

Federalismo: teoria ed evoluzione legislativa italiana - Economia e Politiche Pubbliche - Prof. Leonzio Rizzo

Ci chiediamo in questa lezione perché esistono diversi livelli di governo come Stato, Regioni, Province e Comuni? Perché certi beni pubblici sono forniti da istituzioni che operano a livelli territoriali diversi? Quali schemi di finanziamento sono stati adottati nelle recenti riforme italiane?

1 Le ragioni economiche del decentramento

La funzione di *stabilizzazione* è tipicamente una funzione dello stato centrale. Le aree locali si configurano infatti, rispetto all'economia nazionale, come economie piccole e molto aperte all'esterno. Una misura di politica economica ad esempio a sostegno dell'occupazione potrebbe avere riflessi all'esterno della regione che ne sostiene i costi, inducendo quindi le regioni, che tutelano solo gli obiettivi i cui benefici ricadono sul proprio territorio, a prendere misure inadeguate alle necessità della collettività. La funzione di *redistribuzione*, se decentrata, può generare nelle regioni incentivi ad abbassare il livello di assistenza sociale o sanitaria in modo da indurre i poveri a migrare verso regioni per loro più convenienti. La funzione di *allocazione* è quella di maggior interesse per i livelli inferiori di governo. La motivazione è data dal fatto che esclusi alcuni beni pubblici puri per eccellenza (difesa, sicurezza pubblica e giustizia) esistono una serie di beni pubblici che presentano caratteristiche di non rivalità e non escludibilità a livello locale: servizio di illuminazione pubblica, servizio antincendio. E' tuttavia importante ricordare che tali beni, pur a livello locale, non sono mai beni pubblici puri, poichè, anche se caratterizzati da non rivalità e non escludibilità, sono in genere passibili di forme di congestione: il consumo da parte di altri individui, oltre un certo limite, diminuisce il benessere del soggetto considerato.

E' necessario ricordare che oltre ai beni pubblici locali gli enti locali forniscono anche *beni a domanda individuale*, il cui beneficio, anche se divisibile, è limitato all'area locale; questi sono i trasporti urbani, raccolta rifiuti, fornitura acqua, gas, elettricità.

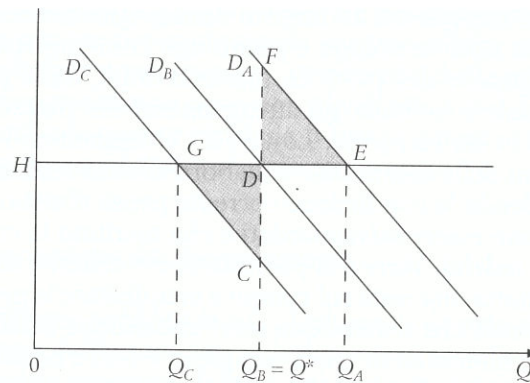
1.1 Il teorema di Oates

La tesi del teorema del decentramento di Oates (1972) è che per un *bene pubblico*, il cui consumo è definito in sottoinsiemi geografici della popolazione totale e per il quale i costi di fornitura di ogni livello del bene sono gli stessi per il governo centrale o i rispettivi governi locali, sarà sempre fornito in modo efficiente dai governi locali.

Tale risultato di forte rilievo dipende da 4 ipotesi. Queste sono 1) la fornitura centralizzata identica del bene nelle due Comunità, 2) costo marginale di offerta del bene pubblico costante, che implica una tecnologia di produzione in cui sono assenti economie di scala, 3) applicabilità del principio di corrispondenza tra territorio da cui è fornito il bene e territorio entro cui la fornitura del bene ha effetto. 3a) E' molto comune che tale ipotesi non sia rispettata a causa dei cosiddetti effetti di traboccamento dovuti alla presenza di esternalità nella fornitura di beni pubblici locali: il sottoinsieme geografico su cui è definito il consumo del bene pubblico non coincide con il sottoinsieme geografico di competenza dell'ente locale, che decide il livello di bene pubblico da fornire. In tal caso la comunità locale fornirà una quantità del bene inferiore a quella che fornirebbe nel caso in cui non ci fosse l'effetto di traboccamento. 4) Le preferenze degli individui di una stessa comunità sono identiche, che da un punto di vista teorico potrebbe essere giustificabile con un meccanismo alla Tiebout: in presenza di eterogeneità delle preferenze gli individui scelgono la comunità in funzione della tipologia e qualità dei servizi pubblici offerte, delle tasse, imposte e tariffe richieste (i cittadini *votano con i piedi*). Si formano così dei sottoinsiemi geografici omogenei su ognuno dei quali è possibile definire un distinto livello di consumo del bene pubblico.

