

CRENOS
CENTRO RICERCHE
ECONOMICHE NORD SUD
Università di Cagliari
Università di Sassari

**L'EFFICIENZA TECNICA DEI DISTRETTI DI CORTE
D'APPELLO ITALIANI: ASPETTI METODOLOGICI,
BENCHMARKING E ARRETRATO SMALTIBILE**

Riccardo Marselli

Marco Vannini

WORKING PAPERS



2004/09

Riccardo Marselli
Università degli Studi di Napoli Parthenope
marselli@uniparthenope.it

Marco Vannini
Università degli Studi di Sassari e CRENoS
vannini@uniss.it

L'EFFICIENZA TECNICA DEI DISTRETTI DI CORTE D'APPELLO ITALIANI: ASPETTI METODOLOGICI, *BENCHMARKING* E ARRETRATO SMALTIBILE*

Riassunto

L'efficienza tecnica dei distretti di corte di appello italiani viene misurata mediante il metodo DEA (*Data Envelopment Analysis*). L'analisi consente di valutare l'efficienza relativa di ciascun distretto come distanza radiale dalla frontiera efficiente, definita dalle prassi migliori, e di individuare le variazioni dei fattori necessarie affinché un distretto inefficiente emuli il proprio *benchmark* (distretto o gruppo di distretti di riferimento collocati sulla frontiera). Viene inoltre effettuata la scomposizione dell'efficienza totale in efficienza tecnica pura ed efficienza di scala e, per ciascun distretto inefficiente, viene calcolato l'arretrato smaltibile. Infine, sulla base di un'analisi multivariata di tipo tobit, si stima l'influenza sull'indicatore di efficienza di una serie di fattori socio-economici del contesto territoriale.

Agosto 2004

* Sebbene frutto di una riflessione comune, i primi tre paragrafi del lavoro sono attribuibili a Marco Vannini ed i rimanenti tre a Riccardo Marselli. Gli autori desiderano ringraziare Daniele Fabbri per i commenti a una precedente versione. Il lavoro è parte dell'attività di ricerca condotta nell'ambito del progetto PRIN2001 su "L'amministrazione della giustizia e la società italiana del 2000"

1. Introduzione

Mentre si infittiscono gli interrogativi e i dibattiti sull'inefficienza della giustizia in Italia, ed episodicamente si adottano provvedimenti – anche importanti - intesi a farvi fronte, il fenomeno non sembra regredire. Com'è facile constatare sfogliando gli interventi ufficiali in occasione dell'inaugurazione degli ultimi anni giudiziari, quello dell'efficienza rappresenta il problema prioritario. Nella relazione più recente del Procuratore Generale presso la Corte Suprema di Cassazione si legge “La giustizia è tuttavia, innegabilmente, ancora in crisi, soprattutto a causa della sua scarsa efficienza e della durata eccessiva dei processi. Il recupero dell'efficienza è l'esigenza fondamentale” (Ministero della Giustizia, 2004a, p. 3). Gli fa eco il Guardasigilli, che dopo aver ricordato il primato dell'Italia come paese più condannato dalla Corte Europea dei diritti dell'uomo per l'eccessiva durata dei processi (aumentata in media più del doppio tanto nel civile quanto nel penale dal 1960 ad oggi) vede il problema principale - subordinato solo al rapporto tra potere legislativo e potere giudiziario – nella lentezza della giustizia, per combattere la quale “ci si deve votare alla cultura dell'efficienza, intesa non in senso lato, ma nella sua accezione strettamente tecnica e scientifica, vale a dire al quoziente tra risultati raggiunti e risorse impiegate” (Ministero della Giustizia, 2004b, p. 4). Risalendo appena indietro, peraltro, devono essere ricordate le ampie relazioni del Consiglio Superiore della Magistratura sul tema generale della crisi della giustizia, e in particolare quella inviata al Parlamento nell'ottobre 2001 (ex art. 28 del R. I., approvata con ventisei voti favorevoli, nessun contrario, quattro astenuti) interamente dedicata alla tutela dei diritti, efficacia e tempi della giurisdizione (CSM, 2001). La relazione muove esplicitamente dalla premessa elementare che “la modernità di un ordinamento dipende non solo dai principi più volte proclamati ma insieme, e soprattutto, dalle strutture e dalla intelligente allocazione delle risorse; sicché nella formulazione delle norme e delle soluzioni interpretative non si può prescindere dal cosiddetto impatto normativo, cioè da tutte le ricadute che inevitabilmente le riforme legislative hanno sulla organizzazione e sulla efficienza del servizio giustizia”. (p. 11).

Come si diceva, negli ultimi anni non sono certo mancati provvedimenti di rilievo, che hanno inciso tanto sulla struttura del sistema giudiziario quanto sulle regole del processo, con effetti importanti, diretti e indiretti, sull'architettura e sulla funzionalità degli

uffici giudiziari. Basti pensare alla riforma del processo civile (legge 26 novembre 1990, n. 353, entrata in vigore nel maggio del 1995), all'istituzione del giudice di pace (legge 21 novembre 1991, n. 374, entrata in vigore nel 1995) e delle sezioni stralcio per i procedimenti basati sul "vecchio rito" (legge 22 luglio 1997, n. 276 e successive modificazioni), alla creazione della figura del "giudice unico di primo grado" (legge 16 luglio 1997, n. 254) con la conseguente unificazione degli uffici di pretura e di tribunale, all'istituzione dei tribunali metropolitani (d.l. 3 dicembre 1999, n. 155), all'inserimento in Costituzione (art. 111) dei principi del giusto processo (legge costituzionale, 23 novembre 1999, n. 2) e alle conseguenti riforme processuali. Malgrado ciò, i segnali positivi sono piuttosto timidi, e non autorizzano certo ad affermare che la giustizia in Italia abbia imboccato stabilmente un sentiero virtuoso.

Secondo i dati esposti dal Procuratore Generale (Ministero della Giustizia, 2004a), nella giustizia civile l'ulteriore aumento delle nuove cause ha sostanzialmente compensato la relativa maggiore produttività del sistema nel primo grado di giudizio, col risultato di una modesta flessione delle pendenze. Ad un esame più approfondito, quest'ultima scaturisce da un andamento opposto delle pendenze presso i Giudici di Pace (in crescita, con conseguente lieve aumento della durata dei processi) e i Tribunali (in diminuzione, con conseguente riduzione della durata media dei processi). Se da un lato si intravede, in relazione al complesso dell'attività processuale dei Tribunali, una positiva inversione di tendenza, dall'altro lato occorre sottolineare che solo per il 58% dei processi definiti con sentenza il verdetto arriva entro i tre anni (p. 7). Nel grado di appello si osserva un lieve incremento delle pendenze indotto da un notevole aumento delle sopravvenienze cui si contrappone un modesto aumento dei processi esauriti. Anche qui, guardando ai due segmenti, i Tribunali¹ e le Corti d'Appello, per i primi si registra una diminuzione della durata media dei processi, con valori che restano tuttavia estremamente elevati (1.013 giorni), mentre per le seconde, i cui valori sono mediamente inferiori (774 giorni), si assiste a un ulteriore aumento. La giustizia civile, osserva il Procuratore Generale (p. 9) "è ancora in crisi, specie per il fatto che non è in grado di affrontare il

¹ Attualmente i Tribunali sono competenti per l'appello contro le sentenze del giudice di pace, mentre continuano a trattare -fino ad esaurimento- le cause d'appello in materia di lavoro e previdenza iscritte prima della riforma del giudice unico.

problema del progressivo aumento dei nuovi procedimenti, avendo scarsi margini di recupero in punto di produttività, soprattutto negli uffici giudiziari più importanti e non essendo perciò in grado di avviare, con l'attuale apparato normativo e organizzativo, un sia pur lento e graduale miglioramento dell'efficienza, che richiederebbe non solo di eliminare progressivamente il lavoro arretrato, destinato a produrre la durata eccessiva dei processi, ma anche di fronteggiare a pieno la sopravvenienza dei nuovi processi”.

Il quadro relativo alla giustizia penale è caratterizzato dalla consistente contrazione dei casi pendenti e dei nuovi casi accompagnato da una riduzione preoccupante nel numero dei procedimenti definiti. In altre parole, la riduzione del carico è imputabile in misura maggiore alle minori sopravvenienze piuttosto che alla produttività dei magistrati. Quest'ultima peraltro non è diminuita, ma nel contempo, osserva il Procuratore (p. 19), si è andata affermando la generale tendenza a protrarre la definizione dei processi, facendo uso di ogni possibile rimedio e ricorrendo a impugnative reiterate. Non stupisce, dunque, che l'aumento della durata media dei processi non accenni ad arrestarsi e investa ormai tutte le fasi successive alle indagini preliminari: “davanti al GIP è passata da 251 a 324 giorni; davanti al tribunale da 321 a 341 giorni e in corte d'appello da 495 a 543 giorni” (p. 20). Si confermano allora le preoccupazioni del CSM (2001) che, rilevando la tendenza del legislatore ad intervenire in modo frammentario e settoriale su singole materie o segmenti procedurali, segnalava i rischi della “coesistenza di modelli processuali tra loro non conciliabili e la sovrapposizione di garanzie proprie del processo inquisitorio ad un sistema accentuatamente accusatorio, così rendendo sempre più farraginoso e lento l'andamento del processo nel suo complesso” (p.37).

Come uscire da questa situazione di stallo? Quale tipo di riscontri, oltre a quelli esistenti, potrebbe aiutare la giustizia italiana ad imboccare la strada giusta?

Il contributo di questo lavoro consiste essenzialmente nell'approfondire una questione tanto semplice quanto trascurata, ma che può essere di grande aiuto nel calibrare le politiche rivolte al settore. In pratica ci chiediamo in che misura i risultati raggiunti in questi anni siano attribuibili a differenze nell'efficienza dei distretti giudiziari. La risposta viene fornita tramite il metodo DEA (*Data Envelopment Analysis*), un metodo di analisi non parametrico particolarmente versatile per lo studio dell'efficienza delle unità produttive operanti nella sfera pubblica,

sempre più utilizzato per l'analisi dei servizi giudiziari e di polizia, ma mai applicato prima d'ora al caso italiano. Con questa metodologia riusciamo non solo a localizzare e quantificare le sacche di inefficienza esistenti ma anche ad individuare le principali modalità per eliminarle. Inoltre, poiché l'inefficienza di ogni singolo distretto viene misurata in rapporto al proprio *benchmark*, ovvero a quel distretto che con la medesima combinazione di fattori riesce a ottenere risultati migliori (o, alternativamente, a quel distretto che ottiene gli stessi risultati con un minor impiego di fattori), siamo in grado di stimare l'entità dell'arretrato smaltibile nell'ipotesi che le unità inefficienti adottino la prassi migliore. Può darsi che tutto ciò non risulti sufficiente a produrre un'apprezzabile riduzione dell'arretrato. Ma in ogni caso, alla fine dell'esercizio, siamo in grado di valutare con maggior cognizione tanto le richieste di risorse aggiuntive da parte dei responsabili di molti uffici quanto l'importanza di altre componenti, come la scala produttiva o certe caratteristiche del contesto ambientale, che potrebbero giustificare sia una revisione delle ripartizioni geografiche tradizionali sia una maggior enfasi sulle politiche deflative incentrate sulla domanda. Ulteriori interessanti spunti potrebbero inoltre emergere con l'applicazione della stessa metodologia a sottoinsiemi più ampi ed omogenei (es. Giudici di Pace, Tribunali) della giurisdizione.

