



Università di Roma "La Sapienza"

Economie di scala e di scopo nel settore bancario: una rassegna teorica

Simona Costagli



Copyright © MMIII ARACNE EDITRICE S.R.L.
00173 Roma, via Raffaele Garofalo, 133 A/B
tel. (06) 72672222 telefax (06) 72672233

www.aracne-editrice.it
info@aracne-editrice.it

ISBN 88-7999-559-7

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

I edizione: settembre 2003

Simona Costagli

Dottore di ricerca presso il dipartimento di Scienze economiche
Università "La Sapienza", Roma.
e-mail: simcostagli@hotmail.com:

Il lavoro presenta una rassegna dei principali contributi all'analisi delle economie di scala e scopo nelle banche e mostra come i risultati ottenuti siano fortemente condizionati dalle definizioni di input e output adottate, dalla considerazione o meno dei canali distributivi e del rischio. Negli anni più recenti, l'attenzione è stata posta sul modo in cui la forma della funzione di costo influisce sul risultato finale. Infine, seppure mai esplicitamente, la normativa sulla regolamentazione sembra aver costituito uno sfondo importante agli studi condotti, quanto meno nell'offrire spunti di riflessione.

A parte gli Stati Uniti, l'Italia è l'unico paese che abbia prodotto un flusso regolare di studi sul tema. Per tale motivo, la rassegna analizza prima la letteratura statunitense, illustrandone la ricchezza dei contributi tecnici e teorici, poi quella italiana, ponendo maggiore attenzione ai risultati raggiunti.

This paper presents a review of the literature on scale and scope economy in banks. It shows that the conclusions of these models are often conditioned by input and output definitions, by the inclusion of bank distributive channels and by the consideration of the role of risk. In the last years the literature has stressed the relevance of the functional form of the cost function. Even the regulatory regime appears to have been relevant at least to stimulate the analyses on these themes.

The USA studies on the theme are more rich and detailed than any others, so in this paper we first consider them: in particular their technical elements; then, in the second section we consider the results of Italian studies, which are interesting even if less careful.

JEL CLASS.: C21; C31; G21

Keywords: modelli di analisi delle economie di scala; costi e dimensioni delle banche.

Simona Costagli

1. Introduzione

Negli Stati Uniti il tema relativo alla relazione tra costi, dimensioni e rete distributiva nel sistema bancario si inserisce nel corso del tempo in un ampio dibattito circa l'opportunità di modificare un sistema caratterizzato da un forte localismo e da una dimensione ridotta della maggior parte degli istituti bancari. Per questo motivo, spesso le conclusioni cui i modelli empirici giungono risultano strumentali a sostenere determinate posizioni circa la direzione da intraprendere per riformare la struttura bancaria del paese. In particolare, gli studi pionieristici sul tema tendevano a valorizzare la piccola dimensione di gran parte delle banche USA rilevando per queste economie di scala positive, al contrario delle banche maggiori in cui invece prevalevano rendimenti decrescenti. Successivamente, l'affinamento delle tecniche utilizzate (sia per la definizione della funzione di costo, sia per la gestione dei dati) ha permesso di rendere più robusti i modelli di analisi, includendo variabili che permettono di tener conto dell'influenza della struttura dei mercati, dei canali distributivi e soprattutto (in epoca recente) dell'impatto che la non neutralità verso il rischio da parte dei manager bancari può avere sulla determinazione dei costi.

Alcuni modelli più recenti hanno quindi in parte modificato le conclusioni più antiche, rilevando economie di scala crescenti anche per le grandi banche. Vi è comunque da sottolineare che, come si vedrà, il diverso impianto tecnico utilizzato nei vari modelli può costituire uno dei principali motivi della diversità dei risultati raggiunti, insieme, ovviamente, alla scelta del significato da attribuire a concetti fondamentali come quelli di input e output bancari.

A parte gli Stati Uniti, l'Italia è l'unico paese che, a partire dagli anni Sessanta, abbia prodotto un flusso considerevole e regolare di studi sul tema delle economie di scala e di scopo nel settore bancario (seppure di portata non paragonabile a quello statunitense). Anche in questo caso, tuttavia, le indicazioni fornite seguono abbastanza fedelmente la traccia indicata nel corso del tempo dalla regolamentazione, e in particolare da quella specifica sulla rete degli sportelli. In Italia, inoltre, raramente le analisi empiriche sono contributi originali al tema, più spesso esse prendono le mosse da analoghi modelli elaborati negli Stati Uniti (che peraltro presentano una continuità di analisi e una ricchezza di spunti di riflessione senza eguali). Per tale motivo, la rassegna analizza nella prima parte la letteratura statunitense, avendo cura

* Ringrazio il professor Roncaglia per la paziente lettura di una versione preliminare del lavoro. Resto comunque l'unica responsabile delle opinioni espresse.

di sottolineare la ricchezza dei contributi tecnici e teorici. Nella seconda parte sono poi presentati i principali contributi italiani, avendo cura di porre in maggiore evidenza i risultati raggiunti.¹

Preliminare a tale rassegna è tuttavia un rapido accenno al concetto di prodotto bancario; a ciò viene dedicato il paragrafo seguente.