Ipotizziamo di avere due comunità C ed A. La comunità C è disposta ad offrire per ogni quantità di bene pubblico locale un prezzo inferiore a quello della comunità A come si nota dalle due funzioni di domanda disegnate nella Figura 1. Il costo marginale di produzione del bene è costante ed è quindi rappresentabile con una retta parallela all'asse delle ascisse. La comunità C se potesse scegliere, dato il costo, sceglierebbe una quantità del bene pari a OQ_C , pagando OQ_CGH , la comunità A sceglierebbe OQ_A pagando, OQ_AEH . Nel caso della fornitura accentrata invece poiché abbiamo detto che lo stato centrale non è in grado (o non vuole) fornire due beni pubblici locali differenti, possiamo ipotizzare che la quantità di bene pubblico fornita sia pari a OQ_B , una combinazione lineare tra OQ_C e OQ_A . In quest'ultimo caso ad ogni comunità verrebbe chiesto di pagare OQ_BDH . La comunità C, data la sua funzione di domanda registrerebbe una perdita pari a CDG e la comunità A, anche se paga meno di quanto pagherebbe nell'ipotesi di fornitura decentrata ($OQ_BDH < OQ_AEH$), non è comunque soddisfatta, poiché pagando di più aumenterebbe il suo surplus netto di EFD .

Figura 1: Teorema del decentramento di Oates



1.2 Dimensione ottima

Nella realtà istituzionale osserviamo diversi livelli di governo che hanno normalmente una base territoriale di diversa ampiezza. Quali sono le ragioni economiche che giustificano enti di dimensioni diverse?

Il problema della dimensione ottima, pur in assenza di economie di scala si presenta nel caso in cui un'istituzione fornisca un bene pubblico ad una collettività.

Nel caso di bene pubblico puro infatti un incremento del numero di individui appartenenti alla comunità ove viene fornito il bene fa diminuire il costo pro-capite che ogni cittadino deve pagare tramite le imposte e/o tasse per finanziare il bene pubblico inoltre tale diminuzione di costo è tanto più rilevante quanto più piccola è la comunità considerata. Infatti, consideriamo un bene pubblico la cui produzione costa 100 e due comunità A e B. La comunità A è composta inizialmente da 1 individuo e la comunità B da 4 individui, quindi il costo pro-capite di produzione del bene pubblico è 100 per la comunità A e 25 per la comunità B. Se si aggiunge un individuo alla comunità A il costo pro-capite di produzione diventa $100/2=50$, se si aggiunge un individuo alla comunità B, il costo pro-capite di produzione diventa $100/5=20$; il risparmio di costo pro-capite di chi era residente nella comunità A prima dell'incremento di popolazione è pari a 50 per ogni cittadino, mentre il risparmio pro-capite di chi era residente nella comunità B prima dell'incremento di popolazione è pari a 5 per ogni cittadino. Quindi il risparmio di costo pro-capite da sostenere per finanziare la fornitura di bene pubblico dovuto ad un incremento unitario di popolazione decresce all'aumentare degli individui della comunità.

Nel caso in cui il bene pubblico fornito non sia puro, ma soggetto a fenomeni di congestione, per un dato livello del bene fornito, un incremento degli individui appartenenti alla comunità genera una diminuzione della qualità del servizio, e quindi del beneficio che ogni individuo ne ottiene, dovuta al costo di congestione. E' ragionevole ipotizzare che l'incremento della popolazione generi una diminuzione del beneficio pro-capite crescente all'aumentare della popolazione: quando in una comunità vi sono pochi abitanti, l'incremento di popolazione diminuisce il beneficio ottenuto dalla fornitura di un bene o servizio pubblico molto di meno che nel caso in cui via sia la stessa variazione di popolazione in una comunità molto affollata.