Per agevolare la lettura da parte dei non specialisti, il lavoro si apre con due sezioni dedicate ai problemi di definizione/misurazione dell'efficienza produttiva e alle applicazioni del metodo *DEA* nel campo dei servizi di giustizia. Seguono due paragrafi dedicati interamente all'indagine empirica, nei quali si illustrano i dati utilizzati, il tipo di elaborazioni effettuate e i principali risultati ottenuti (*benchmarking* dei distretti e calcolo dell'arretrato smaltibile). Infine, nel paragrafo conclusivo, si traggono le implicazioni per le politiche di intervento nel settore.

2. Misurazione dell'efficienza produttiva

Nel descrivere le prestazioni (la *performance*) di un'unità produttiva si è soliti usare, talvolta come sinonimi e quasi mai dandone una definizione esplicita, espressioni quali "produttività", "efficienza", "efficacia". Se è vero che non esiste una definizione oggettiva e/o universalmente accettata per nessuno di essi, è tuttavia raccomandabile chiarire il senso in cui vengono usati. Nel presente lavoro, in sintonia con gran parte della letteratura economica ed ingegneristica, riserviamo al

termine “performance” il ruolo di espressione generale volta a cogliere una vasta gamma di risultati ottenibili da un’unità produttiva. Il termine produttività invece ha un significato molto più specifico, riguardante il rapporto fra l’output e l’input di un processo produttivo. Il calcolo di tale rapporto è immediato nel caso di un singolo output e di un singolo input, mentre diventa più complicato quando l’unità produttiva utilizza una varietà di input per ottenere più di un output; in ogni caso, attraverso opportuni criteri di aggregazione dei diversi elementi alla fine si perviene sempre al rapporto fra due numeri. La produttività è chiaramente un concetto relativo, e poiché esiste discrezionalità nella scelta dei termini da porre al numeratore e al denominatore si possono definire misure di produttività parziale (es. numero di pezzi per addetto) o totale (output in relazione all’insieme dei fattori impiegati)². In generale la produttività varia per tre ragioni: differenze nella tecnologia di produzione, differenze nell’efficienza del processo produttivo, differenze nel contesto entro cui si svolge la produzione (Knox Lovell, 1993, p. 3). L’efficienza dunque è quella componente della produttività che riflette, nell’ambito di una tecnologia di riferimento, quanto bene sono utilizzate le risorse disponibili. Al contrario l’efficacia rappresenta una nozione più sfumata e meno facilmente definibile sul piano quantitativo, che coglie la capacità di una unità produttiva di assicurare col suo output il conseguimento degli obiettivi assegnati³.

2.1. Componenti dell’efficienza produttiva

Vogliamo adesso definire con più precisione la nozione di efficienza produttiva per passare successivamente ai problemi di misurazione del fenomeno col metodo DEA. In riferimento a un’unità decisionale (UD) che trasforma un insieme di input (risorse produttive) in un insieme di output (risultati rispetto ai quali si intende valutare l’unità decisionale),

² L’impiego di misure totali di produttività, benché più complesse da calcolare, è di fondamentale importanza per evitare di attribuire erroneamente guadagni o perdite di produttività ad un fattore piuttosto che ad un altro. Un aumento di produttività risultante da miglioramenti organizzativi potrebbe erroneamente essere attribuito al fattore lavoro se si utilizza esclusivamente una misura parziale come il rapporto output/addetti, e viceversa. In generale, dati i rapporti di complementarità/sostituibilità fra input, tutte le misure parziali di produttività soffrono di questa limitazione.

³ Ad esempio, nel campo della sanità, la capacità di un ospedale di soddisfare la domanda d’interventi chirurgici urgenti. Nella letteratura anglosassone *efficiency* (efficienza) ed *effectiveness* (efficacia) vengono descritte rispettivamente con l’espressione “doing the things right” (fare le cose correttamente) e “doing the right things” (fare le cose giuste).

possiamo definire due diversi concetti di efficienza: efficienza tecnica (ET) ed efficienza allocativa (EA).

- *efficienza tecnica*: attitudine dell'unità decisionale, data la tecnologia corrente, a evitare sprechi, ovvero a produrre il massimo livello di output a partire da una data combinazione di input o, alternativamente, a impiegare la minore quantità possibile di input per ottenere un dato output.
- *efficienza allocativa*: capacità di un'unità efficiente in senso tecnico di produrre un dato output al costo minimo, ossia di combinare in modo ottimale gli input disponibili in funzione dei loro prezzi e delle produttività marginali così da ottenere il livello desiderato di output.

Notare differenze e affinità fra i due concetti. Entrambi sono definiti sullo sfondo di una tecnologia di riferimento prefissata (vedremo più avanti come individuarla, ma in ogni caso l'inefficienza va intesa come distanza da una frontiera ideale). Tuttavia, mentre ET riflette la capacità di trasformare grandezze fisiche e dipende verosimilmente dagli assetti organizzativi e ingegneristici della produzione nonché dalla scala produttiva, EA riflette la capacità di una unità già efficiente in senso tecnico (che non spreca risorse fisiche) di non sprecare denaro (cioè, dati i prezzi dei fattori, di produrre al costo minimo).

Dicevamo che per misurare il grado di efficienza di una UD è necessario disporre di un termine di riferimento. In linea teorica questo è rappresentato dalla frontiera dell'insieme delle possibilità di produzione, detta anche frontiera efficiente. Denotiamo con x il vettore degli input, con y il vettore degli output e con P l'insieme delle possibilità di produzione. Quest'ultimo viene definito come l'insieme di tutte le possibili combinazioni di x e y , ossia di tutte le ricette, o scelte tecniche realizzabili, per ottenere y a partire da x . Formalmente $P = \{(x, y): x \text{ produce } y\}$.

Nel caso di un processo produttivo elementare con un solo input e un solo output, l'insieme delle possibilità di produzione e la sua frontiera efficiente (o funzione di produzione) sono rappresentabili come nella Fig. 1. Più precisamente, l'insieme P coincide con l'area compresa fra l'asse orizzontale e la linea Oe (inclusi i punti sulla linea), mentre la frontiera efficiente è rappresentata dalla sola curva Oe . In altre parole, la frontiera efficiente è il luogo geometrico costituito dai livelli massimi di output ottenibile impiegando un dato livello di input. Si noti che la forma della curva varia a seconda delle ipotesi circa i rendimenti di scala della

produzione, cioè del modo in cui aumenta il prodotto quando aumentiamo l'impiego di tutti gli input in una data misura. Nel caso illustrato, con un solo fattore, poiché muovendosi lungo la curva al crescere di x si osserva un aumento meno che proporzionale di y , i rendimenti di scala sono decrescenti; se invece entrambi aumentassero nella stessa proporzione, come avviene lungo una retta, allora si parlerebbe di rendimenti di scala costanti (va da sé che nel caso di aumenti dell'output più che proporzionali rispetto all'aumento di tutti gli input si parlerà di rendimenti crescenti).

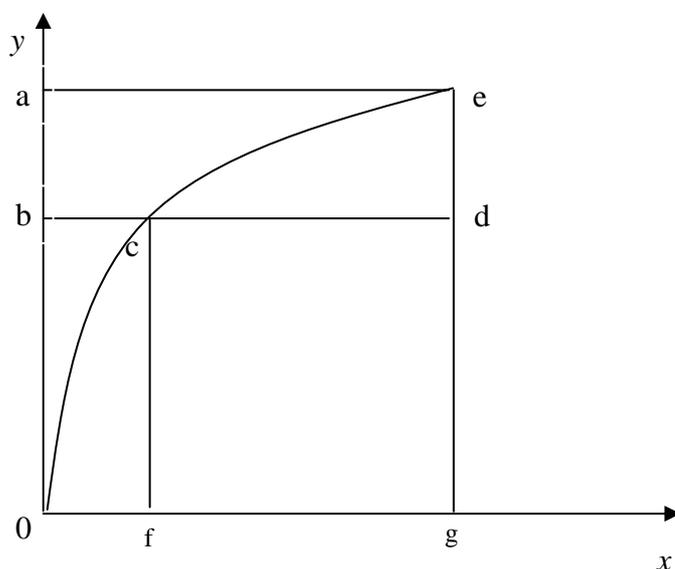


Figura 1 – Frontiera della produzione

Prima di passare alle definizioni formali, occorre sottolineare che poiché in pratica l'efficienza di una UD dipende dalla capacità di quest'ultima di assicurare un output più elevato a parità di input, o di utilizzare meno input a parità di output, è indispensabile stabilire subito in quale area (input o output) si gode di maggiore discrezionalità. Ciò chiaramente dipende dal contesto nel quale avviene la produzione (un'azienda automobilistica può facilmente controllare l'output, il numero di auto prodotte; mentre un ospedale, ad esempio, ha evidentemente un controllo maggiore sugli input). La distinzione è

importante in quanto le misure più appropriate di efficienza variano a seconda del tipo di orientamento. Vediamole concretamente con l'ausilio della Fig. 1. Una UD che produce nel punto d avrebbe potuto ottenere il massimo livello di produzione, e , impiegando lo stesso livello di input g ; alternativamente si sarebbe potuta collocare nel punto c , dove il suo stesso livello iniziale di output può essere ottenuto con un minor impiego di input. Le due misure di efficienza tecnica della nostra UD subordinate all'orientamento sono dunque le seguenti:

ET (orientamento-output): (Ob/Oa) . In pratica, fermo restando il livello di input iniziale, è il rapporto fra il livello di output effettivo e quello massimo ottenibile;

ET (orientamento-input): (Of/Og) . In pratica, fermo restando il livello iniziale di output, è il rapporto fra il livello minimo di input richiesto e il livello di input effettivamente impiegato.

Vediamo ora come è possibile misurare tanto l'efficienza tecnica quanto l'efficienza allocativa in relazione a un processo produttivo con più di un fattore. Consideriamo due input, x_1 e x_2 . La funzione di produzione, come nel caso precedente, indica la quantità massima di output y che si può ottenere impiegando quantità variabili del primo e del secondo fattore. Ma la rappresentazione delle informazioni relative alla funzione di produzione sul piano bidimensionale, avendo ora a che fare con tre dimensioni, obbliga a ricorrere a uno stratagemma basato sulle curve di livello o isoquanti. Si tratta di proiezioni sul piano dei contorni ottenuti sezionando la funzione di produzione a quote diverse, corrispondenti a livelli via via più elevati di produzione. Ognuna di queste curve mostra le combinazioni tecnicamente possibili dei due input che generano un dato livello di output. La forma delle curve di livello dipende dai rapporti di sostituzione esistenti fra gli input. Quella sottesa alla Fig. 2 riflette l'ipotesi che il tasso al quale è possibile sostituire un fattore con un altro, mantenendo inalterato il livello di produzione, sia decrescente (si parlerà di saggio marginale di sostituzione tecnica decrescente).

La rappresentazione completa della funzione di produzione richiede ovviamente una mappa di isoquanti. Anche in questo caso è possibile scoprire l'esistenza o meno di rendimenti di scala guardando a cosa succede (in quale isoquanto si finisce) se si aumentano tutti gli input nella stessa proporzione (cioè muovendosi lungo un raggio che parte

dall'origine degli assi). Nell'esempio ipotizziamo l'esistenza di rendimenti di scala costanti.

