

CICLO

Alfredo Medio

COSTO

Mario Morroni

UTILITÀ

Marina Bianchi



Prima edizione ottobre 1985

© 1985 Editore Boringhieri SpA, Torino, corso Vittorio Emanuele 86
I diritti di memorizzazione elettronica, di riproduzione e di adattamento
totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie foto-
statiche) sono riservati
Stampato in Italia dalla tipografia Gravinese di Torino
CL 74-9220-0 ISBN 88-339-5013-1

COSTO

Mario Morroni

COSTO

Introduzione

Il costo è dispendio o consumo delle risorse necessarie per effettuare un'operazione economica. Il costo di produzione è costituito dalla somma dei prezzi (in unità di conto) dei mezzi di produzione per le rispettive quantità.

Sebbene queste definizioni di carattere generale siano comuni a tutte le teorie economiche, vi sono notevoli differenze sul significato attribuito ai prezzi che misurano il costo delle risorse: per gli economisti classici – Ricardo in particolare – e per Marx i prezzi dei mezzi di produzione corrispondono al *costo reale* che essi identificano con la quantità di lavoro necessaria alla produzione; per alcuni economisti neoclassici, come Marshall e Viner, il prezzo dei mezzi di produzione riflette non solo le condizioni di domanda, ma anche il *sacrificio soggettivo* di coloro che hanno concorso alla produzione; per altri, come ad esempio Wicksteed, Knight, Haberler, il prezzo dei mezzi di produzione è pari al loro *costo opportunità*, inteso come il sacrificio dell'utilità di quei beni che si sarebbero potuti ottenere dalle risorse utilizzate.

Queste diverse concezioni del costo implicano visioni del processo economico radicalmente divergenti. Infatti, gli economisti classici concepiscono il sistema economico come un processo di riproduzione circolare delle merci reso possibile mediante l'attività lavorativa. Il costo, rappresentato dal lavoro necessario alla produzione, è a un tempo sostanza e misura dei valori di scambio attorno a cui oscillano i prezzi di mercato direttamente osservabili. L'obiettivo principale che si pongono gli economisti classici è l'individuazione delle leggi che stanno alla base dell'accumulazione e dello sviluppo economico capitalistico. Il principio del co-

sto lavoro si inserisce come elemento analitico fondamentale nella determinazione delle circostanze che regolano il valore dei beni, la distribuzione del reddito e il processo di accumulazione. Viceversa, gli economisti neoclassici concepiscono il sistema economico come un sistema di scambio in cui i prezzi sono il risultato dell'equilibrio concorrenziale tra la domanda e l'offerta. Il costo di produzione è definito come l'utilità indiretta dei beni a cui si rinuncia o come l'utilità negativa, disutilità, sperimentata dagli agenti intervenuti nella produzione. Il costo, che determina la curva di offerta, è così reso omogeneo, e quindi confrontabile, con l'utilità che è alla base della curva di domanda.

I concetti neoclassici di costo opportunità e di costo reale soggettivo, pur garantendo entrambi la simmetria fondamentale tra domanda e offerta, poggiano su ipotesi diverse circa l'offerta dei mezzi di produzione. Gli economisti che sostengono il costo opportunità suppongono data la quantità disponibile dei fattori produttivi, e di conseguenza rappresentano la produzione come un processo a "senso unico" che va dai fattori ai beni di consumo: se le risorse sono date, il costo di produzione è un indice di scarsità dei fattori produttivi rispetto all'intensità e al livello della domanda ed è regolato, in ultima analisi, dall'utilità dei beni finali. L'economia diviene, in questo contesto, la disciplina che studia il problema dell'ottima allocazione delle risorse scarse. I sostenitori del costo reale soggettivo presuppongono invece che la quantità disponibile delle risorse, a eccezione di quelle naturali, tenda a variare in relazione allo sforzo e sacrificio individuale richiesto per la creazione delle merci prodotte. L'equilibrio tra domanda e offerta è determinato dal bilanciarsi di tendenze opposte e indipendenti: da una parte i bisogni, che esprimono il grado di utilità attribuito alle merci, dall'altra gli sforzi e i sacrifici, ossia la disutilità sperimentata dagli agenti produttivi. In questo secondo caso il compito dell'economista non si limita alla soluzione di un problema di allocazione efficiente, ma riguarda l'analisi delle condizioni di domanda e offerta, ovvero lo studio degli "sforzi che l'uomo compie per soddisfare i propri bisogni, nella misura in cui gli sforzi, e i bisogni, sono misurabili in termini monetari" [Marshall, 122].

Nelle più recenti formulazioni del modello di equilibrio economico generale i concetti di costo reale e di costo opportunità sono però abbandonati: il costo è definito semplicemente come la

somma dei prezzi degli input, calcolati contemporaneamente ai prezzi e alla quantità dei beni prodotti. Nello schema di equilibrio economico generale il sistema dei prezzi è determinato dall'azione congiunta di numerosi elementi: la scarsità delle risorse, le tecniche di produzione, la distribuzione degli obiettivi tra i soggetti e le preferenze individuali.

Anche nel modello di Sraffa del 1960, che segna la ripresa dei concetti, propri dell'economia classica, di riproducibilità, sovrappiù e circolarità del processo produttivo, i costi di produzione, in quanto prezzi degli input, sono determinati assieme ai prezzi dei beni prodotti da un sistema di equazioni simultanee. Sia nel modello di Sraffa che nel modello di equilibrio economico generale, essendo i prezzi degli input calcolati contemporaneamente ai prezzi dei beni prodotti, vi è una interdipendenza reciproca tra costi e prezzi; viene così meno la relazione causale che va dai costi ai prezzi; viene così meno la relazione causale che va dai prezzi ai costi presente nel concetto di costo reale, calcolato in termini di lavoro contenuto e in quello neoclassico di costo reale soggettivo, che è dato dal sacrificio soggettivo. Tuttavia i prezzi, che risultano dal modello di equilibrio economico generale, hanno un significato del tutto diverso dai prezzi calcolati nello schema di Sraffa. Nel primo caso i prezzi dipendono, in ultima analisi, dall'interazione della domanda e dell'offerta, mentre nel secondo caso essi riflettono esclusivamente le condizioni della produzione. Ciò dipende dal fatto che nel modello di Sraffa il volume della produzione è supposto dato e che, di conseguenza, la domanda non può svolgere alcun ruolo nel processo di formazione dei prezzi. L'autolimitazione di non tener conto di mutamenti della produzione, se da una parte consente a Sraffa di concentrare l'analisi sul mutamento dei prezzi relativi e della distribuzione del reddito, dall'altra preclude la possibilità di studiare, nell'ambito dell'analisi stessa, le variazioni dei costi in rapporto alla quantità prodotta.

Il problema dell'andamento dei costi in relazione alle variazioni della produzione, che insieme alla definizione di costo costituisce l'oggetto della presente analisi, è stato affrontato dalle differenti teorie economiche in ambiti e con finalità molto diverse. Negli economisti classici i *costi decrescenti* sono discussi in relazione all'aumento della divisione del lavoro, cioè all'interno della produzione, mentre i *costi crescenti* sono trattati in relazione alla rendita fondiaria e perciò sono connessi allo studio della distri-

buzione. I rendimenti crescenti hanno luogo nell'industria manifatturiera, che può godere pienamente degli effetti della divisione del lavoro; mentre i rendimenti decrescenti si manifestano in agricoltura, a causa dell'espandersi della coltivazione delle terre meno fertili. I costi decrescenti tendono ad accelerare il processo di accumulazione, mentre i costi crescenti in agricoltura portano a un costante aumento delle rendite e a una corrispondente compressione del saggio di profitto generale, che ha per conseguenza il progressivo rallentamento del processo di accumulazione. Nella teoria neoclassica dell'equilibrio parziale, dovuta principalmente alla scuola marshalliana, la legge di variazione dei costi, e in particolare la legge dei costi crescenti, costituisce invece una delle basi per la determinazione del prezzo relativo delle merci. I costi crescenti, all'espandersi della produzione, di una data merce garantiscono, infatti, la simmetria fondamentale tra la domanda e l'offerta di quel bene. Il punto di equilibrio è dato dall'incontro della curva di domanda, determinata dall'utilità marginale decrescente, con la curva di offerta, determinata dal costo marginale crescente.

Nella teoria degli economisti classici le leggi di variazione dei costi dipendono da mutamenti strutturali dell'economia come l'incremento della popolazione, la messa a coltura di nuove terre meno fertili, l'introduzione del progresso tecnico, l'aumento della divisione del lavoro. Tali variazioni dei costi, essendo appunto riferite al *complesso dei beni prodotti*, non possono influenzare il valore relativo delle singole merci. Nell'analisi neoclassica dell'equilibrio parziale le leggi di variazione dei costi sono applicate alla produzione della *singola merce* da parte delle imprese operanti su un dato mercato e quindi influenzano direttamente la formazione del prezzo di offerta di quel bene. Questo radicale mutamento di prospettiva (dall'intero sistema alla singola impresa) comporta un differente modo di considerare la legge dei rendimenti crescenti e decrescenti. Per gli economisti classici l'andamento dei costi al variare dei livelli di attività è legato a due fenomeni perlopiù congiunti: l'aumento della *quantità assoluta* di tutti i mezzi di produzione (l'incremento del numero dei lavoratori, la messa a coltura di nuove terre, l'introduzione di nuove macchine) e il mutamento delle *proporzioni* in cui sono combinati i vari mezzi di produzione (il diverso rapporto tra il lavoro e la terra o tra il lavoro e le macchine). Nelle trattazioni

successive si manifesta sempre più la tendenza a isolare questi due fenomeni concentrando l'attenzione sul secondo, ossia sul mutamento delle proporzioni in cui i vari input sono combinati. Di conseguenza la legge di variazione dei costi, anziché dipendere dalle scelte di accumulazione e sviluppo dei produttori che hanno di fronte una serie (discontinua) di tecniche ordinabili secondo un criterio di efficienza, assume la caratteristica di una "legge naturale" della produzione, di una "legge generale della combinazione dei fattori".

Il mutamento delle proporzioni tra gli input al crescere della quantità prodotta si basa sull'ipotesi che almeno uno di essi sia indivisibile e non aumentabile. Poiché un incremento del volume della produzione può essere ottenuto solo attraverso l'aumento delle dosi degli input variabili, il *costo totale*, in una prima fase, finché gli input fissi rimangono in eccesso rispetto a quelli variabili, cresce meno che proporzionalmente, quindi, superata la combinazione ottima degli input, registra incrementi maggiori a causa della combinazione sempre più inefficiente tra gli input fissi e gli input variabili. Va notato, però, che in concorrenza perfetta l'ipotesi che almeno uno dei mezzi di produzione utilizzati dall'impresa non sia aumentabile presenta alcune difficoltà, se riferita all'analisi statica di lungo periodo; mentre trova riscontro nel breve periodo, in cui l'impresa non è in grado di accrescere la quantità di alcuni input (come ad esempio gli impianti).

Quanto al modello di equilibrio economico generale, le quantità e i prezzi di equilibrio sono determinati simultaneamente, senza necessità di costruire le curve di offerta e di domanda. Poiché tuttavia la quantità delle risorse è data, l'aumento della produzione di un bene implica necessariamente la diminuzione della produzione di un altro o di altri beni. Per determinare se la produzione di un dato bene è soggetta a costi unitari crescenti o costanti è allora necessario sapere quali input, e in che proporzioni, sono stati liberati dalle industrie in declino. Di conseguenza all'interno di uno schema di equilibrio economico generale non è possibile determinare a priori quale sia l'andamento dei costi all'aumentare della produzione di una determinata industria che produce un certo bene.

Gli economisti neoclassici, con l'eccezione di Marshall e di Young, tendono a escludere dalle loro analisi il fenomeno delle economie di scala che è stato messo in luce da Smith e da Marx

e la cui presenza sembra assai rilevante soprattutto all'interno della industria manifatturiera. L'esclusione delle economie di scala dalle teorie neoclassiche è dovuta al fatto che la riduzione dei costi medi al crescere della scala produttiva, ossia all'aumentare dell'utilizzo dei vari input, è incompatibile con qualsiasi idea di equilibrio concorrenziale, sia esso parziale o generale. I rendimenti crescenti, che hanno origine principalmente nell'aumento della divisione del lavoro, nelle economie di dimensione, nel progresso tecnico e nell'apprendimento mediante l'esperienza, possono trovare posto solo in riferimento all'analisi dinamica e a un mercato non perfettamente concorrenziale. L'analisi dell'oligopolio consente di collocare in un contesto analitico coerente tutti questi elementi che stanno alla base delle economie di scala.

Parte prima

1.1 Costi produttivi e improduttivi in Quesnay

François Quesnay ha elaborato per primo uno schema semplificato della produzione in cui si stabiliscono le condizioni che permettono la riproduzione e lo sviluppo del sistema economico. Nel *Tableau économique* egli rappresenta la produzione come un processo circolare di flussi di prodotti annui nel quale le stesse merci compaiono prima come mezzi di produzione e poi, in quantità maggiore, come prodotti. I costi sono costituiti, oltre che dalle spese iniziali dei proprietari fondiari per rendere coltivabile il terreno (anticipazioni fondarie), dai flussi annui dei beni necessari alla produzione. Quesnay ritiene che lo sfruttamento capitalistico delle risorse naturali dia luogo a un prodotto netto o sovrappiù, cioè a una quantità di prodotto maggiore rispetto a quella consumata nella produzione [Quesnay, 233 sg., 259-62]. L'industria manifatturiera e il commercio sono considerati settori improduttivi o sterili, perché, a differenza dell'agricoltura in cui è rilevabile un incremento fisico del prodotto, essi si limiterebbero a trasformare i prodotti senza creare un sovrappiù [*ibid.*, 176 sg., 183-85]. Da ciò deriva la convinzione che solo i costi sostenuti in agricoltura siano produttivi. Tra essi egli menziona le spese per le materie prime, le remunerazioni del lavoro, gli ammortamenti e la manutenzione del capitale fisso, il fondo rischi, le anticipazioni fondarie iniziali. Sono invece improduttivi i costi affrontati in tutte le altre attività economiche. Al riguardo egli cita le spese per le materie prime, le retribuzioni, i costi commerciali di trasporto e di distribuzione.

La distinzione tra costi produttivi e improduttivi è essenziale per gli obiettivi dell'analisi fisiocratica, in quanto permette di di-

stinguere le spese che creano "ricchezza reale", e che quindi consentono lo sviluppo del sistema economico, dalle spese che al contrario comportano una riduzione della base produttiva e della ricchezza [Quesnay, 18 sg., 102-05]. Mentre la rendita percepita dal proprietario fondiario corrisponde al sovrappiù ottenuto dalla coltivazione capitalistica del terreno, nel sistema fisiocratico il profitto dell'affittuario fa parte dei costi, essendo questo formato dalla remunerazione dell'imprenditore, in tutto assimilabile alla retribuzione degli altri lavoratori, e da una quota percentuale sul capitale fisso destinata all'ammortamento, alla manutenzione dei mezzi di produzione e alla costituzione di un fondo rischi contro le perdite dovute a fattori naturali sfavorevoli [*ibid.*, 19-22, 97-98].

1.2 Divisione del lavoro e rendimenti crescenti in Smith

La distinzione fisiocratica tra costi produttivi e costi sterili viene meno in Smith, il quale attribuisce la creazione del sovrappiù non solo all'agricoltura, ma a tutta la produzione sociale. Smith enuncia, contraddittoriamente, due diverse concezioni del costo di produzione riferito al sistema economico capitalistico: in alcuni brani egli sembra mantenere la concezione, relativa alle società precapitalistiche, del costo come quantità di lavoro necessario alla produzione; in altri passi, e più di frequente, egli definisce il costo come la somma dei prezzi "ordinari e medi" del salario, del profitto e della rendita pagati per la produzione [Smith, 32-35, 49-55].

Nel caso in cui il costo si identifica con la quantità di lavoro contenuto, il profitto e la rendita, essendo considerati "detrazioni" o "deduzioni" dal prodotto del lavoro, non fanno parte del costo [*ibid.*, 50 sg., 66]. Il lavoro è considerato da Smith "la misura reale dei valori di scambio" perché (a differenza delle merci che variano nel proprio valore) le quantità di lavoro hanno, "in ogni luogo e tempo", "uguale valore per il lavoratore". Ciò in quanto egli deve sacrificare la stessa quantità del suo riposo, della sua libertà e della sua felicità [*ibid.*, 35]. Nel caso in cui il costo è invece rappresentato dalle tre componenti principali

del prezzo, il profitto e la rendita sono inclusi nei costi alla pari del salario:

Quando il prezzo di una merce non è né più né meno di ciò che è sufficiente a pagare la rendita della terra, i salari del lavoro e i profitti (...), secondo i loro saggi naturali, quella merce - asserisce Smith - verrà venduta per quello che si può chiamare il suo prezzo naturale (...) La merce è allora pagata precisamente per ciò (...) che realmente costa a colui che la porta al mercato. Infatti, sebbene nel linguaggio comune ciò che si chiama il primo costo di una merce non comprenda il profitto di colui che la vende (...), se costui la vende a un prezzo che non gli dà il saggio di profitto che si ottiene nel luogo in cui opera, lavorerà evidentemente in perdita. [Smith, 56 sg.]

Smith non menziona i costi delle materie prime e dei mezzi di produzione giacché, se si risale ai precedenti stadi della produzione, tutte le merci possono essere ridotte a salario, profitto e rendita.

Le due diverse teorie del costo espone da Smith corrispondono a due distinte teorie del valore. Nel definire il costo come la quantità di lavoro necessario alla produzione Smith sembra enunciare, seppure embrionalmente, la teoria del valore lavoro che sarà in seguito sviluppata da Ricardo e da Marx. Definendo, invece, il costo come la spesa per la remunerazione dei fattori produttivi, egli formula una teoria "addizionale" dei valori di scambio, risultante della somma, pura e semplice, delle tre componenti principali dei prezzi (salario, profitto e rendita), ciascuna determinata indipendentemente dalle altre (vedi VALORE). Questa seconda concezione sarà ripresa da Say ai primi dell'Ottocento, sviluppata da Senior e Longfield (1820-30), assieme all'idea del lavoro come sacrificio e pena, nella teoria del costo reale soggettivo, e infine trasmessa a Marshall attraverso la sistemazione del pensiero classico operata da Stuart Mill.

Per quanto riguarda le variazioni del costo al crescere della quantità prodotta, Smith distingue due casi in relazione all'entità della crescita della produzione. Se l'aumento di domanda, che provoca l'incremento della produzione, è leggero, l'offerta si potrà adeguare rapidamente al nuovo livello della domanda con costi di produzione sostanzialmente invariati: dopo un'iniziale e

lieve crescita del prezzo di mercato rispetto al prezzo naturale, determinata dall'eccesso di domanda, il prezzo di mercato tornerà quindi a uguagliare l'originario prezzo naturale. Se invece si verifica un aumento della domanda quantitativamente rilevante e non occasionale, il costo di produzione tenderà a ridursi. Il ragionamento di Smith è molto semplice. Inizialmente, a causa del notevole eccesso di domanda, il prezzo di mercato salirà molto al di sopra del costo di produzione, e quindi del prezzo naturale. Gli alti profitti, determinati dalla differenza tra il prezzo di mercato e il costo di produzione, favoriranno gli investimenti delle imprese esistenti e l'entrata di nuove imprese. Gli investimenti effettuati dalle vecchie e nuove imprese consentiranno, oltre all'introduzione di macchine più efficienti, una migliore divisione del lavoro, che porterà a un aumento della produttività e a un abbassamento dei costi.

La crescita della domanda e l'ampliamento del mercato – che possono essere dovuti a numerosissimi motivi, tra i quali forse i più importanti sono la maggiore facilità di trasporto e la crescita della concentrazione urbana – portano dunque allo sviluppo della divisione del lavoro, che a sua volta determina un incremento della produttività e un abbassamento dei costi. Tale incremento di capacità produttiva è dato dal fatto che la divisione del lavoro consente: 1) l'aumento della destrezza dell'operaio nello svolgere la singola mansione; 2) il risparmio del "tempo che di solito si perde per passare da una mansione all'altra"; e 3) "l'introduzione di nuove macchine che facilitano e abbreviano il lavoro" [Smith, 12]. La divisione del lavoro nella teoria smithiana è riferita non solo alla ripartizione dei compiti tra lavoratori di una singola impresa, ma anche alla specializzazione delle varie imprese di un certo ramo di attività. Ne consegue che l'espansione del mercato permette anche un'elevata specializzazione delle imprese. Smith afferma inoltre che la divisione del lavoro non è originata tanto da innate diversità di intelligenze e talenti individuali (che sono invece l'effetto e non la causa della divisione del lavoro) quanto da una tendenza umana allo scambio e al baratto, senza la quale non vi sarebbe neppure la conseguente tendenza alla specializzazione individuale [*ibid.*, 17-20]. Egli sottolinea il rapporto biunivoco esistente tra accumulazione e divisione del lavoro: da un lato, "l'accumulazione del capitale (...) deve di necessità precedere la divisione del lavoro"; dall'altro, come ab-

biamo visto, la riduzione progressiva delle operazioni di ogni lavoratore a un maggiore grado di semplicità consente l'invenzione e l'introduzione di una serie di nuove macchine, che a loro volta facilitano e abbreviano queste operazioni [Smith, 267 sg.].

Sebbene la teoria della divisione del lavoro fosse stata oggetto di analisi da parte di numerosi filosofi ed economisti, tra i quali Petty, Mandeville, Hutcheson, e avesse addirittura trovato spazio nella voce "Arte" dell'*Encyclopédie* (Diderot, 717), tuttavia il grande merito di Smith è di avere colto con estrema lucidità gli effetti che l'interrelazione tra l'ampiezza del mercato, l'accumulazione del capitale e la divisione del lavoro, ha sulla capacità produttiva e, più in generale, sullo sviluppo economico. La sua analisi rappresenta, come vedremo, la base di tutte le successive elaborazioni sul problema dei rendimenti crescenti.

