

PARTE QUARTA

GLI STRUMENTI ECONOMICI
E DI GESTIONE

ANGELA TESTI ^(*)

MISURE DELL'EFFICIENZA TECNICA E DI SCALA DEI SERVIZI OSPEDALIERI

1. *Introduzione*

L'analisi economica, anche quando condotta a livello microeconomico, con riferimento al singolo ospedale, è finalizzata a studiare l'efficienza della singola struttura non in sé, ma in quanto elemento del sistema economico complessivo. L'oggetto dell'analisi è verificare se le scelte organizzative compiute dalla singola azienda sono efficienti, cioè conducono alla migliore ripartizione possibile delle risorse fra i vari usi alternativi. Se ciò avviene, gli economisti dicono che la soluzione ha le caratteristiche di "ottimo paretiano", perché non è possibile scostarsene per avvantaggiare qualche individuo senza al contempo peggiorare la situazione di qualche altro soggetto.¹

In questa ottica, la scelta fra pubblico e privato, fra sistema mutualistico e solidaristico, fra concorrenza e monopolio non implica nessun giudizio di valore, ma soltanto la scelta tra una soluzione efficiente e una non efficiente dal punto di vista allocativo. Questo modo di vedere le cose riesce, in un certo senso, a comporre l'apparente conflitto tra efficienza e solidarietà (a livello di sistema sanitario) e quello tra efficienza ed etica professionale (a livello di comportamento del singolo operatore). Tale dicotomia è una delle cause che spesso rendono gli operatori sanitari diffidenti e resistenti a considerazioni economiche che interpretano come lesive della propria professionalità. Se, invece, l'aspetto economico viene correttamente con-

^(*) *Facoltà di Economia, Università di Genova.*

¹ L'ottimo paretiano non sempre esiste: in taluni casi l'analisi economica riesce a determinare soluzioni univoche (per esempio, la concorrenza perfetta è meglio del monopolio), ma in altri casi, e il settore sanitario è uno di questi, le cose non sono così semplici e la soluzione efficiente potrebbe anche non esistere.

siderato come un discorso di priorità, ne risultano esaltate sia la professionalità, sia l'eticità del proprio operare. In un contesto di risorse scarse, infatti, poiché compiere un atto medico significa rinunciare a un altro, diventa etico scegliere da un punto di vista professionale l'intervento che dà i risultati migliori per la collettività.

La logica di fondo dell'analisi economica è, in un certo senso, molto semplice perché qualunque sia la problematica affrontata, essa si riduce sempre a un confronto fra benefici e costi. Per costi non si intendono gli esborsi monetari, ma i cosiddetti costi-opportunità, ovvero i benefici delle alternative cui si deve rinunciare. Le domande tipiche che si pone l'economista per determinare la migliore allocazione delle risorse scarse sono sostanzialmente quattro: che cosa produrre? per chi produrre? come produrre? chi prende le decisioni? Per rispondere si segue sempre lo stesso approccio metodologico consistente nel cercare il massimo rapporto benefici-costi. Rilevanti difficoltà, tuttavia, sorgono quando questa analisi viene applicata alla realtà, dove le decisioni sono complesse e coinvolgono più individui e istituzioni che agiscono in un contesto di interdipendenza. L'unico modo è allora quello di procedere per modelli, mettendo in evidenza soltanto un aspetto per volta e considerando dati tutti gli altri. Tali modelli sono costruiti sul presupposto che, se lo scopo finale è quello dell'efficienza allocativa, occorre che in ogni fase dell'attività economica non si sprechino risorse. L'ottimo paretiano richiede, cioè, che anche ai livelli inferiori si verifichino tutta una serie di "efficienze": nello scambio, nella produzione, nella scelta dei beni da produrre.

Il concetto di efficienza che verrà utilizzato in questo capitolo, che è quello più appropriato a livello di ospedale o di unità operativa degenziale, è quello di "efficienza tecnica", che, come si vedrà in seguito, comprende al suo interno anche quello di "efficienza di scala". In sintesi, si tratta di verificare se per ottenere un certo risultato, l'unità produttiva in esame sta utilizzando la quantità minima possibile di risorse. Nei paragrafi successivi tale concetto verrà spiegato dettagliatamente con riferimento a esemplificazioni grafiche. Qui interessa notare come parlare di efficienza tecnica significhi condurre

l'analisi su un aspetto particolare del lato dell'offerta e, pertanto, presupporre che molti problemi siano già risolti. In particolare, rimane fuori dall'analisi la determinazione di quante risorse destinare al settore ospedaliero e la definizione di quali siano le prestazioni efficaci e appropriate che devono essere prodotte con tali risorse.

Da un lato, pertanto, limitarsi soltanto alla valutazione dell'efficienza tecnica può apparire in qualche modo riduttivo, perché il ragionamento deve necessariamente essere completato rispondendo ad altre domande importanti, soprattutto se il produttore è pubblico ed eroga prodotti a favore della comunità, come nel caso del settore sanitario. D'altra parte, il punto di partenza più ovvio per arrivare alla efficienza a livello del sistema economico nel suo complesso è quello di verificare l'assenza di sprechi nelle risorse utilizzate durante il processo produttivo.

La misurazione dell'efficienza tecnica costituisce, cioè, il primo passo indispensabile per qualsiasi tipo di valutazione economica: qualunque sia, infatti, il tipo di unità produttiva di cui si vuole valutare l'efficienza, sia essa privata, pubblica, *for profit* o *non profit*, la prima domanda che ci si deve porre è se le risorse impiegate nel processo produttivo sono bene utilizzate, ovvero se si può far meglio eliminando sprechi eventualmente presenti.

Lo scopo specifico di questo lavoro è, in conclusione, quello di illustrare un modello di valutazione che consenta di verificare che la produzione avvenga in modo da evitare sprechi di risorse.

Si tratta della cosiddetta *Data Envelopment Analysis* (Dea)², basata su un modello di programmazione lineare, che deriva dalle usuali ipotesi della analisi della funzione della produzione di impostazione neoclassica. Il modello viene successivamente utilizzato per valutare la *performance* delle unità degen-

² Tale metodologia, introdotta da Charnes, Cooper, Rhodes nel 1978 [4], sta riscuotendo, in letteratura, un successo crescente con riferimento alla valutazione di unità produttive appartenenti ad amministrazioni pubbliche. Per una esposizione semplice si veda Coelli [5], oltre a due lavori della scrivente [9] e [10], applicati al settore sanitario.

ziali di medicina generale della Regione Liguria (dati 1997)³. Da quanto detto in precedenza, la metodologia di valutazione proposta non serve per fornire indicazioni operative alle singole strutture produttive, ma per monitorare l'erogazione di prestazioni ospedaliere a livello di sistema e garantire una base scientifica rigorosa alle politiche della riallocazione delle risorse e, in generale, delle incentivazioni a livello regionale.

2. L'efficienza tecnica

Efficienza è un termine che può avere diversi significati a seconda del tipo di decisione che si sta considerando.⁴ In questo contesto si utilizzerà il concetto di efficienza "tecnica" di un processo produttivo, cioè della trasformazione di *input* (ad esempio posti letto, ore di lavoro medico, ecc.) in *output* (ad esempio numero di ricoveri ordinari, visite ambulatoriali, ecc.). L'efficienza tecnica si misura facendo il rapporto tra *output* (ciò che esce dal processo produttivo) e *input* (ciò che è entrato nel processo produttivo).⁵

$$(1) \quad \text{efficienza tecnica} = \frac{\text{output}}{\text{input}}$$

³ Si ringrazia per la disponibilità dei dati e la gentile collaborazione il Settore Programmazione sanitaria, patrimonio e investimenti della Regione Liguria. I risultati presentati fanno parte di una complessa ricerca ancora in atto riguardante le specialità degenziali di base (medicina, chirurgia, ortopedia, ginecologia-ostetricia) che sono le più direttamente interessate alla manovra di riorganizzazione dell'offerta ospedaliera in atto nella Regione.

⁴ Zweifel, Breyer [14], per esempio, distinguono tra efficienza produttiva, efficienza clinica, ecc.

⁵ Si noti, inoltre, che il vero processo produttivo da valutare dovrebbe essere quello in cui il prodotto è l'"outcome", cioè il risultato in termini di salute, piuttosto che l'"output", cioè la quantità di prestazioni erogate. La valutazione qui proposta riguarda, invece, un processo produttivo intermedio, quello della produzione di prestazioni sanitarie ospedaliere, le quali sono a loro volta soltanto un input e non il più importante, né necessariamente esistente nella produzione di salute (che rimane il vero prodotto finale). D'altra parte, come si è già detto, anche con riferimento a questo aspetto limitato occorre che non si verifichino sprechi, cioè occorre che la produzione avvenga in modo efficiente.

Una unità operativa è tanto più efficiente (in senso tecnico) quanto più alto è questo rapporto, cioè quanto maggiore è l'*output*, dati gli *input*, oppure quanto minore è l'ammontare di *input* richiesti per ottenere un *output* dato. Nel primo caso, si tratta della cosiddetta valutazione "input-oriented", nel secondo caso di quella "output-oriented". Di seguito ci si riferirà sempre al primo tipo di valutazione, quella orientata agli *input*.

Si noti che molti degli indicatori comunemente utilizzati per la valutazione dell'attività dei reparti ospedalieri (tasso di occupazione dei posti letto, indice di rotazione, ecc.) sono già indici di efficienza tecnica, perché sono rapporti tra *output* e *input*. Il tasso di occupazione dei posti letto è il rapporto fra giornate di degenza effettivamente erogate (*output*) e giornate di degenza disponibili (*input*). Il tasso di rotazione è il rapporto fra numero di ricoveri (*output*) e numero di posti letto (*input*). Tali indicatori, tuttavia, misurano soltanto aspetti parziali del processo produttivo e possono, pertanto, fornire anche informazioni distorte se utilizzati singolarmente. La misura di efficienza che si propone in questo lavoro, calcolata con il modello *Data Envelopment Analysis* (Dea), permette, invece, di tenere conto contemporaneamente di tutti gli *input* e di tutti gli *output*.

Il primo problema da risolvere per calcolare il valore dell'efficienza tecnica, cioè per fare il rapporto fra *output* e *input*, è quello di avere (a numeratore) un unico numero che rappresenta complessivamente gli *output* dell'unità operativa da dividere per un altro numero (a denominatore) che sintetizzi i corrispondenti *input* utilizzati. Un reparto ospedaliero, tuttavia, produce di solito numerosi tipi di prodotti tra loro eterogenei e, pertanto, non sommabili: *output* clinici, (ricoveri, interventi, visite ambulatoriali, ecc.), *output* di tipo alberghiero (giornate di degenza) ed eventualmente anche *output* didattici. Anche gli *input* sono tra loro eterogenei: non si possono, per esempio, sommare posti letto con ore di lavoro.

Per poter calcolare il rapporto (1), occorre, di conseguenza, disporre di una serie di pesi (distinti per gli *output* e per gli *input*). Se, per esempio, si vuole valutare un processo produttivo che con due *input* (posti letto e medici) produce due *output* (ricoveri e visite ambulatoriali), l'espressione (1) per

misurare l'efficienza diventa:

$$(2) \text{ efficienza tecnica} = \frac{\text{ricoveri} \times \text{peso dei ricov.} + \text{visite ambul.} \times \text{peso delle visite}}{\text{posti letto} \times \text{peso dei posti letto} + \text{medici} \times \text{peso dei medici}}$$

Talvolta la scelta dei pesi è abbastanza semplice. Nel caso precedente si potrebbero, per esempio, moltiplicare il numero dei ricoveri e il numero delle visite per le corrispondenti tariffe riconosciute dalla Regione, assumendo implicitamente che le tariffe siano proporzionali all'“importanza” degli *output*. In questo caso, gli *output* vengono resi tra loro omogenei perché sono espressi in termini monetari e possono, pertanto, essere sommati tra di loro.⁶ Non sempre, tuttavia, esistono tariffe (si pensi per esempio agli *output* di tipo didattico), e diventa comunque piuttosto difficile individuare pesi adatti per gli *input*.

