

L'analisi dell'efficienza della giustizia civile: l'approccio della frontiera stocastica*

Rossana Cima

Misurare l'efficienza di una qualsiasi unità produttiva significa, in termini generali, valutare quanto l'organizzazione è abile nell'utilizzare le risorse disponibili. L'analisi, basata sui modelli di frontiera stocastica, propone una stima dell'efficienza nel comparto della giustizia civile. L'efficienza dei distretti di Corte d'appello viene stimata confrontando le performance produttive effettivamente realizzate dall'organizzazione con uno standard ottimale, tenendo conto delle fonti di disturbo non controllabili dal "produttore". L'aspetto innovativo dello studio è l'utilizzo dei dati di spesa come misura di input nonché l'impiego di una nuova misura di output che, diversamente da quelle presenti in letteratura, è direttamente proporzionale al numero di procedimenti esauriti e inversamente proporzionale alla durata dei procedimenti.

Parole chiave: efficienza tecnica, modelli a frontiera stocastica, modelli deterministici, giustizia civile.

La parola efficienza è stata in questi ultimi anni uno dei termini più citati nel dibattito politico e pubblico. È divenuta la stella polare del ministro Brunetta e l'obiettivo di molti provvedimenti adottati per migliorare la macchina della Pubblica Amministrazione (PA).

È chiara l'importanza per qualsiasi unità produttiva di valutare le proprie performance. È importante per le aziende private che devono costantemente migliorare il processo produttivo per reggere la pressione competitiva, e lo è anche per le amministrazioni pubbliche. Ed è forse in questo settore che più si sente la necessità di un'analisi dell'efficienza. È un ambito in cui non vi è né concorrenza, né mercato e scarsi sono dunque gli incentivi e i segnali (prezzi) per migliorare le prestazioni. La PA, ad esempio, grazie ai contributi dello Stato può continuare ad operare nonostante una inadeguata gestione delle risorse, mentre il raggiungimento dei livelli di efficienza è centrale per la sopravvivenza delle imprese private. Ma la qualità dei servizi pubblici è

* La ricerca è stata effettuata con il coordinamento della Fondazione Gorla e grazie al sostegno della Fondazione CRT – Progetto Master dei Talenti della Società Civile. Il lavoro ha beneficiato della supervisione e dei preziosi consigli del prof. Luca Ricolfi, che ringrazio.

Per corrispondenza: Rossana Cima, Dipartimento di Scienze Sociali, Università degli Studi di Torino, via S. Ottavio 50, 10124 Torino (Italia). E-mail: rossanacima@unito.it

comunque un elemento irrinunciabile. Non solo ha ricadute sul benessere della collettività ma è un fattore importante per la competitività del paese, soprattutto in un periodo di crisi economica come quello che stiamo attraversando. Le istituzioni pubbliche con le loro regole del gioco creano l'ambiente in cui le imprese operano, incentivando o disincentivando gli investimenti.

In particolare, è stato più volte sottolineato l'impatto che la qualità del sistema-giustizia può avere sull'economia (Bianco, Giacomelli, 2004; La Porta *et al.*, 1996). Non a caso, uno degli indicatori monitorati dalla Banca Mondiale per classificare i paesi nei quali è più vantaggioso investire è l'efficienza della giustizia civile determinata in base ai tempi e ai costi dei procedimenti per il recupero di un credito. Il quadro che emerge dalla graduatoria 2011 non è affatto confortante per l'Italia. Il nostro paese figura al 157° posto su 183 paesi con 1.210 giorni per la risoluzione di una controversia commerciale. Preceduta in classifica da Kosovo, Pakistan o Indonesia, l'Italia è ben lontana dalla Germania (6° posto), dalla Francia (7° posto) o dal Regno Unito (23° posto). La giustizia in Italia sembra essere lenta e questo costa in termini economici. Secondo recenti analisi della Confartigianato (2009), i tempi lunghi della giustizia civile italiana costano alle imprese circa 2,3 miliardi di euro; una sensibile mole di risorse che un sistema inefficiente sottrae al mercato. In termini generali, stando ad una prima valutazione degli sprechi nella giustizia civile (Ricolfi *et al.*, 2010), il tasso di risorse mal impiegate raggiunge il 31,8% della spesa totale destinata a tale comparto. Inoltre, le differenze di produttività nei distretti di Corte d'appello sono molto pronunciate. Sembra dunque presente un problema di efficienza della giustizia. Ma come possiamo valutare le performance dei diversi distretti di Corte d'appello e individuare margini di miglioramento?

In generale, misurare l'efficienza significa valutare il grado di aderenza del processo produttivo ad uno standard di ottimalità (Nisticò, Prosperetti, 1991). È indispensabile, dunque, disporre di un termine di riferimento, ma la scelta dello standard non è affatto banale. Una soluzione potrebbe essere quella di valutare l'efficienza di ogni singolo distretto di Corte d'appello come scostamento da una frontiera di *best practice*, confrontando le prestazioni non con uno standard di riferimento ideale o con il valore medio conseguito dal campione, ma con i migliori risultati *concretamente* realizzati (Farrell, 1957). In una prospettiva *cross-section*, le unità più virtuose (quelle che rappresenteranno il termine di paragone) non saranno quelle che vantano le migliori prestazioni in senso assoluto, ma saranno le più efficienti fra quelle effettivamente presenti nel campione di riferimento considerato². Ciò permetterà di individuare obiettivi potenzialmente raggiungibili, perché già conseguiti da altre unità produttive (UP) considerate nell'analisi. Questo è, ad esempio, l'approccio utilizzato da alcune recenti ricerche sugli sprechi in alcuni comparti della Pubblica Amministrazione (Ricolfi, 2008; 2009a).

