

REGOLAMENTAZIONE INCENTIVANTE NEL TRASPORTO
PUBBLICO LOCALE FRA TEORIA E PRATICA:
QUALE LEZIONE DALLO STUDIO DI FRONTIERE DI EFFICIENZA?

BENIAMINA BUZZO MARGARI, FABRIZIO ERBETTA, CARMELO PETRAGLIA e
MASSIMILIANO PIACENZA

pubblicazione internet realizzata con contributo della



***Regolamentazione incentivante nel trasporto pubblico locale fra teoria e pratica:
quale lezione dallo studio di frontiere di efficienza?***

Beniamina Buzzo Margari^a
HERMES (Centro di Ricerca sui Servizi Regolamentati)

Fabrizio Erbetta^b
Università del Piemonte Orientale

Carmelo Petraglia^c
Università Federico II di Napoli

Massimiliano Piacenza^{d*}
Ceris-CNR (Istituto di Ricerca sull'Impresa e lo Sviluppo)

Abstract: Il presente lavoro si propone di valutare, su un campione di operatori pubblici italiani di trasporto pubblico locale, l'impatto dei meccanismi di regolamentazione e delle condizioni strutturali del *network* sui divari di efficienza, attraverso la combinazione delle due metodologie DEA (*Data Envelopment Analysis*) e SFA (*Stochastic Frontier Approach*). In un primo stadio sono state calcolate misure DEA di efficienza e valori di *slack* per ciascun input. In un secondo stadio, questi ultimi sono stati regrediti, utilizzando un approccio *panel* SFA (Battese e Coelli, 1992), su un insieme di variabili esogene indicative sia delle caratteristiche del *network* sia del contesto regolatorio. Il focus, in particolare, viene posto sull'efficacia del cambio di regime di regolamentazione da contratti di sussidio di tipo *cost-plus* a modelli più incentivanti di tipo *fixed-price*. I risultati evidenziano una prevalente portata esplicativa degli effetti esogeni e di disturbo statistico nella determinazione dei divari di efficienza, riservando invece un ruolo minore alla componente di *managerial skill* 'pura', specialmente con riferimento all'impiego di lavoro diretto (conduzione dei veicoli) e ai costi operativi per materiali e servizi. Come conclusione di *policy*, ed in accordo con la *New Theory of Regulation* (Laffont e Tirole, 1993), si rileva come il sistema italiano di regolamentazione dei sussidi, attraverso la definizione di appropriati *cap* di costo, si sia incentrato proprio su quei fattori maggiormente sensibili all'intensificazione dell'*effort*, e per i quali si rendeva quindi necessaria una più decisa azione di razionalizzazione.

Codici JEL: C44; D21; L50; L92; R41

^a b.buzzomargari@hermesricerche.it

^b fabrizio.erbetta@eco.unipmn.it

^c petragli@unina.it

^d m.piacenza@ceris.cnr.it

1. Introduzione

Il lavoro propone una valutazione empirica della teoria dei contratti applicata al settore del Trasporto Pubblico Locale (TPL) italiano, attraverso un modello basato sulla *Data Envelopment Analysis* (DEA) che tiene conto sia della cornice regolatoria (in particolare per quanto riguarda i meccanismi di erogazione dei sussidi) e delle caratteristiche ambientali del *network* servito, sia della componente di errore statistico. La principale implicazione di *policy* è relativa al potere incentivante di una regolamentazione dei sussidi basata su meccanismi *fixed-price* rispetto a quelli di tipo *cost-plus*. In particolare, lo scopo è di verificare in quale misura i differenziali di efficienza tra le imprese (calcolati per ciascun input) sono influenzati dal tipo di regolamentazione e da altri fattori “non discrezionali” (caratteristiche del *network*) su cui il management non ha la possibilità di intervenire direttamente. Inoltre, la metodologia di analisi utilizzata permette di superare uno dei principali limiti dei modelli di frontiera di efficienza non stocastici (come la DEA), prendendo anche in considerazione la componente di errore statistico. La volontà di ricavare misure di inefficienza *input-by-input* trova la sua giustificazione nel fatto che i fattori produttivi impiegati dalle imprese di TPL possono essere razionalizzati in maniera differente. Per tale ragione, è prevedibile che cambiamenti di regolamentazione e/o caratteristiche ambientali possano favorire riduzioni di spreco nell’utilizzo di input maggiormente controllabili (come ad esempio il fattore lavoro) rispetto ad altri meno dipendenti dall’impegno profuso dal management (tra cui, ad esempio, il consumo di carburante).

Una questione cruciale nella letteratura volta ad investigare le determinanti dell’inefficienza di produzione (o *X-inefficiency*¹) in un sistema di TPL è il modo in cui le caratteristiche “non discrezionali” inficiano le performance degli operatori. Da un lato, proprio a causa dell’elevato grado di regolamentazione del settore, differenti meccanismi di sussidiazione paiono in grado di ingenerare differenziali di inefficienza anche significativi, a parità di altre condizioni. D’altro canto, tali differenziali possono anche essere legati a caratteristiche esogene del *network* (come ad esempio la velocità commerciale media), che, almeno nel breve periodo, risultano di difficile controllo da parte dell’impresa. Tali considerazioni mettono in evidenza come sia particolarmente importante, nelle analisi di efficienza condotte in settori regolamentati quali il TPL, il ricorso a metodologie di valutazione delle performance in grado di tenere conto dei fattori esterni che influenzano i livelli di inefficienza. Per questi motivi, il presente lavoro adotta un approccio *multi-step* sviluppato di recente (Fried *et al.*, 2002) che, nell’ambito di un contesto deterministico e non parametrico basato sulla DEA, ha il pregio di incorporare la valutazione dell’impatto di variabili

¹ Leibenstein (1966).

esogene (regolamentazione e caratteristiche ambientali) come pure una componente di disturbo statistico, grazie ad un secondo *step* che utilizza l'approccio della *Stochastic Frontier Analysis* (SFA).

La letteratura precedente si è occupata del ruolo svolto dai differenti meccanismi di sussidiatura nello spiegare i differenziali di inefficienza tra gli operatori di TPL, arrivando a stabilire che le previsioni della teoria degli incentivi (Laffont e Tirole, 1993) contribuiscono a spiegare le differenze in termini di efficienza produttiva tra le imprese. Tale filone di ricerca include i contributi di Kerstens (1996) e Gagnepain e Ivaldi (2002a,b) per il caso francese, e di Dalen e Gomez-Lobo (1997, 2003) per quello norvegese. Più recentemente, Piacenza (2006) ha condotto un'indagine sul TPL italiano, tenendo in considerazione anche le caratteristiche ambientali del *network*. Va sottolineato che tutti questi lavori si sono concentrati sulla valutazione dell'efficienza attraverso la stima di funzioni "aggregate" di costo o di produzione, trascurando gli effetti sui singoli fattori. In particolare, i lavori di Gagnepain e Ivaldi (2002a,b) e Dalen e Gomez-Lobo (1997) hanno il pregio di mettere in relazione diretta l'inefficienza con le asimmetrie informative, evidenziando gli effetti positivi della *regulation*; tuttavia, l'utilizzo di modelli strutturali li obbliga a sottostare ad ipotesi abbastanza restrittive come, ad esempio, quella di una tecnologia di produzione Cobb-Douglas o quella di distorsioni nell'utilizzo degli input (ovvero carenze di produttività) a priori legate esclusivamente al fattore lavoro. Dalen e Gomez-Lobo (2003) e Piacenza (2006) abbandonano l'approccio strutturale per testare gli effetti della *regulation* in un ambito meno vincolato (tecnologia flessibile, nessuna ipotesi a priori sugli input fonte di inefficienza produttiva). Anche in tali lavori, l'approccio utilizzato non permette tuttavia di indagare *ex post* su quali input la *regulation* si mostra più efficace.

Il presente lavoro si propone dunque di innovare rispetto alla letteratura precedente sotto tre punti fondamentali. In primo luogo, la metodologia utilizzata permette di combinare gli aspetti positivi di un approccio non parametrico (DEA), che permette di non dover specificare a priori una forma funzionale per la tecnologia, con quelli di un approccio stocastico (SFA), che consente di valutare l'impatto sulle performance d'impresa dei fattori ambientali esogeni e non controllabili e di tener conto anche della componente di errore statistico. Inoltre, grazie alla caratteristica *multi-step* del modello è possibile analizzare in maniera disaggregata l'effetto di tali variabili esogene sui singoli input. Infine, il livello di inefficienza *input-specific* viene scomposto in tre componenti distinte: la prima attribuibile ai fattori esogeni "non discrezionali" (caratteristiche del *network* e *regulation*), la seconda relativa alla capacità manageriale pura e l'ultima legata alla componente di *random noise*.

Il lavoro è organizzato come segue. Il paragrafo 2 presenta una breve trattazione della teoria della regolazione dei contratti e una sintesi del quadro regolatorio italiano. Nel paragrafo 3 viene introdotta la metodologia utilizzata, principalmente ricavata da Fried *et al.* (2002). Dopo la descrizione del campione oggetto di analisi (paragrafo 4), il paragrafo 5 specifica il modello di frontiera utilizzato, esplicitandone le peculiarità. In primo luogo, viene specificata la tecnologia deterministica e non parametrica di riferimento; viene poi chiarita la rilevanza delle caratteristiche ambientali e della regolamentazione dei sussidi utilizzati nella regressione con SFA, fornendo anche alcune indicazioni sui segni attesi per i coefficienti delle variabili utilizzate come *proxy* dei meccanismi di sussidio e degli altri fattori non discrezionali. Vengono infine presentati i risultati dell'analisi (paragrafo 7) e le conclusioni (paragrafo 8).