In tutti i casi, comunque, la scelta dei pesi è un problema cruciale che può determinare valori anche molto diversi dell'efficienza. Uno dei vantaggi principali del modello Dea è che non è necessario scegliere a priori i valori dei pesi, che sono, invece, calcolati come incognite dal modello stesso. È sufficiente, cioè, inserire i dati originari relativi agli *output* e agli *input* e il modello fornisce, insieme alla misura dell'efficienza anche i valori dei pesi che ha assegnato agli *output* e agli *input* per calcolarla.⁷ Si tratta dei pesi che consentono al numeratore di essere il più alto possibile e, contemporaneamente, al denominatore di risultare il più basso possibile. Se una unità operativa, per esempio, sta producendo relativamente molti ricoveri e poche visite ambulatoriali, il peso assegnato ai ricoveri sarà relativamente più alto di quello attribuito alle visite, in modo da valorizzare al meglio la produzione effettuata dal reparto. Se quella stessa unità operativa ha una disponibilità eccedente di posti letto rispetto al numero dei

⁶ Questo è ciò che si farà nella applicazione qui proposta: per semplicità, si considererà un unico *output* pari al valore monetario della produzione espresso in Drg.

⁷ Il modello originario [4] consiste nella massimizzazione del rapporto di efficienza (2) con il vincolo che i pesi determinati non rendano maggiore di uno il rapporto *output/input* di ciascuna delle Dmu del campione.

medici, il modello darà automaticamente un peso molto basso ai posti letto e più alto ai medici, in modo che appaiono meno consistenti le risorse a disposizione. È come se il modello cercasse di minimizzare i punti di debolezza e di massimizzare quelli di forza, in modo da mettere l'u.o. nella migliore luce possibile. Da questo punto di vista il modello è molto comodo. Da un lato il valore dei pesi assegnati dal modello ha un elevato contenuto informativo per il valutatore perché esso implicitamente rivela quali sono gli *output* nei quali la u.o. è relativamente più produttiva (sono quelli a cui sono associati pesi relativamente più alti) e quali sono gli input dei quali ha relativamente sovrabbondanza (sono quelli che presentano pesi relativamente più bassi). Dall'altro, la singola unità operativa non può "protestare" perché si vede assegnata i pesi a lei più favorevoli: essa stessa non potrebbe fare di meglio.

Dopo aver calcolato il rapporto *output/input*, rimane ancora il secondo problema, che consiste nel poter dire se tale rapporto è "buono". Per far questo è necessario disporre di uno *standard* di riferimento. È evidente che la definizione dello *standard* è cruciale per la valutazione del rapporto di efficienza della singola u.o. Anche in questo caso il modello Dea è molto comodo perché non è necessaria una decisione esterna (ad esempio, parametri di legge o parametri decisi discrezionalmente dall'analista) per determinare lo *standard*, ma esso viene individuato automaticamente all'interno del campione. Si tratta dell'u.o. che impiega al meglio le risorse a disposizione. Oltre al vantaggio che non è necessario scegliere a priori lo *standard*, perché è lo stesso modello che seleziona l'unità da copiare (*benchmark*) fra quelle del campione, il Dea ne presenta un altro: l'ideale da copiare non è irraggiungibile, ma quella realtà produttiva che nello stesso contesto operativo concreto (in quel certo ospedale, in quella certa regione, ecc.) sta operando al meglio.⁸

⁸ Non si tratta, di conseguenza di standard assoluti. Quando in seguito, nella verifica empirica il modello sarà applicato a sottocampioni rispetto a quello regionale (individuati per esempio raggruppando le u.o. per bacini di utenza), si vedrà, come in ciascun sottocampione si possano identificare ideali diversi e come la stessa unità operativa possa presentare punteggi diversi a seconda del campione con cui è confrontata.

Gli obiettivi di miglioramento dell'efficienza tecnica che, conseguentemente alle verifiche effettuate, sono suggeriti nei confronti delle u.o. carenti, assumono una caratteristica di effettiva fattibilità, elemento questo significativo in un contesto nel quale è fondamentale operare in termini concreti e rapidi.

La misura dell'efficienza tecnica che risulta dalla valutazione Dea ha, pertanto, la natura di un vero e proprio "punteggio". Per calcolare l'efficienza dell'unità operativa A, infatti, il rapporto *output/input* di A viene confrontato con quello di una unità operativa "ideale" scelta all'interno del campione. Ovviamente per poter applicare il modello Dea è condizione indispensabile che le u.o. del campione operino in ambiti operativi simili (per esempio, reparti di degenza all'interno di uno stesso ospedale, oppure riferiti allo stesso bacino di utenza, ecc.).

$$(3) \text{ punteggio di efficienza tecnica u.o. A} = \frac{\left(\frac{\text{output}}{\text{input}}\right)_{\text{u.o. A}}}{\left(\frac{\text{output}}{\text{input}}\right)_{\text{u.o. "ideale"}}$$

Esistono molte varianti del modello Dea: le due specificazioni illustrate di seguito permettono di rispondere alle due problematiche più rilevanti nella realtà operativa.

La prima è la tipica domanda che si pone la Regione che vorrebbe riuscire ad ottenere un certo risultato produttivo (quantità e tipologia di prestazioni) da una azienda con l'ammontare minimo possibile di risorse.

In questo caso la risposta è fornita dai punteggi di effi-

⁹ Come risulterà più chiaro, in seguito, inoltre, il modello Dea ammette che all'interno del campione di riferimento possano anche esistere diversi "ideali", nel senso che ciascuna unità operativa viene confrontata con quella a lei più simile (come *mix* di prodotti e come *mix* di *input* utilizzati). Si ammette, di conseguenza, che il processo produttivo possa essere più o meno capitalizzato, oppure che la u.o. preferisca specializzarsi in ricoveri piuttosto che in visite ambulatoriali. In sostanza, si lascia una certa libertà di organizzare la produzione in modi anche differenti e di scegliersi in parte i propri obiettivi a seconda delle proprie capacità.

cienza del tipo *input-oriented*¹⁰, che permettono proprio di determinare per ciascuna struttura quante sono le risorse in eccesso rispetto al compito produttivo assegnato e, di conseguenza, di quanto è possibile ridurre la spesa ospedaliera eliminando gli sprechi individuati. La seconda domanda riguarda la determinazione della dimensione ottima. L'inefficienza di una unità operativa può dipendere, infatti, anche da una dimensione sbagliata (troppo grande o troppo piccola) che non consente una organizzazione corretta delle procedure lavorative. In termini più rigorosi si parla in questo caso di valutazione della "efficienza di scala": il modello, in questo caso, consente di determinare se la struttura è sotto o sovradimensionata e, di conseguenza, se le risorse (e il corrispondente compito produttivo) devono essere aumentate o diminuite.

3. *Materiale e metodi*

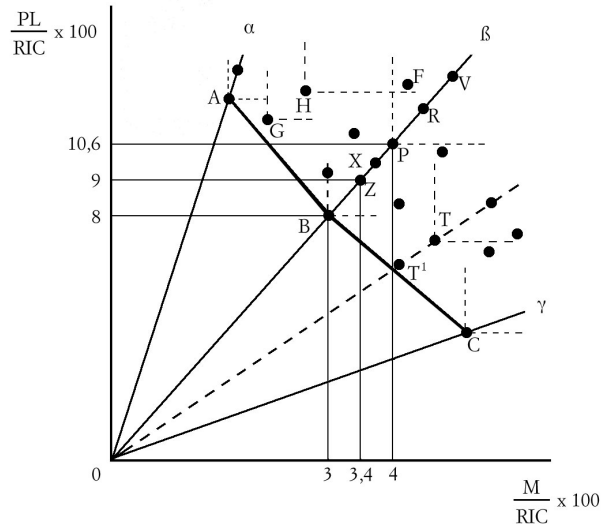
Il calcolo delle misure di efficienza, necessario per poter valutare i risultati ottenuti, si deduce dalle rappresentazioni grafiche riportate di seguito.

3.1. *La determinazione del punteggio di efficienza tecnica*

Nel primo grafico (Fig. 1) si rappresenta un campione di unità operative (u.o.) ospedaliere, di cui si conoscono il numero di ricoveri (Ric) effettuati in un certo periodo di tempo e il numero di posti letto (Pl) e di medici (M). Poiché ciascuna u.o. è identificata da tre numeri (Ric, Pl, M), è necessario normalizzare gli *input* (Pl e M) rispetto all'*output* (Ric) in modo da poter rappresentare il campione su un sistema di assi

¹⁰La valutazione *output-oriented*, che consiste, come si è detto, nel chiedersi quale potrebbe essere l'ammontare massimo di prestazioni che è possibile ottenere con certe risorse date, potrebbe, a sua volta aiutare a rispondere alla tipica domanda che si pone il responsabile di una unità operativa che si pone l'obiettivo di massimizzare il "fatturato" che può ottenere dalle risorse (letti e personale) a disposizione.

FIG. 1 L'individuazione delle unità operative efficienti



cartesiani a due dimensioni: per ciascuna u.o. sull'asse delle ascisse e su quello delle ordinate si indicano, rispettivamente, il numero di medici e il numero di posti letto, necessari per produrre 100 ricoveri. Ciò consente di confrontare tra loro unità operative diverse per dimensione, che diventano invece omogenee (rapportate a 100 ricoveri) quando rappresentate sul grafico.

Si faccia, per esempio, riferimento alle unità operative B e P, caratterizzate dai seguenti dati: B (200 Ric, 16 Pl, 6 M) e P (272 Ric, 29 Pl, 11 M). Nel caso di B i posti letto e i medici per produrre 100 ricoveri sono, rispettivamente, 8 e 3 e nel caso di P, sono 10,6 e 4. Questi numeri sono anche le coordinate dei due punti B e P nella Fig. 1. Le due unità operative pur essendo di dimensioni diverse e producendo una diversa quantità di ricoveri, operano con lo stesso *mix* di *input*, o, come si dice con la stessa "tecnica" produttiva. Esse presentano, infatti, lo stesso rapporto posti letto/medici, pari a 2,6. In entrambi i casi, cioè, ogni medico ha a disposizione 2,6 letti.

Dal punto di vista grafico questo significa che B e P sono allineati sulla stessa semiretta β , che indica, visivamente una tecnica produttiva. Nella Fig. 1 sono evidenziate altre due semirette α e γ , ciascuna caratterizzata da un proprio *mix* Pl/M. La tecnica α è la più “capitalizzata”, cioè è quella che produce ricoveri con una proporzione maggiore di posti letto rispetto ai medici, mentre la tecnica γ è quella meno capitalizzata, nel senso che ricorre a un ammontare di lavoro relativamente più elevato.