1. Efficienza tecnica ed efficienza allocativa

L'efficienza di una organizzazione può essere valutata sotto due diversi punti di vista: tecnico e allocativo. L'efficienza *tecnica* fa riferimento alla capacità di una unità produttiva di produrre la maggior quantità di output a partire dalle risorse produttive disponibili, ovvero di utilizzare una minor quantità di input per realizzare la quantità di output prestabilita. La stima dell'efficienza può essere quindi orientata all'output (*output-oriented*) se l'obiettivo è quello di determinare l'incrementato potenziale della produzione dell'UP a parità di risorse utilizzate o all'input (*input-oriented*) se si vuole valutare la capacità dell'impresa di risparmiare risorse per realizzare la stessa quantità di output. La scelta dell'orientamento è legata alle caratteristiche del contesto esaminato. Dipende, ad esempio, dal potere discrezionale che l'UP ha rispetto alla scelta degli input o dell'output (Kumbhakar, Lovell, 2003). Occorre valutare con attenzione la direzione dell'analisi anche perché le stime di inefficienza possono differire a seconda dell'orientamento adottato. Le due misure convergono solo nel caso in cui la funzione di produzione è a rendimenti di scala costanti³.

Ma i risultati conseguiti dall'azienda, come abbiamo detto, possono dipendere anche da una seconda componente: l'efficienza *allocativa*, ovvero la capacità dei produttori di adottare, fra le combinazioni di input tecnicamente efficienti, quella che minimizza i costi dati i prezzi di produzione. Le due misure di efficienza, quella tecnica e quella allocativa, possono essere combinate in un unico indice di efficienza produttiva totale⁴ (Farrell, 1957). Rimandiamo ai contributi di Kumbhakar e Lovell (2003) e di Coelli *et al.* (2005) per un approfondimento sull'efficienza allocativa. In questo studio ci limiteremo ad esaminare l'efficienza tecnica.

Facciamo un semplice esempio. Immaginiamo N unità produttive che impiegano un solo input (x) per produrre una data quantità di output (y). Possiamo descrivere la relazione che lega i fattori di produzione e il prodotto realizzato da una UP mediante la funzione di produzione

$$y = f(x)$$

dove y è l'output prodotto, x indica il fattore di produzione e f è la tecnologia che descrive la relazione fra l'input e l'output.

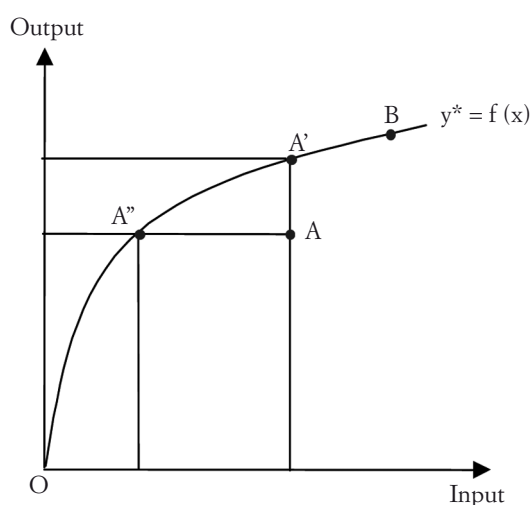
Si assume che una data unità produttiva impieghi una quantità non-negativa di input per produrre una quantità non-negativa di output. Chiamiamo GR l'insieme delle possibilità produttive, ovvero tutte le combinazioni x, y tecnicamente possibili. Ciò corrisponde in figura 1 all'area compresa fra l'asse delle ascisse (x) e la curva individuata da $y^* = f(x)$. Tale curva costituisce la frontiera efficiente e indica il massimo livello di output ottenibile utilizzando una data quantità di input. I rendimenti di scala sono *decreasing*, perché

l'output varia in misura meno che proporzionale rispetto all'input. In caso di rendimenti di scala *crescenti*, y sarebbe variato in misura più che proporzionale rispetto all'input, mentre in un'economia con rendimenti di scala *costanti* – e in presenza di una funzione di produzione omogenea di primo grado – x e y sarebbero aumentati nella stessa proporzione.

Lo scostamento di ciascuna osservazione dalla frontiera efficiente $y^* = f(x)$ fornisce una guida per valutare il grado di efficienza tecnica raggiunto dall'unità produttiva.

Nel caso illustrato, il distretto di Corte d'appello operante nel punto B si rivela pienamente efficiente perché giace sulla frontiera, mentre A può essere definito inefficiente sotto due punti di vista: 1. avrebbe potuto collocarsi nel punto A', ottenendo un più alto livello di output a parità di risorse utilizzate (approccio *output-oriented*); 2. avrebbe potuto minimizzare la quantità impiegata di input producendo lo stesso livello di output posizionandosi nel punto A'' (approccio *input-oriented*).

Fig. 1. Frontiera di produzione



Considerando l'osservazione A e seguendo l'impostazione di Farrell – il primo a misurare empiricamente l'efficienza produttiva ispirandosi ai lavori di Koopmans (1951) e di Debreu (1951)⁵ –, l'efficienza tecnica orientata all'output (TE *output-oriented*) è data dal rapporto fra l'output effettivamente prodotto e quello massimo realizzabile tenendo costante il livello di risorse impiegate.

$${}^{TE} output-oriented = \frac{OA}{OA'}$$

La misura di efficienza tecnica orientata all'input (*TE input-oriented*) è invece uguale al rapporto fra il livello minimo di input richiesto e quello effettivamente impiegato per realizzare una data quantità di output.

$${}^{TE} input-oriented = \frac{OA''}{OA}$$

I due indici di efficienza possono assumere valori compresi fra 0 e 1. Quando TE è uguale a 1, l'UP è pienamente efficiente⁶.

La stima delle diverse componenti dell'efficienza può essere effettuata adottando metodi sia di tipo parametrico che non parametrico. Mentre i primi necessitano di specificare preliminarmente la forma funzionale di produzione e, nel caso di un approccio stocastico, la distribuzione del termine di inefficienza⁷, i secondi si basano su assunti sull'insieme di produzione⁸ e non richiedono la definizione a priori della forma dell'equazione che lega gli input all'output. Il modello di frontiera stocastica (SFA – *Stochastic Frontier Analysis*), che utilizzeremo in questo studio, è un metodo parametrico di tipo stocastico proposto contemporaneamente e in modo indipendente da Aigner *et al.* (1977) e da Meeusen e van de Broeck (1977).