2. Regolamentazione dei sussidi, asimmetria informativa e incentivi all'efficienza

Il basso livello delle tariffe conseguente alla caratteristica di *universalità* del servizio implica che di norma i sistemi di TPL non raggiungano il pareggio di bilancio e rende necessaria l'erogazione di *sussidi pubblici* per consentire di ripristinare le condizioni di economicità della produzione, permettendo il consumo del servizio in quantità e secondo standard di qualità altrimenti non garantiti dall'ordinario funzionamento del mercato.² In tale contesto, il problema dell'efficienza produttiva è un tema molto dibattuto poiché se da un lato le tradizionali politiche passate di ripiano *ex post* dei disavanzi d'esercizio hanno condotto verso uno spreco crescente di risorse, d'altro lato la presenza di significative asimmetrie informative a vantaggio dell'impresa (Agente) interferisce con il controllo del livello dei costi da parte del regolatore (Principale), al quale compete la definizione delle modalità di erogazione dei sussidi: l'operatore di TPL generalmente ha una migliore conoscenza dell'efficienza intrinseca della propria tecnologia³ (fenomeno di *adverse selection*); inoltre, risulta assai difficile per l'Autorità locale valutare gli sforzi compiuti dall'impresa (*effort manageriale*) al fine di ottimizzare l'impiego delle risorse produttive di cui dispone (fenomeno di *moral hazard*).

In base alla *nuova teoria della regolamentazione* (NTR - Laffont e Tirole, 1993), quando i rapporti contrattuali Principale-Agente sono caratterizzati da asimmetrie informative, un regolatore che abbia come obiettivo la massimizzazione del benessere collettivo deve ricorrere a meccanismi che inducano l'impresa a rivelare le maggiori informazioni di cui è in possesso e a mostrare il suo reale livello di efficienza. Partendo dal presupposto che non esiste un unico contratto adatto a tutti i

² Riguardo al ruolo dei sussidi nel TPL si veda, ad esempio, Fabbri (1998) e Hensher e Brewer (2001).

³ L'operatore conosce infatti con maggiore precisione il numero di autobus necessario a servire un certo *network*, i consumi di carburante per veicolo, l'abilità alla guida degli autisti, gli effetti della congestione da traffico sui costi operativi e così via.

tipi di imprese, tale obiettivo può essere perseguito lasciando all'impresa la possibilità di scegliere tra diverse opzioni contrattuali (menù di contratti): ogni operatore preferirà il contratto corrispondente al suo livello di efficienza, per cui l'impresa più efficiente sceglierà lo schema caratterizzato da un elevato potere incentivante (*fixed-price*), in cui l'*effort* manageriale è al livello ottimale di *first best*, mentre l'impresa più inefficiente selezionerà il contratto dotato del minor potere incentivante (*cost-plus*), con un *effort* nullo; tra questi due estremi si collocano poi schemi con un diverso grado di incentivo che verranno scelti da imprese caratterizzate da livelli di efficienza intermedi.

Nell'ambito della regolamentazione del TPL, le Autorità locali possono ricorrere a differenti schemi contrattuali per definire la ripartizione del rischio industriale (legato ai costi) e del rischio commerciale (legato ai ricavi) tra gestore del servizio e regolatore. L'utilizzo di meccanismi *cost-plus* implica che l'entità del sussidio venga determinata *ex-post* in base al risultato economico d'esercizio, per cui l'operatore non sopporta alcun tipo di rischio e non ha quindi incentivi ad impegnarsi nella produzione efficiente del servizio⁴. Gli schemi di rimborso *fixed-price*, al contrario, si basano su una previsione *ex-ante* del quantitativo di sussidi da erogare e sono riconducibili a due principali modalità di ripartizione del rischio: i contratti *gross cost*, in cui il regolatore definisce uno standard di costo e il gestore del servizio sopporta unicamente il rischio industriale, e i contratti *net cost*, che, attribuendo i proventi tariffari all'operatore, fissano uno standard sia per i costi che per i ricavi e gravano l'impresa anche del rischio commerciale⁵.

Le considerazioni teoriche sopra esposte hanno stimolato negli anni recenti alcuni studi empirici nei quali si è cercato di valutare il potere di incentivazione all'efficienza produttiva dei diversi tipi di contratti di regolamentazione utilizzati nella pratica per determinare i trasferimenti alle imprese di TPL. A livello europeo, in particolare, sono stati finora analizzati il contesto norvegese (Dalen e Gomez Lobo, 1997, 2003) e quello francese (Gagnepain e Ivaldi, 2002a,b). Gli autori rilevano in primo luogo che, a causa delle difficoltà pratiche di implementazione del menù di contratti ottimale suggerito dalla NTR, nei casi studiati si osservano solo le due categorie estreme dei meccanismi *cost-plus* e *fixed-price*; il principale risultato di queste analisi è che le imprese regolate da contratti di sussidio con specificazione di parametri di costo vincolanti (*fixed-price*) presentano livelli di inefficienza significativamente inferiori rispetto a quelle soggette a

⁴ Tale forma contrattuale è nota nella pratica della regolamentazione come *management contract* (European Commission, 1998).

⁵ La classificazione è tratta da European Commission (1998). In accordo alla teoria della regolamentazione incentivante discussa brevemente sopra, si possono anche ipotizzare tipologie di contratti intermedie tra le diverse categorie "pure" citate, con rischi parzialmente a carico del regolatore e dell'impresa: per esempio, meccanismi di sussidio con rischio commerciale condiviso fra le parti, quindi intermedi fra un *gross cost* e un *net cost contract*, o con incentivi parziali sui guadagni di produttività conseguiti dall'impresa, intermedi cioè fra un *gross cost* e un *management contract* (Boitani e Cambini, 2002b).

contratti di tipo tradizionale con ripiano *ex post* dei disavanzi di gestione (*cost-plus*), confermando così le previsioni teoriche in merito agli effetti positivi degli schemi ad alto potere di incentivo. Dal punto di vista metodologico, tali lavori ruotano tutti intorno al concetto di *frontiera* di costo, uno strumento che permette di superare le restrizioni implicite nell'ipotesi teorica di minimizzazione dei costi, poco plausibile nel contesto reale delle *public utilities* regolamentate, e di ottenere una misura del grado di inefficienza – inteso come deviazione dalla frontiera di *best-practice* – per ciascuna impresa del campione oggetto di studio.⁶ Già Leibenstein (1966) aveva messo in evidenza l'esistenza di una forma d'inefficienza (nota come *X-inefficiency*) dipendente dall'impegno profuso da manager e lavoratori nell'ambito di un determinato processo produttivo: un ambiente caratterizzato da scarsi incentivi all'*effort* comporta carenze di produttività e conduce a sprechi di risorse, mentre adeguati meccanismi di incentivazione possono stimolare riduzioni significative dei costi operativi. In seguito, con l'affermarsi della NTR, lo schema di analisi Principale-Agente e i modelli microeconomici di informazione asimmetrica vengono ad assumere un ruolo centrale nello studio della performance produttiva delle industrie regolamentate. Ciò è particolarmente evidente nei lavori di Dalen e Gomez-Lobo (1997) e Gagnepain e Ivaldi (2002a), in cui gli autori elaborano modelli “strutturali” in cui l'inefficienza complessiva di costo stimata viene direttamente spiegata in termini di asimmetrie informative tra regolatore e impresa, specificando due parametri non osservabili per l'efficienza tecnologica intrinseca (fattore esogeno) e per l'*effort* manageriale (fattore endogeno, dipendente dal meccanismo di sussidio).⁷ In questi aspetti risiede il fondamento teorico della specificazione econometrica di frontiere di efficienza; tale strumento, oltre a fornire indicazioni circa l'efficacia di meccanismi alternativi di regolamentazione utili per finalità di *policy*, consente quindi anche di valutare la rilevanza empirica dell'apparato concettuale sviluppato dalla NTR.

2.1. Il contesto italiano

In Italia (come del resto in molti altri Stati europei) nel corso degli anni 70 e 80 i costi delle imprese di TPL hanno subito incrementi elevati, assolutamente non compensati da pari incrementi degli introiti. Un primo tentativo volto a razionalizzare l'intervento pubblico nel settore si ebbe con la legge 151/1981, che prevedeva una definizione *ex-ante* dell'ammontare dei sussidi basata sulla

⁶ Per una trattazione esaustiva delle tecniche di analisi dell'efficienza basate sulla stima di frontiere di produzione e di costo si rinvia a Kumbhakar e Lovell (2000).