Le due funzioni di costo pro-capite e beneficio pro-capite dipendono anche dalla quantità di bene pubblico fornita. Il costo pro-capite, poichè preferiamo ragionare in assenza di economie di scala, per un dato livello di popolazione aumenta in modo lineare con la quantità fornita. Il beneficio aumenta anch'esso con la quantità fornita, per un dato livello di popolazione, e tale aumento, in accordo con la caratteristica di una funzione di utilità con rendimento marginale decrescente, è decrescente con il livello della quantità fornita.

In termini analitici utilizziamo il *beneficio pro-capite* e *costo pro-capite* per costruire una funzione obiettivo della comunità pari al *beneficio pro-capite netto*, cioè alla differenza tra beneficio e costo pro-capite. Il beneficio pro-capite B è funzione, sia della popolazione, N , dell'ente locale, sia il bene pubblico, Q . All'aumentare di N i benefici pro-capite diminuiscono, ad un tasso crescente, a causa del verificarsi della congestione; all'aumentare della quantità di bene pubblico i benefici pro-capite aumentano, ma ad un tasso decrescente per il manifestarsi dei fenomeni di saturazione dei bisogni (rendimento dell'utilità marginale decrescente). Formalmente:

$$\frac{\partial B}{\partial Q} > 0, \frac{\partial^2 B}{\partial Q^2} < 0, \frac{\partial B}{\partial N} < 0, \frac{\partial^2 B}{\partial N^2} > 0$$

Ipotizziamo assenza di costi fissi ed un costo totale lineare in Q : $CT = aQ$. Quindi il costo pro-capite risulta:

$$C = C(Q, N) = \frac{aQ}{N}$$

Le Fig. 2a e 2b mostrano l'andamento dei costi e dei benefici pro-capite in funzione della popolazione N e della quantità di servizio prodotta Q . La dimensione ottimale della comunità si ottiene per quei valori di N e di Q che rendono massima la differenza tra beneficio pro-capite e costo pro-capite:

$$\max_{Q, N} \left[B(N, Q) - \frac{aQ}{N} \right]$$

Le condizioni di massimo di primo ordine sono:

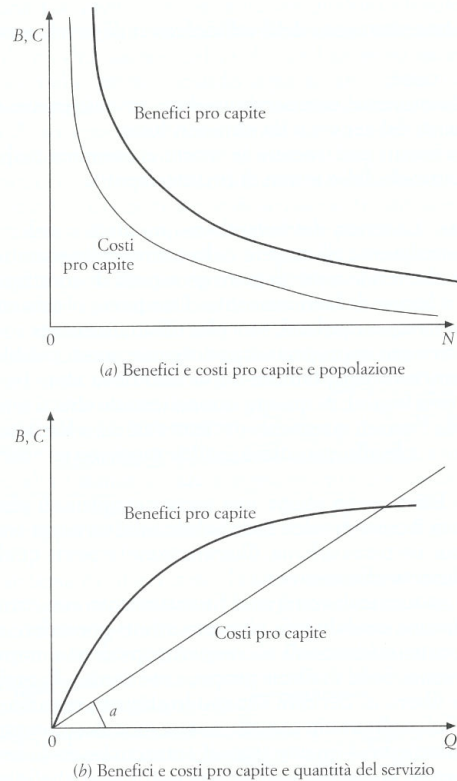
$$-\frac{\partial B}{\partial N} = \frac{aQ}{N^2} \quad (1)$$

$$\frac{\partial B}{\partial Q} = \frac{a}{N} \quad (2)$$

da cui si possono ricavare i valori ottimi di N e Q .