Come si vedrà, il modello Dea per calcolare il punteggio di efficienza di una u.o., la confronta con le altre unità operative caratterizzate dalla stessa combinazione di *input*. Il modello ammette, cioè molte tecniche produttive, anche piuttosto diverse tra loro e all'interno di ciascuna identifica lo *standard* di riferimento. Esistono, di conseguenza, molti “ideali” da copiare e ciascuna u.o. è confrontata con quello più vicino alle sue condizioni operative concrete.

Per calcolare il punteggio di efficienza si procede in due fasi. In un primo momento è necessario identificare gli *standard* di riferimento, cioè le unità operative “ideali” da copiare, a cui si dà per definizione punteggio di efficienza massimo uguale all'unità. A tale scopo si ricorre al cosiddetto criterio di dominanza: ciascuna u.o. domina quelle situate a nord-est, cioè quelle che per produrre gli stessi 100 ricoveri utilizzano una quantità maggiore di almeno uno dei due *input* (posti letto e medici). L'unità operativa H, ad esempio, domina F, ma a sua volta è dominata da A e da G. Le u.o. non dominate sono gli ideali da copiare. Nella Fig. 1 si individuano tre u.o. “ideali”, ciascuna caratterizzata da un diverso mix produttivo. A, B, C pur producendo con processi diversamente capitalizzati, sono tutte ugualmente efficienti con punteggio uguale a uno. Le unità produttive efficienti identificano la cosiddetta “frontiera”. Nella Fig. 1 la frontiera è la linea spezzata che unisce A, B, C. Le unità operative che non stanno sulla frontiera sono dominate e sono inefficienti.

La seconda fase consiste nel misurare la distanza dalla frontiera delle unità operative dominate. Ad esse si assegna un punteggio di efficienza minore di uno. Il punteggio misura la distanza rispetto alla frontiera e sarà, di conseguenza, tanto

più basso quanto più l'u.o. si allontana dalla frontiera. Per calcolare, ad esempio, il punteggio di efficienza di P, il modello ne misura la distanza rispetto a B che utilizza la stessa tecnica produttiva (stesso rapporto posti letto/medici pari a 2,6). Dal confronto risulta che P "spreca" risorse perché per produrre 100 ricoveri con quella tecnica utilizza più letti e più medici. Il punteggio di efficienza di P calcolato dal modello è il rapporto fra i due segmenti OB e OP. Il primo (OP) rappresenta sinteticamente gli *input* che si dovrebbero usare per produrre in modo efficiente 100 ricoveri. Il segmento a denominatore (OB) rappresenta, invece, gli *input* effettivamente utilizzati. In sostanza, la precedente espressione (3) per misurare l'efficienza risulta semplificata, perché entrambe le unità operative (quella che viene valutata e quella ideale), producono 100 ricoveri. Il confronto, pertanto, può avvenire soltanto sulla base degli input utilizzati. L'efficienza tecnica (indicata con il simbolo K) si può determinare più semplicemente con la seguente espressione (4) che è una semplificazione della (3):

$$(4) \text{ punteggio di efficienza tecnica } K = \frac{\text{input che si dovrebbero utilizzare (standard)}}{\text{input effettivamente utilizzati}}$$

Nel caso dell'u.o. P della Fig. 1, utilizzando tale espressione risulta che l'efficienza tecnica dell'u.o. P (K_p) risulta pari a 0,75. E lo stesso calcolo si può ripetere per qualunque u.o. del campione, per esempio per l'u.o. Z, per la quale risulta che $K_z = 0,88$:

$$K_p = \frac{OB}{OP} = \frac{8Pl}{10,6Pl} = \frac{3M}{4M} = 0,75 \quad K_z = \frac{OB}{OZ} = \frac{8Pl}{9Pl} = \frac{3M}{3,4M} = 0,88$$

$K_p = 0,75$ significa che P potrebbe produrre gli stessi ricoveri che attualmente sta producendo (272) utilizzando soltanto il 75% dei suoi *input* attuali, cioè con 22 posti letto e 8 medici. In altre parole, P spreca 7 dei suoi 29 letti e 3 dei suoi 11 medici. In modo analogo si può interpretare il punteggio $K_z = 0,88$, che risulta correttamente più elevato di quello di P perché l'u.o. Z è più vicina alla frontiera.

Molto spesso, infine, l'unità operativa di cui si vuole misu-

rare l'efficienza non sta esattamente sullo stesso raggio di una u.o. della frontiera: è il caso del reparto T della Fig. 1 che utilizza un processo intermedio tra i due processi efficienti β e γ . In questo caso, il significato del punteggio di efficienza non cambia, ma diventa un po' più complessa la sua determinazione. T è confrontata con l'u.o. fittizia (T') che sta sulla frontiera in posizione intermedia tra B e C. Il corrispondente punteggio di efficienza di T sarà, pertanto: $K_T = OT'/OT$. Per conoscere in modo preciso la composizione dell'u.o. T' che costituisce lo standard di T, il modello calcola automaticamente i cosiddetti "coefficienti di attivazione" z_B e z_C . Se, per esempio, $z_B = 0,2$ e $z_C = 0,7$, l'ammontare ideale di risorse, che T dovrebbe usare per operare in modo efficiente (posizionandosi in T'), si determina sommando tra loro il 20% delle risorse di B e il 70% delle risorse di C.

3.2. La determinazione del punteggio di efficienza di scala

In presenza di un punteggio di efficienza tecnica insoddisfacente (minore dell'unità), occorre approfondire l'analisi per evidenziare le responsabilità dello spreco di risorse che risulta dal punteggio: si potrebbe, infatti, trattare soltanto di una "cattiva" gestione da parte dei dirigenti della u.o., ma potrebbe anche dipendere dal fatto che essendo sbagliata la scala operativa (troppo grande o troppo piccola), non si possono applicare in modo corretto tutte le procedure atte a valorizzare al massimo gli *input* esistenti. Oltre al punteggio di efficienza tecnica, è, di conseguenza, necessario determinare anche quello di "efficienza di scala", affrontare, cioè, il problema delle cosiddette "economie/diseconomie di scala". Si tratta della circostanza in cui aumentando la dimensione di una u.o. molto piccola (si assegnano per esempio il doppio, il triplo delle risorse) si verificano delle economie di scala, cioè il prodotto aumenta più del doppio o più del triplo. Ne consegue che conviene aumentare la dimensione di u.o. piccole, perché sfruttando queste economie di scala è possibile diminuire i costi unitari di produzione. Dopo che l'u.o. ha raggiunto una certa dimensione, cominciano, invece, a verificarsi delle dise-

conomie di scala: un ulteriore aumento delle risorse a disposizione genera aumenti meno che proporzionali del prodotto e, di conseguenza, il costo unitario di produzione cresce.

La dimensione ottima è, in conclusione, quella per la quale sono già state sfruttate tutte le economie di scala e non si sono ancora presentate diseconomie. In altri termini, è quella che consente di ottenere contemporaneamente il massimo numero di ricoveri per posto letto e il massimo numero di ricoveri per medico.¹¹ Per questa sua caratteristica di massimizzare la produttività di tutti gli *input* a disposizione, la dimensione ottima viene normalmente definita come la “Most Productive Scale Size” (Mpss). Va rilevato che anche in questo caso il modello Dea è molto flessibile: non esiste, infatti, una sola dimensione ottima. Esiste una dimensione ottima per ciascuna delle tecniche produttive efficienti, cioè di quelle identificate dalla frontiera.¹²

L'efficienza di scala è già compresa nell'efficienza tecnica: quando una u.o. è sulla frontiera essa è completamente efficiente, nel senso che è della dimensione corretta e, al tempo stesso, non spreca i propri *input*. Quando, invece, una u.o. si scosta dalla frontiera, ciò potrebbe avvenire per due motivi che possono anche coesistere. Il primo dipende dal fatto che i responsabili di quella struttura non utilizzano bene gli *input* a disposizione, mentre il secondo si verifica quando anche con un utilizzo efficiente degli *input*, la scala di produzione è comunque “sbagliata”, cioè si potrebbe far meglio aumentando o diminuendo proporzionalmente tutti gli *input*. L'efficienza tecnica K può allora essere scomposta in due fattori: l'efficienza W , che si può chiamare “puramente tecnica” e l'efficienza di scala S :

$$(5) \quad K = W \times S$$

¹¹ In termini più rigorosi, la dimensione ottima è quella per la quale la produttività media di ciascun input è massima ed è, di conseguenza, uguale alla corrispondente produttività marginale, per cui è pari all'unità l'elasticità di scala.

¹² Un ulteriore problema, qui non affrontato, è peraltro quello di classificare le u.o. sulla frontiera, ovvero di determinare l'u.o. più efficiente fra le efficienti. Per una discussione al proposito si veda [11].

gnano in modo diverso, a seconda della loro personalità, delle sollecitazioni provenienti dall'ambiente, delle informazioni di cui dispongono¹³ e che, per ragioni di opportunismo individuale, nella maggioranza dei casi, essi non erogano il massimo sforzo lavorativo.

Una esemplificazione della scomposizione dei punteggi di efficienza nelle due componenti W e S, è indicata nella Fig. 2, in cui sono rappresentate solo alcune delle u.o. del campione di Fig. 1. Si tratta delle u.o. B, P, R, V, X, Z, che nel grafico precedente risultavano allineate lungo la semiretta β , perché utilizzano la stessa tecnica produttiva, caratterizzata da un rapporto $Pl/M=2,6$. Il grafico di Fig. 2 è un modo alternativo di rappresentare il processo produttivo per meglio valutare ciò che avviene all'interno della tecnica β . In questo caso l'*output* (cioè il numero dei ricoveri) è indicato sull'asse delle ordinate, mentre sull'asse delle ascisse si indicano multipli della combinazione produttiva base (2,6/1). Di conseguenza, la combinazione produttiva (letti-medici) del reparto B, che replica 5,7 volte la combinazione base, è indicata dalla coppia (16;6) e quella del reparto B, che la replica circa 11 volte, da (29;11).

Se ci si propone di scomporre le cause di inefficienza, ovvero di determinare quanto dell'inefficienza complessiva è dovuta a opportunismo individuale e quanto è dovuta a una dimensione sbagliata, non è sufficiente individuare una sola frontiera, ma ne sono necessarie due: quella cosiddetta Crs (*Constant returns to scale*), più severa perché conteggia entrambi i tipi di inefficienza, essendo la frontiera che permette di calcolare la misura K, e la cosiddetta frontiera Vrs (*Variable returns to scale*) che dà un giudizio meno severo, perché misura l'inefficienza senza tenere conto della dimensione sbagliata¹⁴ (è

¹³ La teoria della *X-inefficiency* [6] è un tentativo di superare uno dei principali difetti della teoria economica neoclassica, cioè la mancanza dell'analisi dei processi decisionali e operativi che avvengono all'interno dell'impresa. Leibenstein suggerisce al contrario che l'impresa non debba essere vista come una entità decisionale, perché è il singolo individuo che prende le decisioni, valutando di volta in volta il grado di razionalità da applicare (razionalità selettiva).

¹⁴ Ciò significa che si ammette una funzione di produzione caratterizzata da rendimenti variabili di scala.

cioè la frontiera che permette di misurare W). Come risulta dalla (6), la misura dell'efficienza di scala S , si ottiene facilmente dal rapporto tra K e W .