⁷ In tal modo gli autori riescono anche a distinguere nella stima dell'inefficienza globale la componente dovuta a fenomeni di *adverse selection* (non controllabili dall'impresa, almeno nel breve periodo) da quella attribuibile ad azioni di *moral hazard* del manager (su cui agiscono gli incentivi della regolamentazione). A questo riguardo, un risultato di particolare interesse che emerge in Gagnepain e Ivaldi (2002a) è la presenza all'interno del campione di un certo numero di imprese che, sebbene soggette a contratti di tipo *fixed-price*, presentano tuttavia livelli di inefficienza complessiva tra i più elevati: il fenomeno si giustifica con il fatto che in questi casi l'inefficienza intrinseca è talmente marcata che nemmeno uno schema ad alto potere di incentivo riesce a risolvere i problemi strutturali dell'impresa.

differenza tra costi e ricavi stimati, finanziata attraverso uno stanziamento annuale al Fondo Nazionale Trasporti (FNT), che veniva poi ripartito tra le imprese fornitrici del servizio. In tale contesto, lo stanziamento pubblico trovava dunque la sua giustificazione come compensazione economica alle imprese che fornivano un servizio ad un prezzo inferiore al costo e ogni ulteriore deficit avrebbe dovuto essere coperto tramite la vendita del servizio. Nonostante le chiare disposizioni normative, i disavanzi delle imprese sono via via peggiorati anche a causa degli elevati tassi di inflazione di quel periodo (che non hanno trovato adeguato riscontro negli stanziamenti annuali al FNT) e della diminuzione della domanda, costringendo il Governo centrale ad interventi straordinari di ripiano dei disavanzi pregressi. Il manifestarsi di tali situazioni ha però contribuito a vanificare gli effetti di quello che, sulla carta, avrebbe dovuto essere un incentivo a comportamenti efficienti da parte degli operatori: facendo affidamento comunque su un ripiano *ex-post* dei disavanzi d'esercizio, le imprese non hanno attuato particolari politiche di contenimento dei costi e i disavanzi hanno continuato ad aumentare. Pertanto, si può dire che prima del 1996 tutte le imprese di TPL italiane ricevevano in pratica rimborsi di tipo *cost-plus*, con un completo ripiano dei disavanzi di bilancio da parte delle Autorità locali.

Alla fine del 1995, ha preso l'avvio un radicale processo di riforma del settore, il cui intento principale è appunto quello di migliorare le condizioni sia di efficienza sia di efficacia del servizio di TPL.⁸ Tale obiettivo è perseguito in primo luogo attraverso un aumento delle responsabilità finanziarie di tutti i soggetti coinvolti (Autorità locali e imprese). Allo scopo di incoraggiare un utilizzo efficiente delle risorse pubbliche a livello locale, i trasferimenti dal Governo centrale, che passando attraverso il FNT rimanevano vincolati al settore dei trasporti, sono stati eliminati e sostituiti con forme di tassazione regionali, con la conseguente perdita del vincolo di destinazione. Inoltre, la riforma prevede una netta separazione tra le funzioni di programmazione e regolazione (affidate alle Amministrazioni pubbliche) da quelle di gestione industriale (attribuite ad aziende trasformate in società di capitali o in cooperative). Altre importanti novità riguardano poi aspetti di regolazione del mercato, in precedenza del tutto assenti. In primo luogo, viene introdotto l'obbligo di utilizzare meccanismi concorrenziali per l'affidamento dei servizi (in una forma di concorrenza 'per il mercato'), nel tentativo di ridurre i costi di gestione e contemporaneamente incentivare il miglioramento nella qualità dei servizi offerti. Il ricorso a gare per l'affidamento dei servizi concessi in condizioni di monopolio è il principale meccanismo per garantire una forma di concorrenza laddove il mercato libero non è proponibile o è antieconomico: quando ragioni tecniche impediscono l'introduzione della libera competizione in un mercato, l'unica via per ottenere miglioramenti in termini di efficienza degli operatori risiede nel ricorso a forme di

⁸ Per una trattazione della portata innovativa della riforma del TPL italiano si veda Boitani e Cambini 2002a.

concorrenza ‘per il mercato’ (Demsetz, 1968; Laffont e Tirole, 1993). Le esperienze di altri Paesi hanno mostrato come il ricorso a procedure concorsuali per l’affidamento del servizio nel settore del TPL abbia prodotto effetti positivi per quanto riguarda sia gli incrementi di produttività sia i risparmi di costo (Cambini e Filippini, 2003), anche se la definizione della struttura di una procedura di gara nel caso del TPL pone non pochi problemi alle Autorità locali, che devono comunque evitare effetti negativi che si ripercuoterebbero inevitabilmente sugli utenti.

Inoltre, sempre rimanendo in tema di regolazione del mercato, la nuova normativa sancisce l’obbligo di stipulare contratti di servizio tra Enti locali ed aziende produttrici, imponendo che essi siano dotati di certezza e copertura finanziaria per l’intero periodo di validità e che prevedano meccanismi di incentivazione dell’efficienza produttiva (contratti *fixed-price*). Si tratta sicuramente di un’innovazione importante nel panorama del TPL italiano. Tale introduzione ha in primo luogo lo scopo di conferire trasparenza alle relazioni che legano l’Autorità locale (che richiede il servizio) all’impresa (che lo fornisce), esplicitando chiaramente il quadro dei diritti e dei doveri di ciascuna parte e garantendo in tal modo le rispettive aree di autonomia (la pianificazione per l’Ente locale e la produzione del servizio per l’operatore del settore). Le Autorità locali devono in primo luogo definire l’ammontare di servizio necessario a garantire la funzione pubblica del trasporto collettivo (definizione dei ‘servizi minimi’), meritevole della pubblica contribuzione e, in quanto tali, oggetto del contratto di servizio. Inoltre, il risanamento economico del settore cui la riforma mira ha come punti cruciali la richiesta esplicita che tali contratti trovino copertura certa nel bilancio dell’Autorità locale e che contengano meccanismi di incentivo all’efficienza: i trasferimenti dai pubblici bilanci agli operatori di TPL devono infatti essere definiti *ex-ante* nel loro ammontare (contratti *fixed-price*), sulla base dei costi operativi (approccio *gross cost*) o dei deficit (approccio *net cost*) previsti.

Uno degli obiettivi primari del presente lavoro è quello di valutare, attraverso un’analisi di tipo *input-by-input*, se gli operatori di TPL che operano con contratti di tipo *fixed-price* presentano livelli di efficienza più elevati rispetto a quelli soggetti a *cost-plus* grazie ad un maggior *effort* del management, tenendo anche conto delle differenti condizioni del *network*. Inoltre, l’approccio *input-specific* consente di evidenziare più chiaramente su quali input si può concentrare maggiormente il processo di razionalizzazione delle risorse. In tal modo è possibile verificare se i meccanismi di sussidio introdotti dalla riforma hanno sortito l’effetto auspicato di incentivare le performance delle imprese, a conferma di quanto previsto dalla teoria degli incentivi e della regolazione per l’analisi di produzione dei servizi di pubblica utilità.

3. Metodologia di analisi

La metodologia adottata si ispira al lavoro di Fried *et al.* (2002), al fine di arricchire il metodo della DEA con l'analisi degli effetti sulla performance di impresa attribuibili da un lato, a fattori esogeni non discrezionali (schema regolatorio per i sussidi e altre variabili ambientali), dall'altro all'errore statistico. L'obiettivo ultimo delle stime è l'attribuzione dell'impiego inefficiente dei singoli fattori impiegati nel processo produttivo a tre componenti distinte: l'inefficienza manageriale pura, gli effetti causati da variabili esogene (schema regolatorio per i sussidi e contesto ambientale in cui opera l'impresa) e l'errore statistico.

Questo paragrafo descrive la metodologia impiegata, mentre il paragrafo 5 avrà ad oggetto gli aspetti peculiari che emergono dalla sua applicazione al caso dell'industria italiana del TPL, dando conto della specificazione empirica della tecnologia produttiva, dell'introduzione nel modello dei diversi schemi di regolamentazione e della definizione delle altre variabili che definiscono i contorni del contesto ambientale degli operatori dell'industria.

Date I generiche unità di produzione (DMU, *Decision Making Unit*) – per $i = 1, \dots, I$ – impegnate in un processo di trasformazione di N input ($n = 1, \dots, N$) in M output ($m = 1, \dots, M$), il metodo della DEA permette di valutare l'efficienza tecnica caratteristica del processo di trasformazione dell' i -esima DMU, relativamente all'insieme delle I DMU presenti nel campione. L'efficienza tecnica dell' i -esima DMU viene valutata in termini della propria attitudine a produrre il massimo livello di output per una data combinazione di input (*output-oriented* DEA) o, in alternativa, di impiegare la minore quantità possibile di input per ottenere un certo livello di output (*input-oriented* DEA). Naturalmente, la scelta tra le due alternative sull'orientamento della DEA dipende dal grado di controllo che le singole DMU hanno nella scelta dei livelli di input da impiegare o di output da conseguire e, di conseguenza, dalle caratteristiche dell'industria nella quale operano le DMU stesse.

La valutazione dell'efficienza delle singole DMU avviene in termini di distanza delle stesse rispetto ad una comune frontiera "ideale" costruita a partire dai dati osservati di input e output impiegando tecniche di programmazione lineare (Charnes *et al.*, 1978).⁹

L'applicazione del metodo della DEA risulta sia in una misura dell'efficienza tecnica media del processo produttivo delle singole DMU, sia in misure – di tipo *input-specific* – dello spreco di risorse relative all'utilizzo dei singoli input impiegati nel processo produttivo di ciascuna DMU. Nel primo caso, si calcolano degli *score* di inefficienza con valori – compresi tra 0 e 100 – crescenti nel livello di efficienza. Nel secondo caso, si calcolano, per ogni fattore produttivo, i

⁹ La costruzione della frontiera efficiente rispetto alla quale viene valutata l'efficienza relativa delle singole DMU implica l'imposizione di alcune restrizioni teoriche, le più importanti delle quali sono la *free-disposal* e la convessità.

cosiddetti *input slack*, ovvero, valori percentuali crescenti nell'impiego inefficiente dell'input. Entrambe le misure, tuttavia, soffrono di due limitazioni principali, entrambe suscettibili di introdurre delle distorsioni nella valutazione relativa dell'efficienza delle DMU. Infatti, la DEA si basa esclusivamente sui dati osservati di output e input, senza dar conto alcuno di altri fattori osservabili e al di fuori del controllo delle DMU che, contribuendo a determinare un contesto ambientale più o meno favorevole, possono anch'essi influire sul conseguimento di un processo produttivo più o meno efficiente. In secondo luogo, essendo la DEA un modello puramente deterministico, ogni fattore non osservabile viene omesso dall'analisi.