Torniamo però alle nostre misure di efficienza. Poiché vogliamo cogliere anche l'aspetto economico (i costi), nella Fig. 2, insieme all'isoquante, abbiamo rappresentato anche una retta, chiamata isocosto. Questa, come dice il nome, è il luogo geometrico di tutte le combinazioni degli input aventi lo stesso costo (per dati prezzi dei fattori). A ogni livello di spesa, chiaramente, corrisponde una retta di isocosto diversa. La pendenza della retta, tuttavia, è costante, in quanto è uguale per costruzione al rapporto fra i prezzi (fissi) di x_1 e di x_2 . Per non oscurare la figura abbiamo disegnato un solo isoquante e una sola retta di isocosto.

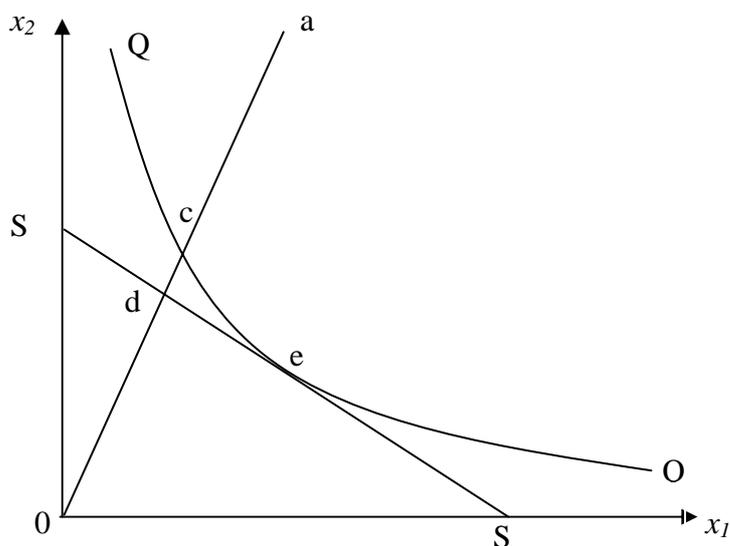


Figura 2 – Efficienza tecnica ed efficienza allocativa

Tutte le combinazioni efficienti in senso tecnico sono collocate sull'isoquante. Questo, infatti, è il grafico dell'insieme dei punti cui corrispondono impieghi minimi di input per produrre il livello prescritto

di output. Il livello indicato, se si vuole, potrebbe essere pari a 1 unità⁴. Teoricamente l'isoquanto dovrebbe riflettere le migliori prassi ingegneristiche. Dunque tutte le UD che si trovano sull'isoquanto QQ sono efficienti in senso tecnico. Per produrre la quantità data al minimo costo, tuttavia, è necessario soddisfare anche il vincolo di bilancio SS. Data la forma dell'isoquanto, esiste una sola combinazione, quella corrispondente al punto e , che permette di ottenere il livello dato di output in modo tecnicamente efficiente e al minimo costo. Questa combinazione (o questa UD se identifichiamo la combinazione con una specifica unità decisionale) è efficiente sotto il duplice profilo tecnico e allocativo.

Si consideri ora una UD che produce il livello di output Q con una combinazione produttiva inefficiente sotto entrambi i punti di vista (tecnico e allocativo). Questa potrebbe essere la UD che produce il livello Q con la combinazione a . Come misurare il grado di inefficienza? Una possibilità, suggerita inizialmente da Farrell (1957) e Debreu (1951), consiste nel considerare la distanza radiale, cioè quella lungo la retta che unisce il punto a all'origine degli assi. Lungo questa retta il rapporto fra gli input rimane inalterato, ma le quantità utilizzate di entrambi variano in misura proporzionale. Le contrazioni/espansioni radiali da un punto inefficiente ai punti sulla frontiera costituiscono la base per misurare l'efficienza nelle moderne analisi della produzione. Infatti, la distanza fra il punto a e l'isoquanto, luogo delle combinazioni tecnicamente efficienti, è proporzionale al rapporto $0c/0a$: quanto più c è vicino ad a tanto più prossimo a 1 (o al 100 per cento) risulta questo rapporto; viceversa esso sarà tanto più piccolo quanto più grande il divario fra c ed a . In modo analogo la distanza fra il punto c e il punto d posto sul vincolo di bilancio (che indica la spesa minima per ottenere il livello di output Q , la stessa spesa del punto e per intenderci) è proporzionale al rapporto $0d/0c$: quanto più vicini c e d tanto più prossimo a 1 (o al 100 per cento) risulta questo rapporto, e viceversa. Pertanto avremo le seguenti misure di efficienza:

efficienza tecnica (ET) = $(0c/0a)$ = rapporto fra combinazione di input tecnicamente efficiente e combinazione di input effettivamente osservata, tenendo costante il rapporto fra gli input

⁴ In questo caso si parla di isoquanto unitario, ragion per cui i livelli di input esprimono le quantità necessarie per produrre una unità di output e la retta di isocosto il costo per produrre 1 unità di output.

efficienza allocativa (EA) = $(0d/0c)$ = rapporto fra costo necessario a produrre l'output dato in condizioni di sola efficienza tecnica e costo minimo (costo nel punto ottimale)

Il richiamo di queste due misure risiede in gran parte nella facile interpretazione cui si prestano. Supponiamo, infatti, che $(0c/0a)$ sia uguale al 70%, ciò significa che una UD che opera nel punto c (dunque con lo stesso rapporto fra input dell'UD operante in a) riesce a produrre lo stesso output dell'impresa che opera nel punto a usando una frazione pari al 70% di ciascun fattore di produzione; alternativamente, ciò significa che con la combinazione di fattori utilizzata dall'impresa operante in a , l'UD tecnicamente efficiente riesce a produrre un livello di output $(0a/0c)$ volte maggiore. Similmente, il costo di produzione dell'output nel punto e (dove tale costo è minimo) sarà una frazione $(0d/0c)$ del costo sostenuto nel punto c .

Se la combinazione produttiva (o la UD) considerata fosse perfettamente efficiente, cioè efficiente sotto il duplice profilo tecnico e allocativo, il costo di produrre l'output dato sarebbe una frazione $(0d/0a)$ del costo effettivamente osservato. Ma questo rapporto, che Farrell (1954, p.255) chiama efficienza globale, è esattamente uguale al prodotto fra ET e EA.

$$\text{efficienza globale} = \text{ET} \times \text{EA} = (0d/0a) = (0c/0a) \times (0d/0c)$$

2.2. Misure non parametriche dell'efficienza

Abbiamo appena visto come in linea di principio sia possibile misurare l'efficienza di una UD rispetto a una tecnologia di riferimento nota. Come rendere operativi i concetti di efficienza illustrati? Il primo passo consiste nella stima della tecnologia di riferimento, ovvero della frontiera efficiente che, incorporando le "prassi migliori" (*best practices*), funge da standard per valutare il posizionamento delle UD osservate.

Al riguardo è possibile utilizzare sia metodi parametrici sia non parametrici. I primi fondamentalmente stimano la frontiera con metodi econometrici che richiedono di postulare preliminarmente la forma funzionale della funzione di produzione (o, ciò che è lo stesso, le caratteristiche dell'insieme di produzione da cui si deriva la frontiera). Nei primi studi di questo tipo la stima veniva effettuata col modello classico di regressione⁵, col risultato che la frontiera stimata risultava una

⁵ L'ipotesi cruciale del modello è che il termine di errore della regressione abbia una distribuzione normale e simmetrica (tipicamente media zero e varianza costante). In altre parole si ipotizzano scostamenti dall'andamento medio sia positivi che negativi.

retta media interpolante la nuvola di osservazioni riguardanti le unità produttive, e il grado di efficienza scaturiva dal confronto con le prestazioni medie del campione di osservazioni piuttosto che in riferimento alle prassi migliori. Recentemente sono emersi strumenti più raffinati, come la SFA (*stochastic frontier analysis*), che con l'introduzione di un termine di errore composito permette di stimare i parametri di una vera frontiera (una linea che racchiude le osservazioni). Restano tuttavia diverse complicazioni in aggiunta a quelle relative alla specificazione della forma funzionale, come il numero elevato di parametri da stimare o la necessità di disporre di informazioni sui prezzi degli output, che si acquisiscono quando si hanno a disposizione poche osservazioni o, come nel caso di molti servizi pubblici, mancano informazioni sui prezzi. Nel caso dei metodi non parametrici, invece, non viene postulata una forma funzionale specifica riguardo alla relazione intercorrente fra input e output, ma vengono fatte ipotesi più o meno restrittive sulle caratteristiche dell'insieme di produzione che portano a costruire la frontiera efficiente in una varietà di modi sulla base delle UD che mostrano le prassi migliori. Rimandando ai saggi raccolti in Fried, Lovell e Schmidt (1993) per un approfondimento su questi metodi, nel seguito ci concentriamo sul metodo DEA.

La tecnica DEA venne sviluppata da A. Charnes, W. Cooper e E. Rhodes (1978) con l'intento di sviluppare e rendere operative le misure di efficienza di Farrell nell'ambito della programmazione lineare. In generale, la costruzione di una frontiera efficiente sulla base di dati statistici riguardanti i processi di trasformazione degli input in output da parte di un gruppo di unità decisionali (a volte useremo l'espressione piani di produzione) richiede delle restrizioni sulle caratteristiche dell'insieme di produzione. La DEA ne postula fundamentalmente due: *free-disposal* e convessità. Per capire cosa ciò implica, sotto diverse ipotesi relative ai rendimenti di scala⁶, si guardi la figura sottostante e ci si concentri sui punti A, B, C ed E corrispondenti alle quattro osservazioni inizialmente disponibili.

⁶ Un'unità produttiva opera in regime di rendimenti di scala costanti se ad una variazione nell'impiego di tutti gli input in una data misura corrisponde una variazione nella stessa misura di tutti gli output. Se invece questi ultimi non variano nella stessa misura degli input i rendimenti di scala si diranno variabili (crescenti oppure decrescenti).

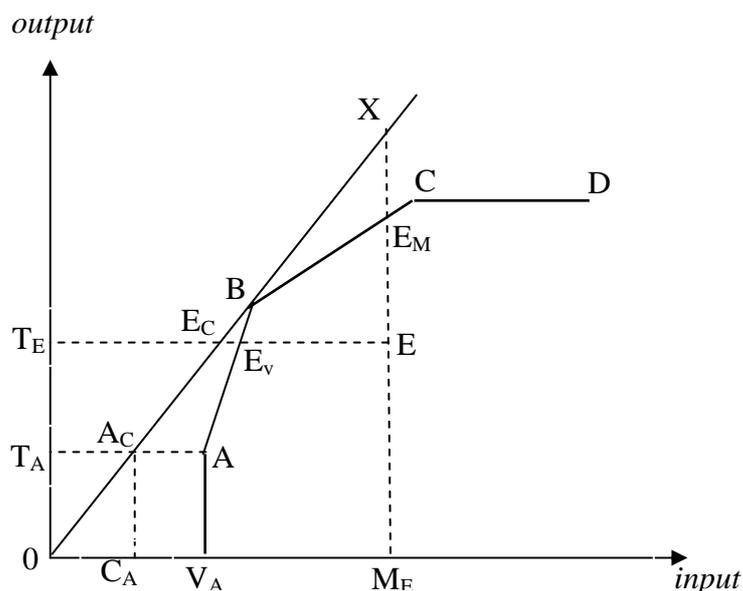


Figura 3 – Frontiere e misure di efficienza

Con rendimenti di scala variabili (RSV), la caratteristica della convessità⁷ garantisce che ogni combinazione lineare fra i punti di partenza (es. A e B) appartiene all'insieme di produzione. Dunque possiamo cominciare a tracciare la frontiera unendo con una spezzata i punti che producono il livello più elevato di output per un dato input, cioè A, B e C. Ma in quale direzione estendere i punti estremi A e C? Qui è l'ipotesi di *free-disposab*⁸ che permette di completare la frontiera con il ramo verticale sotto A e orizzontale a destra di D, essa infatti implica che se il piano A è fattibile lo sarà anche qualunque piano che produce un output inferiore con la stessa quantità di input, mentre se D è fattibile lo sarà anche qualunque piano che produce lo stesso output con un maggiore impiego di input. Analogamente, in regime di rendimenti di scala costanti (RSC), dove qualunque multiplo di un piano di produzione

⁷ Un insieme di punti S nel piano si dice convesso se ogni coppia di punti dell'insieme può essere collegata mediante un segmento che giace interamente all'interno dell'insieme.