Per determinare graficamente la frontiera Crs , è sufficiente riferirsi alla prima definizione di efficienza che è stata data nella (1). Il rapporto *output/input* è, infatti, molto facile da identificare graficamente nella Fig. 2, perché è l'inclinazione del raggio che si ottiene unendo ogni u.o. con l'origine degli assi. Quanto più ripido è questo raggio e quanto più ampio è l'angolo, tanto più alto è il rapporto *output/input* e tanto più, di conseguenza, è efficiente l'u.o. Per trovare la $Mpss$ tra le unità operative appartenenti alla tecnica produttiva β , è sufficiente allora trovare l'u.o. da cui passa il raggio con l'inclinazione massima, cioè quello situato più a sinistra di tutti (B nella Fig. 2). Tale u.o. è la $Mpss$ per la tecnica produttiva β : è, cioè, la u.o. con la dimensione ottimale per quel particolare *mix* di *input* (2,6 Pl per 1 M). La frontiera Crs è il raggio che passa per la $Mpss$.¹⁵

La frontiera Vrs (quella meno severa) si ottiene con il solito criterio della dominanza: in questo caso una u.o. domina quelle che a parità di *input* producono meno *output*, oppure quelle che a parità di *output* utilizzano più *input*, cioè in sostanza tutte quelle che nel grafico si situano a sud-est. Mentre alla frontiera Crs , per definizione, appartiene una sola unità operativa, la $Mpss$, alla frontiera Vrs possono appartenere più u.o. (R e Z, per esempio, nella Fig. 2): la frontiera assume, quindi, un caratteristico andamento a spezzata. Le u.o. che si trovano sulla frontiera Vrs sfruttano nel modo migliore le risorse a disposizione, ma di queste soltanto la B, che appartiene anche alla Crs , ha la dimensione corretta. Le altre, invece, pur lavorando al meglio, sperimentano un certo grado di inefficienza della scala produttiva: la Z è sottodimensionata, mentre la R è sovradimensionata.

¹⁵ Da quanto detto in precedenza, è ovvio che il procedimento appena descritto deve essere ripetuto per le altre tecniche produttive efficienti del campione, ricorrendo a rappresentazioni grafiche analoghe alla Fig. 2. Con riferimento al campione della Fig. 1, occorrerebbero altri due grafici in cui indicare, rispettivamente, la frontiera Crs che passa per la $Mpss$ della tecnica α (u.o. A) e quella che passa per la $Mpss$ della tecnica γ (u.o. C).

Avendo tracciato le due frontiere, è molto semplice controllare visivamente l'efficienza delle varie unità operative. Il confronto è sempre fatto sul segmento orizzontale che unisce l'u.o. alla quale si è interessati con l'asse delle ordinate. La distanza orizzontale dell'u.o. dalla frontiera Crs indica se essa è complessivamente inefficiente, mentre la distanza orizzontale dell'u.o. dalla frontiera Vrs indica se essa è inefficiente perché utilizza male gli input a disposizione. Ne consegue che la distanza orizzontale tra le due frontiere, in corrispondenza della u.o. che si sta valutando, indica il punteggio di efficienza di scala. Un valore minore dell'unità indica inefficienza, ma non ne discrimina la causa che può essere di segno opposto. Occorre, di conseguenza, una ulteriore ispezione: se l'u.o. è situata più in alto rispetto alla Mpss, essa è sovradimensionata (troppo grande), mentre è sottodimensionata (troppo piccola) se è situata più in basso.

Per una maggiore comprensione, si discutono nel dettaglio le situazioni delle u.o. indicate nella Fig. 2, che sintetizzano tutte le possibili combinazioni. Come si è visto, l'u.o. B (l'unica sulla frontiera Crs) è la Mpss, per la quale $K=W=S=1$: la situazione non richiede alcun correttivo. Le altre unità operative sono inefficienti per l'una o l'altra causa, o per entrambe. Nel caso dell'u.o. V si ha soltanto inefficienza puramente tecnica ($K<1$; $W<1$; $W=K$; $S<1$). L'u.o. è correttamente dimensionata e, pertanto, è sufficiente una riduzione di risorse fino a B.

Per le due u.o. Z e R (che si trovano sulla Vrs) l'unica causa di inefficienza è la dimensione sbagliata: troppo piccola per Z e troppo grande per R. In entrambi i casi le u.o. sono caratterizzate da $K<1$; $W=1$ e $S<1$. Nel caso di Z, se ci si basasse unicamente sul punteggio K la strategia richiesta sarebbe quella di ridurre gli *input* da 9 a 8 letti e da 3,4 a 3 medici, posizionandosi in Z'. In questo caso, tuttavia, non sembra avere molto senso ridurre risorse che sono usate al meglio. La strategia più corretta sarebbe quella di aumentare le risorse (16 letti e 6 medici) e contemporaneamente assegnare un compito produttivo di 200 ricoveri per raggiungere la Mpss. Alla Z, cioè, dovrebbero essere assegnate un po' meno del doppio delle risorse attuali per produrre un numero doppio di ricoveri. È ovvio che l'opportunità concreta di seguire

questa indicazione dipende da altri fattori che qui non sono considerati, per esempio, dal fatto che sia appropriato raddoppiare i ricoveri. Nel caso di R, in analogia, la strategia ottima sarebbe quella di ridurre contemporaneamente le risorse assegnate e i ricoveri da effettuare (sempre che ciò fosse possibile), riportandosi in B.

Nel caso delle due u.o. P e X, infine, sono presenti entrambi i tipi di inefficienza: $K < 1$; $W < 1$; $W > K$; $S < 1$, anche se P è troppo grande, mentre X è troppo piccola. Per eliminare l'inefficienza puramente tecnica, le u.o. dovrebbero raggiungere la frontiera Vrs (posizionandosi, rispettivamente in P" e X"), continuando a produrre ciascuna lo stesso numero di ricoveri. Per eliminare anche l'inefficienza di dimensione, nel caso di P (che sta producendo troppo) occorre una ulteriore riduzione degli input a P' con una corrispondente riduzione del compito produttivo assegnato a 200 ricoveri. Nel caso di X, occorre, invece, aumentare gli *input* e aumentare in corrispondenza l'*output*.

4. Risultati

4.1. I punteggi di efficienza e le corrispondenti strategie

Le tre misure di efficienza sopradescritte sono state calcolate per le specialità degenziali di base della Regione Liguria per l'anno 1997.¹⁶ La difficoltà e la complessità del reperimento dei dati ha richiesto un notevole sforzo di omogeneizzazione che può avere comportato errori che potranno essere corretti a seguito di verifiche con le aziende. Si noti, inoltre, che la situazione 1997 non recepisce alcune manovre di riorganizzazione positivamente avviate dalle aziende sanitarie nel corso del '98, in applicazione della programmazione regionale.

¹⁶ È, attualmente, in corso la valutazione della efficienza con dati 1998. Si tratta non soltanto di un aggiornamento dei dati, ma di una attività più articolata che prevede anche momenti di informazione e di condivisione del metodo di valutazione scelto a livello di direttori generali e di definizione più corretta e più omogenea dei dati utilizzati.

TAB. 1 Punteggi medi di efficienza delle specialità di base (Regione Liguria, 1997)

| SPECIALITÀ | NUMEROSITÀ CAMPIONE | NUMERO U.O. SULLA FRONTIERA | K | W | S |
|------------------------|------------------------|--------------------------------|----------------|----------------|----------------|
| Medicina | 40 | 3 | 0.63 (0.16) | 0.78 (0.18) | 0.82 (0.13) |
| Chirurgia | 36 | 6 | 0.83 (0.14) | 0.88 (0.12) | 0.95 (0.09) |
| Ortopedia | 25 | 3 | 0.73 (0.18) | 0.88 (0.12) | 0.83 (0.17) |
| Ginecologia-ostetricia | 17 | 6 | 0.82 (0.18) | 0.93 (0.10) | 0.88 (0.15) |

I dati utilizzati sono stati i seguenti: valore della produzione (calcolata con le tariffe Drg) per l'*output*, numero di posti letto, numero di medici¹⁷ e numero di infermieri per gli *input*. Per chirurgia e ortopedia, inoltre, si è considerato come *input* aggiuntivo, il numero complessivo di sedute operatorie.

La valutazione¹⁸ per le specialità esaminate è riportata sinteticamente nella Tab. 1: per ciascuna di esse si indicano i valori medi dell'efficienza complessiva (K), di quella puramente tecnica (W) e di quella di scala (S).¹⁹ Tra parentesi sono indicate le deviazioni *standard*.

L'analisi dettagliata dei singoli dati viene riportata, invece, soltanto per le unità operative di medicina generale (cfr. Tabb. 2 e 3), per non appesantire troppo l'esposizione.

¹⁷ Il numero dei medici è stato dimezzato nel caso di docenti universitari, che per convenzione sono tenuti a dedicare all'assistenza metà del loro tempo lavoro.

¹⁸ Lo studio complessivo a livello regionale accompagnava queste valutazioni di efficienza tecnica con semplici indicatori di efficienza clinica e di appropriatezza generica, da cui si può notare come esista una certa correlazione tra una buona performance clinica e un buon utilizzo delle risorse. Si tratta di un aspetto che ci si propone di approfondire, perché indicherebbe che non esiste, o per lo meno che non è così rilevante, il *tradeoff* tra efficienza ed efficacia, che spesso si invoca per giustificare situazioni di inefficienza.

¹⁹ È ovvio che questi valori, che sono medie di punteggi riferiti a campioni diversi, non sono del tutto omogenee tra di loro.

Nella Tab. 2, le u.o. per ovvie ragioni di segretezza, non sono indicate con la loro denominazione, bensì con un numero, seguito da una indicazione della tipologia organizzativa (Pr = Presidio; Az = Azienda), della loro localizzazione sul territorio della Regione (Po = Ponente; Le = Levante; Ge = area metropolitana genovese) e, infine, del tipo di direzione (Ssn = Servizio sanitario nazionale, Uni = Università).

Queste sigle saranno utili in seguito per identificare la composizione dei sottocampioni (cfr. Tab. 3 del paragrafo successivo) nei quali si suddividerà il campione complessivo al fine di identificare eventuali cause sistematiche di inefficienza dovute a fattori “esterni”.

Nelle successive colonne E, F, G, si indicano i punteggi di efficienza ottenuti dalla valutazione Dea, utilizzando la terminologia spiegata in precedenza: Eff Crs (K) è, di conseguenza il punteggio più “severo”, cioè la distanza rispetto alla frontiera Crs, mentre Eff Vrs (W) è quello che indica la distanza rispetto alla frontiera meno severa, perché tiene conto dell’inefficienza puramente tecnica. Dall’esame di tali punteggi risulta che la frontiera Crs è formata da tre unità operative (26, 33, 36). Queste sono le unità da copiare, che vengono indicate dal modello come le ideali. Nella colonna H si indica rispettivamente per ciascuna u.o. quale deve essere il suo ideale, o il suo “peer group” di riferimento.²⁰ Si noti come il modello Dea lasci una certa flessibilità nella tecnica produttiva. Le tre unità ideali presentano, infatti, diversi rapporti fra gli *input*. Per esempio, il rapporto posti letto per medico è piuttosto diverso nelle tre *peer units*, ammontando a 8,6, 13 e 6,7, rispettivamente per le u.o. 26, 33 e 36.

La frontiera Vrs, invece, è formata, come ci si poteva aspettare, da un numero maggiore di u.o. Alle precedenti, si aggiungono le u.o. 2, 4, 8, 24, 27, 34. Nella colonna G è indicato il punteggio di efficienza di scala: un valore inferiore all’unità indica soltanto che la dimensione è sbagliata. Dalla colonna Z, si trae l’indicazione per sapere se la u.o. deve essere ridimensionata o ampliata. Drs (*Decreasing returns to scale*) indica

²⁰ Non sono riportate le singole proporzioni, cioè i coefficienti di attivazione, per non appesantire ulteriormente la tavola.