Non considerando nessuna delle suddette possibili fonti di inefficienza, l'applicazione della DEA può risultare in forti distorsioni nella valutazione dell'efficienza delle singole DMU e del loro posizionamento relativo rispetto alle altre DMU presenti nel campione. Seguendo Fried *et al.* (2002), integriamo la DEA con l'approccio SFA, con l'obiettivo di verificare in che misura lo spreco di risorse connesso all'impiego di ciascuno dei fattori produttivi sia effettivamente da imputare all'inefficienza manageriale, e quanta parte sia invece da attribuire agli effetti di variabili esogene e al disturbo stocastico.

L'analisi procede in due *step*. Inizialmente, viene stimato un modello *input-oriented* DEA in ipotesi di rendimenti di scala variabili, risolvendo il seguente problema di programmazione lineare per la *i-esima* impresa (Banker *et al.*, 1984):

$$\min_{\theta, \lambda} \theta \quad [1]$$

sotto i vincoli: $\theta x_i \geq X\lambda$

$$\lambda Y \geq y_i$$

$$\lambda \geq 0$$

$$e^T \lambda = 1$$

dove x_i rappresenta il vettore non negativo ($N \times 1$) degli input della *i-esima* DMU; y_i è il vettore non negativo ($M \times 1$) degli output della *i-esima* DMU; $\lambda = [\lambda_1, \dots, \lambda_I]$ è il vettore ($I \times 1$) dei pesi; $X = [x_1, \dots, x_I]$ è la matrice dei vettori degli input di dimensione ($N \times I$); $Y = [y_1, \dots, y_I]$ è la matrice dei vettori degli output di dimensione ($M \times I$); $e = [1, \dots, 1]$ è un vettore unitario di dimensioni ($I \times 1$). La soluzione al problema di programmazione lineare [1] per ciascuna DMU fornisce i valori ottimali dello scalare $0 \leq \theta \leq 1$, a partire dai quali, si calcolano gli *slack* di inefficienza per ciascun input come segue:

$$s_{ni} = x_{ni} - X_n \lambda; \quad n = 1, \dots, N \quad \text{e} \quad i = 1, \dots, I. \quad [2]$$

Nel secondo *step*, vengono stimate N regressioni del tipo SFA su dati di tipo panel, in ciascuna delle quali la variabile dipendente è l'*n-esimo input slack* e le variabili indipendenti sono z variabili

esogene indicative del contesto in cui opera l'impresa, allo scopo di depurare gli *input slack* ottenuti nel primo *step* dall'impatto dei fattori esogeni e da quello dell'errore statistico, entrambi omessi nel modello DEA. La forma generale delle N regressioni da stimare è la seguente:

$$s_{ni} = f^n(z_i, \beta^n) + v_{ni} + u_{ni} \quad [3]$$

dove $z_i = [z_{1i}, \dots, z_{Ki}]$ è il vettore delle K variabili esogene e β^n sono parametri da stimare. $(v_{ni} + u_{ni})$ è il tipico errore composito dell'approccio SFA, formato dalla componente $v_{ni} \sim (0, \sigma_{vn}^2)$ che rappresenta l'errore stocastico e dal termine u_{ni} distribuito come una normale troncata $-u_{ni} \sim N^+(\mu^n, \sigma_{un}^2)$ – che riflette l'*inefficienza manageriale*. Un'ipotesi alternativa diffusa in letteratura, e imposta nelle nostre stime, è che il termine u_{ni} segua la distribuzione *half-normal* – ovvero, una normale troncata con media zero – per cui $u_{ni} \sim N^+(0, \sigma_{un}^2)$.

Il termine $f^n(z_i, \beta^n)$ viene definito *deterministic feasible slack frontier* e cattura l'impatto dei fattori ambientali sull'*input slack* calcolato con il metodo DEA. Il termine $f^n(z_i, \beta^n) + v_{ni}$, invece, rappresenta la *stochastic feasible slack frontier* e indica il minore *input slack* conseguibile in un contesto stocastico. Date le due ipotesi alternative sulla distribuzione del termine di errore u_{ni} , i parametri da stimare con il metodo della massima verosimiglianza diventano rispettivamente $(\beta^n, \mu^n, \sigma_{vn}^2, \sigma_{un}^2)$ e $(\beta^n, \sigma_{vn}^2, \sigma_{un}^2)$. La componente u_{ni} viene modellata seguendo la specificazione *time-variant* $u_{nit} = u_{ni} \exp[-\eta(t-T)]$, proposta da Battese e Coelli (1992), dove t indica il tempo e T l'ultimo periodo incluso nel panel. Il parametro η è stimato internamente al modello ed indica la dinamica temporale dell'efficienza manageriale. Nella stima di massima verosimiglianza delle singole equazioni [3] la varianza dell'errore composito $(u_{ni} + v_{ni})$ è data da $\sigma_n = \sigma_{un} + \sigma_{vn}$, mentre la proporzione della varianza totale attribuibile ad inefficienza manageriale residua è data dal parametro $\gamma = \sigma_{un}/\sigma_n$, compreso tra 0 ed 1. Un valore nullo indica che il modello è assimilabile ad una funzione media, stimabile attraverso il ricorso al metodo tradizionale OLS, mentre un valore pari ad 1 significa che la frontiera è deterministica, ovvero priva della componente di errore v_{ni} . L'ipotesi $\gamma = 0$ viene verificata attraverso un test LR (Coelli *et al.*, 1998).

4. Descrizione del campione

Il *panel* bilanciato è composto da 42 imprese di TPL italiane (di proprietà pubblica) osservate nel periodo 1993-1999, per un totale di 294 osservazioni. Si tratta di un campione che rappresenta abbastanza bene il panorama italiano e che comprende 6 operatori di piccole dimensioni (con meno di 150 addetti), 19 imprese medie (con un numero di addetti compresi tra i 150 e i 550) e 17 di grandi dimensioni (con oltre 550 addetti). Per quanto riguarda la tipologia di servizio prestato, 15 imprese operano prevalentemente in un contesto urbano, 10 servono contesti extraurbani e le

rimanenti 17 svolgono la loro attività in entrambi i comparti (imprese *miste*). Per distribuzione geografica, 23 operatori sono localizzati nel Nord Italia e 19 nelle Regioni centrali o meridionali. Infine, per quanto concerne il meccanismo adottato per l'erogazione dei sussidi, circa il 24% delle osservazioni riguarda situazioni soggette a contratti di tipo *fixed-price* (71 casi), mentre il restante 76% (223 osservazioni) è ancora legato tipologie di regolamentazione basate su meccanismi *cost-plus*.

Il database utilizzato è stato costruito avvalendosi di fonti informative differenti. La maggior parte dei dati economici e produttivi – come i posti-chilometro forniti annualmente, il numero totale dei dipendenti, il carburante consumato, la dimensione del parco veicoli e il totale dei costi operativi – sono stati ricavati dagli Annuari di ASSTRA, l'Associazione nazionale delle imprese pubbliche di TPL, che nell'anno 2000 contava 165 membri, corrispondenti al 90% dei gestori del trasporto urbano ed al 50% dei gestori nel settore extraurbano. Le informazioni disaggregate sulle differenti caratteristiche tecniche ed ambientali (come ad esempio la ripartizione del personale dipendente tra personale diretto ed indiretto, l'età media della flotta, la velocità commerciale media del *network* o la densità della popolazione nell'area servita), così come la dimensione delle differenti categorie di costo (lavoro, carburanti e materiali e servizi), sono stati ottenuti attraverso la predisposizione di specifici questionari inviati direttamente alle imprese. Anche le informazioni relative al meccanismo di erogazione dei sussidi sono state ricavate attraverso un'apposita domanda nel questionario, in cui è stato chiesto di indicare se l'Ente locale competente (Regione, Provincia o Comune) adottasse forme di rimborso di tipo *cost-plus* o *fixed-price*.

5. Specificazione del modello

5.1. Input e output della frontiera DEA

In linea con la prevalente letteratura DEA e SFA, nel primo *step* DEA viene utilizzata una prospettiva di analisi *input-oriented*: l'ipotesi sottostante è che le imprese di TPL cerchino di minimizzare l'utilizzo degli input fermo restando il quantitativo di output prodotto.

Un momento cruciale di ogni analisi DEA è quello della definizione del set di variabili in grado di rappresentare in maniera corretta ed esaustiva la tecnologia del settore sottostante (Thanassoulis, 2001); è perciò necessario identificare sia l'insieme degli input utilizzati nel processo produttivo sia gli output di tale processo. Come si è detto, in questa prima fase non vengono presi in considerazione i fattori ambientali che possono influenzare la trasformazione degli input negli output: il loro effetto sul grado di efficienza delle singole imprese, sarà valutato nel secondo *step*, attraverso la regressione SFA. Alla luce di tali considerazioni e date le differenze in termini di struttura del servizio e di condizioni ambientali che caratterizzano le diverse tipologie

di servizio, verranno stimate tre distinte frontiere DEA di comparto. Perciò non sarà possibile confrontare direttamente gli *score* di efficienza, anche se attraverso un'analisi della varianza sarà possibile individuare il comparto in cui gli scostamenti medi sono minori e quindi in cui l'efficienza media degli operatori è superiore.