2. Il prodotto bancario nella letteratura: alcuni cenni

2.1. Il punto di vista macro e microeconomico

In letteratura la definizione di prodotto bancario riflette le ipotesi fatte sulle ragioni dell'esistenza e della specificità dell'intermediario banca. In termini generali le interpretazioni prevalenti ricadono in due filoni detti macro e microeconomico.² Nell'interpretazione "macroeconomica", ben riassunta dalla posizione di Goldschmidt (1981), la funzione monetaria svolta dalle banche assume una rilevanza maggiore rispetto a quella creditizia (che si realizza attraverso la concessione di prestiti). Tale visione giustifica l'esistenza dell'operatore banca in virtù della sua capacità di offrire parte dei mezzi di pagamento esistenti nel sistema economico attraverso il processo di moltiplicazione dei depositi. La peculiarità del prodotto bancario risiederebbe quindi nella capacità della banca di generare liquidità (*moneyness*) e poiché questo è un attributo delle sue passività, occorre privilegiare il momento della raccolta su quello dell'impiego dei fondi. I depositi sono quindi l'aggregato più adatto per approssimare l'output bancario.³

Diverso è l'aspetto rilevante nell'ambito dell'approccio microeconomico, in cui viene privilegiato l'aspetto creditizio, ovvero il ruolo delle banche in quanto intermediari che prendono a prestito dalle unità in surplus per offrire risorse a quelle in deficit e che quindi offrono un insieme di servizi finanziari ai creditori e ai prestatori. L'essenza dell'attività bancaria secondo tale visione risiede nell'utilità addizionale che il prodotto bancario arreca al singolo utente; la definizione di output deve pertanto essere derivata da una valutazione empirica del livello di utilità apportato dai depositi e dai prestiti.

Nell'ambito di tale approccio, alcuni autori (Benston 1965, Bell e Murphy 1968a e 1968b, Mullineaux 1978, Gilligan, Smirlock e Marshall

¹ I modelli presentati nel lavoro riguardano solo la parte della letteratura basata sul calcolo di indicatori di scala e di scopo a partire dalla stima di funzioni di costo con stime parametriche e semi non parametriche. Segnaliamo comunque, senza entrare in un'analisi dettagliata di tale letteratura, che esiste un altro filone di ricerca basato su modelli di frontiera a disturbi asimmetrici e sulla *Data Envelopment Analysis*.

² Tali interpretazioni tendono a privilegiare rispettivamente la funzione monetaria e quella creditizia svolta dalle banche.

³ Goldschmidt ritiene che i servizi associati alle voci di attivo siano trascurabili e poco rilevanti per comprendere il ruolo delle banche in un'economia monetaria.

1984 e Berger, Hanweck e Humphrey 1987) definiscono “produzione bancaria” tutti i servizi offerti alla clientela (compresi i depositi) che determinano costi operativi. Sealey e Lindley (1977) invece riassumono la posizione di coloro i quali ritengono che una misura adeguata del prodotto bancario debba essere scelta attraverso una rappresentazione del processo produttivo che colga gli aspetti di trasformazione economica (e non solo tecnica) che la produzione implica. In base a tale visione i depositi appaiono più un input che un output.

Secondo i due autori, per avere una teoria completa delle scelte delle istituzioni finanziarie è necessario superare una descrizione puramente tecnica del processo di produzione introducendo una valutazione in termini di valore. Come sottolinea Frisch (1965), la produzione in termini economici è il tentativo di creare un prodotto che abbia un valore più elevato rispetto agli input originari. Il problema diviene allora quello di chiarire cosa si intenda per valore più elevato e per chi.

Sia i prestatori sia i creditori della banca valutano i servizi bancari in virtù dell'utilità che da questi ricevono; tuttavia è l'impresa stessa che deve considerare output della produzione economica le attività che hanno un valore più elevato rispetto a quello degli input utilizzati e che siano tali che il loro valore sia misurabile su un mercato di riferimento. In altre parole, secondo questa impostazione lo sviluppo di una teoria della produzione bancaria deve partire dal presupposto che ciò che rileva è la valutazione che il mercato dà delle attività, non i benefici che riceve la società.

