.
- Hollingsworth, B., Dawson, P. e Maniadakis, N., 1999, "Efficiency Measurement of Health Care: A review of non-parametric methods and applications", *Health Care Management Science* 2, pp. 161-172.
- Keeler, T.E. e Ying, J.S., 1996, "Hospital Costs and Excess Bed Capacity: A Statistical Analysis", *Review of Economics and Statistics* 78, pp. 470-481.
- Koopmans, T. C., 1951, "An Analysis of Production as an Efficient Combination of Activities", in Koopmans, T.C., a cura di, *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics 13, Wiley and Sons.
- Linna, M., 1999, "The Impact of Health Care Financing Reform on the Productivity Change in Finnish Hospitals", paper presentato al 2° Congresso Mondiale della iHEA, Rotterdam 6-9 giugno, 1999.
- Mapelli, V., 1994, "La sanità: avanti adagio con la riforma del SSN", in Bernardi, L., a cura di, *La Finanza Pubblica Italiana*, Rapporto 1994, Il Mulino, Bologna.
- Mapelli, V., 1995, "La sanità pubblica: la svolta delle tariffe prospettiche", in Bernardi, L., a cura di, *La Finanza Pubblica Italiana*, Rapporto 1995, Il Mulino, Bologna.
- Newhouse, J.P., 1994, "Frontier Estimation: How Useful a Tool for Health Economics?", *Journal of Health Economics* 13, pp. 317-322.
- Nyman, J.A. e Bricker, D.L., 1989, "Profit Incentives and Technical Efficiency in the Production of Nursing Home Care", *Review of Economics and Statistics* 71, pp. 586-94.
- Pestieu, P. e Tulkens, H., 1990, "Assessing the Performance of Public Sector Activities : Some Recent Evidence from the Productive Viewpoint", Discussion Paper n°. 9060, CORE, Université Catholique de Louvain, Belgio.

- Petretto, A. e G. Pisauro, 1995, "La riforma della pubblica amministrazione sotto il profilo dell'analisi economica: i controlli e la ristrutturazione degli uffici", *Politica Economica* 11, pp. 13-59.
- Rebba, V. e L. Rizzi, 1998, "Criteri di finanziamento delle aziende sanitarie: problematiche relative all'integrazione tra trasferimenti regionali per quota capitaria e remunerazione delle prestazioni ospedaliere con tariffe DRG", relazione presentata al IV Workshop di Economia Sanitaria, Torino 4-5 Giugno 1998.
- Sherman, H.D., 1984, "Hospital Efficiency Measurement and Evaluation", *Medical Care* 22, pp. 922-938.
- Siegel, S., 1956, *Nonparametric Statistics for the Behavioural Sciences*, McGraw-Hill, New-York.
- Soderlund, N., Csaba, I., Gray, A., Milne, R. e Raftery, J., 1997, "Impact of the NHS Reforms on English Hospital Productivity: an Analysis of the First Three Years", *British Medical Journal* 315, pp. 1126-1129.
- Taroni, F., 1996, *DRG/ROD e nuovo sistema di finanziamento degli ospedali*, Il Pensiero Scientifico Editore, Roma.
- Valdmanis, V., 1992, "Sensitivity Analysis for DEA Models: An Empirical Example Using Public vs. NFP Hospitals", *Journal of Public Economics* 48, pp. 185-205.
- Zweifel, P. e Breyer, F., 1997, *Health Economics*, Oxford University Press.

## Appendice Statistica

Tabella A.1: Descrizione del campione. Disaggregazione per livello di ospedale.

LIVELLO	1° LIVELLO N°=24	2° LIVELLO N°=7	3° LIVELLO N°=7	AOSP N°=5
Dati medi				
Dimessi*	5.1	17.0	22.1	46.5
Punti DRG*	3.9	14.3	20.4	43.0
Letti	143	451	630	1250
Medici	46	183	278	628
Infermieri	151	557	787	1571
Altro personale	50	178	259	660
% sul totale				
Dimessi*	19%	19%	25%	37%
Punti DRG*	17%	18%	26%	39%
Letti	20%	18%	26%	36%
Medici	15%	17%	26%	42%
Infermieri	17%	19%	26%	38%
Altro personale	16%	16%	24%	44%

Tabella A.2: Descrizione del campione. Disaggregazione per anno e livello di ospedale.

	1994	1995	1996	1997
<b>1° LIVELLO</b>				
Dimessi*	5.0	5.3	5.3	4.7
Punti DRG*	3.8	4.0	4.2	3.8
Letti	156	149	144	122
Medici	51	46	45	44
Infermieri	158	149	150	146
Altro personale	58	53	47	43
<b>2° LIVELLO</b>				
Dimessi*	16.7	18.0	17.9	15.6
Punti DRG*	13.5	15.0	15.3	13.5
Letti	466	489	473	374
Medici	205	179	175	173
Infermieri	575	572	553	528
Altro personale	206	180	173	153
<b>3° LIVELLO</b>				
Dimessi*	22.3	22.6	22.6	20.7
Punti DRG*	19.6	20.9	21.5	19.5
Letti	679	663	629	549
Medici	311	264	271	267
Infermieri	780	797	790	780
Altro personale	260	288	252	237
<b>AOSP</b>				
Dimessi*	46.9	46.8	46.7	45.7
Punti DRG*	40.1	42.8	44.8	44.4
Letti	1375	1298	1245	1084
Medici	668	615	617	612
Infermieri	1539	1558	1601	1586
Altro personale	613	666	684	677