⁸ Formalmente, appartiene all'insieme di produzione qualunque piano di produzione non osservato con livelli di output uguali o minori di quelli di qualche piano di produzione osservato e con una quantità maggiore di almeno un input; o con livelli di input uguali o maggiori di quelli di qualche piano di produzione osservato e con una quantità minore di almeno un output; o con entrambe queste proprietà (vedi Tulkens, 1993, p. 180).

osservato appartiene all'insieme di produzione, le due restrizioni autorizzano a tracciare una frontiera lineare che fuoriesce dall'origine e passa per il punto più efficiente fra quelli osservati,⁹ cioè il punto B. Entrambe le frontiere inviluppano l'insieme delle osservazioni, ma quella con RSV "fascia" questo insieme in maniera più stretta.

L'ipotesi di rendimenti di scala costanti esprime l'idea che la dimensione dell'unità produttiva sia irrilevante per la misurazione della sua efficienza relativa. Ciò significa escludere la possibilità che la crescita dimensionale di una UD favorisca ad esempio una migliore configurazione dei servizi generali o dell'organizzazione interna che dia luogo ad un aumento più che proporzionale dell'output; allo stesso modo, nel caso di una UD di dimensioni elevate, che una ulteriore crescita possa dar luogo a diseconomie di vario tipo. Naturalmente, una tecnologia può essere caratterizzata da rendimenti di scala diversi in corrispondenza di livelli diversi di produzione; ma per la singola UD conviene sempre cercare di operare in corrispondenza della scala ottima: non troppo piccola se i rendimenti di scala sono crescenti, non troppo grande se invece sono decrescenti. Conseguire la scala ottima non è facile e, soprattutto, può dipendere da fattori esterni che la UD non controlla direttamente.

Una volta definite le frontiere corrispondenti alle due ipotesi sui rendimenti di scala, il metodo DEA permette di misurare a) l'efficienza tecnica complessiva delle UD in ciascun regime e b) l'entità della componente relativa all'efficienza di scala. Per ogni UD, il primo indicatore viene calcolato sulla base della distanza dalla frontiera di riferimento, il secondo confrontando il valore dell'indicatore nei due regimi. Con RSV la frontiera fascia i dati più strettamente, per cui ciascuna UD viene confrontata con un sottoinsieme di unità decisionali dalla scala simile; mentre con RSC la frontiera risulta più esterna e ciascuna UD viene confrontata con le migliori UD indipendentemente dalla scala. Pertanto, l'indicatore di efficienza calcolato nel primo caso sarà sempre maggiore di quello calcolato nel secondo caso. Inoltre, sotto l'ipotesi di RSV, esso riflette solo le inefficienze dovute a componenti diverse dalla scala di produzione, col risultato che anche il numero complessivo di UD tecnicamente efficienti sarà maggiore rispetto al

⁹ Notare che l'inclinazione della retta che congiunge l'origine degli assi con un qualsiasi punto del piano è uguale al rapporto fra l'output e l'input del punto considerato, chiaramente rispetto ai punti osservati la retta che passa per il punto B è quella con l'inclinazione più elevata.

regime di RSC. Occorre infine sottolineare che a seconda che l'analisi sia orientata verso gli input o verso gli output, gli indicatori di efficienza assumono in generale valori diversi, così come diverse risultano le UD di riferimento per le unità inefficienti.

In riferimento alla Fig. 3 e all'UD inefficiente E, le diverse misure dell'efficienza tecnica sono calcolabili come segue:

Orientamento	RS	Indicatore	Descrizione
Input	RSV	$I_1 = T_E E_V / T_E E$	Frazione degli input utilizzati necessaria per produrre l'output osservato stando sulla frontiera RSV
Output	RSV	$I_2 = M_E E / M_E E_M$	Frazione dell'output potenziale (quello sulla frontiera RSV a parità di input utilizzati) effettivamente prodotto
Input	RSC	$I_3 = T_E E_C / T_E E$	Frazione degli input utilizzati necessaria per produrre l'output osservato stando sulla frontiera RSC
Output	RSC	$I_4 = M_E E / M_E X$	Frazione dell'output potenziale (quello sulla frontiera RSC a parità di input utilizzati) effettivamente prodotto

Mentre per ciò che riguarda l'efficienza di scala, sempre per il punto E con orientamento verso gli input, l'indicatore viene calcolato dividendo la distanza dalla frontiera RCS e da quella RSV in corrispondenza dell'output prodotto in E, ovvero $T_E E_C / T_E E_V$. Misure analoghe vengono definite nel caso di orientamento verso gli output. Si noti che poiché nel punto B entrambi i tipi di inefficienza sono assenti, si può affermare che l'UD corrispondente a quel punto ha raggiunto la scala ottima. Al contrario, l'UD del punto A sta operando con rendimenti di scala crescenti mentre le UD dei punti C e D sono caratterizzate da rendimenti decrescenti, sono cioè divenute troppo grandi e potrebbero avvicinarsi alla scala ottima del campione riducendo la propria dimensione.

3. Analisi non parametriche sull'efficienza della giustizia

Contrariamente a quanto si potrebbe pensare alla luce della persistenza della crisi della giustizia in molti paesi e al risalto che ne viene dato nei mezzi di informazione, le analisi empiriche sul fenomeno sono ancora molto sporadiche tanto in Italia quanto all'estero. Più precisamente, mentre gli istituti statistici dei diversi paesi compiono notevoli sforzi per raccogliere dati e descrivere sistematicamente

l'evoluzione della litigiosità, della criminalità e della capacità della giurisdizione di dare applicazione concreta alle norme dell'ordinamento giuridico, gli studi empirici che cercano di individuare i fattori che impediscono alla giustizia di operare con efficienza sono piuttosto rari. Ciò dipende in parte proprio dall'incompletezza delle statistiche ufficiali, concepite storicamente più in funzione del controllo immediato dei fenomeni che di studio delle loro determinanti. Inoltre, nel caso specifico dell'efficienza della giustizia, così come di altri beni collettivi, c'è il classico problema della mancanza di riferimenti naturali di mercato per stabilire i "prezzi" dei servizi offerti e dei fattori utilizzati. Nonostante questi limiti, da un po' di anni a questa parte sono apparsi diversi lavori empirici con importanti spunti per le politiche di intervento nel settore.

Fra quelli di tipo non parametrico, sui quali vogliamo concentrare l'attenzione, ricordiamo gli studi di Lewin, Morey e Cook (1982) riguardanti i distretti giudiziari del North Carolina, Kittelsen e Førsund (1992) sui tribunali di primo grado della Norvegia, Tulkens (1993) sul sistema dei giudici di pace del Belgio e, infine, Pedraja-Chaparro e Salinas-Jiménez (1996) sulle sezioni del contenzioso amministrativo dei tribunali di Spagna. Mentre Tulkens impiega il metodo FDH (*Free Disposal Hull*), gli altri autori optano per il metodo DEA. In entrambi i casi l'efficienza viene calcolata come distanza da una frontiera delle prassi migliori ricostruita sulla base delle osservazioni campionarie e delle diverse ipotesi sulle caratteristiche dell'insieme di produzione.

Lewin, Morey e Cook (1982) utilizzano dati annuali relativi a 30 distretti giudiziari (*Criminal Superior Courts of North Carolina*) osservati attraverso due output (il numero di "procedimenti esauriti" e di "casi pendenti per meno di 90 giorni") e cinque input ("casi pendenti e sopravvenuti", "numero di addetti amministrativi e legali", "numero di giorni di udienza", "frazioni dei reati minori sul totale dei casi", "numerosità della popolazione bianca del distretto") selezionati in base ad una batteria preliminare di regressioni dei potenziali input sui potenziali output¹⁰.

L'analisi indica la presenza di 11 distretti inefficienti (27% del totale), per ognuno dei quali viene indicato: il gruppo di riferimento (cioè l'insieme dei distretti comparabili ed efficienti con cui confrontarsi); i cambiamenti richiesti (aumento degli output e/o diminuzione degli

¹⁰ Per avere maggiori gradi di libertà e giungere a una specificazione soddisfacente le regressioni vengono effettuate su dati circondariali (cento *counties* corrispondenti ai trenta distretti) e sotto due diverse ipotesi (lineare e log-log) circa la forma funzionale della relazione.

input) per eliminare l'inefficienza; l'effetto marginale sull'indicatore di efficienza dei mutamenti di ciascun output; l'entità degli "slacks", ossia della riduzione (aumento) ulteriore di un dato input (output) necessaria per raggiungere l'efficienza tecnica dopo che tutti gli input (output) sono stati ridotti (aumentati) in proporzioni uguali per raggiungere la frontiera.

Complessivamente, attraverso l'adeguamento alle prassi migliori, viene calcolato un aumento potenziale dei procedimenti esauriti pari al 4% e un risparmio nel numero di giorni di udienza del 6%. Trattandosi di una delle prime applicazioni del metodo DEA, gli autori confrontano i risultati ottenuti con quelli relativi alla *ratio analysis*, un metodo molto popolare nello studio dell'efficienza fondato sulla costruzione di rapporti di produttività con un singolo output (al numeratore) e un singolo (input) al denominatore. Ovviamente, le graduatorie ottenute sulla base dei dieci rapporti input-output calcolati (5 input x 2 output) sono molto instabili e non permettono di separare i distretti efficienti da quelli inefficienti.

Nell'importante contributo di Kittelsen e Førsund (1992) riguardante la Norvegia, il disegno dell'indagine originaria di Lewin, Morey e Cook (1982) viene sviluppato in diverse direzioni: aumenta la dimensione del campione (107 tribunali distrettuali di primo grado), la profondità temporale delle osservazioni (dati annuali per il periodo 1983-1986), il numero di output considerati (totale procedimenti relativi a sette categorie di reati); vengono introdotte misure esplicite di efficienza della scala produttiva. Per quanto riguarda gli input, invece, l'incompletezza delle informazioni non permette di andare oltre il numero di magistrati in organico e di addetti allo staff del distretto.

Utilizzando i dati medi del periodo per ogni distretto, Kittelsen e Førsund calcolano quattro indicatori di efficienza (fondamentalmente quelli delineati nella sezione precedente) e un indicatore di scala che rivela immediatamente la tipologia dei rendimenti di osservati per ciascuna unità. Le elaborazioni mostrano un numero di unità efficienti molto alto, tanto che secondo gli autori è da escludere, anche tenendo conto delle distorsioni indotte dagli aspetti dimensionali, che l'elevato arretrato esistente sia attribuibile all'inefficienza delle corti.