Per quanto concerne la definizione dell'output, è possibile stimare le condizioni di efficienza in un contesto legato al lato della domanda (misure *demand-oriented*) o dell'offerta (misure *supply-oriented*). Nel caso del TPL, Kerstens (1996) rileva come una misura di tipo *supply-oriented* sia maggiormente indicata dal momento che essa riflette condizioni meglio controllabili da parte del management rispetto alla domanda di trasporto. Andrebbe, a tal proposito, rilevato che le Autorità di regolamentazione impongono agli operatori vincoli abbastanza stringenti in termini di livello di servizio offerto: la definizione dei "servizi minimi" dovrebbe infatti non solo tener conto del flusso di passeggeri, ma anche rispettare la funzione sociale del servizio di TPL. E' comunque ragionevole ipotizzare che tali vincoli alla produzione siano il risultato di un processo di concertazione con le Autorità locali. Per tali motivi, e seguendo Piacenza (2006), nella presente analisi è stata preferita una misura di output orientata al lato dell'offerta.

Convenzionalmente, nel caso del TPL vengono impiegate due tipologie di output di tipo *supply-oriented*: i veicoli-chilometro (*VKM*) ed i posti-chilometro (*PKM*). Il primo rappresenta il numero totale di chilometri percorsi in un anno da tutti i veicoli, ed è quindi calcolato come prodotto tra il numero di vetture e il numero di chilometri medi. La seconda misura di output permette di correggere il precedente indicatore per le differenze in termini di posti offerti in media da ciascuna vettura, essendo definita dal prodotto tra *VKM* e la capacità media dei veicoli della flotta, espressa appunto dai posti medi. Questo indicatore fornisce un'informazione più dettagliata sulla quantità di servizio offerta, dal momento che tiene conto anche delle differenziazioni in termini dimensionali del servizio. Per tali motivi, nella presente analisi si è preferito scegliere *PKM* come misura di output.

Dal lato degli input vengono utilizzati tre fattori, anch'essi ricorrenti in letteratura:¹⁰ lavoro, carburante consumato e costo per materiali e servizi. Alla luce del fatto che il costo per gli autisti risulta avere un'incidenza rilevante sul costo totale del lavoro e dovrebbe essere direttamente connesso alla quantità di servizio prestato, il dato relativo al personale totale è stato suddiviso in due categorie: il personale diretto (*lavDR*) e quello indiretto (*lavIND*). Il carburante (*FUEL*) è misurato dall'ammontare di litri di gasolio consumati nell'anno (o dai kilowatt-ora di energia equivalenti). Poiché la categoria "materiali e servizi" non è al suo interno omogenea, non è possibile identificare una misura fisica adeguata per questo tipo di input. Si è perciò scelto di

¹⁰ Per una rassegna dettagliata degli input utilizzata nel settore dei trasporti pubblici si veda De Borger *et al.* (2002).

utilizzare il corrispondente valore di costo (*CMS*) – opportunamente deflazionato¹¹ – ricavato dai bilanci delle imprese come differenza tra il costo operativo totale e i costi per il personale e per il carburante. Va rilevato come la maggior parte dei costi inclusi in questa categoria riguardino i costi per la manutenzione direttamente effettuata dall'impresa o affidata in *outsourcing*.

La tabella 1 mostra per ciascun comparto una descrizione degli input e dell'output utilizzati nell'analisi.

Tabella 1: Statistiche descrittive delle variabili di input ed output

Tipologia di servizio	<i>PKM</i> (10⁶)^a	<i>lavDR</i>^b	<i>lavIND</i>^b	<i>FUEL</i>(10³ litri)^c	<i>CMS</i>(10⁶ lire)^d
<i>Urbano</i>					
Media	1,166	532	274	5,565	14,242
Deviazione Standard	1,507	602	392	7,446	20,425
Minimo	48	23	4	233	435
1° quartile	246	133	42	1,269	3,943
Mediana	883	344	110	3,865	9,557
3° quartile	1,199	595	307	6,299	14,164
Massimo	6,554	2,758	1,698	33,143	102,436
Indice di variabilità*	129%	113%	143%	134%	143%
<i>Misto</i>					
Media	1,062	507	212	4,999	16,254
Deviazione Standard	973	535	310	4,397	18,939
Minimo	59	30	4	352	1,184
1° quartile	427	173	54	2,168	4,885
Mediana	891	369	97	4,120	9,704
3° quartile	1,102	464	174	5,335	15,822
Massimo	3,909	2,290	1,323	18,103	86,427
Indice di variabilità*	92%	106%	146%	88%	117%
<i>Extraurbano</i>					
Media	820	335	163	4,035	9,285
Deviazione Standard	578	288	258	2,667	6,064
Minimo	148	78	26	1,037	1,924
1° quartile	326	196	60	1,894	4,118
Mediana	735	250	79	3,673	7,753
3° quartile	1,122	342	151	4,775	11,114
Massimo	2,043	1,339	1,307	10,500	24,536
Indice di variabilità*	71%	86%	158%	66%	65%

* L'indice di variabilità è calcolato come rapporto tra la deviazione standard e la media.

^a *PKM* = numero di posti-chilometro offerti nell'anno

^b *lavDR* e *lavIND* = rispettivamente numero di autisti e personale indiretto.

^c *FUEL* = litri di carburante consumato.

^d *CMS* = valore reale delle spese annue per materiali e servizi.

¹¹ Come deflatore è stato utilizzato l'indice dei prezzi alla produzione.

Seguendo Fried *et al.* (2002), il capitale non è stato considerato tra gli input della frontiera DEA. Almeno nel breve periodo, il capitale è un input non discrezionale e l'imposizione di un input quasi-fisso nell'algoritmo di minimizzazione DEA avrebbe portato ad un eccessivo restringimento dell'insieme dei *peers*, enfatizzando eccessivamente l'unicità di ciascuna impresa e di conseguenza restituendo valori di efficienza prossimi all'unità per la maggior parte delle osservazioni.¹² Il capitale è stato comunque tenuto in considerazione nel secondo *step* con l'introduzione di variabili *dummy*, in cui la dimensione è misurata in termini di numero di veicoli della flotta. Tale soluzione permette di evitare le difficoltà e l'inevitabile imperfetta costruzione di una serie temporale dei costi per il rinnovo del capitale, corretto anche per tener conto dell'inflazione e/o delle rivalutazioni. Soprattutto queste ultime avrebbero potuto costituire un serio problema, dal momento che nel corso del periodo osservato la maggior parte delle imprese ha subito una trasformazione da impresa municipalizzata a società di capitali. Come verrà meglio precisato in seguito, l'introduzione di tali *dummy* è primariamente collegato alla possibilità di verificare l'impatto delle specifiche condizioni di sussidiazione dei maggiori *network* (coincidenti di norma con situazione di maggiore congestione) e le conseguenze in termini di sovra-investimento e sprechi di risorse.

5.2. Regolamentazione e variabili ambientali

Come si è detto sopra, nel confrontare i livelli di efficienza non si può prescindere dal contesto specifico in cui l'impresa si trova ad operare. Infatti, le caratteristiche del *network* possono tradursi in vantaggi o svantaggi per l'operatore, influenzandone anche pesantemente il grado di efficienza. In altre parole, un maggiore o minore punteggio di efficienza potrebbe anche essere attribuibile all'azione positiva o negativa di alcune variabili esogene (legate alle caratteristiche ambientali del contesto operativo o regolatorio), piuttosto che ad una reale abilità o incapacità del management. I primi risultati ottenuti con l'analisi DEA devono pertanto essere corretti in qualche modo per tener conto dell'impatto di tali caratteristiche esogene, che non possono essere considerate alla stregua di input o output, poiché non sono controllabili dal management. Vengono di seguito discussi le principali problematiche microeconomiche legate alle variabili incluse nell'analisi, chiarendo di volta in volta gli effetti attesi su ogni input considerato.

Innanzitutto, alla luce della variazione intercorsa nel periodo sotto osservazione in materia di regolamentazione dei sussidi, nel modello è inserita una variabile *dummy* (*REG*) che, per ciascuna impresa, assume valore 1 dopo l'introduzione dei contratti *fixed-price* e 0 altrimenti. Sulla base anche delle evidenze empiriche emerse in altri studi (Kerstens, 1996; Dalen e Gomez-Lobo, 1997,

¹² Per maggiori dettagli sul trattamento degli input quasi-fissi si vedano Coelli *e al.* (1998) e Thanassoulis (2001).

2003; Gagnepain e Ivaldi, 2002a,b; Piacenza, 2006), ci si aspetta che l'introduzione di schemi *fixed-price* incrementi i livelli di efficienza, riducendo gli *slack* nell'uso delle risorse, almeno per quanto riguarda quegli input maggiormente suscettibili di razionalizzazione, come il personale diretto ed i costi per materiali e servizi.