1.3 Ricardo: il costo reale come quantità di lavoro contenuto nelle merci

Ricardo respinge la teoria smithiana del valore di scambio basata sulla somma del prezzo delle tre componenti (salario, profitto e rendita), in quanto spiegherebbe i prezzi con i prezzi stessi, cadendo in un circolo vizioso. Egli rimprovera Smith di aver limitato la validità della teoria del valore lavoro solo alle società precapitalistiche. Per Ricardo, invece, l'impiego di capitale non fa venir meno il principio generale secondo cui il valore di scambio di una merce dipende dal suo costo reale, rappresentato dalla quantità di lavoro, diretto e indiretto, necessario alla produzione nelle condizioni più sfavorevoli nelle quali è necessario effettuare la produzione per ottenere la quantità di prodotto richiesta [Ricardo, 8 sg., 43]. In opposizione a Smith, Ricardo sostiene che "non è perché il capitale si accumula" che il valore di scambio varia:

Esso varia (...) in dipendenza da due sole cause: la minore o maggiore quantità di lavoro occorrente, e la diversa durata del capitale; (...) la prima causa non è mai soppiantata dalla seconda, ma soltanto modificata. (Ricardo *(c)*, vol. 7, 337)
Non possiedo alcun criterio per determinare se una cosa è cara o a buon mercato all'infuori del sacrificio di lavoro sop-

portato per ottenerla (...) Che la quantità maggiore o minore di lavoro impiegata nella produzione delle merci sia la sola causa del loro cambiamento di valore è del tutto chiaro non appena si sia d'accordo sul fatto che tutte le merci sono il prodotto del lavoro e che esse non avrebbero alcun valore se non per il lavoro speso nella loro produzione. (Ricardo (*b*), 397)

Una merce è venduta "al suo valore naturale, quando essa ripaga, attraverso il suo prezzo", il costo di produzione riducibile ai salari e ai profitti. In queste circostanze il valore, in termini di lavoro contenuto, coincide con le spese necessarie alla produzione e vi è quindi proporzionalità tra il costo reale e il costo di produzione. Tuttavia, nel caso di diversa durata del capitale, tale proporzionalità non è assicurata. Le "deviazioni" tra le quantità di lavoro e i costi di produzione hanno origine dal fatto che, quando il rapporto tra il capitale fisso e il capitale circolante non è il medesimo in tutti i processi produttivi, una variazione dei salari comporta un mutamento dei prezzi relativi e quindi del costo di produzione stesso [Ricardo, 20-37].

Dunque, se la distribuzione nel tempo delle quantità di lavoro necessario non è la stessa per tutte le merci, i costi di produzione variano non solo in relazione alla maggiore o minore quantità di lavoro, ma anche per effetto dei valori assunti dalle variabili distributive (vedi DISTRIBUZIONE DEL REDDITO; PREZZO; VALORE). Ricardo giunge alla conclusione che la causa di tali deviazioni debba essere ricercata nella inadeguatezza della misura piuttosto che nella mancanza di validità della teoria del valore lavoro (vedi Sraffa (*e*), xxxvii-xlix). Egli reputa, infatti, che la teoria del valore lavoro sia "la più stretta approssimazione alla realtà, come regola per misurare il valore relativo" (Ricardo (*c*), vol. 8, 279), e che se si riuscisse a trovare una misura perfetta, i prezzi non varierebbero al mutare della distribuzione del reddito (Ricardo (*b*), 373).

Per quanto riguarda la relazione tra costi e quantità prodotta, Ricardo ritiene, come la maggior parte dei suoi contemporanei, che in assenza di progresso tecnico la quantità di lavoro necessaria alla produzione di una data merce, e quindi il suo costo, rimanga in generale costante al crescere del volume della produzione. Come aveva già dimostrato Smith, il meccanismo della

concorrenza garantisce che le risorse (il lavoro, il capitale, le materie prime) si spostino dai settori in declino, meno profittevoli, ai settori in espansione, in cui il profitto è maggiore.

Ricardo pone come oggetto della sua analisi proprio quelle merci su cui agisce la concorrenza e che sono per questo riproducibili "quasi illimitatamente" con "l'impiego della operosità umana". Questi beni, il cui costo rimane costante al variare della produzione, costituiscono la stragrande maggioranza delle merci scambiate giornalmente sul mercato. "La parte di gran lunga maggiore delle merci" è, infatti, prodotta dal lavoro, "eccezion fatta per quella categoria, particolare e poco rilevante", come ad esempio "le statue antiche e i quadri rari, i libri e le monete scarse, i vini di particolare qualità" [Ricardo, 8]. Ora se le singole merci sono riproducibili in condizioni di costi costanti, ciò non vale per l'insieme delle merci che utilizzano un fattore scarso. Nel caso per esempio dell'agricoltura, la coltivazione di uno specifico prodotto può essere accresciuta senza che questo comporti un incremento dei costi di produzione; ma l'aumento della produzione del complesso dei prodotti agricoli di un dato paese accrescerebbe inevitabilmente i costi, giacché "la quantità di terra non è illimitata e la sua qualità non è uniforme [*ibid.*, 40].

La constatazione dei rendimenti decrescenti del suolo porta Ricardo a elaborare la sua teoria della rendita. Essendo il prezzo naturale dei prodotti agricoli pari al costo di produzione nelle condizioni più sfavorevoli, la rendita della terra deriva dal differenziale di produttività tra l'unità produttiva marginale e le altre più efficienti (vedi RENDITA). Ricardo condivide l'idea, efficacemente illustrata da Smith, che il progresso tecnico e la divisione del lavoro avrebbero comportato un aumento della produttività e un conseguente abbassamento dei costi unitari. Tuttavia lo stesso Smith aveva osservato che le possibilità di aumentare la divisione del lavoro erano minori in agricoltura che nell'industria manifatturiera; di conseguenza l'agricoltura avrebbe registrato incrementi di produttività inferiori a quelli dell'industria. Inoltre, all'epoca di Ricardo era diffusa l'opinione che l'abbassamento dei costi derivante dal progresso tecnico non sarebbe stato sufficiente a compensare l'aumento dei costi causato dall'estensione e intensificazione della coltivazione agricola conseguente all'aumento della popolazione. Tale pessimistica valutazione sulle

possibilità del progresso tecnico di contrastare la tendenza verso i rendimenti decrescenti della terra induceva Ricardo a ritenere che il crescente costo dei prodotti agricoli avrebbe comportato una espansione delle rendite e una parallela contrazione dei profitti, con la conseguenza di un progressivo rallentamento dell'accumulazione e dello sviluppo.

Ricardo utilizza, dunque, l'analisi delle variazioni dei costi di produzione per formulare le teorie, strettamente connesse, della rendita, della distribuzione del reddito e dell'accumulazione, e non già, come faranno successivamente gli economisti neoclassici, per determinare i prezzi relativi delle singole merci. Egli infatti concepiva l'aumento della divisione del lavoro non come un fenomeno strettamente dipendente dall'incremento della quantità prodotta, ma piuttosto come un effetto del progresso generale. D'altro canto, il riconoscimento che una maggior produzione agricola portava a un aumento del costo non poteva essere considerato come una causa normale di variazione del prezzo relativo delle merci: l'aumento del costo riguardava infatti il complesso di tutte, o quasi tutte le merci, giacché quasi tutte, in ultima analisi, derivano dalla produzione agricola, e quindi l'azione della produttività decrescente aumentava in proporzione il costo di ciascuna (Sraffa (a), 121).

Va infine menzionata la teoria ricardiana dei costi comparati. Per Ricardo nel caso del commercio internazionale i beni non si scambiano direttamente in proporzione al lavoro in essi contenuto, come accade nell'ambito di un singolo paese, ma in base al principio dei costi comparati (vedi COMMERCIO INTERNAZIONALE). Secondo tale principio ogni paese tende a esportare quei beni per i quali ha un vantaggio comparato maggiore (o un costo comparato minore) e a importare quei beni per i quali il suo vantaggio comparato è minore. Se ad esempio un dato paese è in grado di produrre sia i tessuti che il grano con una quantità di lavoro inferiore a un altro paese, e se il vantaggio nei confronti dell'altro è maggiore per la produzione dei tessuti che per la produzione del grano, esso ha convenienza a esportare tessuti e a importare grano. Infatti, pur avendo un vantaggio assoluto nei confronti dell'altro paese sia per i tessuti che per il grano, esso ha un vantaggio relativo maggiore per i tessuti che per il grano [Ricardo, 92-94].

1.4 Ricardiani e antiricardiani: primi sviluppi del concetto di costo reale soggettivo

Ricardo nega la tesi sostenuta dai suoi contemporanei Say, Lauderdale e Malthus, secondo cui il valore di scambio è regolato dalla domanda, oltre che dal costo di produzione. Per Ricardo la domanda può solo determinare momentanee oscillazioni del prezzo di mercato attorno al prezzo naturale. Infatti, in una situazione di libera concorrenza, gli spostamenti di capitale dagli impieghi meno redditizi a quelli più redditizi fanno sì che la quantità offerta tenda a uguagliare la quantità domandata, così i prezzi di mercato gravitano costantemente attorno ai prezzi naturali, ossia al costo di produzione [Ricardo, 56-59, 290-93] (Ricardo (a), 46). Ma le divergenze tra Ricardo da una parte, e Say, Lauderdale e Malthus dall'altra non sono limitate al diverso ruolo della domanda, ma coinvolgono la stessa definizione di costo di produzione.

Malthus accusa Ricardo "di confondere l'importante differenza tra costo e valore": "Ovunque vi sia profitto - egli osserva - il valore di una merce in rapporto al lavoro (*in exchange for labour*) è generalmente maggiore del lavoro che è stato impiegato in essa." Per Ricardo è facile ribattere che Malthus confonde il lavoro impiegato, che rappresenta il costo complessivo, con il costo del lavoro, che fa parte delle spese di produzione; in realtà, "il valore di scambio delle merci", che "è proporzionale alla quantità di lavoro impiegata", "può essere doppio del costo del lavoro" (*ibid.*, 100-02). In più occasioni Ricardo, rilevando l'assoluta mancanza di chiarezza di Malthus nel definire il costo, distingue il costo reale, costituito dal lavoro necessario, dal costo di produzione, che è pari al prezzo naturale, e che si risolve, secondo le circostanze che regolano la distribuzione del reddito, in salari e profitti (*ibid.*, 34 sg., 273; Ricardo (c), vol. 8, 279).

Molto più netta, rispetto a Malthus, è la posizione di Say che riprende in modo esplicito la teoria delle tre componenti del prezzo di Smith. Sbarazzatosi della teoria del valore lavoro, Say può elaborare una teoria della produzione in cui i tre fattori lavoro, capitale e terra "contribuiscono, ciascuno per la sua parte, alla produzione" [Say, vol. 2, 26]. In questo ambito il costo corrisponde alla somma delle remunerazioni dei fattori produttivi,

le quali dipendono dal contributo di ciascun fattore alla produzione. Si noterà la diversa prospettiva in cui si colloca Say rispetto a Ricardo. La definizione ricardiana di costo, come dispendio di attività lavorativa, si pone dal punto di vista dell'intero sistema economico; mentre la definizione di costo di Say, come somma delle spese implicate nella produzione, si pone dall'esclusivo punto di vista dell'imprenditore.

Negli anni venti-trenta Senior e Longfield, riprendendo la teoria additiva del valore di scambio di Smith e sviluppando il concetto di lavoro come sacrificio, elaborano una dottrina che avrebbe poi assunto il nome di teoria del "costo reale soggettivo". L'idea di Smith del lavoro come sacrificio, e perciò creatore di valore, è estesa da Senior anche al capitale. Considerando l'astinenza dal consumo immediato come il sacrificio del capitalista, Senior può fare del capitale una fonte di produzione nello stesso senso del lavoro. Definito costo di produzione "la somma del lavoro e dell'astinenza" (Senior, 101), egli designa con il termine "astinenza" "quel fattore distinto dal lavoro e dall'opera della natura, il cui concorso è necessario per l'esistenza del capitale e che si trova con il profitto nello stesso rapporto in cui il lavoro si trova con il salario" (*ibid.*, 59). Il profitto e il salario costituiscono rispettivamente la ricompensa del sacrificio affrontato dai soggetti intervenuti nella produzione. Dunque il salario e il profitto sono posti sullo stesso piano in quanto pagamenti di "costi reali". Secondo tale concezione la rendita fondiaria è un sovrappiù residuo, perché non esiste un costo reale, per quanto soggettivamente inteso, che corrisponda all'uso dei doni della natura. Senior afferma che sul piano pratico "la distinzione tra profitto e rendita cade, non appena il capitale (...) diventi, per donativo o eredità, proprietà di una persona alla cui astinenza e attività non è dovuta la sua creazione"; il profitto sarebbe in questo caso "un dono della fortuna, non un risultato di un sacrificio" (*ibid.*, 129). Escludendo dal profitto e chiamando rendita ogni remunerazione "ottenibile senza sacrificio" (*ibid.*, 91), Senior - come è stato osservato - è indotto a considerare rendita "una parte larghissima di ciò che tutti chiamano profitto, dato che nelle moderne comunità, civilizzate e ricche, la proprietà ottenuta per via d'eredità è di gran lunga più cospicua della proprietà acquisita col risparmio dei viventi" (Cannan, 213-15).

Sebbene l'ammissione di Senior ne limitasse la validità generale, la teoria dell'astinenza si diffuse rapidamente. Il principio dell'astinenza segnò una tappa importante nel pensiero economico perché introdusse una concezione soggettiva del costo. Mentre in Ricardo il costo reale consisteva nel dispendio fisico di attività lavorativa, in Senior il costo è rappresentato dalla riluttanza psicologica dei singoli agenti della produzione a lavorare e risparmiare. Senior, assumendo contemporaneamente il punto di vista del singolo lavoratore e del singolo capitalista, completa, rispetto alla concezione di Ricardo, il rovesciamento di prospettiva già presente in Say.

La teoria del costo reale soggettivo è accettata da McCulloch, James Mill e Stuart Mill, che sono i più autorevoli economisti ricardiani. Essi, pur rimanendo formalmente fedeli a Ricardo, ne hanno infatti abbandonato la linea di ricerca per le difficoltà analitiche della teoria del valore lavoro, evidenziate dallo stesso Ricardo, e per l'insoddisfazione di una concezione del profitto come sovrappiù (vedi PROFITTO; VALORE). Stuart Mill utilizza l'espressione costo di produzione in due sensi: talvolta per indicare la somma delle spese implicate nella produzione, le quali costituiscono il valore minimo essenziale affinché una merce sia prodotta in modo permanente; talvolta per esprimere il costo reale soggettivo in termini di penosità, astinenza e rischio, dei singoli agenti intervenuti nella produzione. Mill, ritenendo di completare e migliorare la teoria ricardiana, ne attua, in realtà, un completo, anche se inconsapevole, stravolgimento. Schumpeter osserverà con ragione che Mill ha trasformato la teoria ricardiana del valore, basata sulla quantità di lavoro, nella teoria del costo reale soggettivamente inteso, consegnando così "a Marshall già pronta la teoria dei due fattori del 'costo reale': la disutilità (penosità) sperimentata dal lavoratore (...) e l'astinenza sperimentata dal risparmiatore" [Schumpeter (c), vol. 2, 807].

1.5 Il costo di produzione in Marx

Marx critica aspramente l'idea che il valore abbia origine dal sacrificio del lavoratore e dall'astinenza del capitalista. "Un semplice negativo - egli scrive - in quanto mera negazione, sacrificio ascetico, non crea nulla. Uno può macerarsi, come i monaci

(...), e questa quantità di sacrificio che egli fa non cava un ragno dal buco" [Marx (a), vol. 2, 281]. D'altra parte "se il lavoro per esempio fa piacere all'operaio – come certamente fa piacere all'avarò l'astinenza – il prodotto non perde nulla del suo valore" [*ibid.*, 279]. Marx osserva poi sarcasticamente che per Senior anche il capitalista fa un sacrificio, il sacrificio dell'astinenza, in quanto "si arricchisce, invece di consumare direttamente il suo prodotto"; così "tutte le condizioni del processo lavorativo si trasformano", per Senior, "in altrettante pratiche ascetiche del capitalista". In realtà, conclude Marx, "l'economista volgare non ha mai fatto la semplice riflessione che ogni azione umana può essere concepita come 'astinenza' del suo opposto": "mangiare è astinenza dal digiuno, lavorare è astinenza dall'ozio, oziare è astinenza dal lavoro" [Marx (b), vol. 1, 653]. Secondo Marx il valore dei beni non sta nel sacrificio che si fa per produrli, ma nell'attività creativa del lavoro. È soltanto il lavoro, inteso appunto come attività creativa e positiva, la "sostanza dei prodotti in quanto valori" [Marx (a), vol. 2, 279]. La quantità di lavoro socialmente necessario, vale a dire la quantità di lavoro necessario in media, rappresenta il costo reale e costituisce "la misura dei valori" dei prodotti, perché i prodotti per loro natura "sono lavoro oggettivato" [*ibid.*, 280 sg.].

Marx contrappone il costo di produzione al costo capitalistico, corrispondente alla spesa di capitale anticipato per il pagamento delle materie prime, per il logorio degli impianti e per i salari [Marx (b), vol. 3, 54; (c), vol. 3, 90]. Infatti il lavoro che il capitalista paga per la produzione di una merce, e il lavoro che è necessario per la sua produzione, sono assolutamente differenti [Marx (c), vol. 3, 90]. Tale differenza è dovuta al fatto che il capitalista paga dell'intero "lavoro vivo" o "immediato", svolto nella produzione di un bene, soltanto la parte corrispondente al tempo di lavoro necessario per la produzione del valore della forza lavoro (il salario), mentre la parte di tempo eccedente la produzione del valore della forza lavoro, chiamata pluslavoro, non costa nulla al capitalista [Marx (b), vol. 1, 353; vol. 3, 56].

Dunque il costo capitalistico, ossia il capitale anticipato, è uguale al costo reale, dato dalla quantità di lavoro socialmente necessario, *meno* la quantità di pluslavoro. Poiché in Marx il costo reale corrisponde al valore dei beni, il plusprodotto al plusvalore e il costo capitalistico al prezzo di costo, quest'ultimo

è pari al valore meno il plusvalore. Il prezzo di costo, o costo capitalistico, rappresenta il limite minimo del prezzo di vendita di una merce, giacché se la merce è venduta a meno del suo prezzo di costo, gli elementi consumati del capitale produttivo non possono essere riprodotti integralmente attraverso il prezzo di vendita [Marx (b), vol. 3, 62-64]. Marx sottolinea la differenza tra i costi per la parte distrutta degli strumenti di lavoro e per le materie prime, che si trasferiscono nel valore della merce prodotta, e che per questo egli definisce capitale costante, e i costi rappresentati dai salari che permettono, attraverso l'impiego di "lavoro vivo", la creazione di un nuovo valore, di una eccedenza, e che perciò egli chiama capitale variabile [*ibid.*, 56 sg.]. La somma del capitale costante e del capitale variabile ci dà il capitale totale anticipato che, come abbiamo visto, è pari al costo capitalistico. Agli occhi del capitalista il plusvalore appare come un'eccedenza sul costo capitalistico complessivo e sembra così risultare da tutti gli elementi che concorrono alla formazione del costo capitalistico (appartenenti sia al capitale costante che a quello variabile). Ciò accade perché nella formazione del costo capitalistico non è dato di rilevare la differenza tra il capitale costante, che trasferisce il proprio valore, e il capitale variabile, che crea un nuovo valore [*ibid.*, 63].

Secondo Marx le spese per la realizzazione del valore, ossia i costi puri di circolazione necessari alla trasformazione del capitale merce nel capitale denaro, e le spese di sorveglianza e coordinamento, legate all'esistenza del capitale come rapporto sociale, devono essere considerate "falsi costi" perché non contribuendo alla creazione del valore d'uso, non aggiungono valore alle merci. Tali spese rappresentano, "dal punto di vista dell'intera classe capitalistica, una sottrazione di plusvalore" [Marx (a), vol. 2, 160, 296-99; (b), vol. 2, 153; vol. 3, 348]. Il profitto commerciale, calcolato sul capitale commerciale anticipato, è infatti detratto al plusvalore sociale. Marx dedica molto spazio ai costi puri di circolazione e alla determinazione del profitto commerciale perché vuole confutare l'idea, sostenuta da Proudhon, che il costo capitalistico (o prezzo di costo) costituisca "l'effettivo valore delle merci, e il plusvalore derivi invece dalla vendita della merce a un prezzo superiore al suo valore" [Marx (b), vol. 3, 66]. Punto centrale dell'analisi di Marx è la dimostrazione che il plusvalore ha invece origine nel processo stesso di produ-

zione dei valori d'uso delle merci dalla differenza tra il valore che il lavoro aggiunge al prodotto e il valore della forza lavoro, rappresentato dal salario.

Per quanto riguarda la relazione tra costo e volume della produzione, Marx osserva che "se tutte le circostanze restassero invariate", l'aumento della quantità prodotta di una merce comporterebbe un aumento proporzionale della quantità di lavoro. Di conseguenza il costo reale di produzione, relativo ad ogni unità di prodotto, risulterebbe costante al crescere della produzione [Marx (*b*), vol. 3, 230]. Tuttavia nel sistema capitalistico le condizioni tecniche e sociali del processo lavorativo sono di continuo rivoluzionate per aumentare la "forza produttiva del lavoro". La crescita della "forza produttiva del lavoro" consente di ridurre il tempo di lavoro necessario alla riproduzione del valore della forza lavoro stessa e, a parità di durata della giornata lavorativa, di aumentare il pluslavoro e quindi il plusvalore [*ibid.*, vol. 1, 353 sg.]. L'incremento della capacità produttiva del lavoro è ottenuto attraverso il miglioramento degli impianti, che permette di aumentare la produttività e ridurre gli scarti, e attraverso l'aumento della scala di produzione [*ibid.*, vol. 3, 111-13].

