Economies of scale and scope in the Italian banking industry

Antoine Math

European University Institute, Florence, Italy

Abstract
This study is an attempt to explore cost characteristics in the Italian banking industry. A multiproduct translog cost function is estimated on a representative cross-sectional sample of 1991 bank data. Several cost measures are compared and confronted.

1. Introduction
The study of cost functions in the banking industry is important for its possible implications for the optimum firm size, market structure in banking, and related issues of regulation. To date, there are strong divergences on these issues, although the banking markets have frequently been analysed by means of economies of scale and scope. Until recently, the theory and empirical analysis of the cost structure of banks concentrated only on the cost advantages resulting from returns to scale. The underlying problem of the definition of bank output arises. For a long time, researchers tried to define it as a homogeneous quantity. However, it finally appeared in the 1980s that the diversity of bank services made it unrealistic to deal with any single aggregate output. Today, a consensus has emerged that banks are

---

1. I would like to thank Professors Stephen Martin and Louis Phlips for their comments. I am particularly grateful to Barbara Boehnlein. The usual disclaimer applies.

2. A presentation of the large empirical literature is beyond the scope of this work. For a complete coverage of the literature, see the following complementary surveys: Benston et al. (1982), Gilbert (1984), Cuesta (1985), Clark (1988), Math (1992), Muldur (1992), and Forestieri (1992).
multiproduct firms producing a variety of services. Consequently, when considering cost efficiency in banks, a distinction between cost efficiencies due to economies of scale and those due to economies of scope is essential.\textsuperscript{2}

No consensus has emerged from the empirical analyses, notably on the issue of optimal size. In studies of American institutions, the results are very divergent; a majority concluded that economies of scale are exhausted at a very small size, but amongst the best studies the optimal size is more often found for larger sizes. For European banking markets, most found an optimal size quite below the size of the biggest banks. On this point, the studies of the Italian banking system are marked by a deep contrast since economies of scale are found to be present for all sizes, and if not, the optimal size corresponds to very large banks.\textsuperscript{3} These studies are nonetheless not devoid of weaknesses, especially in the modelling of the cost function and in the tools used for estimating cost economies.\textsuperscript{4}

This paper is an attempt to update the cost analysis of the Italian banking sector, by improving some of the crucial methodological weaknesses of the previous literature. I estimate a multiproduct cost function on a representative sample of Italian banks by employing the flexible translog form. The specification takes into account the restrictions required by theory, and the efficiency of the estimate is enhanced by a simultaneous estimation along with cost share equations. Various cost indicators are calculated. New indicators are developed, expansion path scale economies and expansion path

\textsuperscript{2} In addition, the possibilities of exploiting economies of scope have increased with the deregulatory process that has characterised all European banking markets over the last decade. The range of products and services that financial intermediaries can offer has been widened. At the EC level, the Second Banking directive (89/646) now allows all banks to undertake a very wide range of activities. For a presentation of new national regulatory developments and the tendency to move away from specialised credit institutions to less specialised ones in all OECD countries, see OECD (1992: 50-63). For Italy more particularly, see Monti and Porta (1990), Onado (1992, Chapter 16), and Berlanda and Masera (1993).

\textsuperscript{3} See the Appendix for a survey of this literature.

\textsuperscript{4} There are examples of weaknesses found while surveying these studies (see the Appendix). Firstly, concerning the specification of the estimated cost function: the use of a single aggregate banking output, the absence of input prices, the use of a Cobb–Douglas form in one study, the misspecification of the translog form, the absence of the necessary restrictions on the parameters. On this point, only the model in Conigliani et al. (1991) is well–specified. Secondly, concerning the calculation of cost economies indicators: computation limited to ray scale economies, obtention of product–specific economies of scale and economies of scope, product–specific or global, with very questionable and arbitrary approximations, no indicator based on new developments of the multiproduct firm. Lastly, as for this study, the quality of the data is generally poor because they use accounting data, and the approximation of some variables is arguable.
subadditivity, which compare the cost efficiency of banks that differ in both scale and product mix simultaneously. Therefore, they enable one to compare the efficiency of firms currently represented in the data, as opposed to the standard scale and scope measures. Furthermore, they do not require questionable manipulations for obtaining costs at zero-output levels, and they are derived from new developments of the multiproduct firm analysis. Consequently, they constitute more relevant indicators for examining cost structure and evaluating competitive viability in banking.