I servizi offerti ai creditori e ai prestatori sono allora output in senso tecnico, non economico. I servizi offerti ai depositanti in particolare sono legati all'acquisizione di un input; infatti nell'erogarli l'impresa finanziaria sostiene dei costi senza ottenere ricavi diretti. In un certo senso è come se i depositi fossero un output intermedio. I servizi associati ai conti di deposito sono quindi semplicemente un pagamento per l'utilizzo dei fondi necessari a svolgere l'attività di intermediazione; di conseguenza, i costi che l'impresa finanziaria (banca) sostiene per attrarre i fondi sono in parte impliciti (legati all'utilizzo di lavoro e capitale necessari a offrire i servizi ai depositanti) e in parte espliciti, rappresentati invece dai tassi d'interesse pagati sui depositi.⁴

Secondo questo approccio, un concetto appropriato di output dal punto di vista dell'impresa finanziaria dovrebbe riferirsi esclusivamente ai servizi

⁴ La produzione per una impresa finanziaria è pertanto un processo a più stadi che coinvolge output intermedi e in cui i fondi mutuabili acquisiti dai depositanti e serviti dall'impresa con l'utilizzo di lavoro, capitale e input materiali vengono utilizzati per la produzione di attività che generano reddito. Si tratta di un processo analogo a quello che si verifica nelle imprese manifatturiere verticalmente integrate, dove è possibile che qualche comparto produca e offra output che vengono utilizzati direttamente come input in un altro processo.

offerti ai debitori. Diviene quindi necessario quantificare tali servizi in termini di unità fisiche.

Secondo Sealey e Lindsay (1977), solo i servizi che la banca associa all'erogazione di attività fruttifere (*earning assets*) possono essere considerati prodotto in senso economico, poiché solo per queste il mercato è in grado di realizzare il processo di valorizzazione. In base a tale approccio, pertanto, i depositi sono il prodotto intermedio di un processo produttivo verticalmente integrato. In realtà, più che ai depositi tale definizione dovrebbe applicarsi a tutti i fondi utilizzati per lo stesso scopo dalla banca, come il patrimonio, i fondi rischi, i prestiti ottenuti dalla banca centrale, i fondi acquisiti sul mercato interbancario, le passività subordinate, le obbligazioni e così via. Nel caso specifico di fondi acquisiti tramite il canale dei depositi (debiti verso clientela ordinaria), tuttavia, la controparte per l'acquisizione prevede, oltre al pagamento di un interesse esplicito, un insieme di servizi tra cui gli assegni, le operazioni di pagamento utenze, e così via, che implicano l'impiego di risorse reali da parte della banca. In tal senso vengono quindi considerati input intermedi non solo l'ammontare dei depositi, ma anche tutti i servizi connessi e inscindibili dalla fornitura dei depositi (amministrazione, meccanismi di pagamento, e così via).

Se quindi viene privilegiata la visione che considera la banca trasformatrice di fondi (totale delle passività al netto delle riserve e della parte di patrimonio immobilizzato in macchine e beni immobili) dalle unità in surplus verso quelle in deficit, attraverso una combinazione con gli altri input rappresentati da lavoro e capitale fisico, i fondi messi a disposizione ai debitori dell'impresa bancaria (prestiti alla clientela, titoli e partecipazioni di proprietà della banca e prestiti ad altre banche) rappresentano il vero prodotto.

2.2. *L'approccio della "produzione" e quello dell'"intermediazione"*

In letteratura le riflessioni teoriche sul concetto di prodotto bancario si riferiscono talvolta anche a due diversi approcci detti dell'"intermediazione" e della "produzione", questi, grosso modo, riflettono le caratteristiche micro e macro dell'attività bancaria appena viste. In particolare, l'approccio dell'intermediazione fa riferimento alla funzione di trasformazione delle passività con un elevato grado di liquidità in attività finanziarie con diversa scadenza e rischio creditizio. In base a tale approccio, dai costi (reali e finanziari) sostenuti dalla banca per effettuare tali operazioni è possibile individuare i fattori produttivi e le materie prime impiegate nel processo produttivo, mentre dai ricavi di natura finanziaria è possibile risalire alle diverse tipologie di prodotto. Gli input sono quindi il lavoro, il capitale e l'insieme dei fondi raccolti, mentre gli impieghi, che figurano all'attivo dello stato patrimoniale della banca, e le altre attività in servizi che non risultano stretta-

mente collegate all'intermediazione creditizia rappresentano gli output del processo produttivo.

L'approccio della produzione sottolinea invece l'aspetto tecnico dell'attività bancaria identificando come prodotti tutti i servizi che implicano un qualche utilizzo di fattori produttivi. Tra gli output quindi rientrano non solo le attività che originano ricavi finanziari, ma anche quelle passività che, pur figurando come costo nella determinazione del rischio economico della banca, comportano tuttavia un impiego di lavoro e capitale. In questo senso i depositi non rappresentano la materia prima del processo produttivo ma un prodotto intermedio della funzione di intermediazione che risulta da un processo produttivo di cui occorre valutare l'efficienza.