Come aveva osservato Babbage, le economie di scala hanno la loro origine, oltre che nell'aumento della divisione del lavoro e delle dimensioni delle macchine, anche nella crescita del numero degli addetti che lavorano in uno stesso stabilimento (Babbage, cap. 19). "Con la cooperazione di molti operai (...) si ottiene", da una parte, "un'economia nell'impiego dei mezzi di produzione" derivante "dal loro consumo comune nel processo di lavoro", dall'altra, un incremento della produttività dovuto all'aumento della divisione del lavoro. Inoltre l'aumento delle dimensioni del macchinario permette di realizzare rilevanti economie nei costi di costruzione, installazione ed esercizio [*ibid.*, vol. 1, 366, 371; vol. 3, 111-13]. La tendenza verso l'aumento della capacità produttiva comporta un continuo incremento del volume della produzione che, a sua volta, richiede un'espansione costante delle dimensioni del mercato [*ibid.*, vol. 1, 357]. Tuttavia, se il mercato non si espande allo stesso ritmo degli incrementi della produzione, si possono manifestare crisi di sovrapproduzione. Secondo Marx il sistema capitalistico è dunque caratterizzato da due opposte tendenze, ambedue connesse al feno-

meno delle economie di scala: da una parte la tendenza verso lo sviluppo e l'accumulazione, dall'altra la tendenza verso la sovrapproduzione (vedi CRISI).

1.6 Costo di produzione e scarsità delle risorse produttive: primi sviluppi del concetto di costo opportunità

Per gli economisti antiricardiani, così come per molti seguaci di Ricardo, il costo di produzione era composto da elementi eterogenei: i beni di sussistenza, l'astinenza, la rendita. L'eterogeneità delle componenti del costo di produzione costituiva l'ostacolo principale alla spiegazione dei prezzi basata sul rapporto tra la domanda e l'offerta. È possibile parlare di equilibrio tra domanda e offerta solo se si rendono omogenee, e quindi confrontabili, le forze che stanno alla loro base, vale a dire l'utilità e il costo di produzione (Mill, 143). Questo problema è risolto dai primi economisti neoclassici riducendo il costo, inteso come la somma dei prezzi dei fattori, all'utilità indiretta, cioè all'utilità dei beni di consumo ottenuti con i mezzi di produzione in questione.

La determinazione dei prezzi relativi dei beni intermedi, e quindi delle spese di produzione, in base all'utilità dei beni finali è esplicita nello schema di equilibrio economico generale di Walras pubblicato nel 1874 [Walras, 422]. Essa ha origine dal fatto che si presuppone data la quantità di risorse disponibili e, di conseguenza, si concepisce la produzione come un processo a "senso unico" che va dai fattori di produzione ai beni di consumo. Se la quantità dei mezzi di produzione è limitata, il costo di produzione di un bene è un indice di scarsità dei fattori necessari alla sua produzione rispetto alla distribuzione e intensità delle preferenze individuali.

In polemica con Ricardo e Stuart Mill, Walras nega la validità del concetto classico di riproducibilità dei beni. Se le terre e le capacità personali esistono in quantità limitata, "come potrebbero i prodotti essere moltiplicati senza limiti?" [*ibid.*, 534]. Walras però fraintende la posizione classica. In regime di concorrenza, la riproducibilità non è riferita ai prodotti nel loro complesso, ma è intesa come la caratteristica delle singole merci. Nei

classici, e in particolare in Ricardo, non è dunque l'insieme delle merci che può essere prodotto senza limitazioni, ma qualche singola merce di cui aumenta la domanda. La tendenza verso l'uniformità del saggio di profitto tra i vari settori fa sì che i capitali si spostino nelle produzioni di quei beni la cui domanda è maggiore; la composizione dei mezzi di produzione può essere così adattata a qualsiasi flusso di prodotti (J. Robinson (*a*), 1). D'altra parte la presenza di mezzi di produzione scarsi, come la terra, comporta soltanto una riduzione del sovrappiù distribuito sotto forma di profitti.

La concezione della scuola austriaca è per alcuni aspetti molto vicina a quella di Walras. Secondo il principio dell'imputazione, introdotto nel 1871 da Menger e sviluppato poi dal suo allievo Wieser, è possibile derivare i prezzi relativi dei fattori di produzione, la cui quantità è data, dal contributo marginale che essi arrecano alla soddisfazione dei consumatori. In questo senso i prezzi dei beni intermedi dipendono dal mercato dei prodotti finali e, quindi, dalla struttura e dal livello della domanda. Menger calcola il contributo di ciascun fattore al processo produttivo accertando la perdita di utilità per i consumatori causata dal ritiro di una dose marginale di ogni fattore, ferme restando le quantità di tutti gli altri. Anche nel caso della teoria degli economisti austriaci l'ipotesi di quantità data dei mezzi di produzione, e quindi di scarsità delle risorse, riveste un ruolo essenziale. Menger considera "beni economici, cioè oggetto dell'analisi economica, quei beni la cui quantità è inferiore al bisogno" [Menger, 147 sg.]. Nella concezione della scuola austriaca e di Walras l'economia diviene la disciplina che studia il rapporto tra i bisogni infiniti, ma ordinabili secondo una data scala di preferenze individuali, e le risorse scarse che si prestano a impieghi alternativi. I classici erano interessati allo studio delle cause che determinano l'accumulazione del capitale e la crescita economica. Gli economisti neoclassici sostituiscono alla teoria classica dello sviluppo economico il concetto di equilibrio nell'ambito di una struttura analitica statica. Ai concetti di circolarità del processo produttivo, di riproducibilità delle merci e di sovrappiù, che sono i principali elementi distintivi del pensiero di Ricardo e della concezione classica in generale, gli economisti neoclassici sostituiscono i concetti di unidirezionalità della produzione dai mezzi di produzione ai beni di consumo, di scarsità delle risorse e di

efficienza e razionalità nel comportamento dei produttori e dei consumatori.

L'applicazione della dottrina dell'imputazione e lo studio degli usi alternativi delle risorse date portano Wieser e Böhm-Bawerk a formulare quello che è stato poi chiamato "principio del costo opportunità", secondo l'espressione coniata da D.I. Green nel 1894 (Green, 219). Il principio del costo opportunità spiega il costo di produzione in base al sacrificio dell'utilità di quei beni che si sarebbero potuti ottenere dalle risorse utilizzate per produrre il bene considerato. In questo senso, come ha efficacemente osservato Wicksteed, "il costo di produzione non è altro che l'utilità di qualcos'altro" [Wicksteed, 382]. Secondo gli economisti della scuola austriaca l'importanza delle teorie dell'imputazione e del costo opportunità consisteva nel dimostrare l'erroneità di tutte le teorie del valore basate sul "costo reale" (Wieser, 112).

Le opere di Wicksteed in Inghilterra e di Knight e Haberler negli Stati Uniti contribuiscono a diffondere nei primi tre decenni del Novecento tra gli economisti neoclassici il principio del costo opportunità. Tale diffusione è favorita dalla semplicità analitica e dalla capacità di questa dottrina di offrire una definizione di costo che sia alternativa a quella del costo reale, propria degli economisti classici. Nella Parte seconda ci soffermeremo sulla formalizzazione analitica della teoria del costo opportunità fornita da Haberler nell'ambito dell'ipotesi di equilibrio economico generale.

1.7 Il costo di produzione in Marshall

Il costo reale soggettivo

Marshall non accetta l'idea comune ai tre padri fondatori dell'economia neoclassica – Menger, Walras, Jevons – che i costi di produzione, ovvero i prezzi relativi dei fattori, siano in ultima istanza determinati dall'utilità, e che di conseguenza l'offerta sia un puro riflesso della domanda. Egli definisce costo monetario di produzione la somma dei "prezzi che si devono pagare per suscitare un'offerta adeguata degli sforzi" dei lavoratori e delle "attese" dei risparmiatori "occorrenti per la produzione". L'insieme di questi sforzi e attese costituisce il costo reale di produzione [Marshall, 478]. La simmetria fondamentale tra le forze

che reggono la domanda e l'offerta è qui ottenuta riducendo il costo di produzione, anziché all'utilità indiretta, come nel caso degli economisti della scuola austriaca, al costo psicologico, inteso come disutilità o utilità negativa, degli agenti intervenuti nella produzione.

Contrariamente alla teoria della scuola austriaca, in Marshall il costo, e quindi l'offerta, è del tutto indipendente dalla struttura e dal livello della domanda. All'origine della diversa concezione del costo tra gli economisti austriaci da una parte e Marshall dall'altra sta il fatto che i primi presuppongono data la quantità disponibile dei fattori produttivi, mentre il secondo assume che, a eccezione delle risorse naturali, l'offerta dei fattori vari in relazione allo sforzo e al sacrificio individuale richiesto alla loro creazione. La teoria del costo reale soggettivo di Marshall, basata sul concetto di disutilità, permette di conciliare la dottrina classica del costo di produzione, così come era stata tramandata dall'interpretazione milliana, con la nuova teoria dell'utilità marginale, in modo da poter affermare che i prezzi relativi sono determinati congiuntamente sia dalla domanda che dall'offerta. Secondo questa dottrina "il valore normale di ciascuna cosa (...) si regge come la chiave di volta di un arco, in equilibrio fra opposte forze che premono ai suoi due lati: le forze della domanda premono da una parte, quelle dell'offerta dall'altra" (Marshall (*b*), vol. 2, 593). Alla base della domanda vi è l'utilità marginale sperimentata dal consumatore; alla base dell'offerta vi sono lo sforzo e il sacrificio dell'attesa, sopportati rispettivamente dal lavoratore e dal risparmiatore.

Marshall pensa di poter eludere le critiche di Marx al concetto di astinenza adottando il termine "attesa" proposto da MacVane [Marshall, 351 nota 2] (MacVane, 481). L'attesa, che comporta il sacrificio della dilazione del consumo, non implica la creazione del capitale risparmiato, come nel caso dell'astinenza, ma semplicemente il suo possesso. L'interesse è la remunerazione del sacrificio inerente all'attesa perché pochi risparmierebbero senza interesse; così come si considera il salario la remunerazione del lavoro, perché pochi lavorerebbero senza compenso. Marshall include la rendita fondiaria tra le spese monetarie di produzione, ma nega che essa possa far parte del costo reale. Ciò dipende dal fatto che egli considera le spese monetarie di produzione "dal punto di vista del singolo produttore" e i costi reali soggettivi

"dal punto di vista sociale". Nel primo caso la terra "non è altro che una forma di capitale" e quindi il suo utilizzo comporta una spesa monetaria. Nel secondo caso, quando si considerano i costi "dal punto di vista della nazione nel suo complesso", la rendita *non* appare come remunerazione di un costo reale, perché nessun sacrificio è implicito nell'offerta di terra, che è un bene non riproducibile a differenza "di quegli strumenti di produzione che l'uomo incrementa senza limite" [Marshall, 276 sg., 667].

Come osserva Cannan, è però contraddittorio applicare il principio dell'attesa al capitalista e non al proprietario della terra, giacché quest'ultimo ha le stesse occasioni del capitalista di vendere la sua proprietà e di spendere i ricavi in consumo immediato (Cannan, 190). Queste contraddizioni trovano origine nel fatto che Marshall tenta di conciliare due diverse concezioni: quella soggettiva, su cui poggiano tutte le teorie della domanda e dell'offerta, e quella oggettiva propria di Ricardo. Ricompare in Marshall, anche se in forma diversa, l'antinomia presente in Smith tra la definizione del costo basata sulla dottrina delle tre componenti e quella basata sulla quantità di lavoro necessario alla produzione. Secondo la nuova impostazione, che utilizza il concetto di incremento marginale mutuato dal calcolo differenziale, il salario e l'interesse tendono a coincidere nel lungo periodo con il sacrificio sostenuto dai lavoratori e dai risparmiatori capitalisti non relativamente all'ammontare *totale* del lavoro e del capitale messi a disposizione, come in Senior e Stuart Mill, bensì solo per le loro *unità marginali*.

Marshall ammette "che un certo complesso di sforzi e astinenze diversi non è suscettibile di essere posto a confronto con un altro" (Marshall (*a*), 125 sg.). È necessario perciò misurare il sacrificio del lavoratore e l'attesa del risparmiatore con "il criterio che fornisce la scienza economica", mediante i rispettivi valori di mercato [Marshall, 79-81]. Ma il mutamento rapido delle condizioni di domanda e di offerta, gli attriti, l'ignoranza sulle condizioni presenti e future del mercato, impediscono che tali misure monetarie corrispondano ai sacrifici sostenuti direttamente o indirettamente nella produzione [*ibid.*, 104, 487-91]]. Affinché i costi reali soggettivi coincidano con le spese monetarie di produzione è necessario ricorrere all'ipotesi di stato stazionario di lungo periodo. Secondo Marshall, se le condizioni generali della vita fossero stazionarie per un lasso di tempo abbastanza lungo da

permettere a tutte le forze economiche di produrre appieno i loro effetti, vi sarebbe identità tra prezzi di mercato, prezzi normali, costi monetari normali e costi reali [Marshall, 487 sg.]. Tuttavia lo stato stazionario, in base a cui s'ipotizza di "lasciare trascorrere un tempo sufficiente affinché le cause possano produrre i loro effetti", è una "finzione teorica". "Nel mondo reale tutte le circostanze mutano azione di continuo" e un dato periodo di tempo, per quanto lungo sia, non è mai sufficiente a consentire che le tendenze che si stanno descrivendo si manifestino pienamente, anche perché nel frattempo "le cause stesse possono essere mutate" [*ibid.*, 104, 511 sg.]. La validità della dottrina secondo la quale il prezzo a cui una merce può essere prodotta rappresenta il suo costo reale, è pertanto limitata alla "finzione dello stato stazionario".

Il costo reale in Marshall e nei classici: un'apparente analogia

È utile porre in evidenza le profonde differenze tra la teoria del costo reale di Marshall e quella degli economisti classici e di Ricardo in particolare. Oltre al radicale mutamento della definizione di costo, che nella teoria ricardiana è inteso in termini fisici e in quella di Marshall in termini psicologici, vi è un rovesciamento della catena causale. In Ricardo è il costo di produzione che, in ultima analisi, regola i valori di scambio, quindi la direzione va dal costo ai prezzi rilevabili ogni giorno sul mercato. In Marshall, al contrario, si risale, grazie all'ipotesi-finzione dello stato stazionario di lungo periodo, dai prezzi di mercato, soggetti al continuo e mutevole gioco tra la domanda e l'offerta, fino al costo reale soggettivo, il quale determina, assieme alle condizioni di domanda di lungo periodo, il prezzo normale. Il rovesciamento del nesso causale porta a un mutamento dell'asse analitico rispetto a Ricardo: dal breve periodo (il mondo concreto degli attriti) al lungo periodo (in cui le forze dominanti hanno il tempo sufficiente per superare gli attriti), anziché dal fattore determinante (il costo di produzione) al fenomeno determinato (i prezzi di mercato) (Lunghini, 21 sg.). Marshall, che legge Ricardo attraverso l'interpretazione milliana, non si avvede di quanto la sua concezione sia lontana da quella ricardiana, e addirittura ritiene la sua opera un completamento di quella di Ricardo. Secondo Marshall, Ricardo "fa dipendere il costo di pro-

duzione (...) dalla qualità oltretutto dalla quantità" di lavoro utilizzato nella produzione, "dall'ammontare di capitale accumulato necessario per coadiuvare il lavoro e dalla lunghezza del periodo di tempo durante il quale tale aiuto era richiesto" [Marshall, 674]. Questa interpretazione induce Marshall a ritenere Ricardo il teorico del valore costo anziché del valore lavoro (Becattini, XLIV). Per Marshall, dunque, si tratta di completare la teoria di Ricardo (che aveva attribuito "un rilievo sproporzionato al lato del costo di produzione"), dando alla domanda, sulla scorta delle successive elaborazioni sull'utilità marginale, il ruolo che le spetta nella teoria dei valori di scambio.

Bisogna tuttavia notare che Ricardo ha escluso la teoria della domanda non perché la sua azione fosse "meno oscura di quella del costo di produzione" e neppure a causa della "sua trascurata brevità", come vorrebbe Marshall [*ibid.*, 166, 674], ma perché la domanda non ha alcun ruolo nel processo di formazione dei valori di scambio, dato che le merci sono in generale prodotte a costi costanti. La teoria marshalliana del costo reale soggettivo, piuttosto che il "miglioramento" della teoria ricardiana, appare come lo sviluppo delle teorie della domanda e dell'offerta che nel pensiero classico trovano la loro origine nella dottrina additiva di Smith e il loro punto di arrivo nella teoria milliana, in cui il costo è al tempo stesso la somma delle spese di produzione e il sacrificio soggettivo.

Costi e curve di offerta

Marshall distingue tre possibili andamenti dei costi di produzione di una impresa in rapporto alla quantità prodotta: costi crescenti (rendimenti decrescenti), costi decrescenti (rendimenti crescenti) e costi costanti (rendimenti costanti).

L'andamento dei costi di produzione forma la scheda dei prezzi di offerta della singola impresa. Il prezzo di offerta cresce quando si manifestano costi crescenti, decresce quando prevalgono costi decrescenti ed è costante quando le azioni "dei rendimenti crescenti e decrescenti si compensano a vicenda" [Marshall, 451, 484-87]. Secondo Marshall nell'agricoltura e nelle attività estrattive, dove è preponderante il peso delle risorse naturali, si manifesta la tendenza verso rendimenti decrescenti; mentre nell'industria manifatturiera e nel settore dei trasporti, dove l'attività del-

l'uomo prevale e il costo delle materie prime ha uno scarso peso, agisce la legge dei rendimenti crescenti [Marshall, 275, 450 sg.]. Questi sono attribuibili alle *economie interne ed esterne* che è possibile ottenere aumentando il volume della produzione. Le economie interne dipendono dall'organizzazione e dall'efficienza della singola impresa, mentre le economie esterne sono legate sia allo sviluppo generale dell'industria (ad esempio il progresso nei trasporti, nelle comunicazioni e nelle informazioni), sia all'ampliamento del mercato. L'espansione del mercato consente una maggior localizzazione delle imprese, ovvero l'aumento della concentrazione nella stessa zona di un gran numero di piccole imprese di genere analogo, e anche una maggiore articolazione e suddivisione delle imprese per singole fasi del ciclo produttivo [*ibid.*, 389 sg., 402]. Egli è consapevole dell'incompatibilità dei rendimenti crescenti con l'ipotesi di libera concorrenza, giacché essi potrebbero portare qualsiasi impresa, che ottenga condizioni iniziali favorevoli a conquistare il monopolio dell'intera attività del suo ramo [*ibid.*, 622]. Tuttavia egli ritiene che sia assai difficile che in una singola azienda prevalgano costi decrescenti in un lasso di tempo sufficiente per portare al monopolio. La singola impresa – diversamente dall'industria a cui appartiene, che può godere di rendimenti costanti o crescenti – trova infatti nel corso del suo sviluppo ostacoli che impediscono l'affermarsi prolungato della tendenza verso i rendimenti crescenti. Tali ostacoli sono causati dal concorrere di diversi elementi: *a)* il prevalere delle economie esterne sulle economie interne; *b)* il ciclo vitale che caratterizza la vita di ogni impresa; *c)* le imperfezioni del mercato.

a) Se i rendimenti crescenti sono causati da economie esterne all'impresa, ma interne all'industria, essi sono un fenomeno comune a tutte le imprese che operano in un dato mercato e quindi non possono condurre verso l'eliminazione della concorrenza.

b) Lo sviluppo delle imprese si arresta generalmente a un certo punto per il venire meno dell'originalità e dello spirito di iniziativa del fondatore. Sicché la vita dell'impresa è caratterizzata da una fase iniziale di espansione, seguita, una volta raggiunta l'efficienza massima e quindi i costi unitari più bassi, da una fase di declino e di perdita di efficienza. D'altra parte la trasformazione di alcune imprese in società per azioni, osserva Marshall, sebbene ne eviti il declino connesso al ciclo vitale, ne causa la

perdita di elasticità e ne arresta il progresso. Pertanto la presenza di società per azioni in un dato ramo di attività non impedisce l'esistenza di "imprese rivali più giovani" e di minori dimensioni [Marshall, 413, 443, 448].

c) Infine il terzo ostacolo allo sviluppo delle dimensioni della singola impresa è rappresentato dalle difficoltà di smercio. Le vendite di ciascuna impresa, sostiene Marshall, sono limitate "a quel mercato particolare che l'impresa ha conquistato lentamente (...), e sebbene la produzione possa essere accresciuta molto presto, la vendita non può esserlo". In altre parole, la "curva particolare di domanda" di ciascun produttore "sarà generalmente molto rapida", cosicché "ogni rapido aumento della produzione abbasserà probabilmente in quel mercato il prezzo di domanda in misura sproporzionata alle nuove economie di cui essa godrebbe" [*ibid.*, 414 sg., 620 sg.].

La determinazione del punto di equilibrio della singola impresa è reso dunque compatibile con il caso dei rendimenti crescenti attraverso il riferimento a una situazione *non perfettamente concorrenziale* (curva particolare di domanda per la singola impresa, società per azioni) e l'applicazione dell'*analisi dinamica* (economie esterne dovute al progresso tecnico e ciclo vitale dell'impresa). Per Marshall è però impossibile formulare una teoria dei prezzi relativi (e del costo reale) rimanendo nell'ambito di un'analisi dinamica in cui "le tabelle di domanda e di offerta variano di continuo" a causa del "sorgere di forze perturbatrici", connesse con il trascorrere irreversibile del tempo. Per superare la "casualità" dei prezzi di mercato, continuamente mutevoli, e per far sì che i costi reali soggettivi coincidano con le spese monetarie di produzione, è necessario escludere il mutamento economico ricorrendo alla "finzione dello stato stazionario", in cui l'influenza del tempo non si fa sentire [*ibid.*, 487, 510-13].