Per quanto riguarda invece le caratteristiche non discrezionali, esse includono tanto variabili connesse alla struttura della rete – e cioè la velocità commerciale media e la densità della popolazione – quanto altri fattori ambientali, non direttamente controllabili dal management dell'impresa – come l'età media del parco veicoli, la scala operativa ed il cambiamento tecnologico (Dalen e Gomez-Lobo,2003; Piacenza 2006). La velocità commerciale media (*SPEED*) è comunemente riconosciuta come una caratteristica rilevante nell'influenzare l'efficienza di un *network* ed è solitamente utilizzata come *proxy* dei costi di congestione. Infatti, al diminuire della velocità commerciale media (indice dell'aumentare della congestione di una rete o della una carenza di corsie preferenziali) ci si aspetta un incremento sia delle vetture per chilometro sia del personale diretto. Inoltre, fermate più frequenti e maggiori periodi di sosta provocati da strade troppo affollate influiscono negativamente anche sul consumo di carburante. Infine, a livelli inferiori di velocità del *network* sono solitamente associabili maggiori costi per la manutenzione e per la riparazione dei veicoli, a causa del maggior tasso di incidentalità.¹³ In sostanza, ci si aspetta un effetto positivo sull'efficienza tecnica all'aumentare della velocità commerciale, con una riduzione degli *slack* per il personale diretto, il carburante e i costi per materiali e servizi, anche se non è a priori definibile l'impatto sul personale indiretto.

La densità della popolazione (*DENS*) è stata calcolata come popolazione per chilometro quadrato nell'area servita da ciascun operatore. Analogamente a quanto detto per la velocità commerciale, anche la densità può essere interpretata come misura della congestione della rete e come tale contribuire ad abbassare i livelli di efficienza. D'altro canto, tale variabile può influenzare anche il modo in cui un *network* è configurato, con maggiori o minori fermate lungo le linee, e quindi sortire effetti positivi sul grado di efficienza tecnica. Perciò, seguendo Kerstens (1996), non si hanno aspettative a priori circa l'effetto della densità della popolazione sugli *slack*.

L'età media del parco veicoli (*AGE*) è un dato che riflette l'età media delle vetture utilizzate per la fornitura del servizio. Dato che i veicoli più nuovi hanno una capacità di carico maggiore rispetto a quelli più vecchi, ci si può aspettare che al diminuire dell'età media del parco diminuiscano sia il numero di veicoli necessari alla produzione del servizio sia il numero di autisti. Una minore età media potrebbe, poi, ed in misura più significativa, avere ricadute sui consumi di carburante (in virtù di nuove tecnologie a basso regime di consumo) e sul costo per materiali e

¹³ Infatti, se la velocità del network si riduce, aumenta la probabilità di incidenti, anche se forse diminuisce la gravità degli stessi.

servizi. In quest'ultimo caso, un'età maggiore è ragionevolmente associata a guasti meccanici più frequenti, con conseguenti maggiori costi per riparazioni e manutenzioni. D'altro canto però, come osserva Boame (2004), la presenza di veicoli nuovi può comportare una minore familiarità con essi da parte degli autisti, e tradursi in un più elevato tasso di incidentalità, che aumenta anche i tempi necessari per le riparazioni. In generale dunque, ci si aspetterebbe che al diminuire dell'età media del parco veicoli sia associata una riduzione degli *slack* relativi al carburante ed al costo per materiali e servizi.

Nel secondo *step* è anche incluso un trend temporale (*TR*), in modo da tener conto del cambiamento tecnologico all'interno del settore. Anche se è usuale ipotizzare un progresso tecnico, non vengono poste aspettative a priori, soprattutto dal momento che l'analisi è condotta *input-by-input*. La variabile *TR* cattura il mero spostamento nel tempo della frontiera tecnologica e non i cambiamenti delle performance manageriali, che sono catturati invece dalla componente di inefficienza *one-side* specificata dall'approccio SFA di Battese e Coelli (1992).

La dimensione operativa delle imprese di TPL viene considerata nel modello attraverso l'introduzione di variabili *dummy* (*dBUS1*, *dBUS2*, *dBUS3* e *dBUS4*). I quattro gruppi sono stati definiti sulla base della suddivisione per quartili del numero di veicoli utilizzati nella produzione del servizio. In particolare, la categoria *dBUS1* comprende le imprese più piccole, mentre *dBUS4* include quelle di dimensioni maggiori. La variabile *dBUS1* è stata omessa in modo che i parametri delle altre categorie possano fornire una misura di come gli *slack* di inefficienza si modificano all'aumentare della dimensione operativa delle imprese. Va notato come in situazioni di elevata congestione da traffico le Autorità locali siano maggiormente propense alla promozione dei servizi di trasporto pubblico. E' il caso per esempio dei grandi *network* urbani, in cui si verificano grossi problemi di circolazione soprattutto nelle ore di punta, con ripercussioni sociali, come il peggioramento della qualità della vita causato dall'inquinamento atmosferico ed acustico, che richiedono misure di intervento da parte del Governo locale. E' verosimile che in tali situazioni gli operatori beneficino di sussidi proporzionalmente più elevati per sostenere la necessità di maggiori frequenze di servizio ed investimenti infrastrutturali, che consentano di ovviare ai problemi di circolazione nelle ore più critiche. Conseguenza di ciò potrebbe essere una sovra-dotazione della flotta rispetto alla domanda potenziale di servizio, che si ripercuotono inevitabilmente anche sulle quantità di input immesse nel processo produttivo, soprattutto per quanto riguarda gli input più direttamente correlati alla quantità di servizio offerto (come gli autisti, il carburante e il costo per materiali e servizi). E' perciò interessante valutare l'impatto delle politiche di sussidio agli investimenti sull'uso efficiente degli input, utilizzando come *proxy* le *dummy* relative alla dimensione della flotta.

Infine, il servizio di TPL può essere fornito in un contesto urbano, extraurbano oppure misto (con alcuni tratti qualificabili come urbani ed altri dalle caratteristiche tipicamente extraurbane). Per tener conto delle caratteristiche del comparto servito in questo secondo *step* vengono introdotte delle variabili *dummy* (*dURB* per il trasporto urbano, *dINTC* per quello extraurbano ed infine *dMIX* per i servizi di tipo misto). Come si preciserà meglio nel paragrafo 6.2, a cui si rimanda, tali variabili *dummy* hanno lo scopo di cogliere differenze non tanto nei livelli di efficienza tra comparti, quanto nelle diverse condizioni di variabilità delle performance, all'interno di ciascun segmento dell'industria, nei confronti delle rispettive frontiere.

6. Evidenza empirica

Nel presentare e discutere i nostri risultati, daremo particolare enfasi all'impatto che le variabili di contesto (regolamentazione e caratteristiche del network) hanno prodotto sulle condizioni di efficienza specifiche di ogni input.

Tale scelta, in particolare, è motivata dal fatto che gli studi empirici sul TPL esistenti (Dalen e Gomez-Lobo, 1997 e 2003; Gagnepain e Ivaldi, 2002a e 2002b, Kerstens, 1996; Piacenza, 2006) si sono concentrati sulla valutazione dell'efficienza attraverso la stima di funzioni "aggregate" di costo o di produzione, trascurando gli effetti sui singoli fattori. Gagnepain e Ivaldi (2002a e 2002b) e Dalen e Gomez-Lobo (1997), inoltre, attraverso modelli strutturali che esplicitano la relazione di asimmetria informativa tra regolatore e impresa, hanno individuato a priori nel fattore lavoro la causa primaria dell'inefficienza. Pertanto, un'ulteriore finalità di questo lavoro è quella di verificare empiricamente la rispondenza pratica di tale ipotesi.

6.1. Determinazione degli score di efficienza DEA

La Tabella 2 presenta i valori medi di efficienza DEA del primo *step*, calcolati utilizzando dati *pooled* e separatamente per comparto di attività. Tale analisi preliminare evidenzia un livello medio di efficienza pari a circa il 93% per le imprese urbane, e pari all'85-86% per le imprese extra-urbane e per quelle presenti su entrambi i segmenti dell'industria (miste).

Pare utile, a tal proposito, precisare che, data la stratificazione per comparti del campione, un confronto orizzontale tra condizioni relative di efficienza dei singoli raggruppamenti risulta impraticabile. Piuttosto, i divari di efficienza interni ad ogni gruppo dovrebbero essere attribuiti al grado di variabilità, come confermato dall'osservazione degli indicatori di varianza. Nello specifico, il maggior livello di efficienza dei network urbani trova coerenza nel minore scostamento (medio) delle unità osservate rispetto al *benchmark* empirico.

Tabella 2: Score DEA di efficienza

	Network Urbani DEA	Network Misti DEA	Network Extra-urbani DEA
Efficienza media	93.11	85.45	86.46
Efficienza minima	73.01	64.72	48.20
1° quartile	87.97	77.99	77.33
2° quartile	95.33	83.32	93.27
3° quartile	100.00	97.51	100.00
Varianza	7.13	10.33	15.75
Num. osservazioni efficienti	30	22	19

In ogni caso, pur non conoscendo le posizioni relative delle frontiere specifiche di comparto – e, quindi, non potendo esprimere un giudizio comparativo di efficienza – sembrerebbe emergere la necessità, sia per le imprese miste che per quelle extraurbane, di un più incisivo raffronto comparativo. Attraverso quest’ultimo si renderebbe, infatti, possibile realizzare una omogeneizzazione delle performance, assorbendo le situazioni di elevata inefficienza (si osservino a tal proposito i valori minimi) che in tali contesti vengono a presentarsi.