TAB. 2 I risultati completi della valutazione delle u.o. di medicina generale

| NUMERO UO | ORGANIZZ. | LOCALIZZ. | DIREZIONE | EFF. GRS | EFF. VRS | EFF. SCALA | PEER GROUP | PESO % LETTI | PESO % INF | PESO % INT | DRG (MILD) | LETTI | N° MEDICI | INFERMIERE | LETTI VIRTUALI VRS | MEDICI VIRTUALI VRS | INFERM. VIRTUALI VRS | LETTI VIRTUALI GRS | MEDICI VIRTUALI GRS | INFERM. VIRTUALI GRS | LETTI VIRTUALI MPSS | MEDICI VIRTUALI MPSS | INFERM. VIRTUALI MPSS | DRG MPSS (MILD) | REND. SCALA | DIM. |
|-----------|-----------|-----------|-----------|----------|----------|------------|------------|--------------|------------|------------|------------|-------|-----------|------------|--------------------|---------------------|----------------------|--------------------|---------------------|----------------------|---------------------|----------------------|-----------------------|-----------------|-------------|------|
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | |
| 1 | PR | PO | SSN | 0,52 | 0,66 | 0,79 | 26336 | 10 | 0 | 90 | 8,788 | 64 | 8 | 33 | 42 | 5 | 22 | 33 | 4 | 17 | 23 | 3 | 12 | 5,986 | DRS | |
| 2 | PR | PO | SSN | 0,70 | 1,00 | 0,70 | 26336 | 11 | 0 | 89 | 12,015 | 71 | 10 | 33 | 71 | 10 | 33 | 50 | 7 | 23 | 28 | 4 | 13 | 6,747 | DRS | |
| 3 | PR | PO | SSN | 0,60 | 0,70 | 0,86 | 36 | 100 | 0 | 0 | 6,749 | 40 | 5 | 39 | 28 | 3 | 27 | 24 | 3 | 23 | 20 | 3 | 20 | 5,618 | DRS | |
| 4 | PR | PO | SSN | 0,77 | 1,00 | 0,77 | 33336 | 14 | 86 | 0 | 12,533 | 73 | 7 | 46 | 73 | 7 | 46 | 56 | 5 | 35 | 23 | 2 | 15 | 5,182 | DRS | |
| 5 | PR | PO | SSN | 0,49 | 0,64 | 0,76 | 26336 | 10 | 0 | 90 | 9,441 | 71 | 14 | 38 | 45 | 9 | 24 | 34 | 7 | 18 | 21 | 4 | 11 | 5,766 | DRS | |
| 6 | PR | PO | SSN | 0,68 | 0,93 | 0,73 | 36 | 100 | 0 | 0 | 10,286 | 54 | 10 | 31 | 50 | 9 | 29 | 37 | 7 | 21 | 20 | 4 | 11 | 5,618 | DRS | |
| 7 | PR | PO | SSN | 0,73 | 0,90 | 0,80 | 26336 | 10 | 0 | 90 | 8,598 | 44 | 7 | 23 | 40 | 6 | 21 | 32 | 5 | 17 | 22 | 3 | 11 | 5,903 | DRS | |
| 8 | AZ | GE | SSN | 0,72 | 1,00 | 0,72 | 26336 | 11 | 0 | 89 | 11,927 | 68 | 11 | 32 | 68 | 11 | 32 | 49 | 8 | 23 | 27 | 4 | 13 | 6,642 | DRS | |
| 9 | AZ | GE | SSN | 0,56 | 0,67 | 0,83 | 26336 | 11 | 0 | 89 | 9,252 | 67 | 9 | 32 | 45 | 6 | 22 | 37 | 5 | 18 | 26 | 4 | 13 | 6,253 | DRS | |
| 10 | PR | GE | SSN | 0,73 | 0,82 | 0,88 | 33336 | 0 | 84 | 16 | 6,844 | 50 | 4 | 24 | 41 | 3 | 20 | 36 | 3 | 18 | 27 | 2 | 13 | 5,082 | DRS | |
| 11 | PR | GE | SSN | 0,53 | 0,63 | 0,84 | 36 | 100 | 0 | 0 | 7,179 | 48 | 7 | 29 | 30 | 4 | 18 | 26 | 4 | 15 | 20 | 3 | 12 | 5,618 | DRS | |
| 12 | PR | GE | SSN | 0,65 | 0,67 | 0,96 | 26336 | 10 | 0 | 90 | 5,943 | 33 | 7 | 18 | 22 | 5 | 12 | 21 | 5 | 12 | 20 | 4 | 11 | 5,664 | DRS | |
| 13 | PR | GE | SSN | 0,67 | 0,67 | 1,00 | 26336 | 12 | 0 | 88 | 7,377 | 48 | 8 | 21 | 32 | 5 | 14 | 32 | 5 | 14 | 32 | 5 | 14 | 7,325 | COST | |
| 14 | PR | GE | SSN | 0,41 | 0,45 | 0,91 | 26336 | 13 | 0 | 87 | 4,786 | 55 | 7 | 22 | 25 | 3 | 10 | 23 | 3 | 9 | 40 | 5 | 16 | 8,463 | IRS | |
| 15 | PR | LE | SSN | 0,37 | 0,50 | 0,74 | 26336 | 10 | 0 | 90 | 3,212 | 34 | 7 | 17 | 17 | 3 | 8 | 12 | 3 | 6 | 24 | 5 | 12 | 6,187 | IRS | |
| 16 | PR | LE | SSN | 0,68 | 0,78 | 0,87 | 26336 | 10 | 0 | 90 | 7,307 | 40 | 8 | 21 | 31 | 6 | 16 | 27 | 5 | 14 | 22 | 4 | 11 | 5,877 | DRS | |
| 17 | PR | LE | SSN | 0,59 | 0,60 | 0,98 | 26336 | 10 | 0 | 90 | 6,375 | 42 | 6 | 21 | 25 | 4 | 13 | 25 | 4 | 12 | 24 | 3 | 12 | 6,187 | COST | |
| 18 | PR | LE | SSN | 0,76 | 0,92 | 0,83 | 26336 | 10 | 0 | 90 | 8,252 | 41 | 8 | 21 | 38 | 7 | 19 | 31 | 6 | 16 | 23 | 4 | 12 | 6,028 | DRS | |
| 19 | PR | GE | SSN | 0,63 | 0,88 | 0,72 | 26336 | 10 | 13 | 87 | 3,926 | 30 | 3 | 12 | 26 | 3 | 11 | 19 | 2 | 8 | 31 | 3 | 12 | 6,329 | IRS | |
| 20 | PR | GE | SSN | 0,52 | 0,65 | 0,83 | 26336 | 13 | 0 | 87 | 6,473 | 51 | 7 | 20 | 32 | 4 | 13 | 27 | 4 | 10 | 36 | 5 | 14 | 8,777 | IRS | |
| 21 | PR | LE | SSN | 0,47 | 0,58 | 0,81 | 36 | 100 | 0 | 0 | 7,688 | 58 | 9 | 33 | 34 | 5 | 19 | 27 | 4 | 16 | 20 | 3 | 11 | 5,618 | DRS | |
| 22 | PR | LE | SSN | 0,48 | 0,55 | 0,87 | 36 | 100 | 0 | 0 | 6,970 | 52 | 9 | 33 | 28 | 5 | 18 | 25 | 4 | 16 | 20 | 3 | 13 | 5,618 | DRS | |
| 23 | PR | LE | SSN | 0,46 | 0,75 | 0,62 | 26336 | 12 | 0 | 88 | 3,173 | 32 | 4 | 13 | 24 | 3 | 10 | 15 | 2 | 6 | 38 | 5 | 16 | 8,236 | IRS | |

| U.O. | SITUAZIONE ATTUALE | | | | | | | | | | | | | STRATEGIA VRS | | | | | | STRATEGIA CRS | | | | | | STRATEGIA MPSS | | | | | | DIM. |
|------|--------------------|-----------|-----------|-----------|---------|---------|-----------|------------|--------------|----------|------------|-----------|-------|---------------|----------|--------------------|--------------------|---------------------|--------------------|--------------------|---------------------|---------------------|---------------------|----------------------|----------------|----------------|--|--|--|--|--|------|
| | NUMERO RIF. | ORGANIZZ. | LOCALIZZ. | DIREZIONE | EFF CRS | EFF VRS | EFF SCALA | PEER GROUP | PESO % LETTI | PESO % M | PESO % INF | DRG (MJD) | LETTI | N° MEDICA | INFAMIER | LETTI VIRTUALI VRS | MEDIC VIRTUALI VRS | INFERM VIRTUALI VRS | LETTI VIRTUALI CRS | MEDIC VIRTUALI CRS | INFERM VIRTUALI CRS | LETTI VIRTUALI MPSS | MEDIC VIRTUALI MPSS | INFERM VIRTUALI MPSS | DRG MPSS (MJD) | REND. SCALA | | | | | | |
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | | | | | | |
| 24 | PR | LE | SSN | 0,66 | 1,00 | 0,66 | 36; | 100 | 0 | 0 | 12,294 | 66 | 9 | 41 | 66 | 9 | 41 | 44 | 6 | 27 | 20 | 3 | 12 | 5,618 | DRS | | | | | | | |
| 25 | PR | PO | SSN | 0,59 | 0,60 | 0,98 | 36; | 100 | 0 | 0 | 5,259 | 32 | 9 | 19 | 19 | 5 | 11 | 19 | 5 | 11 | 20 | 6 | 12 | 5,618 | COST | | | | | | | |
| 26 | AZ | PO | SSN | 1,00 | 1,00 | 1,00 | 26; | 14 | 0 | 86 | 11,309 | 60 | 7 | 21 | 60 | 7 | 21 | 60 | 7 | 21 | 60 | 7 | 21 | 11,309 | COST | | | | | | | |
| 27 | AZ | PO | SSN | 0,78 | 1,00 | 0,78 | 36; | 100 | 0 | 0 | 9,613 | 44 | 11 | 33 | 44 | 11 | 33 | 34 | 9 | 26 | 20 | 5 | 15 | 5,618 | DRS | | | | | | | |
| 28 | AZ | GE | SSN | 0,64 | 0,73 | 0,87 | 36; | 100 | 0 | 0 | 7,146 | 40 | 8 | 32 | 29 | 6 | 23 | 25 | 5 | 20 | 20 | 4 | 16 | 5,618 | DRS | | | | | | | |
| 29 | AZ | GE | SSN | 0,48 | 0,51 | 0,93 | 26;36 | 12 | 0 | 88 | 4,973 | 45 | 10 | 20 | 23 | 5 | 10 | 21 | 5 | 10 | 31 | 7 | 14 | 7,163 | IRS | | | | | | | |
| 30 | AZ | GE | SSN | 0,61 | 0,62 | 0,99 | 33;36 | 0 | 88 | 12 | 5,06 | 48 | 4 | 18 | 30 | 2 | 11 | 29 | 2 | 11 | 30 | 2 | 11 | 5,580 | IRS | | | | | | | |
| 31 | AZ | GE | SSN | 0,91 | 0,96 | 0,95 | 33;36 | 0 | 87 | 13 | 6,175 | 46 | 3 | 14 | 44 | 3 | 13 | 42 | 3 | 13 | 37 | 2 | 11 | 5,517 | DRS | | | | | | | |
| 32 | AZ | GE | UNI | 0,57 | 0,69 | 0,83 | 36; | 100 | 0 | 0 | 4,348 | 27 | 4 | 15 | 19 | 3 | 11 | 15 | 2 | 9 | 20 | 3 | 11 | 5,618 | IRS | | | | | | | |
| 33 | AZ | GE | UNI | 1,00 | 1,00 | 1,00 | 33; | 17 | 83 | 0 | 4,813 | 26 | 2 | 15 | 26 | 2 | 15 | 26 | 2 | 15 | 26 | 2 | 15 | 4,813 | COST | | | | | | | |
| 34 | AZ | GE | UNI | 0,70 | 1,00 | 0,70 | 36; | 100 | 0 | 0 | 2,371 | 12 | 4 | 7 | 12 | 4 | 7 | 8 | 3 | 5 | 20 | 7 | 11 | 5,618 | IRS | | | | | | | |
| 35 | AZ | GE | UNI | 0,62 | 0,94 | 0,65 | 36; | 100 | 0 | 0 | 3,457 | 20 | 3 | 11 | 19 | 3 | 11 | 12 | 2 | 7 | 20 | 3 | 11 | 5,618 | IRS | | | | | | | |
| 36 | AZ | GE | UNI | 1,00 | 1,00 | 1,00 | 36; | 100 | 0 | 0 | 5,618 | 20 | 2,5 | 11 | 20 | 3 | 11 | 20 | 3 | 11 | 20 | 3 | 11 | 5,618 | COST | | | | | | | |
| 37 | AZ | GE | UNI | 0,54 | 0,90 | 0,60 | 36; | 100 | 0 | 0 | 3,164 | 21 | 3 | 12 | 19 | 3 | 11 | 11 | 2 | 6 | 20 | 3 | 11 | 5,618 | IRS | | | | | | | |
| 38 | AZ | GE | UNI | 0,74 | 0,94 | 0,79 | 26;36 | 10 | 0 | 97 | 3,424 | 17 | 4 | 10 | 16 | 4 | 9 | 13 | 3 | 7 | 21 | 5 | 12 | 5,828 | IRS | | | | | | | |
| 39 | AZ | GE | UNI | 0,41 | 0,60 | 0,69 | 33;36 | 13 | 87 | 0 | 3,726 | 36 | 4 | 20 | 21 | 2 | 12 | 15 | 2 | 8 | 21 | 2 | 12 | 5,426 | IRS | | | | | | | |
| 40 | AZ | GE | UNI | 0,41 | 0,98 | 0,42 | 36; | 100 | 0 | 0 | 2,067 | 18 | 3 | 10 | 18 | 3 | 10 | 7 | 1 | 4 | 20 | 3 | 11 | 5,618 | IRS | | | | | | | |
| TOT: | | | | | | | | | | | | | 270,4 | 1.744 | 266 | 922 | 1.353 | 204 | 717 | 1.102 | 165 | 580 | 1.014 | 154 | 518 | 246,8 | | | | | | |