Nell'ambito dell'analisi statica egli distingue il breve periodo, in cui si considera fisso "l'ammontare esistente dei mezzi di produzione", e il lungo periodo, in cui si suppone che le forze economiche presenti in dato momento abbiano il tempo sufficiente per poter esplicare pienamente i loro effetti. Nel breve periodo un aumento della produzione porta a una crescita più che proporzionale dei costi a causa dell'impiego, sempre meno efficiente, di una crescente quantità di lavoro su un dato impianto. Ne segue che, in generale, il prezzo di offerta aumenta con l'au-

mentare della quantità prodotta [Marshall, 519-672]. Il prezzo di offerta del singolo produttore è dato, nel breve periodo, dal costo marginale, ossia dal "costo di produzione di quelle merci che (...) non sarebbero prodotte se per esse si fosse atteso un prezzo minore" [*ibid.*, 519]. Se in una particolare situazione di mercato la quantità domandata scende molto, il costo marginale può risultare inferiore al costo unitario totale. Il costo totale è composto dalle "spese generali o supplementari" relative alle spese di ammortamento e dai "costi primi o diretti", i quali comprendono le spese per le materie prime, per il logorio degli impianti e i salari degli operai [*ibid.*, 502 sg.]. Sebbene il singolo imprenditore possa continuare a produrre nel breve periodo per un prezzo di mercato che copra solo i costi variabili, in realtà, quando il prezzo scende al di sotto del costo unitario totale, egli tenderà a offrire una quantità minore di quella che uguaglia il prezzo al costo marginale per paura di guastare il mercato e di ridurre "le possibilità di ottenere in futuro un prezzo migliore dai suoi clienti" [*ibid.*, 520 sg.]. Nel lungo periodo i mezzi di produzione possono essere adeguati in relazione alle aspettative della domanda; le aziende sono in grado di ampliare il volume della produzione utilizzando "ciascun fattore nella debita proporzione" [*ibid.*, 276, 319]. Nell'ambito dell'analisi statica di lungo periodo è possibile considerare le economie che risultano dall'adattamento delle idee esistenti, ma non quelle derivanti da nuove invenzioni [*ibid.*, 623 sg.].

La preoccupazione costante di non staccarsi troppo dalla realtà concreta e, al contempo, la convinzione dell'inadeguatezza dei modelli astratti a cogliere la complessità delle condizioni economiche reali, inducono Marshall a non applicare strettamente l'ipotesi statica e di concorrenza, e a non spingersi fino alla dimostrazione analitica delle relazioni tra le curve dei costi medi e dei costi marginali, dimostrazione analitica che invece sarà effettuata da Edgeworth e Pigou negli anni tra il 1911 e il 1920, sulla base della legge delle proporzioni tra i fattori (Edgeworth, 209-15) [Pigou, 690-99]. Questa legge, che è conosciuta come la "legge delle proporzioni variabili" (Cassels, 106), costituisce l'*estensione alla produzione in generale* della "legge dei rendimenti prima crescenti e poi decrescenti" che Turgot formulò per primo in relazione al caso particolare dell'aumento delle dosi di lavoro su un dato terreno (Turgot, 418). Marshall sottolinea i pericoli di

mettere "in forma definitiva" la teoria statica dell'equilibrio perché più essa "è spinta alle più remote e complicate conseguenze logiche", più "si allontana dalla vita reale". La nozione di relazione funzionale tra produzione e costi in un'analisi rigidamente di breve periodo non presenta difficoltà, ma l'ipotesi di stato stazionario riferito a "periodi relativamente lunghi è pericolosa", poiché è necessario forzare molto la realtà per supporre che le forze che si escludono dalla considerazione (ad esempio il progresso tecnico) rimangano inerti [Marshall, 526, 624].

La teoria statica dell'equilibrio di lungo periodo, se rigorosamente intesa, non è infatti applicabile alle merci che obbediscono alla legge dei rendimenti crescenti. E ciò a causa delle stesse ipotesi di stato stazionario di lungo periodo, nelle quali le variazioni del prezzo di offerta e della quantità prodotta vengono considerate come dipendenti esclusivamente l'una dall'altra, senza cioè alcun riferimento "al periodo di tempo necessariamente occupato da ciascuna impresa nell'estendere" il suo mercato particolare e quindi "la sua produzione" [*ibid.*, 671]. D'altra parte, nel caso dei rendimenti crescenti l'andamento dei costi non può comunque essere rappresentato da una curva statica di lungo periodo a causa della irreversibilità delle economie di scala. Le "larghe economie" dovute agli "sviluppi di applicazioni meccaniche della divisione del lavoro e dei mezzi di trasporto e a miglioramenti organizzativi di ogni genere, una volta conseguiti, non si abbandonano facilmente". Pertanto, se per una qualsiasi ragione la produzione dovesse essere ridotta, dopo un periodo di espansione, "il prezzo di offerta non ritornerebbe per lo stesso cammino (...) ma seguirebbe un percorso meno elevato" [*ibid.*, 1069].

In conclusione e in estrema sintesi, in Marshall risulta evidente una irriducibile contraddizione tra l'analisi statica del "prezzo normale" di lungo periodo, che impedisce di tener conto del fenomeno dei rendimenti crescenti, e l'analisi della realtà industriale, in cui tale fenomeno assume grande rilevanza.

1.8 Relazione tra costo e quantità prodotta: l'analisi critica di Sraffa

John Clapham, in un famoso articolo apparso nel 1922 sull'"*Economic Journal*", pone l'accento sulle rilevanti difficoltà che si incontrano nell'applicare la categoria dei rendimenti decre-

scenti, crescenti e costanti a industrie concrete. Tali categorie sono così rimaste, a suo giudizio, delle "scatole vuote", malgrado costituiscano una parte essenziale dell'apparato teorico della scuola marshalliana. A conferma di ciò, egli ricorda che numerosi brani della *Economics of Welfare* di Pigou iniziano ipotizzando una prevalenza di condizioni di rendimenti decrescenti o crescenti, senza che a tale ipotesi segua mai una sola illustrazione di quali industrie siano effettivamente caratterizzate da rendimenti decrescenti o crescenti (Clapham, 160 sg.).

Nel saggio pubblicato nel 1925 sugli "Annali di Economia" Sraffa dimostra che le "scatole vuote" di cui parla Clapham non sono state "riempite" con industrie concrete, non tanto – come sostiene Pigou in risposta a Clapham – per la mancanza di ricercatori "versati nelle parti più intricate dell'analisi economica e nella teoria statistica moderna" e nello stesso tempo in grado di "condurre un profondo e dettagliato studio di particolari industrie" (Pigou, 465), quanto per la mancanza di un criterio non arbitrario di classificazione delle diverse industrie. Una stessa industria può, infatti, essere considerata a costi crescenti o a costi decrescenti a seconda della definizione di "industria" che si ritiene preferibile adottare (Sraffa (a), 120). Quanto più ampia sarà la definizione di industria, cioè quanto più questa tenderà a includere tutte le aziende che impiegano un dato fattore della produzione (come ad esempio l'agricoltura o la siderurgia), tanto più sarà probabile che in essa prevalgano le forze che operano nel senso della produttività decrescente, poiché il fattore che è caratteristico dell'industria stessa (la terra e il ferro) con l'aumentare della produzione resta, in generale, costante. Al contrario, quanto più restrittiva sarà quella definizione, e quindi quanto più essa tenderà a includere solo quelle aziende che producono un dato tipo di merce (per esempio la frutta e i chiodi), tanto più sarà probabile che in essa predominino condizioni di produttività crescente. Ciò dipende dal fatto che i costi crescenti sono il risultato di variazioni nelle proporzioni in cui sono impiegati i mezzi di produzione, mentre i costi decrescenti derivano da variazioni nella quantità assoluta dell'insieme dei mezzi di produzione (*ibid.*, 152 sg.).

Le due leggi dei rendimenti decrescenti e crescenti hanno avuto origini molto diverse. Come abbiamo visto, infatti, la produttività decrescente era trattata dagli economisti classici in rela-

zione alla rendita fondiaria, ed era perciò connessa allo studio della distribuzione del reddito, mentre la produttività crescente era discussa da questi economisti in relazione alla divisione del lavoro, e compresa quindi nello studio della produzione. "Nessuno fino a tempi relativamente recenti – osserva Sraffa – aveva pensato di fondere quelle due tendenze in una sola legge della produttività non proporzionale, e considerare questa come una delle basi della teoria del prezzo" (Sraffa (a), 120). Solo dopo che gli studi sull'utilità decrescente avevano attirato l'attenzione sul rapporto fra prezzo e quantità consumata, è sorta per analogia l'idea di una connessione fra costo e quantità prodotta. L'importanza delle leggi di variazione del costo agli effetti della determinazione del prezzo delle singole merci è apparsa in conseguenza della dottrina, dovuta principalmente a Marshall, che ha affermato la simmetria fondamentale delle relazioni tra la domanda e l'offerta rispetto al valore. L'incrocio delle curve di domanda e offerta è condizionato dalla non-proporzionalità del costo di produzione rispetto alla quantità prodotta. Infatti, "se il costo di produzione di ogni unità della merce considerata non variesse al variare della quantità prodotta, la simmetria fondamentale sarebbe spezzata, il prezzo sarebbe determinato esclusivamente dalle spese di produzione e la domanda non potrebbe influire su di esso (*ibid.*, 121 sg.).

Lo scopo di Sraffa è di verificare se, ed entro quali limiti, sia possibile una "coordinazione" delle tendenze ai costi crescenti, decrescenti e costanti, in un'unica "legge dei costi non proporzionali" che consenta di costruire una curva statica collettiva di offerta di lungo periodo simmetrica alla corrispondente curva di domanda.

Una prima difficoltà è costituita dall'eterogeneità delle ipotesi che stanno alla base delle tre diverse tendenze. La tendenza ai costi crescenti si fonda sull'ipotesi che mutino le proporzioni in cui si combinano i diversi fattori della produzione. Si deve dunque assumere che almeno uno di essi rimanga costante mentre gli altri aumentano. È questo il caso di una industria esclusiva consumatrice di un dato fattore di produzione. Al contrario la tendenza verso costi decrescenti deriva da variazioni nella quantità assoluta di tutti i fattori. Tale tendenza è dovuta, in primo luogo, al miglioramento dei metodi di produzione – reso possibile soprattutto dall'aumento della divisione del lavoro connesso alla

crescita delle dimensioni aziendali – e, in secondo luogo, alla possibilità di ripartire le spese generali, che col crescere della produzione restano costanti, su un numero maggiore di unità prodotte. I costi decrescenti si presentano in quelle industrie che impiegano mezzi di produzione utilizzati anche in molte altre industrie. Data l'eterogeneità delle due tendenze verso i costi crescenti e decrescenti, è assai improbabile che i costi costanti derivino, come riteneva Marshall, dal compensarsi delle due opposte tendenze. È certamente molto più semplice pensare, osserva Sraffa, che il caso dei costi costanti si verifichi a causa dell'assenza di entrambe le tendenze (Sraffa (a), 150) [vedi Marshall, 451].

Le difficoltà del sistema, che può essere sinteticamente descritto come "l'incrocio delle curve di domanda e offerta", derivano, più che dall'eterogeneità delle diverse tendenze, dal fatto che esso è sottoposto a due condizioni: 1) concorrenza perfetta; 2) indipendenza delle condizioni di produzione della merce considerata da quelle di tutte le altre merci (ipotesi di *ceteris paribus*) (Sraffa (c)). La curva di offerta deve essere indipendente, oltre che dalla corrispondente curva di domanda, dalle curve di offerta di tutte le altre merci, perché essa rappresenta due sole variabili (il costo e la quantità prodotta). È quindi necessario supporre che, secondo l'ipotesi di *ceteris paribus*, con il variare della produzione di una merce, tutte le altre condizioni restino invariate. Inoltre la curva di offerta è valida solo per piccole variazioni della quantità prodotta. Se ci si allontana troppo dalla posizione iniziale di equilibrio, può rendersi necessaria la costruzione di una curva interamente nuova, perché una grande variazione sarebbe incompatibile con la condizione di *ceteris paribus*.

Queste condizioni riducono le possibilità di applicare l'ipotesi di costi crescenti alla curva di offerta di un prodotto limitatamente a quei casi eccezionali in cui la totalità di un fattore è impiegato nella produzione di una sola merce. Ma, in generale, ogni fattore è impiegato da un certo numero di industrie che producono merci diverse. In tal caso una curva di offerta a costi crescenti non è ammissibile. Infatti, se si tratta di un ridotto numero di industrie, e quindi di un ridotto numero di merci diverse, un aumento, anche piccolo, della produzione in un'industria provoca una maggiore utilizzazione del fattore comune e una crescita del suo costo, e ciò non solo nell'industria che ha aumentato la produzione, ma anche nelle altre industrie del gruppo poiché il

fattore in comune entra a formare il costo di ciascuna. Ma tale risultato è contrario alla condizione di *ceteris paribus*. Se invece si tratta di un gran numero di industrie, e quindi di merci prodotte, un piccolo aumento della produzione di una industria non può avere che effetti trascurabili, tanto sul costo della merce stessa, quanto sul costo delle altre del gruppo: l'offerta del prodotto di quell'industria dovrebbe essere considerata a costi costanti.

Altrettanto gravi sono le conseguenze delle condizioni a cui abbiamo accennato sulla curva di offerta a costi decrescenti. Nell'analisi dell'equilibrio parziale è necessario che i vantaggi dell'aumento della produzione nell'industria considerata non siano risentiti in alcun modo dalle altre industrie. Di conseguenza le economie della produzione su larga scala devono essere "esterne" dal punto di vista della singola azienda, ma "interne" dal punto di vista dell'industria. Ma questo è un caso assai raro, giacché nella realtà le economie esterne dipendono dallo sviluppo di grandi gruppi di industrie localizzate nella medesima area o affini per le caratteristiche dei prodotti.

Se nella determinazione degli equilibri particolari delle singole merci, in un sistema statico di libera concorrenza, non è possibile considerare curve di costi crescenti o decrescenti senza introdurre ipotesi che contraddicono la natura stessa del sistema, ne consegue, da questo punto di vista, che le merci siano prodotte in condizioni di costi costanti (Sraffa (a), 159). La vecchia teoria che, in regime di concorrenza, fa dipendere il valore delle merci solo dal costo di produzione – conclude Sraffa – sembra essere ancora la migliore di cui disponiamo (Sraffa (c), 126). Tuttavia, come Sraffa precisa in una lettera del 1926 a Keynes (Sraffa (b)), ciò non implica che nella vita reale prevalgono i rendimenti costanti. Anche se l'ipotesi di Ricardo è la migliore di cui disponiamo per una teoria semplice della concorrenza (cioè, di prima approssimazione), certamente è evidente che nella realtà vi è una connessione fra costo e quantità prodotta. Ma tale connessione non può essere considerata in un sistema di equilibrio parziale, in regime di concorrenza perfetta, come quello concepito da Marshall.

Nell'articolo del 1926, pubblicato sull'"Economic Journal", Sraffa osserva che se si vuole passare a un'ulteriore approssimazione alla realtà, prendendo in considerazione l'ipotesi più realistica di rendimenti non proporzionali, vi sono due possibili soluzioni (Sraffa (c)). La prima, "volendo continuare a proce-

dere sulla via della libera concorrenza", consiste nell'estendere "il campo dell'indagine in modo da considerare le condizioni di equilibrio simultaneo di numerose industrie". A quell'epoca il riferimento andava diretto al modello di equilibrio economico generale di Walras e Pareto. Sraffa tuttavia scarta immediatamente questa soluzione per due ordini di motivi: la sua "complessità", che, "almeno allo stato attuale delle nostre conoscenze", le "impedisce di essere feconda"; e l'impossibilità di far rientrare le circostanze che regolano le economie esterne entro le condizioni statiche necessarie allo schema di equilibrio economico generale. Una seconda via d'uscita, sulla quale Sraffa si sofferma più diffusamente, è indicata nell'abbandono dell'ipotesi concorrenziale e nello studio di quella "zona intermedia" tra i due casi estremi del monopolio e della concorrenza (Sraffa (c), 127 sg.).

Secondo Sraffa è inaccettabile l'idea che, quando la produzione è nelle mani di un gran numero di aziende interamente indipendenti, possono essere adottate le conclusioni proprie della concorrenza, anche se il mercato in cui le merci sono scambiate non è assolutamente perfetto. In realtà molti ostacoli che spezzano quell'unità del mercato, che è la condizione essenziale della concorrenza, non hanno la natura di "atriti", che possono ritardare o modificare leggermente gli effetti della concorrenza, ma sono essi stessi forze attive, che producono effetti permanenti e persino cumulativi (*ibid.*, 128). D'altra parte, prosegue Sraffa, l'esperienza mostra che la generalità delle aziende che producono merci manufatte lavorano in condizioni di costi medi totali individuali decrescenti, e non, come richiederebbe la teoria della concorrenza, a costi crescenti: "Uomini d'affari che pure ritengono di trovarsi in condizioni di concorrenza, troverebbero ridicola l'asserzione che il limite alla loro produzione risiede nell'aumento dei costi al crescere della quantità prodotta." L'ostacolo principale all'espansione della produzione non è nei costi, ma "nella difficoltà di vendere la maggior quantità di merce senza ridurre il prezzo", o senza incontrare sempre maggiori spese di smercio (*ibid.*, 129). Ciò dipende dal fatto che i compratori di una merce non sono indifferenti di fronte ai diversi produttori. Ogni ditta ha un mercato particolare e può influire sul prezzo. Come ha osservato Marshall, quando ogni azienda opera in un mercato particolare dobbiamo accoppiare la curva di offerta di ogni azienda con la sua particolare curva di domanda [Marshall, 612

nota a]. In questo caso, conclude Sraffa, non è possibile sommare curve particolari delle singole aziende in modo da formare una sola coppia di curve collettive di domanda e offerta (Sraffa (c), 131 sg.) [vedi Marshall, 612 nota a].

Ben presto Sraffa si dovette rendere conto che l'analisi delle forme di mercato non concorrenziali non poteva portare a una teoria rigorosa e generale dei prezzi relativi. Nel 1930, in occasione di un simposio sui rendimenti crescenti e l'impresa rappresentativa, egli lascia cadere i suggerimenti sull'abbandono dell'ipotesi concorrenziale e, in polemica con Robertson, ribadisce le sue critiche alla teoria marshalliana: "Tale teoria non può essere interpretata in modo da darle una coerenza interna, e in pari tempo da metterla d'accordo con i fatti che si propone di spiegare. Il rimedio di Robertson è quello di scartare la matematica (...) la mia opinione è che si debba scartare la teoria di Marshall" (Sraffa (d), 620). Così ciò che era sembrata una critica al tentativo degli epigoni di Marshall di ridurre la complessità e la ricchezza dell'analisi marshalliana all'interno di uno schema statico di domanda e offerta, finisce con il sancire l'abbandono della teoria marshalliana per l'impossibilità di fondare su di essa una teoria coerente dei prezzi e dell'impresa.

La ricerca di Sraffa si indirizza verso l'elaborazione di un modello della produzione capitalistica completamente svincolato dalle ipotesi fondamentali neoclassiche, il quale permetta la determinazione simultanea dei prezzi di tutti i beni, mezzi di produzione compresi. Tale modello è stato pubblicato solo nel 1960, sebbene, come egli stesso ci informa, le idee principali fossero già state formulate prima del 1930 [Sraffa, vi]. Le indicazioni contenute nell'articolo di Sraffa del 1926 sull'esistenza di una forma di mercato intermedia, che non può esser assimilata né alla concorrenza perfetta né al monopolio, costituiscono la base degli studi successivi sulle forme di mercato non concorrenziali che vanno dalle elaborazioni della Robinson e di Chamberlin sulla possibilità di un equilibrio stabile e in una situazione di concorrenza imperfetta o monopolistica, alla teoria di Kalecki sul grado di monopolio delle imprese, fino alle più recenti analisi di Sylos Labini e Bain sull'oligopolio (vedi MERCATO). Come si avrà modo di constatare, le analisi sulle forme di mercato non concorrenziali porteranno a una radicale riformulazione della relazione tra costi e quantità prodotta.

2.1 Il costo di produzione nella moderna teoria neoclassica

La riformulazione della teoria marshalliana dei costi di breve e di lungo periodo nell'ipotesi di concorrenza perfetta

Negli anni trenta si assiste a un ampio lavoro di "revisione" dell'analisi marshalliana sotto l'influsso delle critiche di Sraffa e, più in generale, del dibattito apparso sull'"Economic Journal" tra il 1922 e il 1930 sulla relazione tra costo e quantità prodotta in regime di concorrenza perfetta (Del Vecchio, 605). Tra gli economisti neoclassici si diffonde sempre più la convinzione che le difficoltà della teoria marshalliana dipendano dal fatto che essa si fonda sull'analisi dell'equilibrio parziale. "L'esperienza di almeno venti anni di controversie – afferma Lionel Robbins nel '34 – suggerisce che l'uso delle curve di offerta" (proprio dell'analisi dell'equilibrio parziale), "in luogo dell'apparato analitico dell'equilibrio economico generale, (...) comporti dei pericoli che possono intrappolare anche le menti più sottili e acute" (Robbins, 11). Lo schema di equilibrio economico generale appare dunque come l'unico in grado di offrire una soluzione rigorosa al problema della determinazione dei prezzi relativi. Esso però non fornisce una teoria dell'impresa; infatti – come ha osservato Walras – il modello di equilibrio economico generale "può persino prescindere dall'intervento degli imprenditori", considerando che "i servizi produttivi si scambiano in definitiva gli uni contro gli altri" [Walras, 323].

Per analizzare il comportamento dell'impresa, cioè il modo con cui essa tende, in una situazione concorrenziale, a massimizzare il profitto, è necessario ricorrere all'analisi dell'equilibrio parziale, isolando, con la condizione *ceteris paribus*, la relazione tra i costi

e il volume della produzione. Nello schema di equilibrio parziale i prezzi dei mezzi di produzione sono supposti dati e il costo totale (CT) è funzione solo della quantità prodotta (Q):

$$CT = f(Q). \quad [2.1]$$

Sebbene lo studio della funzione del costo di produzione non possa trovare posto in una teoria dell'equilibrio economico generale in cui i prezzi degli input non sono dati a priori, ma sono determinati simultaneamente ai prezzi e alle quantità dei beni prodotti, esso costituisce un'integrazione dello schema di equilibrio economico generale sul comportamento dell'impresa. La riduzione dell'analisi dell'equilibrio parziale entro i limiti della teoria del comportamento dell'impresa spezza l'unità presente nell'analisi marshalliana tra la teoria dei prezzi relativi e la teoria dei costi di produzione.