2. Il modello di frontiera stocastica

L'approccio SFA consente di stimare l'efficienza delle diverse UP in base allo scostamento da una frontiera di *best practice* tenendo conto di quelle fonti di disturbo stocastico in grado di influire sul processo produttivo di ogni unità decisionale.

L'idea che sta alla base del modello SFA è che la deviazione di ciascuna unità decisionale dallo standard efficiente sia dovuta non solo all'incapacità del "produttore" di utilizzare in modo ottimale le risorse disponibili, ma anche a fattori casuali non direttamente controllabili dall'impresa o a errori di misurazione delle variabili. Il vantaggio di tale approccio è la possibilità di catturare ed isolare gli effetti dovuti a shock esterni che altrimenti andrebbero a confondersi con il termine di inefficienza. Questo rischio si corre soprattutto se si adotta un modello deterministico che assume l'assenza di disturbi casuali.

L'output, dunque, è una funzione⁹ degli input e di un termine di disturbo, ε , che può essere scomposto in *due* elementi distinti e indipendenti: il disturbo casuale e l'inefficienza tecnica.

In termini formali, una funzione di produzione stocastica può essere scritta come segue:

$$y_i = f(x_i; \beta)^* \exp(\varepsilon_i) \quad \text{con} \quad \exp(\varepsilon_i) = \exp(v_i)^* \exp(-u_i)$$

o ricorrendo alla forma logaritmica

$$\ln y_i = \beta_o + \sum_n \beta_n \ln x_{ni} + (v_i - u_i)$$

dove y_i è l'output dell'i-esima unità produttiva, x_{ni} rappresenta l'n-esimo input impiegato dall'i-esimo produttore, β_o e β_n sono i parametri della tecnologia di produzione, mentre v_i e u_i sono variabili casuali supposte fra loro indipendenti e distribuite in modo indipendente dagli input x .

Il termine v_i rappresenta la componente di errore dovuta al caso per la quale si assume una distribuzione normale $N(0, \sigma_v^2)$, mentre u_i , che agisce in modo unilaterale secondo una distribuzione asimmetrica positiva, rappresenta l'inefficienza tecnica di produzione. La componente u assume valori positivi e ciò assicura che $y_i \leq f(x_i; \beta)^* \exp(v_i)$.

È possibile specificare diverse distribuzioni per il termine di inefficienza. L'assunzione fatta da Aigner *et al.* (1977) è che u abbia una distribuzione *half-normal*¹⁰ o esponenziale¹¹. Queste due ipotesi distributive sono particolarmente appropriate quando si assume che la maggior parte delle UP abbia bassi livelli di inefficienza, dato che la moda delle due distribuzioni è zero. Scegliendo distribuzioni più flessibili, come la gamma¹² (Greene, 1980) o la normale troncata¹³ (Stevenson, 1980), si potrà avere invece una maggior variabilità (intorno al valore medio) dei punteggi relativi all'inefficienza.

Con la stima dei parametri β della tecnologia è possibile ricavare l'errore complessivo, ε ($\varepsilon_i = v_i - u_i$), come differenza fra il valore di y effettivamente osservato e quello stimato sulla frontiera di produzione:

$$\varepsilon_i = \ln y_i - \left(\beta_o + \sum_n \beta_n \ln x_{ni} \right)$$

Non è possibile, però, osservare direttamente il termine dell'inefficienza (u) che viene invece stimato utilizzando le informazioni su u contenute nel residuo composito ε . Una soluzione al problema della decomposizione del termine di errore totale è stata fornita da Jondrow *et al.* (1982) che propongono, una volta specificata la distribuzione del termine u , di utilizzare la media¹⁴ della distribuzione condizionata della stessa u dato ε . Nel caso di una distribuzione *half-normal*,

$$E(u_i | \varepsilon_i) = \mu_{\varepsilon_i} + \sigma_{\varepsilon_i} \left[\frac{\phi(-\mu_{\varepsilon_i} / \sigma_{\varepsilon_i})}{1 - \Phi(-\mu_{\varepsilon_i} / \sigma_{\varepsilon_i})} \right] = \sigma_{\varepsilon_i} \left[\frac{\phi(\varepsilon_i \lambda / \sigma)}{1 - \Phi(\varepsilon_i \lambda / \sigma)} - \left(\frac{\varepsilon_i \lambda}{\sigma} \right) \right]$$

dove $\mu_{\varepsilon_i} = -\varepsilon_i \sigma_u^2 / \sigma^2$, $\sigma_{\varepsilon_i} = \sigma_u \sigma_v / \sigma$, $\lambda = \sigma_u / \sigma_v$, $\sigma = \sqrt{(\sigma_u^2 + \sigma_v^2)}$, $\phi(\cdot)$ è la funzione di densità di probabilità, $\Phi(\cdot)$ è la funzione di ripartizione.

Ottenuta la stima di u_i il livello puntuale di inefficienza dell' i -esima osservazione sarà dato da

$$TE_i = \exp(-\hat{u}_i)$$

O, alternativamente, secondo la proposta di Battese e Coelli (1988),

$$TE_i = E(\exp\{-u_i\} | \varepsilon_i)$$

uno stimatore¹⁵, quest'ultimo, da preferire quando l'inefficienza non è prossima allo zero¹⁶.

3. L'efficienza della giustizia civile

Un primo problema da affrontare quando ci si appresta a valutare il processo produttivo di un'organizzazione è l'individuazione preliminare di una misura di input (x) e di output (y).