La perdita di efficienza complessiva risulta dell'8-10% ed è in gran parte attribuibile alla scala inadeguata (troppo piccola) di molti tribunali piuttosto che ad inefficienze di natura tecnica. Per la scala ottima viene anche indicato un intervallo di riferimento che va dal 60% a circa il doppio della dimensione del tribunale medio del campione. Come

sottolineano gli autori, dal punto di vista operativo i risultati dell'applicazione DEA sono particolarmente interessanti per il singolo distretto, che può disporre di riferimenti concreti rispetto ai quali calibrare i programmi di miglioramento dell'efficienza.

Il lavoro di Tulkens (1993) sui giudici di pace in Belgio utilizza un metodo di costruzione della tecnologia di riferimento (della frontiera) meno restrittivo rispetto alla DEA, in quanto non richiede l'ipotesi di convessità dell'insieme di produzione. La forma della frontiera costruita con questo metodo, noto come FDH (da *free disposal hull*), è diversa da quella tipica della DEA: con un input e un output l'involuppo dei dati osservati segue un profilo a gradini. La conseguenza più importante è che la fasciatura dell'insieme di produzione è più stretta e la frontiera più "realistica" (nel senso che non può accadere che un'unità di produzione effettiva risulti inefficiente rispetto a un'unità fittizia ottenuta come combinazione di altre osservazioni).

L'applicazione riguarda il sistema dei giudici di pace del Belgio, composto all'epoca da 187 giudici monocratici, assistiti da un piccolo staff, aventi competenza su casi di valore monetario contenuto e di carattere locale (ambito cantonale). Come input viene utilizzato il numero di addetti di ciascun ufficio, mentre la scelta degli output ricade su tre diversi servizi: il numero di procedimenti civili definiti; il numero di udienze su casi di arbitrato nell'ambito delle famiglie; il numero di procedimenti per reati minori definiti.

L'analisi, effettuata iterativamente per i tre anni disponibili, mette in luce la presenza di un elevato numero di unità inefficienti (fra l'82% e l'87%), all'interno del quale prevalgono le unità di dimensione media. Sfruttando da un lato la disponibilità di dati sul numero di casi pendenti a fine anno in ciascun ufficio e, dall'altro, la conoscenza per ciascuna unità inefficiente dell'ufficio esemplare da prendere a riferimento, Tulkens calcola l'ammontare del pregresso smaltibile e non smaltibile: rispettivamente, la quantità di casi pendenti che un'unità potrebbe smaltire adeguandosi alla prassi migliore e quella che potrebbe smaltire solo attraverso un aumento del personale impegnato. Le stime effettuate indicano nel primo caso, in media, un valore del 35%. L'entità dell'arretrato sembra dunque giustificare la richiesta di un aumento di personale.

Nell'applicazione DEA di Pedraja-Chaparro e Salinas-Jiménez (1996) viene sostanzialmente ripercorso lo stesso schema di analisi dei precedenti lavori in riferimento alle 21 sezioni di contenzioso

amministrativo della Spagna, ma con alcune interessanti estensioni riguardanti la significatività dell'ipotesi di omogeneità delle UD analizzate (un'importante ipotesi teorica caratteristica del metodo DEA) e il calcolo di un indicatore di efficienza globale comprensivo del termine "efficienza allocativa". Dal momento che solo 5 su 21 unità sono relativamente efficienti e che l'indicatore medio di efficienza è di poco superiore al 77%, l'analisi mostra l'esistenza di ampi margini di innalzamento dell'efficienza.

4. Dati e metodologia

4.1 La scelta delle unità osservate

Abbiamo visto che il concetto di efficienza di una UD è un concetto relativo, misurabile attraverso il confronto fra la combinazione output/input effettiva di una determinata UD e la combinazione ottimale per quel tipo di UD. Operativamente, con la DEA quest'ultima viene individuata tracciando una frontiera produttiva che incorpora le prassi migliori fra quelle osservate nel campione e che richiede per essere identificata un confronto fra tutte le unità ad esso appartenenti.

La scelta dell'insieme di riferimento è dunque cruciale. Ma non è affatto scontata, perché se da un lato le UD dovrebbero essere sufficientemente omogenee per rendere il confronto significativo, dall'altro dovrebbero comportarsi in modo apprezzabilmente diverso per permettere di discriminare le organizzazioni efficienti da quelle inefficienti. Dato un certo ventaglio di input e output, un aumento della numerosità del campione porta normalmente ad un ampliamento dei casi di UD con comportamenti simili, con conseguente riduzione del valore dell'indicatore di efficienza medio; mentre una diminuzione del numero di imprese, a parità di input e output considerati, tende in media a gonfiare il valore dell'indicatore, in quanto ogni UD che dovesse risultare dominante lungo una particolare dimensione viene decretata efficiente.

Tali questioni sono tanto più serie quando l'analisi riguarda un settore come quello dell'amministrazione della giustizia, dove le unità produttrici –gli uffici giudiziari– forniscono tipicamente una varietà di servizi (si pensi non soltanto alla distinzione tra giustizia civile e penale, ma anche alle diverse materie e alle molteplici modalità di definizione dei procedimenti) e, nell'assolvere queste funzioni, sono influenzati anche da elementi del contesto globale entro cui operano.

Va ricordato in proposito che l'amministrazione della giustizia è caratterizzata da un elevato grado di rigidità (Marchesi, 2003): i principali

fattori di produzione (magistrati e personale amministrativo) non sono fra loro sostituibili e alcuni di essi, i magistrati, sono praticamente inamovibili. Di certo non possono essere spostati da un ufficio all'altro secondo le eventuali necessità.

Lo stesso accade riguardo al prodotto, in quanto, contrariamente a quanto si osserva in altri servizi pubblici ogni unità produttiva è tenuta a servire tutta la domanda che le viene rivolta, con scarsi margini di manovra circa l'ordine di trattazione dei casi e pochi strumenti di razionamento, nonostante i recenti ampliamenti dei poteri direttivi del giudice, diversi dal mero ordine cronologico delle iscrizioni a ruolo.

Infine, anche i differenti uffici¹¹ sono caratterizzati da un accentuato grado di non sostituibilità, giacché l'attività di ciascuno di essi è determinato in modo esclusivo da aspetti quali la materia del procedimento, la gravità del reato, il valore dell'oggetto della controversia, la competenza territoriale.

Tra tutti questi fattori, l'unico sul quale il decisore pubblico può ragionevolmente incidere in tempi brevi, non richiedendo interventi nel campo normativo, riguarda la competenza territoriale. Intervendo su questo aspetto, che ha caratteristiche spiccatamente organizzative, per esempio ridefinendo i confini delle circoscrizioni giudiziarie, si possono redistribuire i carichi di lavoro, commisurandoli alla qualità e all'entità della domanda di giustizia che proviene dal territorio.

Tutte queste considerazioni, insieme ai vincoli derivanti dal tipo di informazioni richieste per svolgere un'analisi DEA completa e, ultimo ma non per importanza, al fatto che i mutamenti nella geografia giudiziaria italiana hanno spesso comportato riallocazioni e spiazamenti fra uffici di uno stesso territorio, ci hanno portato a incentrare questa prima applicazione sui distretti di corte d'appello.

In questo contesto, dopo aver misurato l'efficienza complessiva dei distretti e aver individuato per ciascuno di essi i distretti "virtuosi" da tener presenti per il miglioramento delle proprie prestazioni, abbiamo cercato di quantificare quanta parte dell'inefficienza stimata sia attribuibile a fattori tecnici e quanta al fatto di operare con una

¹¹ Le unità produttive, seguendo l'esempio del mondo della produzione. Dopo la riforma del Giudice Unico di primo grado [2/06/99 per la giustizia civile e 1/01/00 per quella penale], le principali funzioni giudiziarie sono assicurate da sette tipologie di uffici: giudici di pace (848), tribunali (164), corti d'appello (29 + 3 sedi distaccate), Corte di cassazione, Procure della Repubblica presso i Tribunali, Procure Generali della Repubblica presso le Corti d'Appello, Procura Generale presso la Corte di Cassazione.

dimensione sub-ottimale. Sulla base delle informazioni ottenute abbiamo poi calcolato l'entità dell'arretrato smaltibile e non smaltibile. Infine, attraverso un'analisi di regressione dei punteggi DEA su un insieme di grandezze rappresentative dell'area di competenza del distretto, abbiamo cercato di chiarire il legame fra efficienza e contesto operativo.

4.2. *Le variabili utilizzate*

La necessità di comparare unità produttive tra loro potenzialmente molto diverse ci ha indotto a selezionare una gamma molto ristretta sia di fattori di produzione sia di tipologie di prodotti offerti (tavola 1). Questa scelta è peraltro coerente con la regola secondo cui la selezione ideale include il minor numero di output e di input capaci di cogliere adeguatamente tutti gli aspetti essenziali delle operazioni svolte dalle UD oggetto di studio.

Il numero dei procedimenti definiti individua il tipo di prodotto offerto dai distretti. La caratteristica di attività multi-prodotto degli uffici giudiziari, dunque, è stata valorizzata lungo la dimensione esclusiva della materia dei procedimenti, marcando un'unica differenza tra procedimenti civili e penali.

Fra gli input, invece, abbiamo considerato innanzitutto il numero dei magistrati ordinari in servizio presso ciascun distretto, trascurando altre figure professionali –quali per esempio il personale amministrativo– in quanto l'analisi preliminare indicava un grado di correlazione estremamente elevato fra le due variabili. La scarsità delle informazioni disponibili ci ha impedito di utilizzare una qualsiasi misura dello stock di capitale fisico, ma data l'elevata intensità di lavoro dell'attività tipica dei distretti questa lacuna è probabilmente poco influente. Piuttosto, in linea con altri lavori (Lewin, Morey e Cook, 1982) abbiamo considerato un secondo input rappresentato dal numero dei procedimenti pendenti iniziali –divisi per materia civile e penale– che individuano lo stock di procedimenti che ci si attende siano definiti nell'arco dell'anno dagli uffici del distretto.

I procedimenti iscritti all'inizio del periodo, infine, sono stati considerati fra gli input “non controllabili”. Questi procedimenti esprimono la domanda di giustizia proveniente dal territorio e concorrono a definire lo stock dei procedimenti che ci si aspetta siano definiti nell'arco dell'anno. Tuttavia, la loro entità dipende direttamente da fattori quali il livello di litigiosità e criminalità presente nel distretto, piuttosto che da elementi ricadenti sotto il diretto controllo di chi

amministra la giustizia nel distretto (es. capacità delle forze dell'ordine di istruire indagini capaci di condurre all'individuazione dei responsabili di atti illeciti).

L'anno di riferimento dell'analisi è il 2002 e la scelta del periodo campionario è stata dettata soprattutto dalla facilità del reperimento delle informazioni necessarie.

5. L'analisi empirica

5.1. La strategia di analisi

La valutazione dell'efficienza tecnica dei distretti di corte d'appello si sviluppa in diverse fasi. Preliminarmente, adottando un modello con rendimenti di scala costanti, individuiamo quei distretti di corte d'appello che fanno un uso migliore delle risorse a disposizione. Questo tipo di analisi può assumere due differenti prospettive, che risultano speculari data l'ipotesi di rendimenti costanti: selezionare come efficienti quei distretti che rispetto ad altri riescono a definire un maggior numero di procedimenti, pur impiegando la stessa quantità di fattori di produzione (orientamento verso l'output), oppure quei distretti che, per definire lo stesso numero di procedimenti, riescono ad impiegare una quantità inferiore di risorse produttive (orientamento verso gli input).