Tuttavia, i risultati desunti dal primo *step* difficilmente possono essere considerati come effettivamente rappresentativi dell’(in)efficienza manageriale, a causa della mancata considerazione dell’azione dei fattori esogeni che danno specificità ai diversi contesti operativi.

6.2. Impatto della regolamentazione e delle caratteristiche ambientali

La considerazione delle variabili “di contesto”, oggetto delle stime del secondo *step*, può essere inquadrata secondo una duplice finalità. Da un primo, e più importante, punto di vista, essa permette di considerarne l’impatto sulle condizioni di inefficienza separatamente per ogni singolo input, innovando, quindi, rispetto alla letteratura precedente. In secondo luogo, essa consente di scomporre gli *slack* desunti dal primo *step*, separando l’inefficienza indotta da condizioni operative esogene e dalla regolamentazione, oltre che da disturbi stocastici, dall’inefficienza manageriale residuale.

La Tabella 3 presenta i risultati delle stime SFA condotte sui singoli fattori produttivi prescelti per l’analisi DEA. Il coefficiente *SPEED* appare negativo e significativo, ad eccezione dell’input *CMS*. Ciò significa che un incremento della velocità commerciale, e, quindi, una riduzione della congestione del traffico lungo il *network*, è in grado di migliorare le condizioni di gestione degli input. Tale risultato pare di particolare interesse per le autorità locali designate alla progettazione

della viabilità urbana ed extraurbana, dal momento che una maggiore fluidità del traffico è in grado di consentire un recupero di efficienza da parte dei fornitori del servizio.

Oltre a tale caratteristica strutturale della rete, anche il contesto regolatorio (*REG*) si contraddistingue per un impatto positivo sull'efficienza. In effetti, a seguito della progressiva sostituzione di contratti di affidamento di tipo *cost-plus* con contratti incentivanti *fixed-price* si è avuta una sostanziale riduzione degli *slack* di efficienza relativamente a dipendenti diretti (*lavDR*) e costi per materiali e servizi (*CMS*). Tale risultato pare di un certo rilievo, dal momento che, oltre a confermare il ruolo dell'assegnazione del servizio sulla base di nuove forme contrattuali incentivanti, consente di fornire una risposta, conforme alle attese teoriche, in merito a quali siano i fattori maggiormente razionalizzabili. Due sono le considerazioni che ne conseguono. Da un primo punto di vista, si può argomentare che l'introduzione di vincoli di *cap* più stringenti su dipendenti diretti e costi per materiali e servizi possa rappresentare una strategia premiante rispetto ad una generica ed omogenea attribuzione dei *cap* anche su componenti di costo poco o per nulla modificabili. In secondo luogo, tale riduzione degli *slack* parrebbe possibile proprio alla luce di un sovra-dimensionamento iniziale di questi fattori.

Per quanto concerne le variabili *DENS* e *AGE* non si riscontrano risultati significativi. Mentre nel primo caso non si è posta un'aspettativa *a priori* sul segno del coefficiente, nel secondo i parametri segnalano una significatività marginale (*p-value* di poco superiore al 10%) con riferimento all'input *CMS*. In effetti, pare ragionevole assumere che i costi per manutenzione e riparazione si incrementino in funzione dell'innalzamento dell'età media del parco veicoli. Anche per quanto concerne il fattore *FUEL*, il segno appare positivo, coerentemente alle aspettative, ma su un livello di scarsa significatività (*p-value* pari a 0,177).

Tabella 3: Impatto della regolamentazione e delle variabili ambientali sugli *slack* DEA

Variabili	<i>slack-emplDR</i>			<i>slack-emplIND</i>			<i>slack-FUEL</i>			<i>Slack-CMS</i>		
	Parametri	t-stat	p-value	Parametri	t-stat	p-value	Parametri	t-stat	p-value	Parametri	t-stat	p-value
<i>SPEED</i>	-1,898	-2,46	0,014	-1,839	-3,10	0,002	-21,210	-3,33	0,001	-53,956	-1,39	0,164
<i>DENS</i>	-0,007	-1,14	0,255	0,004	1,08	0,279	-0,020	-0,39	0,697	-0,027	-0,08	0,932
<i>AGE</i>	-1,444	-0,87	0,385	0,419	0,36	0,719	20,949	1,35	0,177	138,475	1,48	0,138
<i>REG</i>	-17,817	-2,09	0,036	0,110	0,02	0,987	-95,203	-1,15	0,252	-776,511	-1,95	0,052
<i>TR</i>	7,816	3,54	0,000	-2,159	-1,21	0,227	-53,870	-2,19	0,029	223,943	1,56	0,118
<i>dBUS2</i>	14,460	1,16	0,247	11,786	1,24	0,215	138,483	1,28	0,201	482,460	0,79	0,427
<i>dBUS3</i>	51,739	3,69	0,000	27,187	2,63	0,008	434,319	3,28	0,001	1,922,477	2,90	0,004
<i>dBUS4</i>	12,282	0,83	0,405	31,338	2,78	0,005	421,073	2,97	0,003	1,114,470	1,54	0,123
<i>DINTC</i>	46,039	2,43	0,015	49,606	3,42	0,001	400,607	2,07	0,038	665,922	0,66	0,509
<i>DMIX</i>	67,757	3,80	0,000	39,486	3,29	0,001	852,339	5,44	0,000	2,243,988	2,85	0,004
η	0,270	11,18	0,000	-0,048	-1,27	0,203	-0,077	-2,17	0,030	0,154	2,94	0,003
γ	0,436	4,40	0,000	0,689	9,12	0,000	0,835	17,87	0,000	0,429	3,16	0,000
LR test $\gamma = 0$	249,322		0,000	84,595		0,000	121,198		0,000	89,308		0,000
Wald χ^2 test	52,050		0,000	21,09		0,020	60,02		0,000	24,89		0,006
Osservazioni	294			294			294			294		
Log-likelihood	-1.618,554			-1.490,153			-2.208,112			-2.712,801		

La dimensione del capitale impiegato è stata introdotta nel modello attraverso la considerazione delle variabili *dummy* *dBUS2*, *dBUS3* e *dBUS4*. Come si può osservare, la maggiore dimensione esercita in genere un effetto di segno positivo e significativo sull'inefficienza delle diverse tipologie di input. Per tali *network* pare, dunque, esistere un problema di sovra-investimento e di conseguenti sprechi nella gestione operativa. Più nello specifico, il risultato potrebbe essere attribuito ad un problema di elevata congestione dei grandi *network*, tipicamente operanti nelle maggiori aree urbane, e, per tale motivo, maggiormente sovvenzionati attraverso finanziamenti in conto capitale da parte delle autorità pubbliche locali, interessate a perseguire obiettivi di politica sociale. Tali finanziamenti, assegnati di norma con vincolo di destinazione per l'incremento del parco veicoli, avrebbero ridotto il costo-opportunità del capitale investito, causando una sovra-dotazione di capitale rispetto alle effettive necessità dettate dalla domanda di servizio, con conseguenti sprechi nell'utilizzo dei fattori.

Nelle equazioni stimate nel secondo *step* sono inoltre state inserite *dummy* di comparto (*dINTC* e *dMIX*), al fine di rappresentare le specificità dei diversi segmenti dell'industria. A questo proposito è necessario, tuttavia, procedere ad una puntualizzazione circa il ruolo di tali variabili di categoria nelle stime di frontiera. Dal momento che le stime DEA iniziali sono state condotte separatamente su sub-campioni corrispondenti ai comparti, gli *score* globali così come gli *slack* possono essere unicamente interpretati come inefficienze all'interno dei singoli raggruppamenti. Semmai, sarebbero le condizioni di variabilità dell'efficienza, nei confronti delle rispettive frontiere di *best practice*, a poter essere oggetto in un confronto orizzontale. Dato tale focus, i coefficienti positivi e significativi corrispondenti alle *dummy* qui discusse non dovrebbero essere interpretati in termini di maggiore inefficienza dei comparti misto ed extraurbano rispetto a quello urbano (*dummy* omessa), bensì in termini di maggiore variabilità delle condizioni di efficienza intrinseca di tali segmenti. La necessità di introdurre nel modello tali ulteriori variabili di controllo è stata dettata dall'esigenza di non riversare tale eterogenea variabilità delle stime del primo *step* sulla componente di *noise* dell'errore composito, con conseguente distorsione dei parametri. Coerentemente con questa impostazione, pare dunque necessario porre l'accento sull'impossibilità di interpretare i coefficienti delle variabili *dummy* di comparto come il riflesso di diseconomie legate all'appartenenza a comparti diversi da quello urbano¹⁴.

Infine, è utile dare rilievo alla variabile di trend (*TR*). La tendenza di fondo legata al tempo mostra un progresso significativo per quanto concerne l'utilizzo dell'input carburante (dovuto, per esempio, all'introduzione di tecnologie a più basso consumo) ed un regresso con riferimento alle

¹⁴ Per un'analisi comparativa delle condizioni di efficienza attribuibili al comparto si veda Fraquelli et al. (2004).

condizioni di efficienza nell'impiego del lavoro diretto e nella gestione dei costi per materiali e servizi (anche se in questo secondo caso il dato risulta solo marginalmente significativo).

I risultati presentati in Tabella 3 indicano chiaramente come la variabilità delle condizioni di efficienza tra imprese del TPL sia spiegata in larga misura da differenziazioni delle caratteristiche ambientali dei *network*, oltre che dall'evoluzione del quadro regolatorio nazionale, a cui le varie regioni hanno risposto in tempi e modi diversi.