Per quanto riguarda gli input, coerentemente con la rappresentazione adottata del processo produttivo, questi sono lavoro e capitale nella forma di macchinari e immobili.

3. La letteratura statunitense sulle economie di scala e di scopo

3.1. *Gli elementi alla base dei modelli empirici: le prime analisi e l'uso della funzione Cobb-Douglas*

Il procedimento standard nei modelli di analisi delle economie di scala prevede la costruzione degli indicatori a partire dalla stima di una funzione di produzione. Una tale impostazione richiede scelte oltre che sulla definizione output e input bancari, sul tipo di relazione funzionale con cui rappresentare la funzione di produzione, sull'opportunità di considerare i canali distributivi o il ruolo del capitale proprio come variabili indipendenti insieme ai prezzi degli input. Nel corso del tempo, l'attenzione dei vari studiosi si è concentrata diversamente su ognuno di questi elementi, e ancora oggi non vi è un'opinione condivisa né sulla loro effettiva rilevanza né, soprattutto, sui risultati delle analisi.

In realtà, sin dagli studi pionieristici di Alhadeff (1954) e Horvitz (1963), le analisi sui rendimenti di scala, tranne alcune eccezioni, (Mullineaux 1978) sono state condotte utilizzando al posto della funzione di produzione quella di costo (di più facile gestione) sfruttando la proprietà duale.

L'idea fondamentale alla base dei primi studi sul tema (Alhadeff 1954, Horvitz 1963, Bell e Murphy 1968a e Benston 1972) è che le banche (in particolare le *commercial banks* e le *savings & loan associations*,⁵ tipologie

⁵ Negli Stati Uniti il termine *commercial banks* si riferisce a banche di credito ordinario puro. Esse svolgono operazioni di provvista rimborsabili a vista (o con breve preavviso), operazioni di impiego come anticipazioni o sconti e una serie di servizi accessori come custodia di titoli e valori, servizi di tesoreria e così via. Le *savings & loan associations* sono istituzioni di risparmio private che concedono mutui garantiti da ipoteca di primo grado su immobili ubicati entro un raggio di 50

considerate in queste prime analisi) siano imprese multiprodotto in cui ogni output e servizio offerto può essere realizzato con un processo produttivo separato dagli altri. Si tratta di un'ipotesi fondamentale che giustifica la stima di n distinte funzioni di costo, una per ogni output considerato, per le grandi banche e implica invece la costruzione di un indicatore complessivo dell'output per le banche di minori dimensioni, per le quali l'imputazione dei costi alle varie attività diviene impossibile.⁶

Sia nel modello di Bell e Murphy sia in quello di Benston la rappresentazione della funzione di costo avviene attraverso una forma funzionale di tipo Cobb-Douglas. Nel primo dei due studi l'output è rappresentato da depositi a vista, a risparmio, prestiti rateali, mutui e prestiti immobiliari misurati in termini di numero di conti e la funzione è arricchita da una serie di variabili *dummies* necessarie a rendere omogenea la misura prescelta dell'output (i conti possono infatti differire per dimensione, tipo di servizio reso e composizione) e a tener conto del modo in cui avviene l'espansione della scala di produzione.⁷

Il modello rileva la presenza di economie di scala per quattro delle cinque funzioni stimate; questo, associato alla lettura dei coefficienti delle *dummies* indicative della struttura per filiali, porta gli autori alla conclusione che le economie di scala rappresentano un fattore significativo nella riduzione dei costi delle banche, ma solo quando l'espansione avviene a parità di numero di filiali.

Come si è accennato, in questi primi modelli la forma funzionale adottata per la rappresentazione della funzione di costo è una Cobb-Douglas.⁸ In base a tale relazione funzionale è possibile ottenere stime coerenti delle economie di scala se si assume implicitamente che l'output sia

miglia dalla propria sede, con fondi che si procurano con la vendita di certificati azionari di investimento.

⁶ Tale impostazione non permette evidentemente di considerare le complementarità di costo che si realizzano in un'impresa multiprodotto. Per una trattazione più approfondita di questo punto si veda la critica a questo tipo di impostazione proposta nel paragrafo 4.

⁷ L'aumento della produzione infatti può avvenire sia con un aumento dell'output a parità di numero di sportelli, sia con un aumento del numero delle dipendenze associato all'aumento dell'output; il costo associato all'aggiunta di nuove dipendenze deve pertanto essere distinto da quello connesso all'espansione di una struttura esistente.

⁸ La funzione Cobb-Douglas, come dimostrato da Nerlove (1965) e Walters (1963), è un'equazione in forma ridotta dalla quale possono essere ottenute stime univoche dei parametri strutturali, ovvero attraverso la quale la funzione di produzione è perfettamente identificata, se i coefficienti dei fattori di produzione hanno somma pari a uno.

esogenamente determinato e che l'impresa operi in modo da minimizzare i costi (Bell e Murphy 1968b, Benston 1972).⁹

Nel 1972 il modello elaborato da Bell e Murphy viene riproposto da Benston che individua consistenti economie di scala soprattutto considerando come output i depositi a vista e i prestiti per l'acquisto di beni immobili.