Grazie alle stime ricavate dal bilancio consuntivo fornite dal Ministero della Giustizia abbiamo avuto la possibilità di utilizzare, come input, la spesa per la giustizia civile disaggregata per distretto giudiziario¹⁷. Questo è certamente uno dei principali indicatori per misurare l'efficienza della giustizia in Italia, sebbene in letteratura non siano presenti studi che abbiano adottato misure simili, questo per problemi legati soprattutto alla disponibilità dei dati. Le poche ma interessanti analisi condotte nel nostro paese hanno invece scelto di utilizzare altre misure come, ad esempio, il numero di magistrati in servizio (Marselli, Vannini, 2004; Marchesi, 2008).

Difficile è definire l'output del processo di produzione. Come ha osservato un recente studio (Ricolfi, 2009a), le principali misure adottate per valutare il funzionamento della giustizia in Italia risultano fra loro incoerenti e non superano un test di unidimensionalità¹⁸, ovvero non è possibile riassumerle in un unico indice. Cercando di individuare i fattori di produzione in grado di cogliere adeguatamente il modo di operare delle diverse unità produttive esaminate, si è scelto di definire l'output come (Ricolfi, 2009a):

$$y = \frac{ES}{P}$$

dove E indica i procedimenti esauriti, S i procedimenti sopravvenuti e P rappresenta lo stock dei pendenti, ovvero la media geometrica dei procedimenti iniziali e di quelli finali.

Questa misura, come sottolinea Ricolfi (2009a), è l'unica a rispettare alcuni importanti assiomi: è direttamente proporzionale al numero di procedimenti esauriti a parità di condizioni ed è inversamente proporzionale alla durata media dei procedimenti.

La funzione di produzione stimata esprime dunque l'output così definito in funzione della spesa¹⁹.

$$\ln y_i = \beta_o + \beta_1 \ln spesa_i + (v_i - u_i)$$

La stima delle risorse destinate al funzionamento della giustizia civile si riferisce al periodo 2005-07, mentre i dati sul movimento dei procedimenti civili sono relativi agli anni 2005-08. Le analisi sono state condotte su dati medi disaggregati per i 29 distretti di Corte d'appello²⁰.

Si tratta però di unità d'analisi diverse per dimensione e disponibilità di risorse. Queste ultime, poi, dipendono dalle caratteristiche del territorio, dall'estensione e dall'organizzazione del distretto. Ma se la fonte dell'inefficienza varia in base alla dimensione e alle caratteristiche dell'UP, la componente u potrebbe risultare eteroschedastica (Kumbhakar, Lovell, 2003) e questo potrebbe distorcere non solo le stime dei parametri della funzione di produzione (compresa l'intercetta β_o) ma anche le stime dell'efficienza tecnica. Per questo si è scelto di esprimere la varianza del termine unilaterale u in funzione della spesa e della popolazione di ciascun distretto secondo la specificazione che segue:

$$\sigma_{ui}^2 = \exp (\gamma_o + \gamma_1 \ln z1_i + \gamma_2 \ln z2_i)$$

dove $z1$ rappresenta la spesa per l'amministrazione della giustizia civile (che è anche il nostro input) e $z2$ la popolazione di ciascun distretto²¹.

Il modello così definito, sotto l'ipotesi di una distribuzione *half-normal*²² di u , consente di stimare, mediante il metodo della massima verosimiglianza, i parametri della funzione di produzione e di individuare una misura di efficienza tecnica di ciascuna unità d'analisi al netto dell'errore casuale.

Cominciamo con l'osservare la tabella 1 che presenta le stime dei parametri della funzione di produzione, perché emerge da subito un elemento di una certa importanza. Non solo il parametro β_1 relativo alla spesa è significativo e presenta il segno atteso, ma è anche prossimo a 1 ad indicare che la funzione di produzione è a rendimenti di scala approssimativamente costanti. Dunque, ad una variazione dell'input segue una variazione proporzionale dell'output.

Tab. 1. Stima della funzione di produzione

		Coef.	Std. Err.	z	P > z	Intervalli di confidenza al 95%	
ln y	β_1	1,010	0,054	18,680	0,000	0,904	1,116
	β_0	0,898	0,593	1,520	0,130	-0,263	2,060

Nota: il fit del modello può essere valutato a partire da λ , ovvero dal logaritmo del rapporto di verosimiglianza. Come indicato da Ricolfi (1984), R^2 è infatti uguale a $1-\lambda$ (z/N), dove N è il numero dei casi (29). Nel nostro modello, il valore di R^2 è pari a 0,518.

Il livello medio di efficienza tecnica²³ *input-oriented* dei 29 distretti di Corte d'appello è circa il 63% (tab. 3). Questo significa che se tutte le UP adottassero un'organizzazione più efficiente l'output potrebbe aumentare, nel complesso, di circa il 40% a parità di risorse impiegate, ovvero senza alcun aumento dei livelli di spesa. Sono soprattutto i distretti delle regioni settentrionali a posizionarsi ai vertici della classifica. Registrano buoni risultati Torino, Brescia, Bolzano, Bologna e Milano con livelli di efficienza superiori al 90%. Meno virtuosa è invece Genova, unico distretto del Nord con un valore inferiore alla media nazionale. Le performance calano progressivamente nell'Italia centrale e nelle "regioni rosse". Ma è soprattutto al Sud che si rileva la necessità di un recupero di efficienza. Caltanissetta, che occupa l'ultima posizione in graduatoria, presenta, ad esempio, un punteggio che è circa sei volte inferiore a quello del distretto più efficiente, Torino. Le differenze con il distretto piemontese sono meno pronunciate per gli altri distretti dell'area siciliana. Si nota, però, il risultato positivo di Napoli che con un livello di efficienza pari al 72% è il distretto del Sud più virtuoso.

Un'analisi più approfondita rivela che al crescere della spesa per abitante cresce non solo l'inefficienza dei distretti, ma anche la variabilità di u . L'impatto dei fattori di eterogeneità è infatti significativo. La varianza dell'inefficienza non è dunque costante, ma varia a seconda delle caratteristiche del distretto. Come si vede (tab. 2), si espande di un fattore γ_1 ($\gamma_1 = 3,762$) se aumenta la disponibilità finanziaria e si riduce all'incirca della stessa quantità ($\gamma_2 = -4,195$) se cresce la popolazione.