L'articolo di Jacob Viner del 1931, che rappresenta forse il più noto contributo all'opera di "sistemazione" della teoria marshalliana dell'impresa e della produzione, sviluppa le analisi delle curve di offerta di Edgeworth e di Pigou, basate sull'applicazione della "legge delle proporzioni variabili". I concetti marshalliani di breve e lungo periodo sono depurati da tutti quei riferimenti all'analisi dinamica, presenti nell'esposizione di Marshall, che risultano incompatibili con le ipotesi rigorosamente statiche della teoria dell'equilibrio economico generale. È inoltre abbandonato quello studio dettagliato dei processi produttivi che era caratteristico non solo degli economisti classici, ma anche della stessa analisi di Marshall. I fenomeni della efficienza crescente e decrescente non sono più considerati, come nei classici, il risultato di operazioni economiche "riguardanti il comportamento dei produttori, quali l'introduzione di un ordinamento di efficienza tra le tecniche e la scelta della migliore alternativa tecnica disponibile", ma finiscono con l'apparire il risultato di "leggi tecniche" dipendenti dalle variazioni delle proporzioni in cui sono combinati i mezzi di produzione (vedi Scazzieri (a), 6, 46, 50). Questa posizione è espressa con chiarezza in Viner, il quale rinuncia esplicitamente ad ogni pretesa di fornire una descrizione realistica dell'andamento dei costi al variare del volume della produzione; il suo obiettivo è invece di esporre quei "tipi formali di relazioni tra costo e quantità prodotta" che rendano compatibile la determinazione dell'equilibrio della sin-

gola impresa con le assunzioni di concorrenza atomistica e di comportamento razionale da parte degli agenti economici (Viner (a), 93-96). La riformulazione di Viner dell'analisi marshalliana dei costi di breve e di lungo periodo costituisce la base della esposizione dell'andamento dei costi in relazione alla quantità prodotta tuttora presente nella maggior parte dei libri di testo di microeconomica.

Il breve periodo Nel breve periodo si assumono fissi il lavoro specializzato e le dimensioni dell'impianto, mentre si considerano variabili in relazione alla quantità prodotta il lavoro non specializzato, le materie prime e l'energia. Corrispondentemente sono fissi i costi generali relativi agli ammortamenti degli impianti, agli interessi sui debiti contratti per l'acquisto degli impianti stessi e alle spese di amministrazione, compresi gli stipendi del personale direttivo e il profitto minimo necessario per far sì che l'imprenditore continui la produzione; mentre sono variabili in relazione alla quantità prodotta i costi diretti relativi ai salari degli operai, alle spese per le materie prime, all'energia.

Il costo totale (CT) è pari alla somma dei costi generali (CG) e dei costi diretti (CD):

$$CT = CG + CD. \quad [2.2]$$

La figura 2.2 rappresenta la curva statica dei costi totali, dei costi generali e dei costi diretti in rapporto a valori diversi della produzione. I costi diretti aumentano in relazione alla quantità prodotta, inizialmente in modo meno che proporzionale, successivamente in modo più che proporzionale. Tale andamento si spiega facilmente se si tiene conto della forma che assume la curva del prodotto totale di breve periodo, riportata nella figura 2.1. Se si suppone, secondo la "legge delle proporzioni variabili", che vi sia sostituibilità imperfetta tra i fattori e che i fattori fissi siano, oltre che non aumentabili, neppure divisibili (senza costo), l'incremento della quantità prodotta può essere ottenuto solo attraverso l'aumento delle dosi dei fattori variabili; di conseguenza il *prodotto totale*, in una prima fase, finché i fattori fissi rimangono in eccesso rispetto a quelli variabili, cresce più che proporzionalmente, quindi, superata la combinazione ottima dei fattori, registra incrementi via via minori a causa della combinazione sempre più inefficiente tra i fattori fissi e i

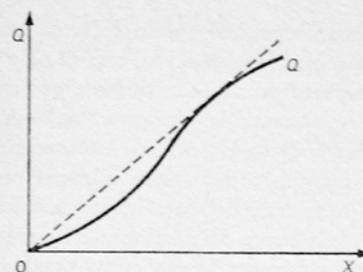


Figura 2.1

fattori variabili. Supponendo, per semplicità, l'esistenza di due soli fattori di produzione omogenei, possiamo indicare la funzione di produzione di breve periodo nel modo seguente:

$$Q = f(X, \bar{Y}) \quad [2.3]$$

dove X indica il fattore variabile, mentre \bar{Y} indica il fattore fisso. La presenza di un fattore fisso consente di riscrivere la [2.3] in forma più sintetica:

$$Q = f(X). \quad [2.4]$$

Poiché il costo diretto è dato dalla quantità del fattore variabile per il suo prezzo (P_x), che è costante, ossia $CD = X \cdot P_x$, la funzione del costo diretto, $CD = f(Q)$, è la funzione inversa della funzione di produzione, $Q = f(X)$. Dalla forma della curva del prodotto totale possiamo dunque ricavare l'andamento del costo diretto rovesciando gli assi cartesiani. Nel caso della curva del prodotto si ha in ascissa il fattore (X) e in ordinata la produzione (Q); invece nel caso della curva del costo diretto, si ha in ascissa la produzione (Q) e in ordinata il costo ($X \cdot P_x$), ovvero la quantità del fattore variabile per il suo prezzo, che è dato (vedi figg. 2.1 e 2.2).

Come si vede chiaramente nella figura 2.2, la curva dei costi totali ha un andamento analogo a quello della curva dei costi diretti, ma è spostata rispetto a questa verso l'alto di un tratto pari all'ammontare di costi generali. In termini analitici la funzione del costo totale di breve periodo ha queste proprietà:

$$f'(Q) > 0 \quad [2.5]$$

$$f''(Q) < 0 \quad \text{prima del punto di flesso} \quad [2.6]$$

$$f''(Q) > 0 \quad \text{dopo il punto di flesso} \quad [2.7]$$

dove la [2.5] indica che il costo marginale, dato dalla derivata prima della funzione del costo, è sempre positivo, e la [2.6] e la [2.7] indicano che esso si presenta prima decrescente e poi crescente. Il costo marginale rappresenta le variazioni del costo quando la quantità prodotta varia di un incremento infinitesimale $dCT/dQ = f'(Q)$. Il costo marginale può essere rilevato geometricamente misurando le variazioni del coefficiente angolare della tangente in ogni punto della curva dei costi totali. Le curve dei costi totali e dei costi diretti hanno la medesima pendenza in ciascun punto; di conseguenza il costo marginale, sia riferito ai costi totali che a quelli diretti, è identico. La figura 2.3 illustra l'andamento della curva del costo marginale e dei costi medi totali, generali e diretti. I costi medi generali sono dati dal rapporto CG/Q . La curva dei costi medi generali ($CMeG$) è un tronco di iperbole equilatera; essa decresce monotonicamente giacché i costi generali si ripartiscono, con l'espandersi della produzione, su una quantità sempre maggiore di unità prodotte. I costi medi diretti ($CMeD$), dati dal rapporto CD/Q , e i costi medi

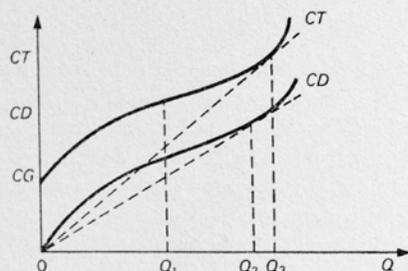


Figura 2.2

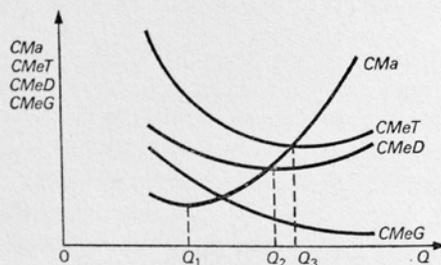


Figura 2.3

totali ($CMeT$), dati dal rapporto $(CG+CD)/Q$, hanno andamenti prima decrescenti e poi crescenti all'aumentare del volume della produzione. In termini geometrici l'andamento delle curve dei costi medi, totali, diretti e generali è rilevabile misurando le variazioni dei coefficienti angolari delle rette che uniscono l'origine O degli assi cartesiani ad ogni punto delle rispettive curve di costo medio. Nella figura 2.3 riscontriamo che la curva del costo marginale (CMe) interseca sia quella del costo medio diretto ($CMeD$) che quella del costo medio totale ($CMeT$) nei corrispondenti punti di minimo. Il punto di minimo dei costi totali corrisponde alla combinazione più efficiente tra i fattori variabili e quelli fissi, ed è per questo chiamato punto di ottimo tecnico.

Il fatto che nel breve periodo i costi marginali siano crescenti oltre un certo ammontare della produzione consente di determinare esattamente le dimensioni di equilibrio dell'impresa. In condizioni di concorrenza perfetta il prezzo è indipendente dalla produzione della singola azienda ed è quindi rappresentabile da una retta parallela all'asse delle ascisse. Ora, ciascun imprenditore massimizzerà il profitto espandendo la produzione fino al punto in cui l'ultima unità prodotta fornisce un ricavo pari al relativo costo aggiuntivo, cioè fino al punto in cui la retta del prezzo interseca la curva ascendente del costo marginale. Se il prezzo corrisponde al punto di minimo dei costi medi totali, esso ricopre tutti i costi compreso il profitto medio; se il prezzo è superiore al punto di minimo l'impresa realizza un sovraprofitto; se invece il prezzo è inferiore a tale punto l'impresa produce in perdita. Nel breve periodo, affinché l'impresa possa continuare a operare, il prezzo deve essere superiore almeno ai costi medi variabili. Le curve dei costi marginali delle singole imprese variano nella inclinazione e nella lunghezza a seconda delle specifiche caratteristiche tecniche di ogni impresa. La somma dei dati relativi alle curve marginali di ciascuna azienda ci dà la curva di offerta di breve periodo dell'industria nel suo complesso.

Il lungo periodo Nell'analisi statica di lungo periodo si assume che tutti i fattori produttivi possano aumentare al crescere della quantità prodotta, tranne la capacità organizzativa imprenditoriale, che è considerata data e non trasferibile [Hicks, 89 sg.; Koopmans, 89] (Arrow e Debreu, 267). La capacità imprendito-

riale è un fattore scarso nell'ammontare disponibile non per qualche limitazione naturale, come ad esempio la terra di un certo tipo, ma per una caratteristica speciale della produzione dell'azienda, perché soltanto una persona può svolgere quel determinato lavoro (Kaldor (a), 311). Tale capacità è fissa per la singola impresa, ma non per l'intera industria, altrimenti quest'ultima sarebbe composta da una singola azienda o da un numero fisso di aziende. L'ipotesi di offerta di capacità imprenditoriale data è essenziale per la determinazione della produzione d'equilibrio della singola impresa in regime di concorrenza. Affinché i costi marginali siano crescenti, e quindi si intersechino nel punto di equilibrio con la retta orizzontale del prezzo, è necessario infatti che anche nel lungo periodo si possa applicare la legge delle proporzioni variabili e che quindi almeno un fattore sia costante. In base a questa ipotesi, la curva dei costi medi di lungo periodo assume un andamento a U: nel primo tratto, grazie alle *economie di scala*, i costi medi sono decrescenti; nel secondo tratto, oltrepassata la "dimensione ottima di scala", i costi medi divengono crescenti per il *prevalere delle diseconomie organizzative*.

Numerosi economisti hanno collegato anche i rendimenti crescenti di lungo periodo alla presenza di qualche fenomeno di *indivisibilità* degli input (Kaldor (a), 310; Hahn (a), 131 sgg.) [Koopmans, 187-89]. Koopmans afferma in proposito di non aver "trovato un esempio di rendimenti di scala crescenti in cui non vi fosse un qualche bene indivisibile". Secondo Koopmans anche "l'aumento della specializzazione" (dei lavoratori e delle macchine) e il passaggio a "unità di fattori qualitativamente differenti e tecnologicamente più efficienti", che sono stati considerati da Chamberlin le due più importanti cause dei rendimenti crescenti indipendenti dal fenomeno dell'indivisibilità (vedi Chamberlin, 181 sgg.), denunciano la caratteristica che può essere adottata per definire un bene indivisibile:

In entrambi i casi vi è un bene, il lavoro e il capitale, al quale non si può (...) associare un parametro esprimente il suo ammontare, in maniera tale che possa essere ridotto al di sotto di una unità naturale senza cambiamento qualitativo, cioè senza variazioni nei rapporti fattori/prodotti in almeno uno dei processi produttivi nei quali esso è impiegato. [Koopmans, 187 sg.]

Tuttavia, come ha ammesso di recente Kaldor, l'indivisibilità non può essere riconosciuta come la *sola* causa dei rendimenti crescenti. Ad esempio il processo inarrestabile e graduale di acquisizione di nuove conoscenze attraverso l'esperienza, le cosiddette economie di scala dinamiche, non ha nulla a che vedere con il fenomeno dell'indivisibilità. Ma trascurando, per ora, le "economie dinamiche", che non possono essere considerate nell'ambito dell'analisi statica di lungo periodo, e rimanendo nel campo delle "economie statiche e reversibili", vi sono rendimenti crescenti dovuti alla tridimensionalità dello spazio, cioè al fatto che nel caso dei contenitori (cisterne, tubazioni, forni) il costo cresce approssimativamente in proporzione all'area della superficie delle pareti, mentre la capacità produttiva aumenta in proporzione alla capacità cubica (ad esempio una caldaia di diametro e superficie doppi ha una capacità produttiva quattro volte maggiore). "Tubazioni e oleodotti - osserva Kaldor - non sono caratterizzati da alcuna indivisibilità; da un punto di vista tecnico, la costruzione di tubi con diametro relativamente grande può essere non meno agevole di quella di tubi con diametro relativamente piccolo, e tra le due può esistere una gamma continua di dimensioni intermedie." L'esistenza di rendimenti crescenti dipende dalla natura dello spazio, e lo spazio in quanto tale non ha nulla di indivisibile (Kaldor (c), 93).

Passiamo ora alla costruzione della curva dei costi medi e marginali di lungo periodo. Non potendo considerare il progresso tecnico nell'ambito dell'analisi statica, supponiamo che le dimensioni degli impianti varino secondo un "libro delle possibilità tecnologiche", che registra, in un dato momento, "lo stato delle conoscenze tecniche". Se il numero degli impianti di diverse dimensioni, che è possibile progettare, è molto elevato, siamo in grado di tracciare una curva dei costi medi di lungo periodo formata dai tratti delle curve dei costi medi di breve periodo, in cui il costo medio per produrre una data quantità di merce è il minore dei costi medi che si devono sostenere in corrispondenza dei vari impianti (Chamberlin, 176-78). Nella figura 2.4 la curva dei costi medi di lungo periodo (*CMeL*) appare costruita da una serie di "lamelle" corrispondenti a piccoli tratti delle dodici curve dei costi medi di breve periodo (*CMeI*), relative a dodici singoli impianti. Nella stessa figura è disegnata anche la curva discontinua del costo marginale di lungo periodo (*CMaL*), risultante dai tratti

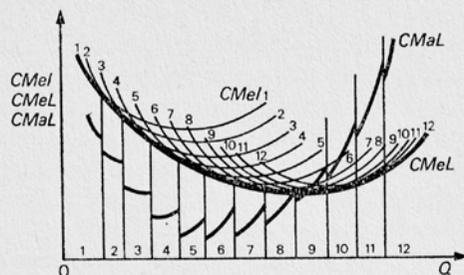


Figura 2.4

delle curve del costo marginale di breve periodo relative a quell'ambito di quantità prodotta per il quale la rispettiva curva di costo medio di breve periodo contribuisce a formare la curva di costo medio di lungo periodo. Se supponiamo che il numero delle dimensioni di impianto sia talmente elevato da rendere la curva dei costi medi di lungo periodo priva di punti angolosi, un aumento della quantità prodotta implica un cambiamento sia dell'impianto sia della quantità dei mezzi di produzione impiegati. In questo caso la curva dei costi medi di lungo periodo rappresenta l'involuppo delle curve dei costi medi di breve periodo e la curva dei costi marginali di lungo periodo diventa continua (vedi fig. 2.5). Ogni curva dei costi medi di breve periodo è tangente in un solo punto la curva dei costi medi di lungo periodo, e tale punto di tangenza coincide con il livello minimo dei costi medi di breve periodo solo nel caso della curva corrispondente all'impianto più efficiente (nella figura 2.5 la curva $CMeL_2$). La curva dei costi marginali di lungo periodo interseca la curva dei costi medi di lungo periodo nel punto di massima economia che, come si è appena visto, è anche il punto di minimo della curva del costo medio totale dell'impianto più efficiente. La curva dei costi medi totali individuali riveste particolare importanza perché, indicando il costo medio per unità di prodotto al crescere della produzione, è un indice di efficienza dell'azienda. Nel lungo periodo la retta del prezzo (P), che coincide con la curva di domanda per il singolo produttore, è sempre tangente alla curva dei costi medi totali nel punto di massima economia. In altre parole, nel lungo periodo ciascuna impresa produce e vende al minimo costo unitario di produzione (vedi fig. 2.5).

Si è visto in precedenza che nell'ambito dell'analisi di equilibrio parziale i costi medi di lungo periodo di una *industria nel*

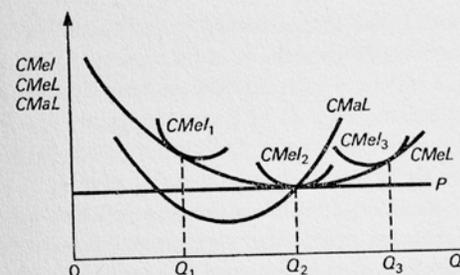


Figura 2.5

suo complesso non possono che essere costanti, fatta eccezione per due casi molto particolari, in cui è possibile considerare costi medi crescenti e decrescenti. I costi crescenti sono ammissibili nel caso assai raro che l'industria considerata utilizzi totalmente l'offerta di un fattore. Mentre i costi decrescenti sono ammissibili nel caso, altrettanto raro, che essi dipendano da economie esterne all'impresa e interne all'industria. Se, come avviene in generale, le varie industrie utilizzano solo parte dell'offerta dei loro mezzi di produzione e godono di economie esterne ad esse, l'unico caso logicamente valido nel lungo periodo, in regime di concorrenza perfetta, è quello dei costi costanti. La curva di offerta collettiva corrisponde pertanto a una retta orizzontale formata dalla somma delle quantità prodotte da un numero crescente di aziende operanti a costi medi di lungo periodo minimi. L'aumento della quantità offerta da parte dell'industria ha luogo mediante l'incremento del numero di imprese presenti sul mercato. Se la curva di offerta dell'industria è orizzontale, la domanda ha rilevanza nel determinare la quantità prodotta e venduta, ma non ha alcuna influenza sul livello del prezzo, che corrisponde ai costi medi minimi delle singole imprese operanti nell'industria.

Costo di produzione ed equilibrio economico generale

Il modello di equilibrio economico generale descrive l'interazione tra venditori e acquirenti in tutti i mercati attraverso un sistema di equazioni la cui soluzione determina simultaneamente i prezzi delle merci prodotte, i prezzi degli input (e quindi i costi monetari di produzione) e le quantità prodotte e scambiate (vedi EQUILIBRIO).

Affinché esista un sistema di prezzi di equilibrio che sia unico,

stabile e capace di soddisfare le condizioni paretiane di ottima allocazione delle risorse tra le imprese e, a un tempo, di ottima distribuzione dei prodotti tra i consumatori in base alle preferenze individuali, ipotizziamo che: 1) vi siano condizioni di concorrenza perfetta nei mercati dei beni e degli input, cioè che nessun agente economico sia in grado di influenzare i prezzi; 2) gli input e le merci prodotte siano divisibili e omogenei; 3) gli input siano sostituibili tra essi con rendimenti decrescenti (convessità degli isoquanti di produzione); 4) le funzioni di produzione siano continue, derivabili, non esibiscano rendimenti di scala crescenti e siano indipendenti (assenza di economie e diseconomie esterne); 5) l'obiettivo di ciascuna impresa sia la massimizzazione del profitto e l'obiettivo di ciascun consumatore sia la massimizzazione dell'utilità. Assumiamo inoltre dati il numero degli agenti economici (imprese e consumatori), la disponibilità di risorse e la loro distribuzione tra gli agenti economici, la tecnologia e i gusti dei consumatori.

In un modello di equilibrio generale il sistema dei prezzi dipende dall'azione congiunta di numerosi elementi: dalla scarsità delle risorse, dalle tecniche di produzione, dalla distribuzione degli obiettivi tra gli individui e infine dalle preferenze individuali (Malinvaud, 122). Come ha osservato Pareto, da questo punto di vista risulta erronea, o quanto meno parziale, l'idea sostenuta da molti economisti neoclassici che il prezzo degli input, cioè il costo monetario di produzione, derivi in ultima istanza dal grado finale di utilità, e quindi dalla struttura e dal livello della domanda [Pareto, 58].

L'ipotesi di massimizzazione del profitto implica, come abbiamo già avuto modo di considerare, che ogni impresa produca fino al punto in cui il costo aggiuntivo di una unità di prodotto non superi il ricavo marginale che se ne può trarre. In altre parole, l'impresa massimizza il profitto se produce una quantità tale da uguagliare il costo marginale della merce prodotta al suo prezzo. Considerando due imprese, la prima che produce il bene A e la seconda che produce il bene B , avremo quindi:

$$CMa_A = P_A \quad [2.8a]$$

$$CMa_B = P_B \quad [2.8b]$$

dove CMa_A , CMa_B , P_A , P_B rappresentano rispettivamente il costo marginale e il prezzo dei beni A e B .