Tra gli studi di questo periodo uno dei pochi a non rappresentare la funzione di costo con una forma di tipo Cobb-Douglas è quello di Schweiger e McGee (1961) che definiscono una funzione di regressione multipla in cui le variabili indipendenti sono una serie di fattori che influenzano i costi:

$$C_i/A_i = b_0 + b_1 M_i + b_2 D_i + \sum_j b_j E_{ji} + b_3 G_i + \sum_k b_k O_{ki};$$

dove:

C_i/A_i sono i costi operativi totali sulle attività totali della banca i -esima; M_i è una variabile dimensionale posta una volta uguale ai depositi e un'altra alle attività totali; b_1 è un coefficiente di scala (se negativo indica la presenza di economie di scala); D_i ed E_j rappresentano rispettivamente i depositi a risparmio sul totale dei depositi e un insieme di attività fruttifere; G_i è la percentuale di crescita delle attività e O_k racchiude altre variabili di struttura.

Studi impostati su tali semplici funzioni di regressione vennero presto messi da parte dagli altri analisti e criticati poiché a una tale impostazione non era sottesa alcuna specifica teoria della produzione. La motivazione data è che il principio che permette lo studio delle economie di scala attraverso l'analisi della funzione di costo è dovuto alla corrispondenza tra quest'ultima e la funzione di produzione in un'impresa che minimizza i costi, cosicché se le economie di scala sono una proprietà della funzione di produzione lo sono anche della funzione di costo associata. Ciò significa che la funzione di costo specificata deve essere coerente con il tipo di struttura organizzativa della produzione dell'impresa, condizione che risulta pertanto violata anche dai primi lavori di Alhadeff (1954), Horvitz (1963), Gramley (1962), Schweiger e McGee (1961), Greenbaum (1967) e Powers (1969). Bell e Murphy (1968a e 1968b) e Benston (1972), d'altro canto, pur considerando funzioni di costo (e quindi di produzione) ben definite, le stimano indipendentemente l'una

⁹ Quest'ultima ipotesi viene considerata valida soprattutto nel caso di istituzioni finanziarie private. Benston non spiega il motivo, ma ritiene che vi possano essere problemi nel caso delle *savings and loan associations*. La seconda ipotesi, ovvero che un modello uniequazionale sia sufficiente a tracciare la curva dei costi solo se l'output risulti esogeno al modello è rispettata se si suppone: che l'output sia determinato dalla domanda del mercato locale di riferimento; che la tecnologia utilizzata nell'industria bancaria sia comune a tutte le banche (che possono quindi usarla ai costi più bassi); che vengano esclusi dalla definizione di costi operativi quelli affrontati per spostare la curva di domanda: spese per interessi e per la pubblicità. Ciò chiaramente nell'ipotesi che i depositi fossero parte dell'output bancario.

dall'altra e ciò impedisce loro di tener conto delle interrelazioni esistenti tra i vari output; così, ad esempio, considerare i prestiti per l'acquisto di beni immobili, i mutui aziendali e i prestiti rateali mutuamente indipendenti non permette loro di considerare la tecnologia della gestione di portafoglio che li lega insieme.

Nel 1972 Schweitzer effettua uno studio su un campione di banche costituite nella forma di *holding company*¹⁰ in cui per la prima volta la spesa per interessi viene esplicitamente considerata parte dei costi operativi.¹¹ Tale scelta non indeboliva l'ipotesi di base che l'output fosse esogeno al modello, poiché le banche statunitensi erano all'epoca di dimensioni piuttosto piccole e il loro legame con i clienti così forte che difficilmente manovre sui tassi d'interesse, peraltro rigidamente vincolati dalla *Regulation Q*,¹² avrebbero potuto spostare significativamente la curva di domanda, che quindi poteva essere considerata "data", come il livello dell'output, senza bisogno di inserire funzioni esplicite che ne spiegassero il livello.

L'ulteriore passo in avanti rispetto agli studi precedenti consiste nell'utilizzare un'unica funzione di costo in cui l'output viene sostituito da un indicatore dello stesso. Per costruire tale indicatore l'autore segue l'approccio di Greenbaum (1966), assumendo i rendimenti correnti dei prestiti come base per il calcolo. Tale approccio si basa sull'idea che il rendi-

¹⁰ Date le caratteristiche del campione utilizzato, nello studio non si fa accenno al confronto tra banche monosportello o con filiali, ciò poiché le *bank holding companies* nacquero come strumento per superare il divieto all'apertura di filiali soprattutto in stati diversi da quello di origine.