Una volta riportate le differenti unità osservate su un medesimo piano in termini di contesto di riferimento (ambientale e "politico"), il peso dell'inefficienza manageriale pura, catturata dal termine u_{ni} , risulta particolarmente basso, come si desume dall'osservazione del parametro γ . Più nello specifico, si può argomentare che, per quanto concerne il lavoro diretto e le spese per materiali e servizi, ovvero proprio quelle voci che maggiormente incidono sul costo operativo complessivo e che risultano maggiormente sensibili a processi di ristrutturazione, solo il 43% della variabilità complessiva dell'errore composito è attribuibile a fattori manageriali puri.

Per quanto la componente manageriale in senso stretto presenti, per i due fattori analizzati, uno scarso peso sulla variabilità totale delle performance, essa tende, comunque, a ridursi nel tempo, come indicato dal segno positivo e significativo del parametro di direzione η del modello di Battese e Coelli (1992). I parametri stimati sono pari a 0,270 per il fattore lavoro diretto e a 0,154 per i costi di materiali e servizi. Ciò potrebbe essere in parte attribuito al cambiamento, stemperato nel tempo ma progressivo, delle forme di *governance*. In sintesi, gli interventi normativi degli inizi degli anni '90 hanno individuato una direzione precisa nel passaggio da aziende municipalizzate ad aziende speciali, fino alla costituzione di vere e proprie S.p.A. (anche se sempre nell'ambito della proprietà pubblica). Il passaggio, più o meno diretto a seconda dei casi, ha condotto ad un allontanamento degli Enti Locali da forme di coinvolgimento diretto nella gestione a favore di situazioni di governo più autonome. In aggiunta, tale processo si è spesso realizzato seguendo forme di concertazione tra dimensione politica e manager, così da rendere il passaggio graduale e non inatteso, stimolando in tal modo il grado di *effort* da parte delle strutture dirigenziali chiamate a gestire il cambiamento.

7. Conclusioni

Il presente lavoro innova rispetto alla letteratura precedente sull'analisi dell'efficienza nel settore del TPL e, più in generale, nei settori soggetti a regolamentazione, per almeno tre ordini di ragioni.

Innanzitutto, l'approccio utilizzato, derivato da Fried *et al.* (2002), consente di utilizzare la metodologia DEA – la quale, in linea di principio, possiede il vantaggio di non specificare *a priori* una forma funzionale della tecnologia – evitando, tuttavia, gli inconvenienti tipici di tale metodo,

ovvero, la difficoltà computazionale ed operativa di includere fattori esogeni e non controllabili (regolamentazione e caratteristiche ambientali del *network*), e la mancanza di una componente di *noise*, che conduce a confondere effetti di disturbo statistico con inefficienza produttiva.

In secondo luogo, la caratteristica *multi-step* del modello consente di estrarre informazioni disaggregate per ciascun input circa l'effetto di tali variabili esogene, laddove gli studi precedenti sull'impatto della regolamentazione hanno sempre stimato relazioni "aggregate" di produzione o di costo.

Infine, il modello presentato consente di disarticolare l'inefficienza complessiva, relativa ad ogni singolo fattore di produzione, in una componente attribuibile a fattori esogeni di tipo strutturale (caratteristiche del *network*) e politico (cornice regolatoria), in una componente di *managerial skill* pura, e, da ultimo, in un termine di *random noise*.

Nel presente lavoro, dato l'interesse principale incentrato sul tema della regolamentazione incentivante, si è indirizzato il focus verso gli aspetti legati all'efficacia dei nuovi strumenti contrattuali introdotti durante gli anni '90 per l'affidamento del servizio TPL

I risultati principali evidenziano l'incidenza positiva, tra i fattori esogeni, della velocità commerciale – quale *proxy* del grado di congestione del *network* – e dell'introduzione di contratti incentivanti di tipo *fixed-price*, sul grado di efficienza, separatamente per ciascun input. In particolare, il lavoro diretto ed i costi operativi per acquisto di materiali e servizi, rappresentano le variabili maggiormente sensibili al cambiamento del quadro regolatorio, in accordo con i modelli teorici precedentemente richiamati. Ciò sembrerebbe indicativo di un sistema efficace di regolamentazione dei sussidi nel TPL in Italia, in quanto incentrato sull'attribuzione di forme di *cap* proprio su quelle componenti di costo che rivestono un ruolo fondamentale nel processo di ottimizzazione della performance produttiva.

Inoltre, il complesso dei fattori esogeni di tipo strutturale e politico risultano spiegare larga parte della variabilità osservata delle performance nell'industria, relegando in posizione secondaria l'effetto delle competenze manageriali pure. Pur con le cautele del caso, dovute all'esigenza di riconoscere l'uniformità del dataset analizzato, interamente composto da operatori di proprietà pubblica, pare emergere con chiarezza la necessità di non trascurare gli effetti di specificità ambientali e regolatorie nella valutazione dei differenziali di efficienza fra imprese. Pertanto, ogni analisi condotta nell'ambito del TPL, e più in generale di settori regolamentati, che non tenga conto di tali fattori d'impatto rischierebbe di attribuire misure di efficienza falsate dall'azione di elementi al di fuori del controllo del management.

Riferimenti Bibliografici

- Banker R. D., A. Charnes e W. W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, 30(9), pp. 1078-1092
- Battese G.E. e T. Coelli (1992), "Frontier production function, technical efficiency and panel data: with application to paddy farmers in India", *Journal of Productivity Analysis*, 3(1-2), pp. 153-169
- Boame A. K. (2004), "Technical Efficiency of Canadian Urban Transit Systems", *Transportation Research - Part E: Logistics and Transportation Review*, 40(5), pp. 401-416
- Boitani A. e C. Cambini (2002a) "Il trasporto pubblico locale in Italia. Dopo la riforma i difficili albori di un mercato", *Mercato concorrenza regole*, n. 1, pp. 45-72
- Boitani A. e Cambini C. (2002b), "Regolazione incentivante per i servizi di trasporto locale", *Politica Economica*, 18(2), pp. 193-225.
- Cambini C. e M. Filippini (2003), "Competitive Tendering and Optimal Size in the Regional Bus Transportation Industry: An Example from Italy", *Annals of Public and Cooperative Economics*, 74(1), pp. 163-182
- Charnes A., W. W. Cooper e E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2(6), pp. 429-444
- Coelli T. J., D. S. P. Rao e G. E. Battese (1998), "An Introduction to Efficiency and Productivity Analysis", Boston, Kluwer Academic Publishers
- Dalen D. M. e A. Gomez-Lobo (1997), "Cost Functions in Regulated Industries Characterized by Asymmetric Information", *European Economic Review*, 41(3-5), pp. 935-942
- Dalen D. M. e A. Gomez-Lobo (2003), "Yardsticks on the Road: Regulatory Contracts in the Norwegian Bus Industry", *Transportation*, 30, pp. 371-386
- De Borger B., K. Kerstens e A. Costa (2002), "Public transit performance: what does one learn from frontier studies?", *Transport Reviews*, 22(1), pp.1-38
- Demsetz H. (1968), "Why Regulate Utilities?", *Journal of Law and Economics*, 11, pp. 55-65
- European Commission (1998), "Improved structure and organisation for urban transport operations of passengers in Europe (ISOTOPE)", Transport Research, Fourth Framework Programme, Urban Transport VII-51
- Fabbi D. (1998), "Public Transit Subsidy: from the Economics of Welfare to the Theory of Incentives", in *Transport Networks in Europe*, a cura di K. Button, Nijkamp P. e Priemus H., pp. 224-247
- Fraquelli G., M. Piacenza e G. Abrate (2004), "Regulating Public Transit Networks: How Do Urban-Intercity Diversification and Speed-up Measures Affect Firms' Cost Performance?", *Annals of Public and Cooperative Economics*, 75(2), pp. 193-225.
- Fried, H.O., C.A.K. Lovell, S.S. Schmidt e S. Yaisawarng (2002), "Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis", *Journal of Productivity Analysis*, 17(1/2): 157-74
- Gagnepain P. e M. Ivaldi (2002a), "Incentive Regulatory Policies: The Case of Public Transit Systems in France", *RAND Journal of Economics*, 33(4), pp. 605-629
- Gagnepain P. e M. Ivaldi (2002b), "Stochastic Frontiers and Asymmetric Information Models", *Journal of Productivity Analysis*, 18(2), pp. 145-159

- Hensher D. e Brewer A. (2001), “Transport: An Economics and Management Perspective”, Oxford University Press, Oxford
- Kerstens K. (1996), “Technical Efficiency Measurement and Explanation of French Urban Transit Companies”, *Transportation Research- Part E: Logistics and Transportation Review*, 30(6), pp. 431-452
- Kumbhakar, S. C. e C. A. K. Lovell (2000), “Stochastic Frontier Analysis”, Cambridge University Press, Cambridge
- Laffont J. J. e J. Tirole (1993), “A Theory of Incentives in Procurement and Regulation”, MIT Press, Cambridge
- Leibenstein H. (1966), “Allocative Efficiency versus «X-Efficiency»”, *American Economic Review*, 56, pp. 392-415
- Piacenza M. (2006), “Regulatory Contracts and Cost Efficiency: Stochastic Frontier Evidence from the Italian Local Public Transport”, *Journal of Productivity Analysis*, 25(3), pp. 257-277
- Thanassoulis E. (2001), “Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software”, Kluwer Academic Publishers, Norwell