Dividendo la [2.8a] con la [2.8b] otteniamo che, in equilibrio, il rapporto tra i prezzi di due beni qualsiasi è pari al rapporto tra i loro costi marginali:

$$\frac{P_A}{P_B} = \frac{CMa_A}{CMa_B} \quad [2.9]$$

Si consideri ora la possibilità di variare l'impiego di un solo input, mantenendo fisso quello di tutti gli altri. In questo caso è possibile definire il concetto di costo marginale di un input, ovvero di costo marginale del prodotto rispetto alle variazioni dell'utilizzo di un solo input. Si noti che il costo marginale di un input è un concetto completamente diverso da quello di costo marginale del prodotto, dove tutti gli input variano. L'ipotesi di massimizzazione del profitto fa sì che le imprese aumentino l'impiego di ogni mezzo di produzione finché l'incremento dei ricavi, ottenuto utilizzando un'unità aggiuntiva dell'input considerato, è superiore al relativo incremento di costo. Avremo quindi che, in equilibrio, il costo marginale dei diversi input impiegati nella medesima produzione è uguale al prezzo del prodotto. Se per semplicità si suppone che esistano solo due input X e Y , utilizzati da entrambi i produttori, avremo

$$CMa_A^X = CMa_A^Y = P_A \quad [2.10a]$$

$$CMa_B^X = CMa_B^Y = P_B \quad [2.10b]$$

dove CMa_A^X , CMa_A^Y , CMa_B^X , CMa_B^Y rappresentano i costi marginali dei due input X e Y rispettivamente nella produzione di A e B .

Il costo marginale di un input è dato dal rapporto tra il suo prezzo e la sua produttività marginale fisica. Si può quindi riscrivere la [2.10a] e la [2.10b] nel modo seguente:

$$\frac{W}{\partial A / \partial X_A} = \frac{R}{\partial A / \partial Y_A} = P_A \quad [2.11a]$$

$$\frac{W}{\partial B / \partial X_B} = \frac{R}{\partial B / \partial Y_B} = P_B \quad [2.11b]$$

dove W e R indicano i prezzi dei due input X , Y e $\partial A / \partial X_A$, $\partial A / \partial Y_A$, $\partial B / \partial X_B$ e $\partial B / \partial Y_B$ sono le produttività marginali fisiche dei due input X , Y rispettivamente nelle produzioni A e B .

Ora, poiché in un mercato perfettamente concorrenziale il

prezzo di un dato input è il medesimo nelle diverse produzioni, uguagliando per W e R la [2.11a] e la [2.11b] avremo

$$W = \frac{\partial A}{\partial X_A} \cdot P_A = \frac{\partial B}{\partial X_B} \cdot P_B \quad [2.12a]$$

$$R = \frac{\partial A}{\partial Y_A} \cdot P_A = \frac{\partial B}{\partial Y_B} \cdot P_B. \quad [2.12b]$$

Se si indica con $PMaF$ la produttività marginale fisica degli input possiamo riscrivere la [2.12a] e la [2.12b] come segue:

$$W = PMaF_A^X \cdot P_A = PMaF_B^X \cdot P_B \quad [2.13a]$$

$$R = PMaF_A^Y \cdot P_A = PMaF_B^Y \cdot P_B. \quad [2.13b]$$

La [2.13a] e la [2.13b] ci dicono che, in equilibrio, la produttività marginale in valore di un input (ottenuta moltiplicando la produttività marginale fisica per il prezzo del prodotto) è pari al prezzo dell'input stesso in tutte le diverse produzioni. Le imprese dunque, utilizzano gli input fino al punto in cui la loro produttività in valore uguaglia la loro remunerazione.

Riassumendo i risultati fin qui raggiunti possiamo affermare che in una situazione di equilibrio economico generale il prezzo di ogni prodotto è uguale al suo costo marginale [2.8a, 2.8b], il costo marginale dei diversi input nella medesima produzione è il medesimo e corrisponde al prezzo del bene prodotto [2.10a, 2.10b], la produttività marginale in valore di un input in produzioni diverse è la medesima e corrisponde al prezzo dell'input stesso [2.13a, 2.13b].

Costo privato marginale e costo sociale marginale

Quando vi è uguaglianza tra il prezzo e il costo marginale di ogni bene prodotto il sistema economico raggiunge la condizione di ottimo paretiano nell'allocazione delle risorse tra le imprese e nella combinazione del prodotto tra gli individui. All'inizio si è assunto che le funzioni di produzione siano indipendenti le une dalle altre; si è ipotizzata, cioè, l'assenza di effetti esterni sia positivi che negativi. Nel caso contrario, in cui si abbiano economie e diseconomie esterne, non è possibile raggiungere nella situazione

di equilibrio il punto di ottimo paretiano perché viene meno l'uguaglianza tra prezzi e costi marginali. Le economie o diseconomie esterne, o più in generale gli effetti esterni, sono interdipendenze tra due o più agenti economici non rispecchiate dai prezzi e quindi fuori dal mercato. Si parla in questi casi di incapacità del mercato di tener conto degli effettivi costi o benefici che alcune azioni economiche comportano. Si hanno diseconomie esterne di produzione quando un'impresa paga solo parte dei costi che il processo produttivo implica. Se ciò avviene, il *costo privato marginale*, a cui corrisponde il prezzo del bene prodotto, è inferiore al *costo sociale marginale*, che può essere definito, secondo il "principio di compensazione" introdotto da Hicks e da Kaldor, come la somma necessaria a ristabilire il precedente livello di utilità che si è perduto a causa della produzione del bene in questione (Hicks, 696; Kaldor (b), 549). L'esempio forse più frequente di diseconomie esterne è l'inquinamento. Si verificano, invece, economie esterne di produzione quando un'impresa attraverso la sua azione arreca ad altri, imprese o consumatori, un beneficio non ricompensato. Tra i molti casi di economie esterne di produzione è spesso citato quello dell'addestramento da parte di un'impresa di lavoratori che poi passano ad altre imprese, permettendo a queste ultime di risparmiare sui costi di addestramento.

La disuguaglianza tra costo marginale sociale di un dato bene e il suo prezzo, dovuta all'incapacità del mercato di tener conto degli effetti esterni, fa sì che l'aggiustamento reciproco dei prezzi e delle quantità non permetta di raggiungere nella posizione di equilibrio l'ottimo paretiano. Ad esempio, nel caso di una produzione inquinante la quantità prodotta sarà tale da uguagliare, secondo l'ipotesi di massimizzazione del profitto, il prezzo al costo marginale privato, anziché al costo marginale sociale; di conseguenza la quantità prodotta sarà maggiore rispetto a quella richiesta per raggiungere l'ottimo paretiano. Viceversa, nel caso di una produzione che implica economie esterne la quantità prodotta sarà minore della quantità ottima.

Il costo opportunità

La riesposizione moderna del principio del costo opportunità è dovuta a Gottfried Haberler (Haberler (a), 175-82). Egli fu il

primo ad applicare tale principio alla teoria del commercio internazionale, esprimendo la legge dei costi comparati, anziché in termini di quantità di lavoro incorporate nei beni, in termini di costi alternativi o costi opportunità (vedi COMMERCIO INTERNAZIONALE).

La formulazione di Haberler del concetto di costo opportunità è, come egli stesso ha sottolineato, una versione semplificata di un modello di equilibrio economico generale, di cui conserva le ipotesi fondamentali (concorrenza perfetta, offerta data dei fattori produttivi, assenza di effetti esterni e di rendimenti crescenti) (Haberler (*b*), 224 nota). Egli illustra il concetto di costo opportunità con l'aiuto di una curva di trasformazione che indica le diverse combinazioni di due beni che è possibile produrre utilizzando appieno ed efficientemente le risorse disponibili. Nella figura 2.6a è tracciata una linea di trasformazione *ba* tra due beni qualsiasi *A* e *B*. Il punto *a* rappresenta la quantità ottenibile dal bene *A* se si dedicasse ad esso l'intera dotazione di fattori produttivi; il punto *b* indica la quantità prodotta di *B* nel caso opposto in cui tutti gli input disponibili siano utilizzati per la produzione di *B*. I punti intermedi rappresentano combinazioni diverse di quantità dei due beni prodotti. L'inclinazione della linea di trasformazione *ba*, in ogni suo punto, cioè la derivata di *B* rispetto ad *A*, misura il saggio marginale di trasformazione di un bene nell'altro:

$$SMT_{A,B} = -\frac{dB}{dA} \quad [2.14]$$

Il saggio marginale di trasformazione indica quale ammontare di uno dei due beni è necessario sacrificare per ottenere un'unità aggiuntiva dell'altro. Esso rappresenta il *costo opportunità* che si deve affrontare per ottenere un'unità aggiuntiva di un bene, sotto forma di rinuncia a una data quantità dell'altro bene (Haberler (*a*), 177). Il saggio marginale di trasformazione è sempre negativo perché, essendo le risorse date, la quantità prodotta dei due beni considerati deve variare in maniera inversa. La linea di trasformazione è retta, come nella figura 2.6a, e il saggio marginale di trasformazione costante, quando si verificano rendimenti costanti, cioè quando l'ammontare di *B* a cui si deve rinunciare per produrre un'unità aggiuntiva di *A* rimane costante al crescere della quantità di *A*. La linea di trasformazione è concava verso

l'origine, come nella figura 2.6b, e il saggio marginale di trasformazione crescente (partendo da sinistra verso destra) in presenza di rendimenti decrescenti. I rendimenti decrescenti si esprimono con il fatto che l'ammontare di *B*, a cui si deve rinunciare per produrre un'unità aggiuntiva di *A*, cresce man mano che la quantità di *A* aumenta.

Il saggio marginale di trasformazione è costante e la linea di trasformazione retta, se si ipotizza l'esistenza di un unico input omogeneo non riproducibile, e quindi scarso (ad esempio il lavoro), il quale entra nella produzione di tutti i beni (Malinvaud, 118). Nel caso invece che gli input scarsi siano più d'uno, l'interpretazione della forma della curva di trasformazione, anche in presenza di rendimenti di scala costanti, non è univoca.

Vi sono almeno due diverse spiegazioni della conformazione della curva di trasformazione basate su differenti ipotesi circa le caratteristiche degli input. Haberler suppone che i vari input possano essere distinti in non specifici e specifici. Gli input non specifici, come ad esempio il lavoro non qualificato, possono spostarsi liberamente da un'attività all'altra, mentre gli input specifici sono destinati a determinate produzioni e non sono quindi trasferibili (Haberler (*a*), 177 sg.). Se, come afferma Haberler, per industria si intende il settore che utilizza totalmente l'offerta di un certo input (specifico), si avranno rendimenti decrescenti e curva di trasformazione concava verso l'origine: considerando due beni qualsiasi *A* e *B*, appartenenti a due industrie caratterizzate dall'uso di input specifici differenti, la crescita della produzione di uno di essi, poniamo *A*, a scapito della produzione di *B*, comporta incrementi decrescenti di prodotto a causa dell'aumento degli input non specifici applicati agli input specifici nella produzione di *A*. Se invece per industria si intende l'insieme di

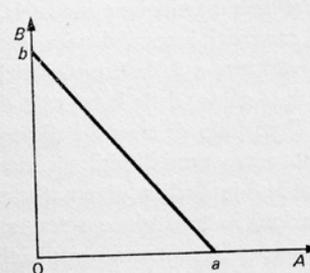


Figura 2.6a

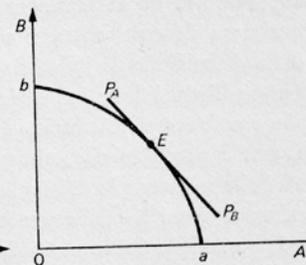


Figura 2.6b

imprese che producono un certo bene, può accadere che un'industria utilizzi un input adottato non da tutte le industrie, ma da alcune soltanto. Sicché un dato fattore può essere trasferibile (e quindi non specifico) all'interno di un certo numero di industrie ma non trasferibile (e quindi specifico) all'esterno di questo gruppo. Allora, per conoscere l'andamento dei costi al crescere della produzione di una certa industria è necessario sapere quali sono gli input utilizzati dall'industria (o dalle industrie) che contemporaneamente riducono la produzione. In questo caso la forma della linea di trasformazione non può essere determinata a priori.

Diversa è la spiegazione dell'andamento della curva di trasformazione fornita da Samuelson. Egli suppone che gli input siano omogenei, sostituibili tra essi con rendimenti decrescenti, perfettamente trasferibili da una produzione all'altra. L'andamento rettilineo o concavo della linea di trasformazione è legato alle proporzioni in cui gli input sono combinati nelle produzioni dei due beni considerati. Se gli input sono utilizzati in proporzioni diverse nelle due industrie, il passaggio da una produzione all'altra, lungo la curva di trasformazione, comporta inevitabilmente combinazioni degli input sempre meno favorevoli; di conseguenza si avranno rendimenti decrescenti e la linea di trasformazione sarà concava verso l'origine. Viceversa, se i vari input sono combinati nelle medesime proporzioni nelle due produzioni, si avranno rendimenti costanti e la linea di trasformazione sarà retta (Samuelson, 181-86). Anche in questo caso non è possibile stabilire se un'industria è caratterizzata da costi costanti o crescenti, al crescere della quantità prodotta, finché non si conosce quali fattori sono stati liberati dalla diminuzione della produzione dell'altra industria (o delle altre industrie) (J. Robinson (b), 65).

La possibilità di determinare se una particolare industria è caratterizzata da costi costanti o crescenti è condizionata, oltre che dalla definizione di industria adottata e dalle caratteristiche degli input liberati dalle industrie in declino, dalla lunghezza del periodo considerato. Più lungo è il periodo di tempo a disposizione, più è possibile che, almeno entro certi limiti, gli input liberati dalle industrie la cui domanda è in declino, siano adattati alle esigenze tecniche delle produzioni in espansione attraverso corsi di riqualificazione professionale e il reinvestimento delle quote di ammortamento in nuove macchine. Pertanto nel lungo

periodo la tendenza verso i rendimenti decrescenti, presente in alcuni casi, dovrebbe risultare sensibilmente attenuata.

Il costo marginale di una data quantità di merce *A*, in termini di costo opportunità, è dato dalla quantità di merce *B* alla quale si deve rinunciare per ottenere un'unità addizionale della merce *A* (Haberler (a), 177). È infatti facile dimostrare, utilizzando la curva di trasformazione, l'uguaglianza nel punto di equilibrio tra il saggio marginale di trasformazione e il rapporto inverso tra i costi marginali dei due beni considerati:

$$SMT_{A,B} = -\frac{dB}{dA} = \frac{CM_{aA}}{CM_{aB}}. \quad [2.15]$$

Tutte le combinazioni produttive di una curva di trasformazione sono ottenute attraverso l'impiego della stessa quantità di input il cui prezzo è il medesimo nelle due produzioni alternative, di conseguenza la variazione del costo totale provocata dalla variazione della quantità prodotta del bene *A* (dC_A) deve essere pari e di segno opposto alla variazione del costo totale dovuta alla variazione corrispondente della quantità prodotta del bene *B* (dC_B):

$$dC_A = dC_B. \quad [2.16]$$

Ora, il costo marginale dei due beni è dato dal rapporto tra l'incremento del costo e l'incremento corrispondente della quantità prodotta:

$$CM_{aA} = \frac{dC_A}{dA} \quad [2.17a]$$

$$CM_{aB} = \frac{dC_B}{dB}. \quad [2.17b]$$

Sostituendo le relazioni [2.17] nella [2.16] otteniamo l'uguaglianza [2.15] in cui il saggio marginale di trasformazione è pari al rapporto inverso dei costi marginali.

Il rapporto tra i prezzi di mercato dei beni considerati è rappresentato graficamente da una retta. L'equilibrio della produzione è espresso dalla condizione di tangenza tra la curva di trasformazione e la retta dei prezzi, cioè dall'uguaglianza del saggio marginale di trasformazione con il rapporto tra i costi marginali

(fissati in termini di costo opportunità) e con il rapporto tra i prezzi. Avremo quindi

$$SMT_{A,B} = -\frac{dB}{dA} = \frac{CMa_A}{CMa_B} = \frac{P_A}{P_B}. \quad [2.18]$$

Se il saggio marginale di trasformazione è costante, la linea di trasformazione assume andamento rettilineo. Di conseguenza la retta dei prezzi non può che coincidere con la linea di trasformazione. I prezzi relativi sono così determinati esclusivamente dalle condizioni tecniche della produzione, mentre la domanda determina solo le quantità prodotte. Invece nel caso in cui il saggio marginale di trasformazione sia crescente e la curva di trasformazione sia concava verso l'origine, la retta dei prezzi è tangente alla curva di trasformazione in un solo punto. Ne segue che i prezzi e le quantità sono determinati congiuntamente dalle condizioni di produzione e dai dati relativi alla domanda (vedi figg. 2.6a e 2.6b).

La [2.18] è identica alla condizione di uguaglianza [2.9], riportata in precedenza nell'esposizione delle relazioni tra costi e prezzi nello schema di equilibrio economico generale. Anche se il principio del costo opportunità deriva i propri presupposti teorici dal modello di equilibrio economico generale, non deve sfuggire la diversa prospettiva in cui esso si pone. Il principio del costo opportunità sottolinea l'uguaglianza tra i costi marginali dei beni considerati e le alternative produttive a cui si rinuncia, mentre lo schema di equilibrio economico generale pone l'accento più semplicemente sull'uguaglianza tra costi marginali e prezzi. Per quanto riguarda i mezzi di produzione, il loro costo, in termini di costo opportunità, può essere considerato come il pagamento necessario per sottrarli agli impieghi alternativi che seguono immediatamente in ordine di redditività. In questo caso, il costo opportunità di un qualsiasi input nella produzione di una data merce è il massimo ammontare che esso produrrebbe nel suo impiego alternativo migliore (Stigler, 104 sg.). Va ricordato che in una situazione di equilibrio concorrenziale la produttività marginale e il prezzo di ogni servizio produttivo sono uguali in tutti gli utilizzi alternativi. Di conseguenza il costo opportunità di un dato bene è sempre pari al suo costo storico, cioè alle spese monetarie effettivamente sostenute per la sua produzione.

Viner, commentando criticamente la definizione di Haberler, osserva che il principio del costo opportunità, sebbene utile nel calcolare il costo di quei fattori il cui utilizzo non comporta disutilità, "sfugge alle difficoltà connesse con le preferenze tra il lavoro e il tempo libero e le diverse occupazioni" (Viner (b), 520-24). In altre parole, la sostanza della critica di Viner è che l'approccio di Haberler, presupponendo date le risorse produttive, non tiene conto della variazione nell'offerta dei servizi produttivi, come invece fa la teoria del costo reale.

Il problema è stato definitivamente risolto da Walsh alla metà degli anni cinquanta, considerando il tempo libero come una merce qualsiasi. Walsh passa dalla consueta curva di trasformazione tra due beni a un grafico tridimensionale in cui il tempo libero compare come il terzo prodotto (Walsh, 253). Sul piano AB della figura 2.7a, è riportata la curva di trasformazione ba tra due beni, mentre sull'asse L è misurato il tempo libero. La superficie lba , concava verso l'origine degli assi, rappresenta le possibili combinazioni tra i due prodotti fisici A e B e il tempo libero L . Ad esempio, per un dato ammontare di tempo libero pari a Ol_2 la curva di trasformazione tra A e B sarà a_2b_2 . Se proiettiamo sul piano OAB le diverse sezioni orizzontali corrispondenti ai vari livelli di tempo libero, otteniamo una mappa di curve di trasformazione tra i beni A e B (vedi fig. 2.7b). Ogni curva di trasformazione indica le possibili combinazioni tra i due beni fisici per ogni dato livello di tempo libero.

Jaroslav Vanek, sviluppando argomentazioni molto simili a

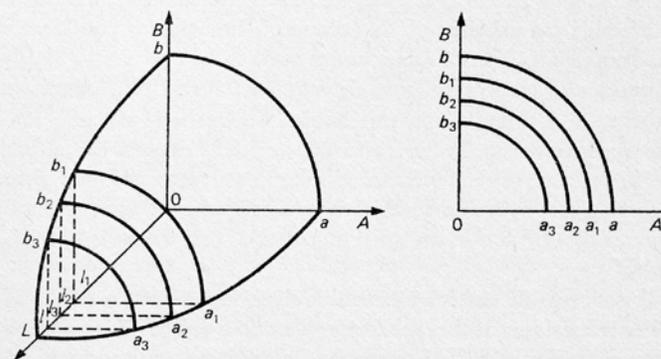


Figura 2.7a

Figura 2.7b

quelle di Walsh, ha posto fine alla disputa sul costo reale soggettivo / costo opportunità dimostrando la compatibilità dei due diversi approcci in determinate condizioni. Infatti, se si considera il tempo libero come una componente del reddito, il costo relativo all'offerta di lavoro può essere espresso indifferentemente sia in termini di costo opportunità che in termini di costo reale soggettivo. Il costo opportunità pone l'accento su ciò a cui si rinuncia (il tempo libero), mentre il costo reale soggettivo sottolinea il sacrificio o la disutilità, di ciò che si sceglie (il lavoro) (Vaneč, 198).

Va notata, tuttavia, una differenza fondamentale tra il costo reale e il costo opportunità. Il costo reale, sia esso espresso in termini oggettivi o soggettivi, influenza la determinazione dei prezzi relativi, mentre nel caso del costo opportunità non vi è un rapporto di causa ed effetto ma di semplice identità tra i costi (alternativi) e i prezzi. In uno schema di equilibrio economico generale i costi marginali delle merci (fissati in termini di opportunità) sono pari ai prezzi dei beni considerati; in tal senso, i costi opportunità sono il riflesso, e non la "spiegazione", dei prezzi. Per questa ragione il principio del costo opportunità è stato definito da Dobb, riprendendo un'espressione di Durbin, un "truismo aritmetico" [Dobb (*a*), 148]. Altri, come Samuelson, ritengono che il principio del costo opportunità sia "un'inutile complicazione che non presenta le condizioni di equilibrio in modo molto diretto" [Samuelson, 219 sg., 435 nota 32]. In effetti, tale principio richiede, rispetto a uno schema di equilibrio economico generale, la condizione che i servizi produttivi siano trasferibili e indifferenti tra usi diversi. Se invece gli input sono considerati non trasferibili, il costo opportunità varia in base al grado di aggregazione al quale scegliamo di valutare i mezzi di produzione considerati. Un input non trasferibile, utilizzato in un solo settore concorrenziale, ha un costo opportunità, rispetto al settore che lo utilizza, pari al prezzo necessario per sottrarlo alle altre imprese dello stesso settore; viceversa, rispetto all'industria nel suo complesso, non avendo usi alternativi, non ha costo opportunità.