¹¹ Un'importante caratteristica di molte analisi sui costi bancari di questo periodo (che sono confinate al solo studio delle economie di scala, senza alcun riferimento a quelle di scopo) è quella di escludere la spesa per interessi dal computo dei costi operativi: la motivazione reale di tale scelta, come si è visto, derivava dalla necessità di rendere l'output esogeno al modello. Affinché ciò si verificasse era necessario supporre che questo fosse funzione della domanda del mercato di riferimento, e che questa fosse data; per renderla tale vennero esclusi dal modello gli elementi che potevano influenzarla: spese per la pubblicità e, appunto, la spesa per interessi. La giustificazione data per argomentare tale scelta era che la spesa per interessi era una spesa "finanziaria" e non reale per la società, e in quanto tale non idonea a misurare l'efficienza sociale. L'obiezione più immediata è però che questa spesa rappresenta un costo reale per le banche che la sostengono e non tenerne conto significa ignorare tutto il processo di "tecnologia finanziaria" con cui i depositi vengono trasformati in prestiti e che viene riassunto nella funzione di produzione. Poiché inoltre l'identificazione dei parametri che costituiscono la funzione di produzione attraverso quella di costo è possibile solo se vale l'ipotesi che i banchieri minimizzino i costi, è evidente che questi terranno conto dei costi che sono "reali" per loro, e non per la società.

¹² La *Regulation Q* fu introdotta dopo i fallimenti bancari dei primi anni Trenta e vietava agli istituti di deposito la remunerazione dei depositi a vista ponendo tetti alle altre tipologie. Venne abolita definitivamente solo nel 1986.

mento di un'attività finanziaria sia una buona prima approssimazione delle risorse necessarie a servire un'unità monetaria di attività.¹³ Per quanto riguarda le attività detenute sotto forma di titoli, si assume che il rendimento sia dato e pari a R_S ; l'ipotesi è infatti che l'offerta di titoli sia perfettamente elastica e che non esistano imperfezioni del mercato che possano modificare tale assunzione. I pesi degli output diversi dai titoli vengono stimati attraverso un'equazione deflazionata del tipo: $(R_L/A) = a + \sum_i r_i (L_i/A) + u$; con $R_L = \sum_i r_i L_i$,

dove:

L_i sono i prestiti, r_i i tassi d'interesse, A sono le attività totali della banca (per cui R_L/A è il rendimento medio sulle attività) e u è un termine di errore con media pari a zero e varianza costante.

L'equazione stimata viene quindi arricchita con variabili *dummies* per tenere conto della collocazione geografica (in particolare per distinguere le banche che operano in aree metropolitane, più competitive, da quelle che operano nelle aree rurali). Il valore stimato di r (r^\wedge) viene quindi utilizzato come peso per definire l'output: $Y = \sum r^\wedge L_i + R_S$.

Il modello proposto da Schweitzer viene quindi completato dall'introduzione di una serie di variabili strutturali che rilevano se la banca faccia o meno parte del *Federal Reserve System* o se sia o meno parte di una *holding company*. Queste vengono riassunte in una funzione che, combinata con una funzione dell'output, in modo da ottenere una funzione di produzione omotetica con isoquanti omotetici, dà luogo (grazie al teorema di Shepard sulle funzioni di produzione omotetiche) a una funzione di costo che, espressa in termini logaritmici, assume una forma del tipo:

$$\log C = \log a + (1/\gamma) \log Y + \beta_1 \log H_i + \beta_2 \log F_j + \log e$$

dove:

H_i indica una serie di variabili *dummies* inserite per tenere conto se le banche siano affiliate a *holding* grandi o piccole o se facciano parte del *Federal Reserve System*; F_j rappresenta un insieme di *dummies* inserite per catturare l'impatto sui costi dei diversi livelli salariali dovuti alla collocazione geografica delle banche; Y è un indicatore dell'output definito sopra.

Lo studio rileva la presenza di economie di scala per banche con attività totali inferiori ai 3,5 milioni di dollari, rendimenti di scala costanti per banche con attivo tra i 3,5 e i 25 milioni di dollari e rendimenti decrescenti per quelle con un attivo superiore. È interessante notare come mentre per banche con attivo inferiore ai 3,5 milioni di dollari e superiore ai 25 milioni non si rileva una relazione significativa tra efficienza e affiliazione, per le

¹³ L'approssimazione non è perfetta perché può riflettere imperfezioni del mercato e soprattutto non rispecchia gli effetti di lungo periodo legati ai guadagni o alle perdite in conto capitale sui rendimenti effettivi.

banche della fascia intermedia l'appartenenza a una *holding company* diviene una possibile fonte di riduzione dei costi. L'affiliazione alle *holding company* poteva quindi rappresentare una via che le banche monosportello avevano per sfruttare i benefici (legati alla maggiore articolazione territoriale) altrimenti usufruibili solo da banche situate in stati con legislazione più permissiva in termini di apertura di sportelli.