Il risultato principale di questa analisi è rappresentato dal calcolo degli indicatori di efficienza, o *efficiency score* (punteggio), cioè di una misura radiale della distanza di ciascun distretto da quello che, con la medesima combinazione di fattori, riesce a conseguire risultati migliori o, alternativamente, un medesimo risultato con un minor impiego di risorse. I distretti efficienti sono caratterizzati da un indicatore pari ad uno, mentre quelli inefficienti da un indicatore inferiore all'unità.

Il passo successivo consiste nel valutare, per ciascun distretto efficiente, quanti distretti risultano inefficienti al suo cospetto. In questo modo si selezionano, tra i distretti efficienti, quelli più rappresentativi in termini di organizzazione ottimale delle risorse. Infatti, quando un distretto è valutato come efficiente se confrontato con un elevato numero di distretti è probabile che rappresenti un caso autentico di impiego ottimale delle risorse produttive, di organizzazione e gestione delle attività di definizione dei procedimenti. Al contrario, quando l'insieme di riferimento è molto ristretto, è probabile che il distretto risultato efficiente presenti una combinazione di input oppure di output talmente particolare da non poterlo assumere come punto di riferimento organizzativo.

Nel caso dei distretti inefficienti, invece, è possibile misurare –a parità di numero di procedimenti definiti- la riduzione proporzionale degli input necessaria per diventare efficienti (orientamento verso gli input), oppure l'aumento proporzionale del numero dei procedimenti definiti –a parità d'impiego di risorse di produzione- che consentirebbe di essere classificati come efficiente (orientamento verso l'output). In alcuni casi, può succedere che anche dopo il raggiungimento della frontiera efficiente (attraverso una riduzione equiporzionale di tutti i fattori nel caso ad esempio dell'analisi orientata verso gli input) si renda necessaria un'ulteriore riduzione di qualche fattore di produzione (non per andare oltre la frontiera, bensì per muoversi verso un punto fattibile dove si consuma ancor meno un dato input). In questo caso si parlerà di presenza di *slack* relativamente a quell'input.

Un'importante caratteristica della DEA è quella di permettere di individuare due importanti fonti di inefficienza. In particolare, è possibile decomporre l'inefficienza complessiva in inefficienza tecnica genuina, ovvero nell'incapacità di combinare in modo ottimale i fattori di produzione e di ottenere il numero di procedimenti definiti più elevato possibile per quel dato impiego di risorse, ed inefficienza di scala, dovuta cioè essenzialmente a una dimensione operativa inadeguata. In pratica si ricalcolano gli indicatori di efficienza sotto l'ipotesi di rendimenti di scala variabili. Le unità che sotto quest'ultima ipotesi ottengono un punteggio più elevato di quello corrispondente all'ipotesi di rendimenti costanti sono unità la cui inefficienza dipende anche da fattori di scala.

Infine, effettuando un'analisi della varianza per accertare la significatività delle differenze tra gruppi, è possibile indagare in che misura gli indicatori di efficienza sono influenzati da variabili di sfondo, riguardanti le condizioni economiche e sociali dell'area di competenza del distretto, e vadano dunque opportunamente corretti.

5.2. *I risultati dell'analisi*

I dati riportati nella tavola 2 mostrano che a fronte di 12 distretti efficienti ne esistono 17 inefficienti, con un valore medio dell'84% che si colloca fra il valore massimo del distretto di Napoli (circa il 99%) ed il minimo di Sassari (circa il 61%). In base a questi numeri, possiamo affermare che, per esempio, il distretto di Milano ha un grado d'efficienza pari al 90% circa rispetto al distretto che esprime la prassi migliore.

Il grado di rappresentatività dei distretti efficienti è molto variabile. Il distretto di Ancona presenta una combinazione di input/output comparabile a quella di altri 12 distretti, ma rispetto a questi è capace di ottenere lo stesso numero di procedimenti definiti impiegando una quantità inferiore di fattori di produzione: per questi 12 distretti, quello di Ancona rappresenta un esempio di gestione ottimale delle risorse da emulare. Un discorso analogo vale per i distretti di Bolzano, Brescia e l'Aquila, che ricorrono spesso come distretti da emulare. All'estremo opposto c'è il distretto di Lecce, per il quale non è possibile individuare alcun altro distretto che possa essere ad esso comparato; in questo caso, la combinazione di input/output è troppo singolare, e l'organizzazione del lavoro da esso prodotto non può costituire un utile riferimento per alcuno degli altri distretti presenti nel campione.

Per i distretti inefficienti abbiamo indicato, in ordine decrescente d'importanza, i modelli di efficienza con i quali dovrebbero confrontarsi per migliorare le loro prestazioni.

La figura 4 presenta in forma sintetica questa informazione, e dalla sua osservazione emerge in modo netto che –fatta eccezione per quelli calabresi- i distretti efficienti si localizzano nel Nord-Est e lungo la fascia adriatica, ovvero in quelle regioni dove si è affermato un modello di sviluppo incentrato sull'esistenza dei c.d. distretti industriali (Becattini, 1979; Brusco e Paba, 1997). La letteratura di economia industriale ha variamente caratterizzato l'importanza dei distretti, ma ai fini del nostro lavoro è importante ricordare alcune caratteristiche salienti di questo tipo di configurazione produttiva: fitta rete di relazioni tra attori locali incentrate su fiducia, reciprocità e circolazione delle informazioni; elevato livello di cooperazione. Tutti questi elementi, com'è noto, concorrono all'accumulazione e rigenerazione di un livello elevato di capitale sociale che si manifesta anche in un'ampia condivisione delle norme e una conseguente maggiore fluidità e coesione del tessuto sociale.

Riprendendo l'esame della tabella contenente i valori degli indicatori di efficienza, a titolo di esempio possiamo considerare il distretto di Genova, che presenta un grado di efficienza pari all'84% circa del proprio insieme di riferimento, del quale fanno parte i distretti di Bolzano, Ancona, Brescia, Trento e Venezia. Nel determinare le variazioni necessarie per migliorare l'efficienza, un ruolo determinante spetta a Bolzano. Il confronto con questo distretto porta a concludere che nel caso di Genova, organizzando diversamente l'attività degli uffici,

si potrebbe conseguire lo stesso numero di procedimenti definiti di quelli ottenuti nel distretto di Bolzano, ma impiegando il 16% in meno di magistrati. Inoltre, il carico iniziale dei procedimenti civili dovrebbe ridursi di circa il 16% e quello penale di circa il 32%.

In base a questi riscontri, perché Genova possa raggiungere la frontiera efficiente è necessario che, in aggiunta a una riduzione equiproporzionale di tutti gli input, il numero dei procedimenti pendenti iniziali in materia penale scenda di una ulteriore percentuale, segnalando così l'esistenza di uno *slack* (ulteriore sacca di inefficienza) in relazione a questo specifico fattore di produzione.

Uno sguardo d'insieme a questa parte della tabella suggerisce che, almeno nel caso dei distretti della ripartizione meridionale, l'inefficienza sia legata all'esistenza non solo di un numero eccessivo di magistrati, rispetto alla quantità di procedimenti che si riescono a definire, ma anche di un consistente carico iniziale di procedimenti, sia in ambito civile sia penale. Nel resto del paese, invece, l'inefficienza dei distretti è condizionata in modo quasi esclusivo dalla rilevanza del carico di procedimenti penali che non sono stati smaltiti nei periodi precedenti.

Al contrario, la domanda di giustizia che proviene dal territorio, misurata dal numero di procedimenti iscritti, sembra esercitare un'influenza ridotta sui risultati di efficienza, anche se con alcune notevoli eccezioni, come nel caso del distretto di Torino¹² o di quello di Messina¹³.

L'esistenza di un rilevante carico di procedimenti emerge come uno dei fattori determinanti del risultato d'inefficienza. E' opportuno quindi chiedersi in che misura questo carico possa essere ridotto conformandosi al comportamento del distretto che nel proprio insieme di riferimento esprime la prassi migliore. La tavola 3 presenta i risultati di quest'indagine.

Per ciascun distretto inefficiente sono riportati l'indicatore d'efficienza (colonna [a]), il numero di procedimenti pendenti in materia civile e penale (colonne [b] e [c]) e il numero di procedimenti definiti, in materia civile e penale (colonne [d] e [e]), dal più importante distretto di riferimento. Moltiplicando i valori di queste due ultime colonne per il grado di efficienza, otteniamo una misura del numero ottimale di

¹² Il distretto di Torino potrebbe realizzare un risultato simile a quello del distretto efficiente al quale è comparabile se il numero di procedimenti iscritti in materia penale si riducesse del 23% circa.

¹³ Nel caso del distretto di Messina, una riduzione del 13% circa del numero di procedimenti iscritti in materia civile farebbe migliorare l'efficienza del distretto.

procedimenti che ciascun distretto potrebbe definire se si conformasse alle *best practices* del proprio distretto di riferimento (colonne [f] e [g]). L'arretrato dei procedimenti smaltibile grazie ad una diversa organizzazione del lavoro è data dal confronto tra le colonne [b] e [f] per la materia civile, e [c] e [g] per quella penale. L'arretrato non riducibile (colonne [l] e [m]), invece, indica il numero di procedimenti pregressi che potrebbero essere smaltiti soltanto grazie ad un incremento degli input di produzione. Le ultime due colonne, infine, riassumono il pregresso smaltibile in percentuale al numero dei procedimenti pendenti.

Nel complesso questa tavola rivela che, in aggregato, solo il 10% dell'arretrato in materia civile, e il 14% di quello penale, potrebbe venire smaltito attenendosi alle *best practices* dei distretti più efficienti. Scendendo nel dettaglio, però, si può notare che, in genere, per i distretti meridionali la percentuale sale a circa il 20% per la materia civile ed è ancora maggiore per quella penale. Nel resto del paese, i distretti di Perugia, Potenza e Messina rappresentano i casi più critici.

Resta ora da chiedersi in che misura i risultati riflettano un'effettiva incapacità di organizzare l'attività degli uffici distrettuali, piuttosto che un'inadeguatezza della scala alla quale questi operano. Una valutazione di tale aspetto può essere fatta confrontando i risultati di efficienza del modello con rendimenti di scala costanti (CRS) con quelli del modello a rendimenti variabili (VRS). E' opportuno ricordare che, nel primo caso, l'efficienza di un distretto è valutata in rapporto al comportamento di tutti gli altri distretti presenti nel campione; mentre nel secondo caso il confronto è fatto solo con quei distretti che operano sulla medesima scala: di conseguenza, i risultati d'efficienza del modello con rendimenti variabili sono non inferiori a quelli del modello a rendimenti di scala costanti.

La tavola 4 mostra la decomposizione degli indicatori d'efficienza in base alle diverse ipotesi circa i rendimenti di scala. Il modello a rendimenti costanti permette di misurare l'efficienza complessiva; mentre l'efficienza tecnica pura è misurata dal modello a rendimenti variabili. Il rapporto tra la prima e la seconda, invece, misura l'efficienza di scala: quando questa è pari ad uno (100%) vuol dire che il distretto opera ad una dimensione ottimale (MPSS, *most productive scale size*); se è inferiore ad uno (100%), allora una variazione della dimensione delle sue operazioni può consentire un guadagno d'efficienza.

Per sapere se le dimensioni devono essere aumentate o ridotte, occorre confrontare la misura di efficienza del modello a rendimenti

variabili con quella a rendimenti non crescenti (NIRS): se la seconda è inferiore alla prima, siamo in presenza di rendimenti di scala crescenti (IRS, *increasing returns to scale*), e dunque un distretto potrebbe migliorare la propria efficienza operando su una dimensione maggiore; se invece è uguale, saremmo in presenza di rendimenti di scala decrescenti (DRS, *decreasing returns to scale*), nel qual caso i guadagni di efficienza si potrebbero conseguire solo riducendo la dimensione delle attività del distretto.