che l'u.o. è troppo grande e deve essere ridimensionata; Irs (*Increasing returns to scale*) suggerisce che l'u.o. deve essere ampliata per sfruttare le economie di scala; Crs (*Constant returns to scale*) indica, infine, che la dimensione è corretta. Su 37 u.o. inefficienti, 22 u.o. (cioè circa il 60%) appaiono sovradimensionate. Si noti, peraltro, come la dimensione ottima (M_{pss}) dipenda dalla tecnica produttiva. Nel caso della u.o. 26, la dimensione ottima si raggiunge con 60 posti letto e una produzione di 11.309, mentre nel caso delle altre due *peer units* le economie di scala risultano sfruttate al massimo per dimensioni più contenute: 26 posti letto per la 33 e 20 per la 36.

Nelle colonne I, J, K si riportano in percentuale i pesi relativi degli *input* produttivi. Si è detto che le u.o. sono portate a pesare gli *input* in misura tanto più elevata quanto più essi sono scarsi e viceversa. Di conseguenza, da un esame delle colonne I, J, K, è possibile concludere come i posti letto siano relativamente l'*input* più scarso (con peso maggiore), mentre gli *input* relativamente sovrabbondanti sono i medici, dato che il modello ha assegnato pesi molto bassi, con valori spesso molto vicini a zero (indicati con zero, per semplicità nella tavola).

Nelle colonne L, M, N, O sono riportate le situazioni effettive delle u.o. esaminate, cioè il valore prodotto e le risorse a disposizione. Nelle successive colonne si indicano, invece, gli *input* virtuali, ovvero quelli necessari per riportare, l'u.o., rispettivamente sulla frontiera Vrs (colonne P, Q, R), oppure sulla frontiera Crs (colonne S, T, U). Le rispettive variazioni, ovviamente, si ottengono per differenza. Nelle successive quattro colonne (V, W, X, Y) si indica, infine, quale dovrebbe essere la configurazione delle singole u.o. per sfruttare appieno tutte le economie di scala senza incorrere in diseconomie, cioè per operare in una situazione di rendimenti di scala localmente costanti. Si noti che mentre nelle precedenti situazioni virtuali (Vrs e Crs) le modifiche riguardano soltanto gli *input*, a parità di *output* (questa è la caratteristica dei modelli *input-oriented*), le strategie di perseguimento della dimensione ottima (ridimensionamento o ampliamento) richiedono non soltanto variazioni negli *input*, ma anche nell'*output*.

La somma delle colonne indica l'effetto complessivo che le strategie suggerite dalla valutazione Dea prevedono a livello

della rete ospedaliera regionale. I 1744 letti attualmente disponibili dovrebbero ridursi a 1353 se si vuole recuperare efficienza puramente tecnica e a 1014 se si volessero dimensionare correttamente tutte le unità operative. In quest'ultimo caso, si deve anche programmare di ridurre la produzione complessiva, che dovrebbe passare da un valore di 270.4 a un valore di 246.8 miliardi.

Si noti, peraltro che non è sufficiente ridurre soltanto i posti letto, ma che è necessario ridurre anche gli altri *input* e che tali strategie sono differenziate a seconda delle singole situazioni delle u.o. In alcuni casi, per esempio, si deve prevedere un aumento di posti letto e in altri casi una riduzione. La differenza 1744-1014 è soltanto una somma algebrica. Si noti anche come le strategie volte a riportare le u.o. alla dimensione corretta possono anche essere conflittuali rispetto a quelle intese a riportare le u.o. semplicemente sulla frontiera Vrs o Crs, a parità di valore della produzione. Si veda, per esempio, la situazione della u.o. n. 23, in cui seguendo i punteggi Vrs e Crs gli esistenti 32 letti dovrebbero essere ridotti a 24 e a 15, rispettivamente, mentre la dimensione ottima si potrebbe raggiungere con 38 letti (cioè con un aumento rispetto agli attuali). In questo ultimo caso, tuttavia, il valore dei Drg dovrebbe passare da 3.173 a 8.236 (cioè più che raddoppiare), il che potrebbe condurre a situazioni inefficienti se si prevede che le prestazioni aggiuntive siano inappropriate.

I risultati della Tab. 2 possono essere facilmente interpretati con riferimento alla Fig. 2.²¹ Le tre unità operative efficienti sono del tipo B, cioè non richiedono alcun intervento correttivo. Ci sono poi 10 unità operative che sono efficienti soltanto dal punto di vista puramente tecnico. Queste u.o. con *W* uguale o molto vicino all'unità²² sono sulla frontiera Vrs: sei

²¹ Per la precisione, occorrerebbe disegnare tre grafici del tipo della Fig. 2, uno per ciascuna delle tre unità operative efficienti: si tratta, infatti, di tre tecniche produttive differenti, ciascuna caratterizzata da un proprio *mix* produttivo.

²² Si considerano efficienti anche le u.o. con *W* molto vicino all'unità: si tratta delle unità operative: 2, 4, 8, 24, 27, 31, 34, 35, 38, 40, che sono sulla frontiera Vrs.

(le u.o. 2, 4, 8, 24, 27, 31) sono sovradimensionate, ovvero si trovano in un tratto di rendimenti decrescenti di scala in posizione simile alla u.o. R della Fig. 2; quattro sono, invece, sottodimensionate e, pertanto, producono con rendimenti crescenti di scala come l'u.o. X della Fig. 2.

Si possono poi individuare quattro u.o. (13, 17, 25, 30) che sono inefficienti soltanto dal punto di vista puramente tecnico, ma che sono di dimensione corretta. Sono unità operative del tipo V sulla Fig. 2. Le rimanenti ventitre unità operative sono inefficienti sia dal punto di vista puramente tecnico, sia di scala. Possono essere suddivise in due sottogruppi: quattordici sono sovradimensionate (1, 3, 5, 6, 7, 9, 10, 11, 12, 16, 18, 21, 22, 28) del tipo P in Fig. 2 e richiedono, di conseguenza successive riduzioni di *input* (e anche di *output*) per riportarle alla Mpss. Le ultime nove u.o. (14, 15, 19, 20, 23, 29, 32, 37, 39) sono, invece, sottodimensionate del tipo X della Fig. 2 e, come si è già visto, prevedono strategie di recupero di efficienza che possono essere conflittuali.

4.2. *Confronto tra frontiere e analisi dei differenziali di efficienza*

Le considerazioni fin qui fatte sono quelle immediatamente possibili sulla base della valutazione Dea. È possibile, tuttavia, procedere nell'analisi per individuare in modo ancora più dettagliato le ragioni dei differenziali di efficienza riscontrati. Si tratta, cioè, di verificare in che misura l'efficienza delle unità produttive esaminate possa essere influenzata da ulteriori fattori che non sono omogenei all'interno del campione. Si fa riferimento alla influenza delle variabili cosiddette "esterne" perché, a differenza di quelle utilizzate finora, non sono sotto il controllo delle unità decisionali. Tra le variabili esterne si possono individuare sia variabili categoriche, sia variabili continue. Tra le prime, che indicano la presenza o l'assenza di una certa caratteristica, si possono immaginare, per esempio, il tipo di proprietà (pubblica o privata), l'appartenenza o no a una area metropolitana, la presenza di strutture didattiche, ecc. Le seconde potrebbero far riferimento ad alcune caratteristiche della popolazione, quali l'età, il reddito, e così via.

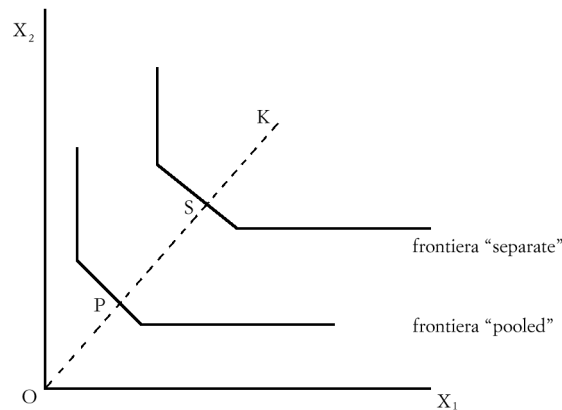
Per includere le variabili esterne nell'analisi dell'efficienza, esistono tre metodi principali.