Tab. 2. Stima della varianza di u_i e deviazione standard di v

		Coef.	Std. Err.	z	P > z	Intervalli di confidenza al 95%	
ln σ^2_u	γ_1	3,762	1,546	2,430	0,015	0,731	6,792
	γ_2	-4,195	1,593	-2,630	0,008	-7,317	-1,072
	γ_0	-12,045	7,393	-1,630	0,103	-26,535	2,445
	σ_v	0,091	0,083			0,015	0,546

Poiché i parametri della varianza per il termine u sono simili in modulo (il valore è circa 4), possiamo riscrivere più sinteticamente la formula nel modo che segue:

$$\sigma_u^2 = \exp(\gamma_o)^* \left(\frac{Spesa}{Popolazione} \right)^{\cong 4}$$

Questa riscrittura più compatta ci permette di vedere ancor più chiaramente che la varianza dell'inefficienza è funzione diretta della spesa e indiretta dell'ampiezza della popolazione.

Soffermiamoci ancora sulla relazione fra inefficienza e spesa procapite. Sembrerebbe che la qualità della performance diminuisca proprio dove lo Stato si dimostra più generoso. A ben vedere, però, i territori con un maggior numero di risorse in rapporto alla popolazione sono anche quelli in cui l'indice di litigiosità²⁴ risulta più elevato, un fattore questo che correla negativamente con l'efficienza tecnica dei distretti.

Abbiamo visto come è possibile stimare l'efficienza nel comparto della giustizia civile utilizzando un approccio parametrico di tipo stocastico. Ma cosa succede se si sceglie un'altra strada per valutare le performance dei distretti giudiziari?

Facciamo un semplice esercizio e confrontiamo questi risultati (modello A) con quelli emersi in un precedente studio (Ricolfi *et al.*, 2010) in cui si è seguito un approccio di tipo sostanzialmente deterministico (modello B). Nel modello B è stata adottata la specificazione più semplice, assumendo la presenza di rendimenti di scala costanti. La frontiera di *best practice* è costituita dai 9 distretti più produttivi, ossia Torino, Brescia, Bolzano, Bologna, Milano, Trento, Trieste, Firenze e Venezia. La produttività è stata calcolata come rapporto tra l'output effettivamente prodotto e la spesa sostenuta per l'amministrazione della giustizia. In termini formali:

$$\pi_i = \frac{y_i}{x_i}$$

Per stimare il tasso di sottoproduzione, si è calcolato lo scarto fra quanto gli uffici giudiziari dovrebbero realizzare se producessero quanto la media dei distretti virtuosi e l'output effettivo.

Non sono state fatte ipotesi sulla distribuzione dell'inefficienza né si è distinto l'effetto del disturbo stocastico da quello di u , così lo scostamento dalla frontiera è stato attribuito alla sola inefficienza tecnica.

Consideriamo, per cominciare, l'assunto fatto nel modello B sui rendimenti di scala della funzione di produzione. Questa era un'ipotesi suggerita dalle rilevanze empiriche che di fatto risulta confermata dai dati ottenuti con il metodo SFA. Abbiamo infatti visto che β_1 (tab. 1), ossia il coefficiente della

funzione di produzione stocastica associato alla spesa, è molto vicino a 1 ($\beta_1 = 1,010$) e questo indica che la relazione tra la variazione dei fattori di produzione e la variazione del suo output è (quasi perfettamente) costante.

Tab. 3. L'efficienza tecnica output-oriented nei 29 distretti giudiziari

Distretto	Modello stocastico (A)		Modello deterministico (B)	
	Rank	Efficienza tecnica	Rank	Efficienza tecnica
Torino	1	0,953	1	1,171
Brescia	2	0,946	2	1,082
Bolzano	3	0,937	3	1,076
Bologna	4	0,921	4	1,021
Milano	5	0,905	5	0,999
Trento	6	0,874	6	0,934
Trieste	7	0,857	7	0,915
Venezia	8	0,853	9	0,851
Firenze	9	0,828	8	0,879
Ancona	10	0,734	11	0,762
Napoli	11	0,721	10	0,776
Perugia	12	0,689	12	0,718
Genova	13	0,557	13	0,589
Salerno	14	0,553	14	0,586
Bari	15	0,512	15	0,531
L'Aquila	16	0,507	16	0,530
Roma	17	0,486	17	0,522
Lecce	18	0,482	18	0,502
Campobasso	19	0,455	19	0,480
Taranto	20	0,414	20	0,434
Catanzaro	21	0,407	21	0,432
Potenza	22	0,337	22	0,355
Catania	23	0,335	23	0,352
Sassari	24	0,312	24	0,326
Palermo	25	0,287	25	0,305
Messina	26	0,284	26	0,300
Cagliari	27	0,272	27	0,282
Reggio Calabria	28	0,235	28	0,251
Caltanissetta	29	0,163	29	0,173
Media dei 29 distretti		0,627		0,682

Nota: il livello *medio* di efficienza calcolato sui 29 distretti giudiziari è pesato per la spesa dell'amministrazione della giustizia civile.

I livelli medi di efficienza nei due modelli si differenziano di poco. Nel modello B la media è pari a 0,68, mentre nel modello A è 0,63. Cambia, invece, il *range* dei valori anche perché nel modello deterministico l'indice di efficienza non è vincolato ad assumere valori compresi fra 0 e 1 come nel modello stocastico²⁵: nel modello B oscilla fra un minimo di 0,173 ad un massimo di 1,171, mentre con la SFA varia fra 0,163 e 0,953. Ma quello che più è interessante è il grado di correlazione fra le due misure di efficienza che è pari a 0,993. Si registra solo una inversione fra Venezia e Firenze e fra Ancona e Napoli. Eccezion fatta per queste due differenze, l'ordinamento dei distretti giudiziari risulta identico.