I risultati, sinteticamente riportati anche nella figura 5, evidenziano che 12 sono le unità efficienti secondo la dimensione di scala, mentre 11 operano con rendimenti decrescenti e 6 con rendimenti crescenti. In media, il punteggio d'efficienza di scala è del 96% circa, superiore al punteggio medio d'efficienza globale che si assesta sul 91% circa. Inoltre, la variabilità dei punteggi d'efficienza è maggiore nel caso dell'efficienza globale che in quello di scala. Nel complesso, sulla base di quest'evidenza sembra di poter affermare che esistono certamente problemi legati alla scala alla quale operano i diversi distretti, ma i maggiori guadagni d'efficienza possono essere sicuramente conseguiti modificando l'organizzazione interna delle attività degli uffici o riducendo il carico iniziale dei procedimenti, fattori questi che -come abbiamo già visto- condizionano i risultati degli indicatori d'efficienza globale.

A completamento di quest'indagine, gli indicatori di efficienza sono stati utilizzati come variabile dipendente di un modello tobit nel quale come regressori sono state utilizzate alcune variabili di sfondo. La variabile dummy, che seleziona i distretti della ripartizione meridionale (Sud), distingue quei distretti per i quali si ritiene che i problemi di efficienza siano più gravi. L'indice di litigiosità (Litigiosità), invece, intende cogliere la pressione che la domanda di giustizia esercita sulla funzionalità dei distretti. Il numero di procedimenti da trattare in appello in rapporto al totale (Appello)¹⁴, invece, tiene conto, nelle nostre intenzioni, di un importante elemento qualitativo, ovvero della possibilità che la maggiore durata dei procedimenti registrata in alcuni distretti sia il riflesso di un'attività di definizione dei procedimenti che avviene secondo standard condivisi: l'ipotesi di un *trade-off* negativo tra qualità e

¹⁴ Il numero dei procedimenti iscritti in appello è misurato in rapporto al numero dei procedimenti definiti nei gradi precedenti nello stesso anno, in mancanza di informazioni circa il numero di procedimenti definiti in periodi precedenti. Questa procedura introduce un chiaro errore di misurazione, probabilmente poco importante solo nell'ipotesi in cui i procedimenti definiti nei gradi precedenti non siano molto variabili nel tempo.

durata dei procedimenti ci induce ad ipotizzare un segno negativo per il parametro di questa variabile.

Infine, l'ultima variabile tiene conto di un fattore che, nel dibattito recente, è emerso come rilevante per spiegare la differente durata dei procedimenti. Secondo Marchesi (2003), soprattutto nel caso della giustizia civile, la lunghezza dei tempi di risoluzione dei procedimenti sarebbe condizionata dall'esistenza di incentivi perversi, che agiscono in modo particolare dal lato della domanda. La determinazione del tasso di interesse legale, di molto inferiore ai rendimenti di mercato; la ripartizione delle spese in giudizio, raramente a carico della sola parte soccombente; le stesse regole di svolgimento del processo forniscono incentivi ad un prolungamento delle cause. Tra questi, un ruolo importante spetta alla formula di determinazione dell'onorario degli avvocati. Marchesi (2003), confrontando l'esperienza di diversi paesi dell'Unione Europea, trova che quelli caratterizzati da una più celere definizione dei procedimenti, e anche da un più frequente ricorso a procedure di conciliazione tra le parti che evitino il giudizio, sono anche quelli che prevedono tariffe a quota fissa. Al contrario, laddove l'ammontare dell'onorario dei legali viene determinato secondo una tariffa oraria, o addirittura commisurato al numero di attività svolte per l'assistenza del cliente, non ultimo il numero di udienze, si osserva una durata dei procedimenti maggiore della media.

Per cercare di misurare l'entità di questo effetto, non disponendo del dato relativo al numero di avvocati per distretto, abbiamo inserito tra le variabili esplicative una proxy rappresentata dalla media degli immatricolati nelle Facoltà di Giurisprudenza negli ultimi tre anni (Matricole). Nella tavola 5 sono riassunti i risultati della stima. L'unica variabile non significativa è l'indicatore di litigiosità; le altre variabili, invece, sono statisticamente significative, anche se a differenti livelli di significatività.

L'evidenza empirica raccolta sembra effettivamente suggerire che la funzionalità degli uffici giudiziari sia condizionata dalla competizione tra i soggetti coinvolti nella definizione dei procedimenti, così come dal comportamento di particolari figure professionali con forti incentivi all'allungamento dei procedimenti. Esiste, poi, un problema di efficienza legato alla localizzazione geografica ma, poiché il ruolo assunto dalla domanda di giustizia che proviene dal territorio non risulta statisticamente significativo (coerentemente con i risultati dell'analisi non

parametrica), questo sembra sempre collegato alla mancanza di norme condivise e all'esistenza di comportamenti opportunistici.

6. Conclusioni

La tecnica DEA applicata al settore della giustizia consente di misurare una serie di indicatori di efficienza che hanno un importante ed immediato risvolto. Nel breve periodo, quando tutti gli input sono fissi, l'eliminazione di inefficienze tecniche dal lato dell'output può aiutare a risolvere l'esistenza di uno stock rilevante di procedimenti pregressi; nel medio periodo, quando invece gli input sono variabili ma la domanda di giustizia è predeterminata dalla ampiezza del territorio di competenza degli uffici, la misura di efficienza rilevante si basa un risparmio nell'impiego di fattori di produzione; nel lungo periodo, infine, l'efficienza può essere valutata anche in relazione alla scala di operazione.

L'esercizio che abbiamo proposto permette, pur con i limiti che abbiamo più volte sottolineato, di affermare che:

- l'inefficienza tecnica osservata è consistente: 17 su 29 distretti sono inefficienti, con un grado medio di inefficienza intorno al 10%;
- i distretti esemplari sono in genere localizzati nel Nord-Est e lungo la fascia adriatica
- l'inefficienza va addebitata in particolare all'esistenza di un rilevante carico di procedimenti pendenti, soprattutto in materia penale; marginale, invece, è la pressione esercitata dalla domanda di giustizia, misurata dal numero di procedimenti sopravvenuti;
- l'arretrato è solo in minima parte smaltibile con un adeguamento dei distretti alle prassi dei distretti più efficienti, tranne che nei casi di Perugia, Potenza e Messina;
- per alcuni distretti esiste un problema di efficienza di scala: 11 distretti sono troppo grandi, mentre 6 distretti sono troppo piccoli

Nel complesso, una maggiore efficienza nell'amministrazione della giustizia può essere perseguita solo risolvendo il problema dello stock di arretrato da smaltire. A tal fine, gli interventi da intraprendere devono riguardare –a seconda dei casi- misure di deflazione dei procedimenti, adeguamento degli organici, ridefinizione degli ambiti di competenza territoriale. Da non trascurare, poi, un intervento istituzionale che riformi soprattutto il rito civile, eliminando i forti incentivi che oggi spingono le parti interessate a comportamenti che inducono l'allungamento dei tempi del processo.

Bibliografia

- Becattini, G. (1979) "Dal settore industriale al distretto industriale. Alcune considerazioni sull'unità di indagine dell'economia industriale", *Rivista di Economia e Politica Industriale*, 1.
- Brusco, S., S. Paba (1997) "I distretti industriali italiani", in F. Barca, (a cura di) *Storia del capitalismo italiano*, Donzelli, Roma
- Cammelli, M. (2002) "Alcune cose da condividere a proposito di giustizia", in *il Mulino*, n. 5, settembre-ottobre, pp. 844-853.
- Consiglio Superiore della Magistratura (2001) *Relazione al Parlamento sullo stato della giustizia. Tutela dei diritti, efficacia e tempi della giurisdizione*, Quaderni del CSM, n. 120 (<http://www.csm.it/pages/quaderni.html>).
- Cooper, W., L. Seidorf, K. Tone (2000) *Data Envelopment Analysis*, Boston, Kluwer.
- Dakolias, M. (1999) "Court Performance Around the World: A Comparative Perspective", vol. 2., pp.
- Di Iulio, J. J. Jr (1993) "Measuring Performance When There Is No Bottom Line", in Department of Justice, Bureau of Justice Statistics *Performance Measures of the Criminal Justice System*, pp. 142-155.
- Førsund, F., N. Sarafoglou (2002) "On the Origins of Data Envelopment Analysis", *Journal of Productivity Analysis*, v. 17, pp. 23-40.
- Fried, H. O., C. A. K. Lovell, S. S. Schmidt (1993) *The Measurement of Productive Efficiency*, Oxford, Oxford University Press.
- Gravelle H. (1990) "Rationing Trials by Waiting: Welfare Implications", *International Review of Law and Economics*, v. 10, pp. 255-270.
- Gravelle, H. (1990) "Rationing Trials by Waiting: Welfare Implications", *International Review of Law and Finance*, v. 10, 255-270.
- ISTAT ((2001) *La giustizia civile in Italia: funzionalità e processo di riforma*, Informazioni, n. 3.
- Kittelsen, S. e F. Forsund (1992) "Efficiency Analysis of Norwegian District Courts", *The Journal of Productivity Analysis*, n. 3, pp. 277-306.
- Knox Lovell, C. A. (1993) "Production Frontiers and Productive Efficiency", in (a cura di) Fried, H. O., C. A. K. Lovell, S. S. Schmidt (1993).

- Lewin, A., Morey, R. e T. Cook (1982) "Evaluating the Administrative Efficiency of Courts", *OMEGA The International Journal of Management Science*, n. 4, pp. 401-411.
- Marchesi, D. (2002) "L'inefficienza della giustizia civile fra domanda e offerta", in *il Mulino*, n. 5, settembre-ottobre, pp. 854-862.
- Marchesi, D. (2003) *Litiganti, avvocati e magistrati. Diritto ed economia del processo civile*, Bologna, il Mulino.
- Ministero della Giustizia (2004a) *Relazione sull'amministrazione della giustizia nell'anno 2003 del Dott. Francesco Favara, Procuratore Generale della Repubblica presso la Corte Suprema di Cassazione*, (www.giustizia.it/uffici/inaug_ag/relaz_index2004.htm)
- Ministero della Giustizia (2004b) *Intervento del Ministro della Giustizia sen. Roberto Castelli, Inaugurazione dell'anno giudiziario 2004* (www.giustizia.it/uffici/inaug_ag/relaz_index2004.htm)
- Orlandi, R. (2002) "Giustizia penale e ruolo dello Stato: un rapporto in crisi", in *il Mulino*, n. 5, settembre-ottobre, pp. 863-873.
- Pedraja-Chaparro, F. e J. Salinas-Jiménez (1996) "An Assessment of the Efficiency of Spanish Courts using DEA", *Applied Economics*, n. 28, 1391-1403.
- Pritchard, A. (2002) "Measuring Productivity Change in the Provision of Public Services", in *Economic Trends*, n. 582, May, pp. 20-32.
- Spottiswoode, C. (2002) "Improving Police Performance. A new Approach to Measuring Police Efficiency. Technical Annexes", Public Service Productivity Panel, UK Government.
- Torre, A. (2003) "The Impact of Court Delays on the Prosecutor and the Defendant: An Economic Analysis", in *European Journal of Law and Economics*, v. 16, pp. 91-111.
- Tulkens, H. (1993) "On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit", in *Journal of Productivity Analysis*, n. 4, pp. 183-210.
- Vereeck, L., M. Mühl (2000) "An Economic Theory of Court Delay", *European Journal of Law and Economics*, v.10, n.3, pp. 243-268
- Wilson, J. Q. (1993) "The Problem of Defining Agency Success", in Department of Justice, Bureau of Justice Statistics, *Performance Measures of the Criminal Justice System*, pp. 156-166.