Il primo consiste in una analisi a due stadi: si risolve il modello, come è stato fatto finora, includendo soltanto gli *input* e gli *output* tradizionali e, successivamente, si regrediscono i punteggi di efficienza ottenuti rispetto alle variabili esterne ritenute importanti. Questo metodo, oltre a richiedere che il campione sia sufficientemente numeroso per rendere significativa la regressione, è più adatto con variabili esterne non categoriche. Un secondo metodo consiste nella variante del modello Dea proposta da Banker e Morey [2] che permette di confrontare la performance di una u.o. soltanto con quella delle unità che, nello stesso campione, presentano un valore della variabile esterna minore o uguale. In questo modo si evita un confronto con unità oggettivamente differenti riguardo a una determinata variabile.

Qui si preferisce seguire un terzo metodo, relativamente più semplice [11] [12], che consiste nel calcolare l'efficienza della stessa unità operativa con riferimento sia al campione complessivo, sia a un sottocampione e poi nel confrontare i due punteggi così ottenuti. La logica sottostante è verificare se la frontiera identificata per il sottocampione (frontiera "separate") si scosta da quella del campione complessivo (frontiera "pooled"). In caso affermativo, ciò significa che le u.o. comprese nel sottocampione producono con una funzione di produzione sistematicamente diversa dalle altre e ciò viene interpretato come segnale della influenza della variabile esterna, cioè di una condizione non controllabile da parte del responsabile della unità operativa.

Anche in questo caso per comprendere meglio il significato della scomposizione, si propone un esempio grafico (Fig. 3), dove per la generica u.o. K , appartenente al sottocampione è possibile calcolare due punteggi di efficienza, in base a due frontiere di riferimento. Il metodo di calcolo consiste, come al solito, nel rapportare gli *input* che sarebbero sufficienti con quelli effettivamente utilizzati. Indicando con Eff_{pool} e Eff_{sep} , l'efficienza calcolata con riferimento rispettivamente alla frontiera *pooled* e a quella *separate*, risulta: $Eff_{pool} = OP/OK$, e $Eff_{sep} = OS/OK$.

FIG. 3 Identificazione di cause "esterne" di inefficienza



Poiché lo scopo è quello di identificare eventuali effetti di variabili esterne, ciò che conta è, soprattutto, lo scostamento tra le due frontiere, misurato dal rapporto fra i due punteggi, che identifica un nuovo punteggio di efficienza, la cosiddetta efficienza "between" (Effbet). Nel caso della figura in esame:

$$(6) \quad \text{Effbet} = \frac{\text{Effpool}}{\text{Effsep}} = \frac{\frac{OP}{OK}}{\frac{OS}{OK}} = \frac{OP}{OS}$$

Quanto maggiore è la distanza fra le due frontiere (e conseguentemente il valore di Effbet) tanto più è probabile che esistano fattori sistematici che rendono diverse le tecnologie utilizzate dal singolo sottocampione, oppure fattori esterni che ne rendono diversa sistematicamente l'efficienza, rispetto al campione complessivo. In conclusione, tanto minore è Effbet, tanto più le frontiere si scostano, cioè tanto più il sottocampione è inefficiente all'interno del campione generale (nel senso che mediamente usa più input delle altre u.o.). Se è uguale all'unità le due frontiere coincidono.

Questo terzo metodo è stato utilizzato per verificare all'interno del campione delle u.o. di medicina liguri l'eventuale influenza di tre variabili esterne: la localizzazione geografica, la struttura di incentivi di tipo organizzativo e la presenza di attività di insegnamento e di ricerca. La prima variabile esterna riflette l'estrema eterogeneità del territorio: la zona di ponente è vicino al confine e risente della presenza di strutture francesi piuttosto efficienti; l'estrema zona di levante è molto più vicina, dal punto di vista delle comunicazioni, a regioni piuttosto ricche e ben servite, quali la Toscana e l'Emilia-Romagna; la zona di Genova è densamente abitata e costituisce un buon esempio di area metropolitana, sulla quale insistono strutture di ricovero piuttosto differenziate (presidi, aziende, istituti scientifici, ecc.). Con riferimento soltanto ai presidi ospedalieri, si sono raggruppate le u.o. in tre sottocampioni "ponente", "levante" e "genova", per evidenziare l'eventuale presenza di differenziali sistematici di inefficienza.

La seconda variabile esterna si riferisce a una disaggregazione per tenere conto che le unità operative differiscono tra loro anche per una diversa struttura organizzativa. Esistono, infatti, nel campione delle unità degenziali esaminate, alcune unità operative che fanno parte di "aziende ospedaliere" e altre che, invece, appartengono a "presidi". Alcune unità operative, poi, oltre ad appartenere ad aziende, sono anche a conduzione universitaria. Come noto, le unità appartenenti alle aziende ospedaliere hanno un grado più accentuato di autonomia. In tal caso, infatti, risultano più chiaramente distinti il ruolo di produttore e di finanziatore dei servizi, mentre nel caso dei presidi, accade che i due ruoli siano accentrati nella stessa azienda, che agisce come acquirente per conto del paziente, ma al tempo stesso compra da se stessa, dato che i "presidi" sono articolazioni interne delle Asl.

La terza variabile esterna si riferisce, infine, alla individuazione di un ulteriore sottocampione, quello delle unità operative universitarie all'interno di una azienda ospedaliera, dove coesistono con unità ospedaliere.

Si tratta in questo caso di verificare se svolgere assistenza con scopi anche istituzionali e di ricerca modifica in modo strutturale il processo produttivo.

Tab. 3 L'effetto delle variabili esterne

| | PONENTE | LEVANTE | GENOVA | PRESIDI | AZIENDE | UNIVERSITÀ |
|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| <i>Pooled</i> | 0,640 (0,108) | 0,559 (0,136) | 0,602 (0,097) | 0,599 (0,115) | 0,687 (0,197) | 0,666 (0,221) |
| <i>Separate</i> | 0,908 (0,122) | 0,797 (0,184) | 0,929 (0,122) | 0,840 (0,153) | 0,687 (0,197) | 0,666 (0,221) |
| <i>Between</i> | 0,703 (0,044) | 0,703 (0,047) | 0,647 (0,046) | 0,713 (0,035) | 1,000 (0,000) | 1,000 (0,001) |

Nella Tab. 3 sono riportate per ciascun sottocampione e per ciascun tipo di efficienza (*pooled*, *separate* e *between*) le medie dei punteggi e le corrispondenti deviazioni *standard*. Nella prima riga, le medie si riferiscono all'efficienza *pooled*, calcolata cioè con riferimento al campione complessivo delle 40 unità operative (i cui punteggi dettagliati sono quelli riportati nella Tab. 2). Con riferimento a questi punteggi di efficienza, fra loro confrontabili, perché riferiti alla stessa frontiera produttiva risulta che le aziende sono mediamente più efficienti rispetto ai presidi e che all'interno delle aziende il fatto di avere una direzione universitaria lascia sostanzialmente invariato il punteggio di inefficienza (0,666 rispetto a 0,687).²³ Per quanto riguarda poi l'effetto localizzazione all'interno dei presidi, quelli del levante sono decisamente i meno efficienti (0,559).

Nella seconda riga sono indicate le medie dei punteggi delle u.o. valutate ciascuna all'interno del proprio sottocampione (*Effsep*). I punteggi compresi in questa riga non sono, pertanto, tra loro confrontabili, perché si riferiscono a frontiere diverse. Sono, invece, sostanzialmente omogenei²⁴ all'in-

²³ Per rendere più corretto il confronto occorrerebbe inserire tra gli *output* anche la formazione e la ricerca. In tale caso, tuttavia, andrebbero modificati anche gli *input* di lavoro medico, che nella presente analisi sono stati considerati, per gli universitari pari al 50% del tempo di servizio effettivo. È, pertanto, probabile, che il risultato non cambierebbe molto.

²⁴ Il confronto è soltanto in parte corretto, perché essendo il Dea una metodologia del tipo *best practice*, il valore del punteggio di efficienza dipende anche dalla numerosità del campione. In particolare risulta che minore è la numerosità del campione, maggiore è, a parità di altre condizioni, il punteggio di efficienza di una unità operativa.

terno di ciascuna colonna, perché calcolati per lo stesso gruppo di u.o. Leggendo dall'alto verso il basso, si ha una idea di come l'efficienza di ciascuna u.o. migliori quando viene valutata all'interno del proprio sottocampione, cioè con u.o. molto più simili, invece che con riferimento a quello complessivo. Il miglioramento più marcato riguarda le u.o. genovesi, che partendo da un punteggio piuttosto basso (0,602) raggiungono una efficienza vicina all'unità (0,929).

L'efficienza *between* riportata nella terza riga quantifica questo eventuale miglioramento o, in altri termini, la distanza fra le due frontiere, che sta a indicare che esistono cause sistematiche di inefficienza, legate a quel gruppo e di cui le u.o. non sono responsabili. Nel caso in esame, questo scostamento è evidente per i presidi (sia che li si consideri congiuntamente, sia suddividisi territorialmente) rispetto alle aziende. Con riferimento alla quarta colonna, sottocampione di tutti i presidi, la grave inefficienza mediamente riscontrata (Effpool = 0,559) indica che queste u.o. utilizzano quasi il doppio degli *input* che sarebbero necessari: ciò, tuttavia, non è imputabile soltanto a una *performance* inefficiente delle singole strutture, ma in parte anche al fatto di essere organizzati come presidi all'interno di una Asl. Se si scorpora questa causa di inefficienza (di tipo sistematico), l'efficienza di tali u.o. aumenta, diventando in media pari a 0,840.

L'efficienza *between* è uguale all'unità nelle ultime due colonne, perché ciascuna unità operativa presenta lo stesso punteggio sia che essa sia valutata con riferimento al campione complessivo o al suo gruppo. Ciò accade perché le *peer units* del campione complessivo appartengono proprio a questi sottocampioni.²⁵ In altri termini, dato che le frontiere dei sottocampioni "aziende" e "università" coincidono con quella del campione complessivo, il fatto di essere azienda o di essere u.o. a direzione universitaria non comporta in sé alcun motivo di inefficienza.

²⁵ Come si può dedurre dal valore piuttosto elevato della deviazione standard, si tratta, peraltro, di gruppi in cui esiste una certa variabilità dell'efficienza: unità operative efficienti appartenenti alla frontiera coesistono con altre che presentano punteggi molto bassi (cfr. anche Tab. 2).

5. *Discussione*

Per valutare l'efficienza tecnica e di scala delle unità operative ospedaliere, si è proposto il modello *Data Envelopment Analysis*, che, pur basandosi sulle usuali ipotesi della funzione di produzione neoclassica, risulta più flessibile delle tradizionali stime econometriche. Esso fornisce, inoltre, informazioni più precise sulla entità e sulle ragioni degli scostamenti delle unità produttive valutate rispetto alla frontiera efficiente di produzione. Tale modello è stato applicato all'attività delle unità degenziali di medicina della Regione Liguria nel 1997. Per la prima volta, a quanto risulta, la disaggregazione è stata spinta fino alle singole unità operative e alla singola specializzazione produttiva, consentendo così un confronto corretto tra strutture molto omogenee, con compiti produttivi simili, che hanno accesso alla stessa tecnologia produttiva e operano sullo stesso territorio.

I risultati raggiunti devono, comunque, essere letti con molta cautela per due motivi. Il primo dipende dalle caratteristiche del modello: i punteggi Dea sono, infatti, molto sensibili alla qualità dei dati utilizzati e alle specificazioni prescelte degli *output* e degli *input* [7]. Si tratta, inoltre, di un modello deterministico che non prevede un termine di errore e quindi i risultati risentono in misura notevole di eventuali *input* o *output* omissi [8]. Rispetto ai modelli stocastici, d'altra parte, il modello Dea ha una maggiore capacità informativa e si può applicare anche a campioni non particolarmente numerosi.²⁶ Un ulteriore vantaggio è che la frontiera è del tipo *best practice*, cioè, propone come *standard* realtà produttive che stanno già agendo al meglio nello stesso contesto operativo.