4. Per concludere

L'analisi dell'efficienza tecnica ha permesso di capire se e dove vi sono margini di miglioramento per potenziare la macchina della giustizia civile italiana, valutando le performance dei distretti giudiziari non tanto rispetto ad una condizione ideale, ma a standard *concretamente* realizzabili. Lo si è fatto scegliendo una misura di output diversa da quelle che comunemente vengono utilizzate in letteratura – come ad esempio la durata media dei procedimenti²⁶ o il tasso di ricambio²⁷ –, dato che questi indicatori presentano risultati non sempre coerenti fra loro e addirittura non superano un test di unidimensionalità. Si è preferito, invece, adottare una misura in grado di soddisfare due assiomi fondamentali per l'analisi dell'efficienza nel campo della giustizia civile. L'output, infatti, dovrebbe essere direttamente proporzionale al numero dei procedimenti esauriti e inversamente proporzionale alla durata del procedimento (Ricolfi, 2009a).

Per quanto riguarda i fattori di produzione della giustizia, è stata scelta la spesa, un'importante misura di input che non è mai stata impiegata nei precedenti studi in questo campo, soprattutto nelle analisi basate sulla frontiera stocastica.

Per analizzare l'efficienza dei 29 distretti di Corte d'appello abbiamo fatto ricorso ad un modello a frontiera stocastica comparando poi i risultati ottenuti con quelli stimati mediante un approccio deterministico. Questo ci ha consentito di verificare quanto le stime dell'inefficienza e l'ordinamento dei distretti sono stabili al variare del metodo prescelto.

Dal punto di vista metodologico, il quadro che emerge da queste analisi è abbastanza interessante. Il modello di frontiera stocastica e l'approccio deterministico hanno fornito risultati molto simili: i livelli medi di efficienza si attestano in entrambe le analisi intorno al 60% ed il *ranking* finale dei 29 distretti giudiziari risulta, ad eccezione di due minime inversioni, invariato. Questo suggerisce che quando l'errore casuale è contenuto²⁸ la scelta del metodo adottato (deterministico *vs* stocastico) non influisce in modo

significativo sui risultati ottenuti. Non c'è dubbio, però, che l'eventuale presenza di errori stocastici non trascurabili debba essere verificata. Per questo è vantaggioso l'utilizzo dei modelli SFA dato che, come detto precedentemente, queste tecniche consentono di stimare l'efficienza al netto degli eventuali fattori non direttamente controllabili dal produttore.

È pur vero, però, che i modelli SFA presentano alcuni limiti. La costruzione di una frontiera stocastica comporta problemi non banali come la scelta della distribuzione della componente u o della forma funzionale della funzione di produzione. Ed errori nella specificazione del modello potrebbero condurre ad interpretazioni errate. Per questo è preferibile testare diversi modelli per verificare la stabilità dei risultati.

Vi è però da dire che il modello a frontiera stocastica ha permesso di verificare ipotesi a priori sulla tecnologia di produzione, consentendo di far luce sulla relazione fra input e output. In particolare, le stime hanno confermato l'ipotesi del modello deterministico sulla costanza dei rendimenti di scala. E questo è un risultato per nulla ovvio dal momento che studi precedenti (Marchesi, 2008) condotti su unità produttive più ristrette – i tribunali – hanno evidenziato la presenza di rendimenti di scala variabili: se consideriamo i soli tribunali, la produttività di un magistrato tende ad aumentare al crescere delle dimensioni della UP e a diminuire quando la dimensione diventa troppo elevata. Ma nelle corti d'appello, secondo le nostre stime, output ed input variano in modo proporzionale indipendentemente dall'estensione del distretto²⁹.

Veniamo ora alla parte più interessante dello studio: la valutazione delle performance dei 29 distretti di Corte d'appello.

L'analisi ha messo in luce come in Italia vi siano sensibili divari di produttività fra i territori considerati, con tassi di sottoproduzione particolarmente accentuati al Sud. I risultati confermano quanto già conosciuto³⁰. Si ripropone, dunque, una dicotomia fra regioni del Nord e del Mezzogiorno, anche se non mancano le eccezioni come Napoli, che risulta il distretto del Sud più efficiente, e Genova, che si colloca invece a metà classifica.

Rimangono da capire le ragioni di tali differenze. È plausibile che ciò dipenda anche da un'organizzazione poco razionale del lavoro dei giudici. L'approvazione del progetto "Strasburgo" – che mira all'esaurimento dei processi a "rischio Pinto"³¹ –, varata dall'allora Presidente del Tribunale di Torino Barbuto, mostra come una ristrutturazione del lavoro del giudice possa effettivamente migliorare l'efficienza del tribunale visti i buoni risultati conseguiti del tribunale piemontese (Corrado, Leonardi, 2007). Forse, però, il processo di riorganizzazione dovrebbe essere accompagnato anche da un'efficiente allocazione delle risorse che dovrebbero essere assegnate non tanto in base all'arretrato, ma in base alla domanda di giustizia o in base ai risultati conseguiti dal distretto (Ricolfi, 2009b).

NOTE

¹ Dati aggiornati al 1° giugno 2010 (World Bank, 2010).

² Se si cambia la popolazione di riferimento, cambierà anche il grado di efficienza di ciascuna unità produttiva.

³ Con il termine *rendimenti di scala* ci si riferisce alla relazione che esiste fra la variazione di tutti i fattori di produzione impiegati dalla UP e la variazione dell'output prodotto. La funzione di produzione è a rendimenti di scala costanti quando l'output varia in modo proporzionale rispetto all'input.

⁴ Calcolato come prodotto fra l'efficienza tecnica e l'efficienza allocativa.

⁵ Si deve a Koopmans (1951) una prima definizione di efficienza tecnica, ma è Farrell (1957) a misurare l'efficienza produttiva empiricamente, ispirandosi al lavoro dello stesso Koopmans e di Debreu (1951).

⁶ L'efficienza tecnica *input-oriented* è uguale all'efficienza tecnica *output-oriented* solo in presenza di rendimenti di scala costanti (Färe, Lovell, 1978).