Per una *data quantità*, la dimensione ottimale della popolazione è quella ove il beneficio marginale derivante dalla riduzione del costo pro-capite dovuto alla partecipazione di un soggetto in più al finanziamento del servizio uguaglia la riduzione marginale di benessere derivante da un aumento della congestione (1). Per un *dato livello di popolazione*, la quantità ottimale del servizio è quella ove il beneficio marginale di un'unità addizionale di servizio uguaglia il costo marginale del servizio supposto costante (2). In questo modello quindi per definire la dimensione ottimale dell'ente è necessario determinare simultaneamente sia il livello della popolazione, che il livello del bene pubblico.

Figura 2: Dimensione ottimale dei governi locali



2 Forme di finanziamento dei livelli inferiori di governo

Le forme di *finanziamento autonomo* dei livelli inferiori di governo possono essere di due tipi. Tariffe e tasse ed imposte. Per una parte rilevante di servizi pubblici locali trovano appropriata applicazione *prezzi pubblici e tariffe*: questi sono infatti beni, sì di interesse e rilevanza sociale, ma rivali nel consumo e da cui i fruitori possono essere esclusi. Un'altra categoria di beni offerti localmente sono finanziati invece con le *tasse*. Questi sono la scuola, il servizio antincendio, la raccolta rifiuti. Tutti questi sono servizi che oltre a dare beneficio al cittadino che direttamente richiede il servizio, danno anche beneficio ai cittadini che vivono nella stessa città di colui il quale consuma il servizio: è quindi opportuno che colui il quale consuma il servizio non paghi l'intero costo, ma una parte tramite la tassa e il resto sia pagato con una quota di gettito da imposta locale, cui i cittadini sono chiamati a partecipare in base alla propria capacità contributiva. Le *imposte* oltre a fornire le risorse per la copertura dei servizi in parte pagati con le tasse, servono soprattutto a finanziare i servizi pubblici locali che presentano carattere di indivisibilità dei vantaggi, come viabilità, opere pubbliche ed illuminazione.

Le *entrate tributarie proprie locali* possono essere di diversi tipi: tributi propri, addizionali e compartecipazioni a imposte di livello superiore di governo. Per i *tributi propri (imposte e tasse)* il governo locale definisce in modo autonomo l'aliquota ed in alcuni casi anche la base imponibile. L'*addizionale* è uno strumento con il quale si distribuisce il gettito derivante da una stessa base imponibile fra diversi livelli di governo. L'ente di governo inferiore definisce l'aliquota d'imposta di sua competenza che si aggiunge a quella fissata dal governo centrale sullo stesso imponibile. Nel sistema tributario italiano il gettito derivante dall'istituzione di addizionali è stato spesso utilizzato per finanziare Regioni e Enti locali: addizionale comunale e regionale Irpef, addizionale regionale di consumo sul gas metano. . . La *base imponibile* è l'espressione quantitativa del presupposto dell'imposta, ovvero della particolare situazione di fatto alla quale la legge ricollega l'obbligo di pagare l'imposta. Alla base imponibile si applica l'aliquota d'imposta al fine di ottenere l'imposta dovuta. La base imponibile Irpef, ad esempio, corrisponde al possesso da parte di una persona fisica di un reddito imponibile in denaro o in natura: per un soggetto in possesso di solo reddito da lavoro dipendente, la base imponibile corrisponderà al reddito annuale al netto di eventuali deduzioni. La *compartecipazione* è una quota fissa percentuale sul totale del gettito erariale. Gli enti locali compartecipano secondo la quota stabilita ad un gettito derivante da un'imposta spettante allo Stato. La compartecipazione Iva, ad esempio, è utilizzata per finanziare la sanità regionale. Esempio: gettito Iva totale 1000, quota di compartecipazione 40%, Compartecipazione regionale all'Iva= $0.4 \cdot 1000 = 400$.

In tutti i sistemi finanziari federali esistono *trasferimenti* dal centro necessari a finanziare l'intera spesa sostenuta dagli enti locali. Infatti i tributi propri non sono in genere mai sufficienti a finanziare l'intera fornitura dei beni pubblici locali. Il motivo di tale asimmetria tra servizi forniti e il loro finanziamento è dovuto alla minaccia di concorrenza fiscale.