In secondo luogo, una certa cautela è necessaria perché i risultati ottenuti devono essere considerati soltanto una prima informazione grezza da completare con ulteriori approfondimenti. In particolare, per pervenire alla individuazione del fabbisogno a livello regionale, l'analisi effettuata, che riguarda esclusivamente il lato dell'offerta, dovrebbe essere accompa-

²⁶ È infatti, sufficiente che il numero delle u.o. sia pari a due-tre volte la somma degli *input* e degli *output*.

gnata da quella della domanda e dell'appropriatezza delle scelte cliniche. Si tratta, peraltro, di integrazioni non facili da effettuare.

Per quanto riguarda la domanda, l'unica possibilità concreta per una sua valutazione è quella di basarsi sulle scelte effettuate dai pazienti nel passato, cioè di tenere conto della domanda effettiva e delle matrici di mobilità. Questa analisi dovrebbe comunque essere condotta "a parità di condizioni" e, in particolare, a parità di gravità dei pazienti che si presentano alla struttura stessa. Ciò può essere fatto, per esempio, definendo una popolazione *standard*, ovvero una struttura di utenza identica per tutte le strutture con cui ponderare i diversi indicatori relativi alla popolazione dei diversi campioni. Soltanto con questa correzione è possibile utilizzare i punteggi di efficienza all'interno di schemi incentivanti per premiare le strutture produttive più efficienti.²⁷

Per quanto riguarda l'appropriatezza, esistono già alcuni indicatori ufficiali, quali, ad esempio, i "ricoveri di un giorno" e i "ricoveri oltre la soglia"²⁸, ma è evidente che essi non sono sufficienti, perché non è automaticamente assicurata la corrispondenza fra prestazione e risultato in termini di salute. Occorrono, di conseguenza, procedure valutative *ad hoc*. In altri termini, all'interno di ogni ospedale, le valutazioni di efficien-

²⁷ È evidente, cioè, come la pressione di domanda di per sé non sia sufficiente per rivelare la qualità delle prestazioni fornite, dato che i servizi erogati hanno sostanzialmente la natura di *experience goods*. Potrebbe comunque essere interessante verificare se un eccesso di domanda (rilevabile, per esempio, dalla consistenza della lista di attesa) induce un aumento o una diminuzione dell'efficienza produttiva.

²⁸ Un numero eccessivo di ricoveri di un giorno è indice di un fenomeno anomalo, che dovrebbe essere corretto sostituendo tali ricoveri con interventi alternativi, per esempio ricorrendo al *day hospital*. Nel caso dei ricoveri oltre la soglia (cosiddetti *outliers*) si tratta di quei ricoveri che superano la soglia di degenza prevista dal tariffario nazionale Drg. È un indice che, soprattutto nell'area medica, potrebbe evidenziare l'assorbimento da parte degli ospedali della quota significativa di lungodegenza che il servizio sanitario regionale, attraverso la riorganizzazione della funzione ospedaliera, sta attualmente cercando di trasferire all'area dei servizi territoriali. Per evidenziare meglio questo fenomeno, è più opportuno calcolare tale indicatore soltanto con riferimento alla popolazione sopra i 65 anni di età.

za, di competenza della direzione generale (o della regione), dovrebbero integrarsi con quelle di efficacia e di appropriatezza, a livello di dipartimento o di area funzionale, basate su linee guida, protocolli, *auditing*, ecc.

Pur in mancanza di un'analisi completa la disponibilità di misure in grado di descrivere e quantificare nei dettagli le cause di inefficienza disaggregate per specialità costituisce, comunque, un primo strumento indispensabile per stimolare verso comportamenti efficienti. Come si è detto, la metodologia proposta non serve per fornire indicazioni operative alla singole strutture, ma a livello di sistema per ridurre lo svantaggio informativo di cui soffre il regolatore-finanziatore (Regione) con riguardo alle condizioni di erogazione del servizio, conosciute, invece, dai singoli agenti (responsabili di unità operativa o di ospedale). In tale situazione, con un sistema di pagamento prospettivo, quale quello attuale dei cosiddetti Drg o *Diagnosis Related Groups*²⁹, ci si deve necessariamente attendere che gli ospedali (o le singole u.o.) mettano in atto tutta una serie di comportamenti impropri³⁰ per massimizzare le proprie entrate e/o le proprie dotazioni di posti letto e personale, entrando così in conflitto con gli obiettivi del loro principale (Regione).

Le possibili soluzioni, suggerite dall'analisi economica, sono quelle di:

- controllare l'operato degli agenti;
- collegare la loro retribuzione ai risultati raggiunti.

Si tratta di provvedimenti particolarmente difficili da implementare in ambito sanitario, dei quali, tuttavia, il modello Dea è in qualche modo, capace di garantire la fattibilità concreta.

In primo luogo, la conoscenza dei punteggi di efficienza permette di risparmiare sull'attività di controllo, che può, addirittura, diventare inutile. Come insegna, infatti, la cosiddetta teoria dei "tornei" [3], il confronto fra i risultati ottenuti

²⁹ Per una analisi approfondita al riguardo si veda Vanara [13].

³⁰ Si parla in questo caso di "azzardo morale". Con tale termine si indica il comportamento opportunistico messo in atto dal singolo agente, che approfitta del suo vantaggio informativo per massimizzare la propria utilità piuttosto che quella del principale su delega del quale dovrebbe agire.

da strutture degenziali nelle stesse condizioni operative permette di scorporare dalla valutazione gli effetti di stati di natura sfavorevoli: tali circostanze negative sono, infatti, comuni a tutte le u.o. e le differenze di risultato sono dovute unicamente ai diversi comportamenti delle unità operative interessate.

In secondo luogo, le misure di efficienza sono esse stesse indicatori utili per strutturare lo schema retributivo complessivo degli agenti (premi, incentivi, penalità, riallocazione di risorse, ecc.) in modo tale da indurre implicitamente l'agente ad erogare il massimo sforzo produttivo. In questo senso le valutazioni Dea appaiono particolarmente utili, perché non sono soltanto misure ordinali con le quali costruire classi di merito (eccellente, media, ecc.) ma possono facilmente interpretarsi anche come misure cardinali di efficienza [7]. I punteggi di efficienza ottenuti potrebbero, di conseguenza, servire direttamente per redistribuire le risorse (posti letto, personale, *budget* del dipartimento), per inserire incentivi o penalità nelle retribuzioni, per correggere la remunerazione per tariffe della struttura in modo da tenere conto anche della sua efficienza, e così via.

I risultati della valutazione evidenziano comportamenti nel complesso piuttosto inefficienti e, di conseguenza, la possibilità di un notevole risparmio di risorse.

Una parte dell'inefficienza riscontrata non è ascrivibile al comportamento dei responsabili delle u.o., ma piuttosto a una non corretta programmazione regionale o aziendale. Le situazioni sono piuttosto differenziate tra loro e sembrano confermare quanto ci si potrebbe aspettare a livello teorico, ovvero che nella pubblica amministrazione si verifichi una moltiplicazione impropria di u.o., con la conseguenza che ciascuna di loro è di dimensione troppo piccola. Ciò è vero soltanto per le unità operative universitarie, mentre, al contrario, nel caso dei presidi ospedalieri, di solito, l'inefficienza deriva da un sovradimensionamento.

Un'altra parte dell'inefficienza risulta, invece, dipendere, dal fatto che lo sforzo erogato non è ottimale: si tratta di quei casi in cui opportuni incentivi potrebbero indurre una modifica dei comportamenti, per eliminare l'area "inerte" [6] derivante dall'opportunismo individuale. Sono, inoltre, le si-

tuazioni in cui occorrerebbe un miglioramento dell'aspetto organizzativo anche a livello aziendale.

Una ulteriore quota di inefficienza deriva, infine, dalla presenza di cause esterne, che non sono sotto il controllo dei responsabili delle u.o. Tra quelle prese in considerazione (localizzazione, compiti didattici e struttura organizzativa), il fattore che sistematicamente sembra comportare un uso più elevato di risorse (maggiore inefficienza) è l'appartenenza a un presidio ospedaliero. Le unità operative inserite in aziende ospedaliere risultano, infatti, mediamente più efficienti rispetto a quelle che fanno parte di aziende sanitarie territoriali. Questo risultato, se confermato da ulteriori valutazioni, significa che molto probabilmente gli incentivi all'efficienza presenti in soluzioni regolamentative del tipo "mercati "interni" sono più potenti di quelli impliciti in soluzioni del tipo "managed competition". In altre parole sembrerebbe preferibile un modello di tipo contrattuale con una netta separazione fra produttore e acquirente dei servizi. Si avrebbe, cioè, una indicazione empirica a proseguire verso la deintegrazione del sistema, così come previsto dai decreti 502/517, che appare la via più corretta per recuperare efficienza. Rimane, ovviamente, aperta la questione di come il modello contrattuale possa condurre a soluzioni che siano anche efficaci e appropriate.

Riferimenti bibliografici

- 1) Banker R. D., Charnes A., Cooper W. W.
1984 "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, 30: 1078-92.
- 2) Banker R. D., Morey R. C.
1986 "Efficiency analysis for exogenously fixed inputs and outputs", *Operational Research*, 34: 513-521.
- 3) Brosio G.
1995 *Introduzione all'economia dell'organizzazione*, La Nuova Italia Scientifica, Roma.

- 4) Charnes A., Cooper W. W., Rhodes E.
1978 "Measuring the Efficiency of Decision Making Units",
European Journal of Operational Research, 12 (2): 429-444.
- 5) Coelli T., Rao D., Battese G.
1998 *An introduction to efficiency and productivity analysis*,
Kluwer, Boston.
- 6) Leibenstein H.
1992 "Empirical Estimation and Partitioning of X-Inefficiency:
A Data-Envelopment Approach", *American Economic
Review, Papers and Proceedings*, 82 (2): 428-433.
- 7) Magnussen J.
1996 "Efficiency Measurement and the Operationalization
of Hospital Production", *Health Services Research*,
31(1): 21:37.
- 8) Newhouse J. P.
1994 "Frontier estimation: How useful a tool for health eco-
nomics?", *Journal of Health Economics*, 13: 317-322.
- 9) Testi A.
1996 "La valutazione non parametrica dell'efficienza nella
produzione di ricoveri ospedalieri: aspetti metodologici
e applicazione empirica", *Economia Pubblica*, XXXVI
(3): 81-121.
- 10) Testi A.
1997 "Public versus Private Hospitals: an Assessment of
Technical Efficiency", *Ricerca Operativa*, 27(81-82):
121-137.
- 11) Testi A.
1997 "Ranking Efficient Decision Making Units: an applica-
tion to Hospital Departments", *Discussion Paper* n.9,
ottobre '97, Istituto di Economia, Genova.
- 12) Valdmanis V.
1992 "Sensitivity Analysis for DEA Models. An Empirical
Example Using Public vs. NFP Hospitals", *Journal of
Public Economics*, 48 (2): 185-205.

- 13) Vanara F.
1996 “Sistemi di pagamento e reazioni delle strutture di offerta: indicazioni della letteratura teorica”, *Mecosan*, 19: 116-122.
- 14) Zweifel P., Breyer F.
1997 *Health Economics*, Oxford University Press.