⁷ Data una generica funzione di produzione, l'inefficienza – solitamente indicata con u – rappresenta lo scostamento fra l'output potenzialmente realizzabile e quello osservato. Come verrà meglio illustrato in seguito, u è una variabile casuale con distribuzione asimmetrica positiva. La stima del termine u ci consente di misurare l'efficienza delle UP. Infatti $TE_i = \exp(-\hat{u}_i)$ o alternativamente $TE_i = E(\exp\{-u_i\} | \varepsilon_i)$.

⁸ L'insieme di produzione, ricordiamo, è costituito da tutte le combinazioni di input e di output potenzialmente realizzabili.

⁹ Lineare nei parametri.

¹⁰ $u_i \sim \text{iidN}^+(\sigma, \sigma^2)$.

¹¹ $u_i \sim \text{iidExp}(\lambda)$, dove λ è la media.

¹² $u_i \sim \text{iidG}(\lambda, m)$, dove m sono i gradi di libertà.

¹³ $u_i \sim \text{iidN}^+(\mu, \sigma^2)$.

¹⁴ Può essere utilizzata anche la moda della distribuzione condizionata di u dato ε . Ma è più frequente l'uso della media.

¹⁵ In questo caso, ipotizzando una distribuzione *half-normal* di u ,

$$E(\exp\{-u_i\} | \varepsilon_i) = \left[\frac{1 - \Phi(\sigma_\varepsilon - \mu_\varepsilon / \sigma_\varepsilon)}{1 - \Phi(-\mu_\varepsilon / \sigma_\varepsilon)} \right] \exp\left\{-\mu_\varepsilon + \frac{1}{2} \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}\right\}$$

dove, ricordiamo, $\mu_\varepsilon = -\varepsilon_i \sigma_\varepsilon^2 / \sigma^2$, $\sigma_\varepsilon^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$, $\lambda = \sigma_u / \sigma_v$, $\sigma = \sqrt{(\sigma_u^2 + \sigma_v^2)}$, $\phi(\cdot)$ è la funzione di densità di probabilità, $\Phi(\cdot)$ è la funzione di ripartizione.

¹⁶ In entrambi i casi, però, la stima dell'inefficienza tecnica non è consistente, perché la variabilità associata alla distribuzione di u condizionata al termine di errore ε è indipendente da i e non tende a 0 al crescere della numerosità del campione. Viene stimata in modo consistente nel caso di dati panel per T che tende a ∞ , dove T è il numero di osservazioni nel tempo di ogni UP (Kumbhakar, Lovell, 2003).

¹⁷ I dati di spesa comprendono le retribuzioni al personale (magistratura, personale giudiziario e amministrativo) e la spesa sostenuta per i beni mobili e immobili. Ringraziamo il dott. Giuseppe Belsito e la dott.ssa Maria Raffaella Rancan per aver messo a disposizione le stime delle risorse economiche destinate all'amministrazione della giustizia civile.

¹⁸ L'analisi fattoriale effettuata su sei indicatori (tasso di ricambio, quoziente di estinzione totale, determinazione delle controversie, velocità *vs* inverso durata, produttività nominale, velocità standardizzata) spesso utilizzati in letteratura per valutare l'amministrazione della giustizia in Italia indica la presenza di tre fattori distinti (Repr = 33%; sequenza autovalori: 3,525; 1,170; 0,936; 0,227; 0,130; 0,013).

¹⁹ Secondo una specificazione di tipo Cobb-Douglas.

²⁰ Oltre ai 26 distretti di Corte d'appello sono state considerate le sezioni distaccate di Bolzano, Sassari e Taranto.

²¹ L'eteroschedasticità potrebbe essere presente anche in v o contemporaneamente in u e v . Nel primo caso, si otterranno stime distorte sia dell'intercetta β_0 (che verrà sottostimata) che dell'efficienza tecnica. Nel secondo, il termine di errore complessivo potrebbe essere sottostimato. Sono state prese in considerazione entrambe le possibilità, ma il modello più soddisfacente è quello che controlla gli effetti dell'eteroschedasticità in u . Quest'ultimo modello è l'unico fra quelli considerati che fornisce: *a*) una stima dell'inefficienza robusta e una graduatoria dei territori piuttosto stabile rispetto alle diverse specificazioni della forma funzionale del termine u ; *b*) coefficienti di impatto (relativi alla spesa e alla popolazione) sulla varianza di u significativi.

²² Abbiamo postulato una distribuzione del termine di inefficienza *half-normal* perché è la specificazione più adottata in letteratura. Va comunque osservato, come è stato sottolineato da Kumbhakar e Lovell (2003), che la graduatoria di efficienza risulta piuttosto robusta rispetto alle diverse specificazioni del termine di inefficienza. Per verificare la stabilità dei risultati, è stato stimato un ulteriore modello ipotizzando una distribuzione esponenziale del termine di inefficienza. I risultati sono molto simili a quelli ottenuti dal modello da noi prescelto. La correlazione fra i livelli di efficienza con distribuzione *half-normal* ed esponenziale è pari a $r = 0,994$, $p < 0,001$.

²³ La stima di TE di ciascun distretto è ottenuta mediante il software STATA che permette di stimare l'efficienza mediante la specificazione di Battese e Coelli (1988). Dunque, $TE_i = E(\exp\{-u_i\} | \varepsilon_i)$.

²⁴ L'indice è stato costruito nel modo seguente: $(EC-ES)/E$, dove EC sono i procedimenti esauriti con sentenza, ES sono i procedimenti esauriti senza sentenza ed E i procedimenti esauriti totali. Come spiega l'ISTAT, l'indice può essere utilizzato per misurare il grado di litigiosità (valori positivi) o conciliabilità delle controversie (valori negativi). La correlazione fra l'indice di litigiosità e l'efficienza è pari a $r = -0,736$, $p < 0,001$.