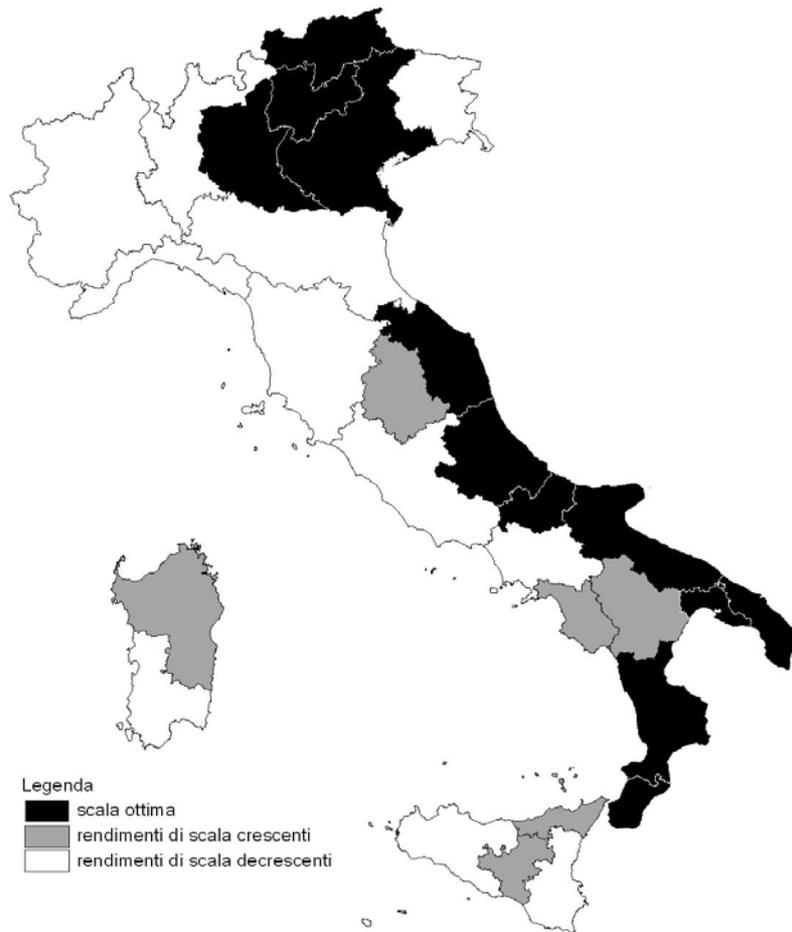


Figura 5. La natura dei rendimenti di scala.

Tavola 1 Le variabili utilizzate

	min	max	media
Magistrati	54	911	274
Pendenti civili	10676	649569	144200
Pendenti penale	12322	973174	196903
Iscritti civile	6217	155001	41999
Iscritti penale	16908	359671	108374
Definiti civile	6993	202948	48922
Definiti penale	17371	358138	107053

Tavola 2. Gli indicatori di efficienza

Nr ord	Distretto	Efficienza	Numero di distretti per i quali si è un riferimento	Numero di distretti di riferimento	Distretti di riferimento	Riduzione percentuale fattori di produzione				
						Magistrati	Pendenti Iniziali Civile	Pendenti Iniziali Penale	Iscritti Civile	Iscritti Penale
2	BS	100	10		BS	0	0	0	0	0
4	BZ	100	11		BZ	0	0	0	0	0
5	TN	100	7		TN	0	0	0	0	0
6	VE	100	4		VE	0	0	0	0	0
12	AN	100	12		AN	0	0	0	0	0
14	AQ	100	7		AQ	0	0	0	0	0
15	CB	100	2		CB	0	0	0	0	0
18	BA	100	2		BA	0	0	0	0	0
19	LE	100	0		LE	0	0	0	0	0
20	TA	100	3		TA	0	0	0	0	0
22	CZ	100	4		CZ	0	0	0	0	0
23	RC	100	4		RC	0	0	0	0	0
16	NA	99,39		4	AN, TA, CZ, BS	-0,6	-0,6	-46,3	0	-8,3
9	BO	98,19		4	BZ, AN, BS, CZ	-1,8	-1,8	-51,3	0	-7,2
1	TO	97,39		4	BZ, TN, BS, RC	-2,6	-2,6	-41,5	0	-22,6
10	FI	96,8		5	BZ, BS, TN, CZ, AQ	-3,2	-3,2	-3,2	0	-0,9
13	RM	96,66		4	AN, TA, BS, CZ	-3,3	-4,8	-3,3	0	-6,7
11	PG	93,66		4	BS, AN, BA, TA	-6,3	-6,3	-22,5	0	-1,5
7	TS	90,81		4	BZ, TN, BS, VE	-9,2	-9,2	-31,9	-1,4	0
3	MI	90,27		3	BZ, VE, AN	-9,7	-9,7	-50,7	-0,2	0
24	CL	84,4		3	TN, RC, AQ	-47,2	-15,6	-34	0	-16

Tabella 2. Continuo

8	GE	83,82	5	BZ, AN, BS, TN, VE	-16,2	-16,2	-31,7	0	0
27	PA	82,62	3	CB, AN, AQ	-44,1	-26,8	-17,4	-7	0
26	ME	79,53	3	AN, BZ, RC	-20,8	-57,7	-0,5	-12,8	0
28	CA	75,81	3	BZ, AQ, AN	-24,2	-27,3	-24,2	-3,2	-11,4
17	SA	71,95	3	BS, AQ, BA	-28	-34,2	-35,4	-9,2	0
21	PZ	69,41	5	AN, BZ, CB, AQ, RC	-30,6	-45,3	-30,6	0	0
25	CT	64,45	4	AN, BZ, TN, AQ	-37,4	-35,6	-35,6	-3,1	0
29	SS	60,56	5	BZ, BS, TN, AN, VE	-39,4	-39,4	-49	0	0
		84,45							

Tavola 3 L'arretrato smaltibile dei procedimenti giudiziari

Nr ord	Distretto	Efficienza	Procedimenti pendenti		Distretto di riferimento Procedimenti definiti		Definizione ottimale procedimenti		Arretrato smaltibile		Arretrato non smaltibile		civile	penale	
			civile	penale	civile	penale	civile	penale	civile	penale	civile	penale			
		(a)	(b)	(c)	(d)	(e)	(f=a*d)	(g=a*e)	(h=min(b,f))	(i=min(c,g))	(l=b-h)	(m=c-i)	h/b	i/c	
16	NA	99,39	649569	973174	AN	33197	64018	32994	63627	32994	63627	616575	909547	0,051	0,065
9	BO	98,19	166942	504664	BZ	8897	19691	8736	19335	8736	19335	158206	485329	0,052	0,038
1	TO	97,39	140959	413752	BZ	8897	19691	8665	19177	8665	19177	132294	394575	0,061	0,046
10	FI	96,8	174798	288411	BZ	8897	19691	8612	19061	8612	19061	166186	269350	0,049	0,066
13	RM	96,66	529092	541266	AN	33197	64018	32088	61880	32088	61880	497004	479386	0,061	0,114
11	PG	93,66	48505	99335	BS	44723	145044	41888	135848	41888	99335	6617	0	0,864	1,000
7	TS	90,81	45394	106904	BZ	8897	19691	8079	17881	8079	17881	37315	89023	0,178	0,167
3	MI	90,27	221708	456965	BZ	8897	19691	8031	17775	8031	17775	213677	439190	0,036	0,039
24	CL	84,4	30495	46349	TN	8583	26258	7244	22162	7244	22162	23251	24187	0,238	0,478
8	GE	83,82	114648	206895	BZ	8897	19691	7457	16505	7457	16505	107191	190390	0,065	0,080
27	PA	82,62	158468	126133	CB	6993	17371	5778	14352	5778	14352	152690	111781	0,036	0,114
26	ME	79,53	135194	47275	AN	33197	64018	26402	50914	33041	47275	102153	0	0,244	1,000
28	CA	75,81	67528	84866	BZ	8897	19691	6745	14928	6745	14928	60783	69938	0,100	0,176
17	SA	71,95	138932	188873	BS	44723	145044	32178	104359	32178	104359	106754	84514	0,232	0,553
21	PZ	69,41	63308	46971	AN	33197	64018	23042	44435	23042	44435	40266	2536	0,364	0,946
25	CT	64,45	131401	149729	AN	33197	64018	21395	41260	21395	41260	110006	108469	0,163	0,276
29	SS	60,56	38999	81735	BZ	8897	19691	5388	11925	5388	11925	33611	69810	0,138	0,146
			2855940	4363297						291361	635272			0,102	0,146

Tavola 4. La dimensione di scala.

Nr ord	Distretto	Rendimenti di scala			Efficienza tecnica globale	Efficienza di scala	Natura dei rendimenti
		costanti CRS	variabili VRS	non crescenti NIRS			
1	TO	97,39%	100,00%	100,00%	97,39%	97,39%	drs
2	BS	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
3	MI	90,27%	100,00%	100,00%	90,27%	90,27%	drs
4	BZ	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
5	TN	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
6	VE	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
7	TS	90,81%	93,77%	93,77%	90,81%	96,84%	drs
8	GE	83,82%	86,38%	86,38%	83,82%	97,04%	drs
9	BO	98,19%	100,00%	100,00%	98,19%	98,19%	drs
10	FI	96,80%	100,00%	100,00%	96,80%	96,80%	drs
11	PG	93,66%	100,00%	93,66%	93,66%	93,66%	irs
12	AN	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
13	RM	96,66%	100,00%	100,00%	96,66%	96,66%	drs
14	AQ	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
15	CB	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
16	NA	99,39%	100,00%	100,00%	99,39%	99,39%	drs
17	SA	71,95%	74,05%	71,95%	71,95%	97,16%	irs
18	BA	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
19	LE	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
20	TA	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
21	PZ	69,41%	78,47%	69,41%	69,41%	88,45%	irs
22	CZ	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
23	RC	100,00%	100,00%	100,00%	100,00%	100,00%	MPSS
24	CL	84,40%	100,00%	84,40%	84,40%	84,40%	irs
25	CT	64,45%	73,92%	73,92%	64,45%	87,19%	drs
26	ME	79,53%	100,00%	99,53%	79,53%	79,53%	irs
27	PA	82,62%	95,86%	95,86%	82,62%	86,19%	drs
28	CA	75,81%	76,99%	76,99%	75,81%	98,47%	drs
29	SS	60,56%	71,89%	60,56%	60,56%	84,24%	irs
media		90,89%	94,87%		90,89%	95,58%	
st dev		0,12	0,10		0,12	0,06	
min		60,56%	71,89%		60,56%	79,53%	
max		100,00%	100,00%		100,00%	100,00%	
eff dmu		12	21		12	12	
MPSS							12
drs							11
irs							6

Tavola 5 L'influenza dei fattori di sfondo sulla funzionalità dei distretti

	coeff	std err	Z	P> z
Appello	-0.001202	0.0006911	-1.74	0.087
Litigiosità	-6.442825	5.414139	-1.19	0.246
Matricole	-0.001022	0.0005269	-1.94	0.063
Sud	-7.970125	4.35526	-1.83	0.079
Pseudo R2	0.2335			
Prob > chi2	0.02277			