Il concetto di costo opportunità rivela un notevole grado di indeterminazione anche se si considerano periodi di tempo di diversa lunghezza. Ciò è comprensibile perché, col passare del tempo, varia il grado di trasferibilità degli input. Gli usi alterna-

tivi di una risorsa sono generalmente maggiori nel lungo periodo che nel breve periodo. Ad esempio, un lavoratore può acquisire, col passare del tempo, una diversa qualifica che gli permette di accedere a nuovi lavori. Analogamente, nel caso di un particolare macchinario, col passare degli anni, le quote di ammortamento possono essere reinvestite in altre forme di capitale (Haberler (*a*), 181; Stigler, 107). Queste difficoltà analitiche sembrano difficilmente superabili. Comunque il concetto di costo opportunità, così come quello di costo reale, non sono affatto necessari alla teoria dei prezzi basata sullo schema di equilibrio economico generale; in un tale schema i costi sono definiti semplicemente come i prezzi degli input utilizzati nel processo produttivo.

2.2 Il costo di produzione nella teoria dei prezzi di Sraffa

La critica alle "leggi dei costi crescenti e decrescenti" e alla "simmetria fondamentale" della teoria marshalliana dei valori di scambio, svolta da Sraffa negli scritti del 1925 e del 1926, porta il suo autore alla elaborazione di una teoria dei prezzi relativi che segna l'abbandono dei concetti di "curve di domanda e di offerta" e, al tempo stesso, una ripresa dei concetti, propri dell'economia classica, di riproducibilità, sovrappiù e circolarità del processo produttivo. Sraffa elabora uno schema della produzione in cui le merci sono prodotte mediante merci e lavoro. "L'indagine riguarda esclusivamente – egli avverte – quelle proprietà di un sistema economico che sono indipendenti da variazioni nel volume della produzione e nelle proporzioni fra i 'fattori' impiegati" [Sraffa, v]. Questa autolimitazione consente a Sraffa di esaminare un sistema economico in cui nulla muta tranne i prezzi relativi e la distribuzione del reddito. Se il sistema economico è in grado di produrre sovrappiù e se il salario supera il livello di sussistenza, i prezzi relativi possono essere determinati simultaneamente a una delle variabili distributive (i salari o il saggio di profitto), qualora siano note la tecnologia e l'altra variabile distributiva (Roncaglia, 6). Poiché Sraffa non considera alcun cambiamento nel volume della produzione, il problema della relazione tra costi e quantità prodotta non sussiste e "la questione se i

rendimenti siano costanti o variabili non sorge nemmeno" [Sraffa, v].

La separazione della determinazione dei prezzi da quella delle quantità prodotte esclude qualsiasi ruolo della composizione della domanda nella formazione delle ragioni di scambio. Mentre nella teoria neoclassica i prezzi dei mezzi di produzione dipendono dall'interazione delle forze di domanda e offerta, in Sraffa, al pari dei classici, riflettono, come tutti i prezzi, le condizioni della produzione. Nello schema di Sraffa non vi è però una relazione causale che va dai costi ai prezzi, relazione che è invece presente sia nella dottrina del valore lavoro di Ricardo, sia nel principio del costo reale soggettivo di Marshall. Sraffa preferisce evitare l'espressione "costo di produzione" proprio perché essa è legata all'idea che i prezzi dei prodotti possano essere determinati dai costi, cioè dalla quantità e dai prezzi degli input direttamente richiesti per produrli; il che sarebbe ammissibile solo nel caso in cui i costi siano "grandezze suscettibili di essere misurate indipendentemente dalla determinazione dei prezzi dei prodotti, e anteriormente ad essa", circostanza questa che si verifica esclusivamente nel caso delle merci non necessarie alla produzione (merci non base). Viceversa per tutte le merci che direttamente o indirettamente entrano nella produzione (merci base) vi è interdipendenza reciproca tra costo e prezzo, poiché "i prezzi dei mezzi di produzione" di una merce base "dipendono dal suo stesso prezzo non meno di quanto quest'ultimo dipenda da quelli" [*ibid.*, 11].

L'autolimitazione di non tener conto dei mutamenti del volume della produzione, se da una parte consente l'elaborazione di una teoria rigorosa dei prezzi relativi, dall'altra esclude la possibilità di studiare, nell'ambito della teoria stessa, le variazioni dei costi in rapporto alla quantità prodotta e, più in generale, il comportamento dell'impresa.

Negli ultimi decenni numerosi studiosi hanno sviluppato l'analisi dei costi di impresa in un contesto non neoclassico. Come vedremo nel prossimo paragrafo, l'elemento caratterizzante la loro analisi è l'abbandono del modello statico di concorrenza perfetta a causa della sua incapacità sia di prevedere i reali comportamenti dell'impresa, sia di tener conto dei numerosi elementi che influenzano l'andamento dei costi e quindi le scelte aziendali.

2.3 Costo di produzione e struttura del mercato

Nell'analisi neoclassica l'impresa è caratterizzata da costi medi e marginali, sia di breve che di lungo periodo, con andamento prima decrescente e poi crescente. Tale andamento non è giustificato empiricamente, ma risponde a esigenze teoriche dettate dall'ipotesi concorrenziale. Come sappiamo, in regime di concorrenza il costo marginale deve crescere oltre un certo ammontare del prodotto perché le dimensioni della singola azienda siano determinate dall'uguaglianza tra il costo marginale e il prezzo. Inoltre, l'ammontare del prodotto in corrispondenza del quale il costo marginale inizia a crescere deve essere piccolo abbastanza da lasciare sul mercato un numero sufficiente di imprese.

Il principio del costo marginale crescente non è essenziale solo alla teoria dell'impresa in regime di concorrenza, ma su esso poggia gran parte dello schema teorico neoclassico. Ad esempio la relazione inversa tra salari e occupazione e, più in generale, la teoria neoclassica della distribuzione del reddito sono conseguenza dell'ipotesi di costi marginali crescenti (vedi *DISTRIBUZIONE DEL REDDITO*). Nel *breve periodo* l'andamento a U dei costi medi e marginali è spiegato dalla "legge delle proporzioni variabili": supponendo dato l'impianto, l'aumento delle materie prime e dei lavoratori applicati al macchinario fisso comporta dapprima rendimenti crescenti e, oltre un certo punto, rendimenti decrescenti. Se tale legge, per quanto riguarda i rendimenti decrescenti, può avere qualche fondamento in agricoltura, perché è possibile, entro certi limiti, aggiungere e sottrarre lavoratori e fertilizzanti a un terreno, essa non può essere estesa anche al caso dell'industria, dove il fattore costante è il macchinario anziché la terra. Infatti, come era noto ai classici e come fu sottolineato da Babbage, in un'impresa industriale ad ogni macchina deve essere assegnato, in generale, un numero fisso di lavoratori (Babbage, 212). Si può dunque affermare che l'aumento o la diminuzione della produzione, entro i limiti della massima capacità produttiva dell'impianto, comporta un utilizzo maggiore o minore, ma non migliore o peggiore, del macchinario (Sylos Labini, 56). L'esperienza insegna che il costo totale di breve periodo della singola

azienda industriale cresce in proporzione alla quantità prodotta fino al punto di massima capacità produttiva dell'impianto; se si aumenta ulteriormente la quantità di lavoro e di materie prime impiegate è ancora possibile espandere la produzione ma a costi rapidamente crescenti.

Il costo totale di produzione (CT) è rappresentato graficamente da una retta ascendente fino al punto Q_3 , che indica la quantità prodotta utilizzando appieno la capacità produttiva dell'impianto esistente. Da Q_3 in poi, il costo totale assume inclinazione crescente (vedi fig. 2.8). Nel primo tratto, finché la curva del costo totale (CT) ha inclinazione costante, il costo medio diretto ($CMeD$) e il costo marginale (CMa) coincidono e sono rappresentati da una retta parallela all'asse delle ascisse, mentre il costo medio totale ($CMeT$) ha andamento decrescente; nel secondo tratto, in cui l'inclinazione del costo totale è crescente, il costo medio diretto e il costo marginale assumono andamento divergente e crescente (vedi fig. 2.9).

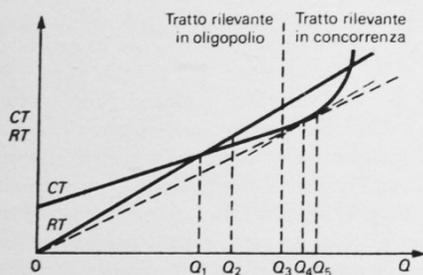


Figura 2.8

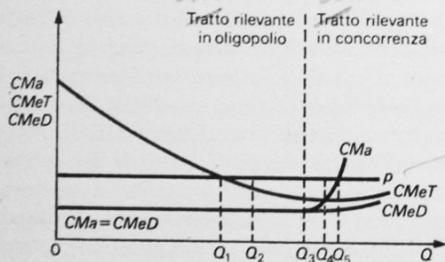


Figura 2.9

Gli studi empirici che sono stati effettuati in numero crescente, soprattutto a partire dagli anni quaranta, in Gran Bretagna e negli Stati Uniti (Staehle; Andrews; vedi Dean) hanno confermato l'idea, suggerita da Sraffa nel 1926 e ripresa da Kalecki in un saggio del 1939, che le aziende industriali operano normalmente in condizioni di costi marginali costanti e di costi medi totali decrescenti (Sraffa (b); Kalecki). Il fatto che nel tratto rilevante per l'imprenditore il costo marginale sia costante indica che quello della concorrenza è un caso limite raramente realizzato, mentre l'oligopolio è la forma di mercato più frequente nella realtà industriale (Sylos Labini, 57). In concorrenza la quantità prodotta è sempre limitata dai costi marginali crescenti: l'imprenditore, trovandosi di fronte a una domanda che al prezzo corrente di mercato è perfettamente elastica, cioè che assorbe qualsiasi quantità di merce, ha convenienza a utilizzare l'impianto oltre la sua "massima capacità produttiva", fino al punto in cui i costi marginali crescenti uguagliano il prezzo. Per tutti i livelli inferiori di produzione l'ultima unità prodotta fornirebbe un ricavo maggiore del relativo costo aggiuntivo e vi sarebbe quindi convenienza ad accrescere la produzione e le vendite. Pertanto, in concorrenza, nel tratto rilevante per l'imprenditore i costi marginali sono sempre crescenti.

Viceversa in oligopolio il limite all'espansione della produzione è dato non dai costi, ma dalla domanda, le cui dimensioni sono limitate dall'ampiezza del "mercato particolare" di ciascuna impresa (vedi MERCATO). In oligopolio le imprese progettano impianti di dimensioni tali che la loro "massima capacità produttiva" ecceda il volume medio previsto dalle vendite, in modo da garantirsi una riserva di capacità. Nelle figure 2.8 e 2.9, Q_2 corrisponde alla capacità normalmente utilizzata che, in generale, è il 70-80 per cento della "massima capacità produttiva". Una riserva di capacità produttiva è essenziale per assicurare flessibilità al processo produttivo, e in tal modo poter fronteggiare temporanei incrementi della domanda o riparare alcuni macchinari senza interrompere completamente la produzione. Dunque in un mercato oligopolistico l'imprenditore non aumenta la produzione, sebbene produca con costi marginali costanti, perché non riesce a vendere di più a quello stesso prezzo. Per aumentare la domanda ed espandere il proprio "mercato particolare" sarebbe necessaria

una riduzione del prezzo tale da non poter essere compensata dall'incremento delle vendite.

Se si abbandona il modello di concorrenza perfetta, che implica piena utilizzazione delle risorse, e se si suppone che, salvo situazioni congiunturali eccezionali, le imprese producano a un livello inferiore alla loro "massima capacità produttiva", e che nel sistema economico vi siano normalmente risorse non utilizzate, il concetto di costo opportunità di un bene non ha alcun significato. L'incremento della produzione di una data merce può non voler dire "rinunciare alla produzione di un altro bene", ma soltanto impiegare le risorse non utilizzate.

Per quanto riguarda il *lungo periodo*, come abbiamo visto, i costi medi e marginali crescenti, postulati dall'economia neoclassica, dipendono dall'ipotesi che la capacità organizzativa imprenditoriale sia data per la singola azienda. Tuttavia questa ipotesi, oltre a essere logicamente debole, non ha alcun riscontro con la realtà aziendale. Sul piano logico va notato che, in condizioni di equilibrio perfettamente statico, le funzioni imprenditoriali sono limitate all'ambito della pratica corrente, giacché si devono escludere dalla considerazione tutte quelle funzioni connesse con l'esigenza di rispondere in modo coordinato al cambiamento, nelle quali prevale la componente dinamica. È difficile dunque concepire le funzioni di mera routine come non divisibili (Marris, 363). Inoltre, dal punto di vista della realtà aziendale, va osservato che la capacità di coordinamento del singolo imprenditore può ampliarsi con la crescita dell'azienda grazie all'adozione di diversi sistemi di controllo e all'assegnazione di molti compiti amministrativi e organizzativi a personale specializzato (Penrose, 74-78). Pertanto, se non si può dubitare che, almeno fino a un certo punto, operino le economie di scala tecniche derivanti dal passaggio a impianti di dimensioni ed efficienza maggiori, viceversa, in concorrenza non trova alcun fondamento il motivo addotto per giustificare il prevalere di diseconomie di scala organizzative da un certo punto in poi della produzione (Sylos Labini, 57). È vero, però, che in un dato momento esiste un limite massimo alle dimensioni dell'impianto dipendente dallo stato della tecnologia. Se l'impresa vuole produrre più di ciò che le consente l'impianto di dimensioni massime progettabili, deve moltiplicare il numero degli impianti. Ciò significa che oltre questo limite il costo medio e quello marginale di lungo periodo divengono costanti e che ci

troviamo di fronte non a una curva, ma a una serie di punti discreti (Sylos Labini, 57 sg.).

Il termine "curva dei costi di lungo periodo" è ingannevole giacché tale curva rappresenta, nell'ambito dell'analisi statica, non un adattamento nel tempo (reale o cronologico) delle dimensioni della singola azienda, ma varie alternative (teoriche) nella dimensione degli impianti realizzabili in un dato momento, secondo le conoscenze tecniche disponibili. Perciò è preferibile sostituire, quando è possibile, all'espressione tradizionale "curva dei costi medi di lungo periodo" l'espressione più precisa "curva delle dimensioni di scala". Si può dunque concludere che in concorrenza (nel lungo periodo) i costi medi e marginali relativi alle *diverse dimensioni di scala* non hanno un andamento a U, ma a L, sono cioè prima discendenti e poi costanti. Come appare nella figura 2.10, nel primo tratto discendente il costo medio ($CMeDS$) è superiore al costo marginale ($CMaDS$), mentre nel secondo tratto il costo medio e quello marginale coincidono e sono rappresentati geometricamente da una serie di punti discreti giacenti su una retta parallela all'asse delle ascisse. Il punto in cui i costi medi cessano di diminuire (Q_{sme}) è chiamato "punto di scala minima efficiente" o "dimensione minima ottima di scala".

In concorrenza il fenomeno delle economie di scala deve essere limitato, nel senso che la quantità prodotta corrispondente alla scala minima efficiente deve essere per ogni impresa una quota molto piccola della capacità produttiva dell'intera industria. Se la curva delle dimensioni di scala ha andamento a L, anziché a U, le dimensioni di equilibrio dell'impresa concorrenziale rimangono indeterminate nell'ambito dell'analisi di lungo periodo. In regime di concorrenza perfetta è possibile, dunque, stabilire la quantità prodotta d'equilibrio della singola impresa solo nel caso dell'analisi di breve periodo, in cui sono ammissibili costi marginali crescenti secondo la "legge delle proporzioni variabili".

Oltre il punto corrispondente all'impianto di dimensioni massime è possibile moltiplicare il numero degli impianti. Quindi, al di là di tale punto, non ci troviamo di fronte a una curva continua, ma a una serie di punti discreti. Nella realtà industriale è molto frequente che la serie di impianti minori, corrispondenti a diverse tecnologie non sia continua. In questo caso non si può propriamente parlare di una curva continua neppure nel tratto che precede il limite segnato dall'impianto massimo, giacché non

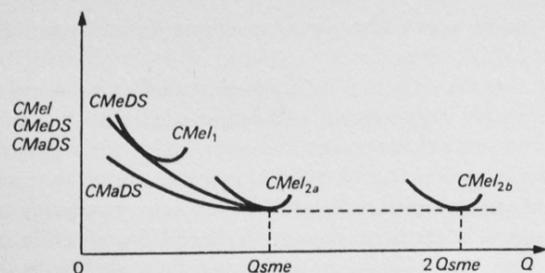


Figura 2.10

è possibile passare da un impianto all'altro per piccoli incrementi. Paolo Sylos Labini ha evidenziato che la presenza di un numero finito di tecnologie adottate dalle diverse imprese è una delle caratteristiche fondamentali del mercato oligopolistico. In oligopolio pertanto non esiste una vera e propria curva delle dimensioni di scala. È invece possibile, rimanendo nell'ambito dell'analisi statica, tracciare una serie di curve dei costi corrispondenti a diversi impianti e a diverse tecnologie, disponibili in un dato istante (Sylos Labini, 58, 64 sg.). Come vedremo nel prossimo paragrafo, l'analisi dell'oligopolio permette il superamento dell'ipotesi statica necessaria alla teoria neoclassica dei costi di breve e lungo periodo, e quindi consente di considerare l'influenza del progresso tecnico sui costi.

Rimanendo nell'ambito dell'analisi statica, supponiamo, per semplicità, che le aziende operanti in un dato ramo di attività possano adottare tre diverse tecniche. Indichiamo la quantità massima producibile con \bar{Q} , e la quantità media normalmente prodotta da ciascuna tecnica con Q^* . Naturalmente sia

$$0 \leq Q^* \leq \bar{Q}. \quad [2.19]$$

Seguendo la formalizzazione fornita al riguardo da Renzo Bianchi (Bianchi, 71-76), ordiniamo le tecniche produttive disponibili, in ordine crescente, secondo il valore della quantità massima producibile:

$$\bar{Q}_1 < \bar{Q}_2 < \bar{Q}_3. \quad [2.20]$$

Supponiamo che i costi medi diretti ($CMeD$) siano costanti entro i limiti della capacità produttiva massima e che le tecniche

con capacità produttiva maggiore siano caratterizzate: a) da costi generali (fissi) più elevati; b) da costi medi diretti inferiori; c) da costi medi totali inferiori sia al livello della quantità normalmente prodotta ($CMeT^*$) che a quello della quantità massima producibile ($CMe\bar{T}$). Si ha quindi

$$CG_1 < CG_2 < CG_3 \quad [2.21]$$

$$CMeD_1 > CMeD_2 > CMeD_3 \quad [2.22]$$

$$CMeT_1^* > CMeT_2^* > CMeT_3^* \quad [2.23]$$

$$CMe\bar{T}_1 > CMe\bar{T}_2 > CMe\bar{T}_3. \quad [2.24]$$

Alle figure 2.11, 2.12 e 2.13 sono riportati i diversi andamenti dei costi totali, medi diretti, medi totali e marginali, corrispondenti alle tre diverse tecniche disponibili. Poiché abbiamo supposto che entro il limite della massima capacità produttiva il costo medio diretto sia costante, esso è rappresentato da una retta parallela all'asse delle ascisse e coincide con il costo marginale. La riduzione del costo medio totale che si ottiene, in corrispondenza della quantità normalmente prodotta, passando a impianti di maggiori dimensioni, è dovuta a quelle che sono definite le economie di scala di impianto o economie di scala tecniche. Una volta raggiunta la dimensione massima progettabile dell'impianto, l'impresa ha esaurito la possibilità di realizzare delle economie di scala tecniche. Tuttavia, un'ulteriore crescita della capacità produttiva, attraverso l'aumento del numero degli impianti, può consentire all'impresa il raggiungimento di economie di scala finanziarie, organizzative, commerciali e di economie nel prezzo di acquisto degli input. L'arresto del processo di espansione delle

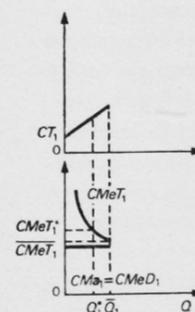


Figura 2.11

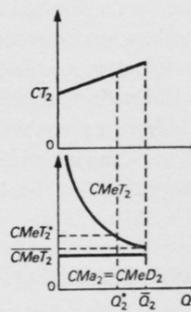


Figura 2.12

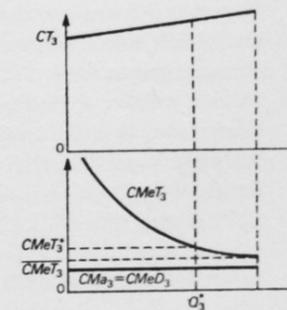


Figura 2.13

dimensioni aziendali e quindi della realizzazione delle economie di scala può essere dovuto alla ristrettezza del mercato o al sorgere di diseconomie di scala. Le diseconomie di scala consistono in un innalzamento dei costi medi totali al crescere della capacità produttiva. Esse dipendono da limitazioni nell'offerta degli input (quantità disponibile data o prezzo crescente all'aumentare della domanda) e quindi dal mutamento delle proporzioni in cui essi sono impiegati nella produzione. Le diseconomie di scala hanno dunque un'origine diversa rispetto alle economie di scala che sono determinate dall'aumento contemporaneo di tutti gli input nelle proporzioni necessarie alla produzione.

In conclusione, se non si verificano limitazioni nell'offerta dei mezzi di produzione, le imprese con capacità produttiva più elevata producono a costi medi totali inferiori, sebbene abbiano costi generali superiori a quelli delle aziende più piccole; questo perché esse possono dividere i costi generali tra un numero maggiore di unità prodotte, e anche perché godono di costi medi diretti minori grazie all'applicazione di metodi tecnici e organizzativi più efficienti. Generalmente il maggior costo medio generale è più che compensato dal minor costo medio diretto. Costi medi totali più bassi consentono alle imprese di dimensioni più grandi una maggiore influenza nel processo di fissazione del prezzo rispetto alle imprese più piccole.