²⁵ Nel modello A, i valori di TE variano nell'intervallo $[0,1]$ perché $y_i \leq f(x_i; \beta) \cdot \exp(v_i)$ dato che, per definizione, u_i è maggiore o uguale a zero. Nel modello B, invece, l'output osservato può essere superiore all'output ottimale e le UP non sono vincolate a giacere al di sotto della frontiera di *best practice*. Ecco perché Torino, Brescia, Bolzano e Bologna presentano punteggi superiori a 1.

²⁶ La durata media dei procedimenti viene calcolata facendo il rapporto fra la media dei procedimenti iniziali e finali e la media dei procedimenti sopravvenuti ed esauriti moltiplicato per il numero di giorni dell'anno.

²⁷ Il tasso di ricambio è uguale al rapporto fra i procedimenti esauriti e quelli sopravvenuti moltiplicato per 100.

²⁸ Il valore minimo di v è $-0,07$, mentre il valore massimo è $0,13$. Il termine di inefficienza u , invece, va da un minimo di $0,05$ a $1,81$.

²⁹ Va però precisato che i risultati ottenuti da Marchesi non sono potenzialmente in contrasto con le nostre stime, perché diverse sono le unità di analisi considerate. Le UP da noi esaminate sono i 29 distretti di Corte d'appello che sono unità giudiziarie più estese del tribunale. Può dunque succedere che variando il campione di riferimento, cambino le proprietà della tecnologia impiegata per realizzare l'output nonché il grado di efficienza di ciascuna unità produttiva.

³⁰ Cfr. ad esempio Marselli e Vannini (2004).

³¹ La legge Pinto (legge 24 marzo 2001, n. 89) prevede una "equa riparazione" a chi è stato coinvolto in un processo per un periodo di tempo considerato troppo lungo.

RIFERIMENTI BIBLIOGRAFICI

Aigner D., Lovell C. A. K., Schmidt P.

1977 *Formulation and Estimation of Stochastic Frontier Production Function Models*, in "Journal of Econometrics", 6, 1, pp. 21-37.

Battese G. E., Coelli T.

1988 *Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data*, in "Journal of Econometrics", 38, pp. 387-99.

- Bianco M., Giacomelli S.
 2004 *Efficienza della giustizia e imprenditorialità: il caso italiano*, in “Economia e Politica Industriale”, 124, pp. 89-113.
- Coelli T., Rao D. S. P., Battese G. E.
 2005 *An Introduction to Efficiency and Productivity Analysis*, Springer, New York.
- Confartigianato
 2009 *Giustizia sempre più lenta: i ritardi costano alle imprese italiane 2,3 miliardi e oltre 450 milioni in Lombardia*, in www.confartigianato-lombardia.it.
- Corrado D., Leonardi M.
 2007 *Così l'efficienza entra in Tribunale*, in www.lavoce.info.
- Debreu G.
 1951 *The Coefficient of Resource Utilization*, in “Econometrica”, 19, pp. 273-92.
- Färe R., Lovell C. A. K.
 1978 *Measuring the Technical Efficiency of Production*, in “Journal of Economic Theory”, 19, pp. 150-62.
- Farrell M. J.
 1957 *The Measurement of Productive Efficiency*, in “Journal of the Royal Statistical Society”, 120, 3, pp. 253-90.
- Greene W. H.
 1980 *Maximum Likelihood Estimation of Econometric Frontier Functions*, in “Journal of Econometrics”, 13, 1, pp. 27-56.
- Jondrow J., Lovell C. A. K., Materov I. S., Schmidt P.
 1982 *On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model*, in “Journal of Econometrics”, 19, 2-3, pp. 233-8.
- Koopmans T. C.
 1951 *An Analysis of Production as an Efficient Combination of Activities*, in T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, 13, Wiley, New York.
- Kumbhakar S. C., Lovell C. A. K.
 2003 *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.
- La Porta F. et al.
 1996 *Law and Finance*, NBER Working Paper, n. 5661.
- Marchesi D.
 2008 *L'enforcement delle regole. Problemi di efficienza della giustizia civile, riforme intraprese e riforme possibili*, in “I Temi dei Rapporti Trimestrali ISAE”.
- Marselli R., Vannini M.
 2004 *L'efficienza tecnica dei distretti di Corte d'appello italiani: aspetti metodologici, benchmarking e arretrato smaltibile*, Working Paper CRENOS.
- Meeusen W., van den Broeck J.
 1977 *Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error*, in “International Economic Review”, 18, 2, pp. 435-44.

- Nisticò A., Prosperetti L.
 1991 *Produzione e produttività*, in G. Marbach (a cura di), *Statistica Economica*, UTET, Torino.
- Ricolfi L.
 1984 *Modelli dell'attore e analisi dei dati*, Giappichelli, Torino.
 2008 *Profondo rosso. Italia 2005 fra ripresa economica e crisi dei conti pubblici*, Guerini e Associati, Milano.
 2009a *L'output della giustizia civile: una proposta per misurarlo*, in "Polena", 2, pp. 73-83.
 2009b *Processi civili, perché sono eterni?*, in "Panorama", 11 giugno.
 2010 *Il sacco del Nord. Saggio sulla giustizia territoriale*, Guerini e Associati, Milano.
- Ricolfi L., Rancan M. R., Cima R.
 2009 *Sottoproduzione e sprechi nella giustizia civile: un'analisi empirica*, relazione tenuta al Convegno "La risultante organizzativa: i suoi diversi significati in ambito giudiziario", 22 maggio, Vicoforte (CN).
 2010 *Sottoproduzione e sprechi nella giustizia civile: un'analisi empirica*, paper presentato al XXIV Convegno SISP, Università IUAV di Venezia, 16-18 settembre 2010, Venezia.
- Stevenson R. E.
 1980 *Likelihood Functions for Generalized Stochastic Frontier Estimation*, in "Journal of Econometrics", 13, pp. 57-66.
- World Bank
 2010 *Doing Business 2011. Making a Difference for Entrepreneurs*, in www.doingbusiness.org.