Se si esclude il ricorso a una misura complessiva dell'output, che ha permesso la stima di un'unica funzione di costo, gli studi di questo periodo (inizio anni Settanta) non hanno aggiunto granché alla conoscenza dell'efficienza operativa delle banche rispetto agli approcci pioneristici di Alhadeff (1954) e Horvitz (1963). Secondo Mullineaux (1978) ciò sarebbe dovuto al fatto che questi studi facevano riferimento solo a ciò che egli definisce "efficienza tecnica" (ovvero alla capacità di un'impresa di produrre più output a parità di input, rispetto ad altre imprese) mentre sarebbe stato utile considerare anche un'"efficienza di prezzo" (allocativa) che consiste invece nell'eguagliare il valore del prodotto marginale di ogni input al suo costo.

La novità dello studio di Mullineaux consiste nell'uso della funzione di profitto¹⁴ per lo studio di entrambi i tipi di efficienza: allocativa e tecnica.

L'approccio basato sulla funzione di profitto per lo studio dell'efficienza gode di una serie di proprietà; tra queste il fatto che il livello dell'output non compare come variabile nella funzione di profitto. Questo secondo Mullineaux rappresenta un vantaggio notevole,¹⁵ inoltre, poiché la funzione di profitto considera esplicitamente i prezzi, essa permette di definire un concetto più ampio di efficienza rispetto agli studi precedenti sui costi bancari che considerano solo quella tecnica.¹⁶

Una tappa obbligata, anche in un modello basato sulla funzione di profitto, consiste poi nella scelta di una forma funzionale per la rappresentazio-

¹⁴ La funzione di profitto esprime il profitto massimo che può essere ottenuto da un'impresa che opera in un contesto concorrenziale in funzione dei prezzi dell'output e degli input (entrambi variabili) e delle quantità di fattori produttivi fissi. Essa è non-negativa, convessa, non decrescente nel prezzo dell'output e nelle quantità di fattori fissi e non crescente nei prezzi degli input, omogenea di grado uno nei prezzi dell'output e degli input. Nel definire la funzione di profitto si assume che l'impresa operi in modo da massimizzare il profitto, che sia *price-taker*, sia nel mercato dell'output sia in quello degli input e che la funzione di produzione sottostante sia concava negli input.

¹⁵ La definizione dell'output bancario è infatti uno dei problemi che caratterizza gran parte degli studi sulla presenza di economie di scala basati sulla stima della funzione di costo.

¹⁶ Vi è tuttavia da sottolineare che nel modello di Mullineaux il profitto è espresso, oltre che in funzione dei prezzi degli input e della quantità dei fattori (fissi) della produzione, anche dei prezzi dell'output. Ciò significa che, se pure non inserito esplicitamente nella funzione, una definizione di output è comunque necessaria, e ciò ripropone gli stessi problemi che l'autore sosteneva di aver superato.

ne della funzione da stimare. La soluzione adottata da Mullineaux è particolare. Egli infatti utilizza una funzione ibrida: translog nei prezzi dell'input lavoro e Cobb-Douglas nei prezzi dell'output, degli altri input e delle quantità di fattori produttivi fissi.

$$\ln \pi = a_0 + \sum_i a_i \ln q_i + \sum_j b_j \ln p_j + \sum_m s_m \ln v_m + 1/2 \sum_m \sum_j h_{mj} \ln v_m \ln v_j + \sum_k c_k z_k$$

in cui

q_i ($i = 1, \dots, m$) sono i prezzi degli m output; p_j ($j = 1, \dots, n$) sono i prezzi degli n input diversi dal lavoro;

v_m ($m = 1, \dots, t$) sono i prezzi dell'input lavoro e z_k ($k = 1, \dots, w$) le quantità di fattori fissi della produzione.¹⁷

Lo studio di Mullineaux considera quindi l'efficienza delle *one-bank holding companies* (OBHC) e delle *multi-bank holding companies* (MBHC) (società finanziarie che controllano una o più banche rispettivamente) rispetto alle banche non costituite sotto forma di *holding*.¹⁸ La variabile dipendente, i profitti bancari, è definita come ricavi operativi meno spese operative al netto dei costi di locazione.¹⁹ Le variabili esplicative considerate sono:

¹⁷ Il vantaggio della "componente" translog è di rendere meno restrittive le ipotesi sulle possibilità di sostituzioni tra input e output rispetto alla Cobb-Douglas.