Numerosi studi empirici sulle economie di dimensione (Florence; Bain; Johnston; Pratten; Dean), malgrado le differenze relative alla metodologia adottata e alle fonti statistiche utilizzate (vedi Walters; Gold) sembrano concordare sulla presenza di economie di scala nella maggior parte delle industrie esaminate. Non solo il fenomeno delle economie di scala interessa numerosi settori dell'industria manifatturiera, ma in molti casi esso assume notevoli proporzioni. Pratten, ad esempio, ha rilevato che su 44 attività, relative a 25 diverse industrie britanniche prese in considerazione, la scala minima efficiente per un singolo impianto corrisponde per 7 casi al 100 per cento o più dell'intera produzione del Regno Unito, ed è superiore al 25 per cento in altri 10 casi (Pratten, 269-77).

2.4 Economie di scala e analisi dinamica

Come si è avuto modo di considerare nelle pagine precedenti, il fenomeno delle economie di scala è difficilmente conciliabile

con una teoria dei prezzi relativi basata sull'equilibrio concorrenziale tra domanda e offerta (Hahn e Matthews, 91 sg.). In concorrenza perfetta la quantità prodotta dalla singola impresa deve essere abbastanza piccola da non influire sul prezzo di mercato; quindi le economie di scala associate a grandi volumi della produzione non sono ammissibili (Arrow e Debreu, 91 sg.; vedi Hahn (b), 12 sg.). Inoltre il progresso tecnico e l'apprendimento mediante l'esperienza, che sono strettamente connessi alla crescita delle dimensioni aziendali e al fenomeno delle economie di scala, non possono essere inclusi nell'analisi statica di breve e di lungo periodo. Non stupisce pertanto che il tema delle economie di scala, sebbene fosse stato sviluppato da Smith sin dalla fine del diciottesimo secolo, abbia trovato, a parte alcune eccezioni (Marx, Marshall, Young), scarsissimo spazio tra gli economisti fino a tempi relativamente recenti, quando l'analisi dell'oligopolio ha fornito un contesto analitico in cui inserire coerentemente il fenomeno dei rendimenti crescenti. Come era stato intuito da Marshall, questi ultimi possono trovare posto solo in riferimento all'analisi dinamica e a un mercato non perfettamente concorrenziale.

Nell'analisi dinamica dell'oligopolio è possibile tenere conto di tutti gli elementi, compresi il progresso tecnico e l'apprendimento mediante l'esperienza, che sono legati allo sviluppo delle dimensioni aziendali e che stanno alla base delle economie di scala. Queste ultime consistono, come si è visto, in una riduzione dei costi medi totali di produzione all'aumentare della capacità produttiva. L'incremento della capacità produttiva può essere ottenuto accrescendo le dimensioni dei singoli impianti (economie di scala di impianto o tecniche), moltiplicando il loro numero (economie di scala d'impresa o *multi-plant economies*) e, infine, attraverso la formazione di un gruppo formato da diverse imprese collegate (economie di scala di gruppo). Nel caso delle economie di scala d'impianto possiamo distinguere tra le economie derivanti dal passaggio a impianti di dimensioni maggiori che producono *un solo bene* (*product-specific economies*) ed economie derivanti dal passaggio a impianti di dimensioni maggiori che producono *più beni* (*multi-product economies*). Le economie di scala d'impresa e di gruppo si basano principalmente sulla riduzione dei costi di trasporto, sulle economie di scala manageriali e finanziarie, sulla riduzione dei costi transazionali

ottenibile internalizzando alcune operazioni svolte dal mercato, ossia sostituendo i meccanismi di mercato con il "controllo gestionale" della produzione e della distribuzione (vedi IMPRESA). Le economie di scala traggono origine da diversi fattori, tra i quali i più importanti sono:

a) La possibilità di ottenere gli input a prezzi più bassi man mano che cresce la scala produttiva e, conseguentemente, il potere dell'impresa sul mercato degli input (economie pecuniarie).

b) Le grandi dimensioni aziendali consentono spesso una migliore amministrazione e organizzazione (economie amministrative e organizzative).

c) L'ampliamento del volume della produzione permette, oltre che la specializzazione del personale e delle macchine (economie di specializzazione), anche la riduzione dei tempi morti di utilizzo di "fattori fondo", ossia degli input che non operano di continuo nel processo produttivo in cui sono impiegati (Georgescu-Roegen, 5 sg.; Scazzieri (b), 26).

d) Le aziende di grandi dimensioni possono reperire finanziamenti con maggior facilità. Inoltre esse sono spesso in grado di mantenere una propria rete di vendita risparmiando sui costi di commercializzazione e assistenza (economie finanziarie e commerciali).

e) È possibile ripartire i costi indipendenti dalla scala di produzione (ad esempio il costo di ricerca, di progettazione e di disegno di un nuovo prodotto, il costo della prima copia di un libro o il costo di una campagna pubblicitaria) tra un numero maggiore di unità prodotte al crescere della capacità produttiva.

f) L'aumento delle dimensioni delle macchine consente spesso di ridurre i costi di costruzione e gestione degli impianti. Le economie di dimensione si verificano quando, passando a impianti di dimensioni maggiori, i costi di costruzione e installazione dell'impianto aumentano in maniera meno che proporzionale rispetto alla capacità produttiva. È questo il caso, già considerato, dei contenitori come le cisterne, le turbine, i forni. Nel caso, ad esempio, di una cisterna, se lo spessore delle pareti non è influenzato dalle dimensioni, allora il costo cresce approssimativamente in proporzione all'area della superficie delle pareti, mentre la portata, e quindi la capacità produttiva, aumenta in proporzione alla sua capacità cubica (E. A. G. Robinson; Pratten). In alcune produzioni l'aumento delle dimensioni degli impianti comporta, ol-

tre che un minor costo di investimento per unità di prodotto, anche minori costi unitari di esercizio grazie alla minore dispersione di calore che si ottiene riducendo il rapporto superficie/volume degli impianti.

g) L'utilizzo di grandi quantità di risorse (materie prime, macchine, capitale circolante) dà origine a notevoli economie nella gestione delle scorte (economie di *massed reserves*). Come è noto, secondo il teorema del limite centrale (o legge dei grandi numeri) lo scarto fra la media del campione e quella dell'universo si riduce all'ingrandirsi del campione stesso. Ora, più elevato è il numero delle risorse trattate da un'impresa e minore è, in proporzione, la quantità di scorte necessarie per far fronte a eventuali imprevisti. Ad esempio, con l'aumento del numero delle macchine di uno stesso tipo utilizzate dalla medesima impresa, si ottiene un risparmio nel numero dei pezzi di ricambio che è necessario immagazzinare per riparare possibili guasti (E. A. G. Robinson, 26 sg.).

h) Innovazioni tecniche. È molto difficile isolare, nella realtà dei fatti, le economie di scala dalle economie derivanti dal progresso tecnico, perché spesso un aumento delle dimensioni favorisce l'adozione di nuove tecniche, così come l'introduzione di nuove tecniche consente spesso un aumento della scala di produzione. Il progresso tecnico conduce a un miglioramento degli impianti esistenti e/o all'ampliamento delle dimensioni massime progettabili. Il miglioramento degli impianti esistenti ha per conseguenza l'abbassamento delle curve dei costi medi totali relative ai vari impianti, mentre l'installazione di un nuovo impianto di dimensioni ed efficienza maggiori comporta la costruzione di una nuova curva dei costi medi totali. L'introduzione di nuove tecnologie nei processi produttivi non ha un effetto univoco sulla scala di produzione. In alcuni casi i progressi tecnologici favoriscono un ampliamento della scala di produzione, rendendo possibile la costruzione di impianti di dimensioni sempre più grandi; in altri casi le innovazioni tecnologiche rendono possibili produzioni su piccola scala caratterizzate da un elevato grado di specializzazione e di flessibilità. Le tecnologie che favoriscono la grande scala produttiva sono state definite "tecnologie centripete", mentre quelle che favoriscono le piccole dimensioni "tecnologie centrifughe" (Blair, cap. 6).

i) L'apprendimento mediante l'esperienza. Una volta intro-

dotto un nuovo impianto si verifica un progressivo miglioramento nell'utilizzazione delle nuove apparecchiature; il che comporta un incremento nel tempo della produttività e un costante, anche se lieve, abbassamento dei costi unitari. Da questo punto di vista, come è stato posto in luce da Alchian e da Arrow, il costo è funzione non solo del volume della produzione per unità di tempo, ma anche del volume della produzione complessiva progettata ed effettivamente realizzata (volume cumulativo) (Alchian, 47; Arrow (a), 32).

La connessione tra economie di scala e sviluppo economico, presente nelle analisi di Smith, Marx e Young, trova un posto rilevante nei lavori di Kaldor arricchita, oltre che dagli studi sulle forme di mercato non concorrenziali, anche dalle acquisizioni dell'analisi keynesiana sul rapporto tra crescita degli investimenti e crescita del reddito (Kaldor (c)). Kaldor considera lo sviluppo economico come il risultato di un processo di interazione tra aumenti della domanda indotti da incrementi dell'offerta e incrementi dell'offerta generati in risposta a incrementi di domanda. Le economie di scala, che egli considera acquisibili prevalentemente nel settore manifatturiero, si inseriscono in questo processo cumulativo ampliandolo. Il processo di sviluppo sarà tanto più rapido quanto più è alta la propensione al consumo del sistema e maggiore la tendenza a utilizzare gli incrementi di reddito per l'acquisto di prodotti industriali nella cui produzione è appunto possibile godere di economie di scala.

Riferimenti bibliografici

- ALCHIAN A. *Costi e prodotti*, in W. Breit e H. M. Hochman (a cura di), "Problemi di microeconomia" (Etas Libri, Milano 1971) vol. 2. Trad. P. V. Bondonio e G. Brosio di: *Costs and Outputs*, in M. Abramovitz e altri, "The Allocation of Economic Resources" (Stanford University Press, Stanford 1959).
- ANDREWS P. W. S. *Manufacturing Business* (Macmillan, Londra 1949).
- ARROW K. J. (a) *Le implicazioni dell'apprendere per esperienza*, in B. Jossa (a cura di), "Progresso tecnico e sviluppo economico" (Angeli, Milano 1974). Trad. di: *The Economic Implications of Learning by Doing*, Rev. econ. Stud. (1962).
- ARROW K. J. (b) *The Firm in General Equilibrium Theory*, in R. Marris e A. Wood (cura di), "The Corporate Economy" (Macmillan, Londra 1971).
- ARROW K. J. e DEBREU G. *Existence of an Equilibrium for a Competitive Economy*, *Econometrica* (1954).
- BABBAGE C. *On the Economy of Machinery and Manufactures* (Knight, Londra 1832). Trad. G. Piatti e L. Casini: *Sulla economia delle macchine e delle manifatture* (Gabinetto scientifico-letterario Vieusseux, Firenze 1834).
- BAIN J. *Barriers to New Competition* (Harvard University Press, Cambridge, Mass. 1956).
- BECATTINI G. *Invito a una rilettura di Marshall*, Introduzione ad A. Marshall e M. Paley Marshall, *Economia della produzione*, trad. it. (Isedi, Milano 1975).
- BIANCHI R. *Economie di scala e barriere all'entrata: un modello di oligopolio*, *Note econ.* (1974).
- BLAIR J. M. *Economic Concentration* (Harcourt Brace Jovanovich, New York 1972).
- CANNAN E. *Rassegna della teoria economica*, in C. Arena (a cura di), "Storia delle teorie" (Utet, Torino 1937). Trad. R. Fubini di: *A Review of Economic Theory* (King, Londra 1929).

- CASSELS J. M. *On the Law of Variable Proportions* (1936), in W. Fellner e B. F. Haley (a cura di), "Readings in the Theory of Income Distribution" (Allen & Unwin, Londra 1961).
- CHAMBERLIN H. E. *Proporzionalità, divisibilità ed economie di scala* (1948), in *Verso una più generale teoria del valore* (Utet, Torino 1960). Trad. M. Giorda di: *Proportionality, Divisibility and Economies of Scale*, in *Towards a More General Theory of Value* (Oxford University Press, New York 1957).
- CLAPHAM J. H. *Le scatole vuote dell'economia*, in G. Zanetti (a cura di), "Contributi per un'analisi economica dell'impresa" (Liguori, Napoli 1980). Trad. di: *On Empty Economic Boxes*, Econ. J. (1922).
- DEAN J. *Statistical Cost Estimation* (Indiana University Press, Bloomington 1977).
- DEL VECCHIO G. (a cura di), *Economia pura*, "Nuova collana di economisti", vol. 4 (Utet, Torino 1937).
- DIDEROT D. *Art*, in D. Diderot e J. D'Alembert (a cura di), "Encyclopédie" (Parigi 1751) vol. 1.
- EDGEWORTH F. Y. *Contributions to the Theory of Railway Rates*, sez. 4, Econ. J. (1913).
- FLORENCE P. S. *The Logic of British and American Industry* (Routledge & Kegan, Londra 1953).
- GEORGESCU-ROEGEN N. *The Economics of Production*, Pap. Proc. Am. econ. Rev. (1970).
- GOLD B. *Changing Perspectives on Size, Scale and Returns: an Interpretative Survey*, J. econ. Lit. (1981).
- GREEN D. I. *Pain Cost and Opportunity Cost*, Q. J. Economics (1894).
- HABERLER G. (a) *The Theory of International Trade* (Hodge, Londra 1936).
- HABERLER G. (b) *Some Problems in the Pure Theory of International Trade*, Econ. J. (1950).
- HAHN F. H. (a) *Proportionality, Divisibility and Economies of Scale*, Q. J. Economics (1949).
- HAHN F. H. (b) *On the Notion of Equilibrium in Economics: an Inaugural Lecture* (Cambridge University Press, Cambridge 1973).
- HAHN F. H. e MATTHEWS R. C. O. di), *La teoria dello sviluppo economico*, in F. Caffè (a cura di), "Il pensiero economico contemporaneo" (Angeli, Milano 1974) vol. 2. Trad. G. Pala di: *The Theory of Economic Growth: a Survey*, Econ. J. (1964).
- HARROD R. F. *The Law of Decreasing Cost*, ibid. (1931).
- HICKS J. R. *The Foundations of Welfare Economics*, ibid. (1939).
- JANNACCONE P. *Il costo di produzione* (Utet, Torino 1956).
- JOHNSTON J. *Statistical Cost Analysis* (McGraw-Hill, New York 1960).
- KALDOR N. (a) *L'equilibrio dell'impresa*, in G. Zanetti (a cura di), "Contributi per un'analisi economica dell'impresa" (Liguori,

- Napoli 1980). Trad. di: *The Equilibrium of the Firm*, Econ. J. (1934).
- KALDOR N. (b) *Welfare Comparisons of Economics and Interpersonal Comparisons of Utility*, ibid. (1939).
- KALDOR N. (c) *La irrilevanza delle teorie dell'equilibrio economico*, in M. D'Antonio (a cura di), "La crisi postkeynesiana" (Boringhieri, Torino 1975). Trad. Studio editoriale Esse di: *The Irrelevance of Equilibrium Economics*, Econ. J. (1972).
- KALECKI M. *Salari monetari e salari reali: parte prima* (1939), in *Studi sulla teoria dei cicli economici: 1933-1939* (Il Saggiatore, Milano 1972). Trad. C. Boffito di: *Money and Real Wages, Part I*, in *Studies on the Theory of Business Cycles* (Blackwell, Oxford 1966).
- LUNGHINI G. *La crisi dell'economia politica e la teoria del valore* (Feltrinelli, Milano 1977).
- MACVANE S. M. *Analysis of Cost of Production*, Q. J. Economics (1887).
- MALINVAUD E. *Lectures on Microeconomic Theory* (North-Holland, Londra 1972).
- MARRIS R. K. *Perché l'economia ha bisogno di una teoria dell'impresa*, in G. Zanetti (a cura di), "Contributi per un'analisi economica dell'impresa" (Liguori, Napoli 1980). Trad. di: *Why Economics Needs a Theory of the Firm*, Econ. J. (1972).
- MARSHALL A. (a) *Mill's Theory of Value* (1876), in "Memorials of Alfred Marshall", a cura di A. C. Pigou (Kelley & Millman, New York 1956) vol. 2.
- MARSHALL A. (b) *Principles of Economics* (1890), nona edizione (Variorum) a cura di C. W. Guillebaud, 2 voll. (Macmillan, Londra 1961).
- MILL J. S. *Notes on N. W. Senior's Political Economy*, *Economica* (1945).
- PENROSE E. *La teoria dell'espansione dell'impresa* (Angeli, Milano 1973). Trad. M. T. Bruno di: *The Theory of the Growth of the Firm* (Blackwell, Oxford 1959).
- PIGOU A. C. *Empty Economic Boxes: a Reply*, Econ. J. (1922).
- PRATTEN C. F. *Economies of Scale in Manufacturing Industry* (Cambridge University Press, Cambridge 1971).
- RICARDO D. (a) *Notes on Malthus* (1820), in "The Works and Correspondance of David Ricardo", a cura di P. Sraffa con la collaborazione di M. Dobb (Cambridge University Press, Cambridge 1951-73) vol. 2.
- RICARDO D. (b) *Pamphlets and Papers, 1815-1823*, ibid., vol. 4.
- RICARDO D. (c) *Letters, 1816-1821*, ibid., voll. 7 e 8.
- ROBBINS L. *Remarks upon Certain Aspects of the Theory of Costs*, Econ. J. (1934).
- ROBERTSON D. H. *Gli alberi della foresta*, in "La produttività crescente e l'impresa rappresentativa: simposio", in G. Del Vecchio

- (a cura di), "Economia pura" (Utet, Torino 1937). Trad. P. Baffi di: *The Trees of the Forest*, in "Increasing Returns and the Representative Firm", Econ. J. (1930).
- ROBINSON E. A. G. *The Structure of Competitive Industry* (1931) (Cambridge University Press, Cambridge 1964).
- ROBINSON J. (a) *Rising Supply Price*, *Economica* (1941).
- ROBINSON J. (b) *Eresie dell'economia*, cap. 4 "Rendimenti crescenti e decrescenti" (Etas Kompass, Milano 1972). Trad. G. Ricoveri di: *Economic Heresies* (Basic Books, Londra 1971).
- RONCAGLIA A. *Sraffa e la teoria dei prezzi* (Laterza, Roma e Bari 1975).
- SAMUELSON P. A. *International Factor Price Equalisation once again*, Econ. J. (1949).
- SCAZZIERI R. (a) *Efficienza produttiva e livelli di attività* (Il Mulino, Bologna 1981).
- SCAZZIERI R. (b) *Scale and Efficiency in Production Models*, in M. Baranzini (a cura di), "Advances in Economic Theory" (Blackwell, Oxford 1982).
- SENIOR N. *An Outline of the Science of Political Economy* (1836) (Allen & Unwin, Londra 1951).
- SHOVE G. F. *L'impresa rappresentativa e la produttività crescente*, in "La produttività crescente e l'impresa rappresentativa: simposio", in G. Del Vecchio (a cura di), "Economia pura" (Utet, Torino 1937). Trad. P. Baffi di: *The Representative Firm and Increasing Returns*, in "Increasing Returns and the Representative Firm", Econ. J. (1930).
- SRAFFA P. (a) *Sulle relazioni fra costo e quantità prodotta* (1925), in G. Zanetti (a cura di), "Contributi per un'analisi economica dell'impresa" (Liguori, Napoli 1980).
- SRAFFA P. (b) *Lettera a J. M. Keynes* (Milano, 6 giugno 1926), pubbl. parz. in A. Roncaglia, *Sraffa e la teoria dei prezzi* (Laterza, Roma e Bari 1975).
- SRAFFA P. (c) *Le leggi della produttività in regime di concorrenza*, in G. Lunghini (a cura di), "Valore, prezzi ed equilibrio economico generale" (Il Mulino, Bologna 1971). Trad. P. Baffi di: *The Laws of Returns under Competitive Conditions*, Econ. J. (1926).
- SRAFFA P. (d) *Una critica*, in "La produttività crescente e l'impresa rappresentativa: simposio", in G. Del Vecchio (a cura di), "Economia pura" (Utet, Torino 1937). Trad. P. Baffi di: *A Criticism*, in "Increasing Returns and the Representative Firm", Econ. J. (1930).
- SRAFFA P. (e) Introduzione a *The Works and Correspondence of David Ricardo*, a cura di P. Sraffa con la collaborazione di M. Dobb (Cambridge University Press, Cambridge 1951-73) vol. 1.
- STAEHLE H. *The Measurement of Statistical Cost Functions: an Appraisal of Some Recent Contributions*, Am. econ. Rev. (1942).

- STIGLER G. J. *The Theory of Price* (Macmillan, New York 1966).
- SYLOS LABINI P. *Oligopolio e progresso tecnico* (1956) (Einaudi, Torino 1972).
- TURGOT A. R. J. *Observations sur les Mémoires de M. De Saint-Péray en faveur de l'impôt indirect*, in "Oeuvres de Turgot", a cura di M. E. Daire (Guillaumin, Parigi 1844) vol. 1.
- TURVEY R. *Marginal Cost*, Econ. J. (1969).
- VANEK J. *An Afterthought on the 'Real Cost/Opportunity Cost' Dispute and Some Aspects of General Equilibrium under Conditions of Variable Factor Supplies*, Rev. econ. Stud. (1959).
- VINER J. (a) *Curve di costo e curve di offerta*, in G. Lunghini (a cura di), "Valore, prezzi ed equilibrio economico generale" (Il Mulino, Bologna 1971). Trad. A. Stagni di: *Cost Curves and Supply Curves*, Z. NatÖkon. (1931).
- VINER J. (b) *Studies in the Theory of International Trade* (1937) (Allen & Unwin, Londra 1955).
- WALSH V. C. *Leisure and International Trade*, *Economica* (1956).
- WALTERS A. A. *Production and Cost Functions: an Econometric Survey*, *Econometrica* (1963).
- WIESER F. *The Austrian School and the Theory of Value*, Econ. J. (1891).
- YOUNG A. *Rendimenti crescenti e progresso economico*, in S. Lombardini (a cura di), "Teoria dell'impresa e struttura economica" (Il Mulino, Bologna 1973). Trad. di: *Increasing Returns and Economic Progress*, Econ. J. (1928).