¹⁸ I dati utilizzati sono quelli del *Functional Cost and Profit Analysis Programme* del Federal Reserve System. A tale proposito, vi è da dire che i diversi risultati cui gli studi giungono possono essere in parte imputati alle diverse basi dati cui fanno riferimento. Le due fonti più utilizzate: il *Functional Cost and Profit Analysis Programme* (FCA) del Federal Reserve System e il *Call Report and Financial Statement Data* della Federal Deposit Insurance Corporation della Federal Savings and Loan Insurance Corporation e del National Credit Union Share Insurance Fund, inoltre, non sono tali da garantire una elevata rappresentatività dell'intero universo delle banche USA. In particolare, la prima tra le due fonti (anche per frequenza con cui è stata impiegata nei vari studi) ha il vantaggio di essere costruita con criteri contabili che permettono di isolare i costi per ciascuna funzione svolta dalla banca. Il problema è che la partecipazione al programma FCA è volontaria: la mancanza di criteri oggettivi per la presentazione dei dati fa sì che le cifre fornite siano spesso frutto di manipolazioni di tipo "*window dressing*" operate dalle banche per mettere in risalto aspetti particolari della loro gestione. Inoltre, il campione della FCA è sbilanciato verso le banche di minori dimensioni con depositi solitamente inferiori ai 200 milioni di dollari.

Il *Call Report* contiene informazioni su una serie più ampia di istituzioni, e richiede condizioni standard di segnalazione da parte delle banche che vi partecipano, cosa che facilita l'estensione dei risultati all'intero sistema. Mancano tuttavia tra le informazioni fornite il numero e l'ammontare medio delle relazioni di deposito e di finanziamento, e ciò rappresenta una notevole limitazione al campo di applicabilità di tali dati.

¹⁹ Seguendo McFadden (1966), dal cui modello prende spunto, l'autore esclude i costi di locazione dal calcolo dei profitti in quanto costi fissi (si veda Mullineaux 1978, p. 265).

i prezzi degli output (tasso sui prestiti per l'acquisto di beni immobili, sul credito al consumo, sui prestiti commerciali e per l'agricoltura e canone d'affitto sulle cassette di sicurezza); i prezzi degli input (tasso salariale dei funzionari e degli impiegati, tasso sui depositi a vista, su quelli a risparmio, sui certificati di deposito, affitto per i computer – questa variabile viene inserita come indicatore della diversa produttività delle macchine); le quantità di fattori fissi (sportelli pienamente operativi, sportelli ad attività limitata, stazioni di pagamento e deposito, dimensione media degli sportelli a piena operatività); lo *status* organizzativo (l'inserimento di queste due variabili serve a misurare l'influenza della struttura organizzativa sulla dimensione dei profitti); la struttura di mercato (numero relativo equivalente).

La presenza o meno di economie di scala è rilevata attraverso la stima dei coefficienti dei fattori fissi.²⁰ Questa tuttavia può essere influenzata dalla regolamentazione in vigore nei vari stati, in particolare da quella relativa agli sportelli. Per tenere conto di questa intuizione nel modello viene inserita una variabile che misura l'impatto della struttura di mercato sui profitti: si tratta del "numero equivalente" (NE), che misura il numero di ipotetiche banche della stessa dimensione che genererebbero lo stesso grado di concentrazione esistente in quel mercato.²¹

I risultati della stima di quest'ultimo modello indicano una forte influenza della regolamentazione nella misura delle economie di scala, in particolare potenziali economie di scala sono rilevabili soprattutto nelle banche commerciali operanti negli stati a regolamentazione monosportello; tuttavia, come si è visto, la regolamentazione impedisce loro di sfruttare questo vantaggio attraverso l'ampliamento del numero di sportelli. Le economie di scala presenti nelle banche degli stati che permettono un libero insediamento delle filiali sono consistenti ma di intensità inferiore.²²

In generale, le economie di scala individuate con lo strumento della funzione di profitto appaiono più consistenti rispetto a quelle stimate utilizzando la funzione di costo.

²⁰ In particolare, se la somma di tali coefficienti è pari all'unità si può concludere a favore dell'ipotesi di economie di scala costanti nel settore bancario. Ciò discende dalla condizione necessaria e sufficiente affinché la funzione di produzione sottostante a quella ibrida considerata sia omogenea di grado k (Mullineaux 1978, p. 268).

²¹ Il NE cattura informazioni sia relativamente al numero sia alla dimensione dei concorrenti presenti su un mercato ed è pertanto migliore di un semplice indice di concentrazione da utilizzare come misura *proxy* del grado di concorrenza di un mercato. Un valore consistente di NE suggerisce un elevato grado di concorrenza.

²² Lo studio di Mullineaux sembra quindi confermare l'utilità della costituzione delle MBHC come strumento per superare le limitazioni all'insediamento di nuovi sportelli: tale struttura organizzativa permetterebbe infatti alle banche monosportello lo sfruttamento di economie di scala altrimenti precluse dal divieto di aprire dipendenze.