As in previous studies, some major problems have not been solved. First, the lack of transparency of banking sectors has made it impossible to find a flawless database. As in all work of this kind, some variables have had to be obtained through approximations from accounting data. Second, other conceptual issues, such as the hypothesis of exogeneity of the outputs and the appropriate level of disaggregation of the outputs, are not discussed. Last, it is a cross-sectional analysis with 1991 data. In a now moving banking environment, the results are not constant over time, and the important impact of technological changes on cost economies is not captured. Therefore, the results, as for all banking cost studies, must be taken with great caution.

The following section discusses the theoretical approach of banking production. Section 3 presents the specification of the model and the estimation procedure. Section 4 focuses on the tools employed for analysing the cost structure in banking. The data and the variables are described in Section 5. The results are presented in Section 6, and Section 7 draws the conclusions.

2. The definition of banking output

As already pointed out, there is now a consensus on the multiproduct nature of banking activity, and it may appear surprising that a cost analysis still requires the preliminary determination of what a bank produces. Actually, the most controversial issue bears precisely on what constitutes inputs and outputs for the banking sector. Unlike the analyses of manufacturing firms, the definition of the output of a bank has always been ambiguous because of the different underlying theoretical assumptions and the implicit nature of the various banking revenues. Two basic approaches

---

5 For a thorough discussion, see Math (1992).
6 Originally, the controversy was also related to the measure of the output, either in physical terms—the number of accounts or transactions, or in monetary terms—the value of these services. The literature has now fully converged to the second option.
No consensus has yet come about and certainly will not come about between both approaches since they are complementary and both reflect part of banking reality. Most studies have noted that the evidence does not indicate significant differences between the results obtained. The choice of the most appropriate approach is not straightforward and must be made according to the financial institutions examined. The intermediation approach is valuable if banks can easily determine the size of their total deposits. This case is more likely to occur when long periods are considered and when the deposits markets are competitive, and certainly less in a cross-section analysis because of the high rigidity of deposits observed in real markets. On the contrary, when banks operate as a deposit-taking and loan placement network, then the production approach is preferred. In some cases, a mixed solution, demand deposits as outputs and time deposits as inputs, could be supported.\footnote{This results from a recent pragmatic approach, labelled "user cost". A banking activity is considered as an input or output according to the sign of the derivative in an estimated profit function. Thus, a kind of deposit will be considered as an input if it influences the profit negatively, and if not, as an output.}

This study follows the production approach. The deposits cannot be considered as an input when one considers the reality of banking markets such as the Italian one. Banks cannot extend easily the deposits, the level of the latter is still essentially determined by the size of the branching network, that is, by the quantity of real resource inputs, labour and physical capital, and deposits are rather the result of a productive process. This argument is particularly relevant in a cross-sectional analysis. The treatment of output chosen here is thus in line with a representation of the bank based on the production approach, treating both loans and deposits as outputs.

3. The specification of the functional form

Consider a multiproduct firm producing an $m$-vector of outputs $Y_i$ with technology

$$F(Y_1, \ldots, Y_m; X_{1}, \ldots, X_n; A) \quad (1)$$

where $X_i$ are the $n$ inputs and $A$ is a vector of control variables, e.g., for banks, the number of branches $B$ or the organisational form.

Assuming that technology and input prices are given and that banks are cost minimising, the duality theorem between cost and production ensures that for every function (1) there exists a dual cost function of the form
The common properties, called regulatory conditions, required for the cost function are: positive for all \( Y \) and non-decreasing in \( Y \), non-decreasing, concave and continuous, and linearly homogenous in \( P \) (see Nadiri, 1981).

This is often referred to as the BBM model, for Benston (1965), Bell and Murphy (1968), who first employed this method in bank cost studies.

It was proposed for production functions by Christensen et al. (1972). Benston et al. (1982) introduced the translog cost function in banking cost studies.

The first one of the proposed functions is the translogarithmic functional form. Translog functions are appealing because of their immediate detection of overall scale economies. Their major drawback comes from their bad behaviour in the neighbourhood of zero–output level. This is particularly
problematic for the computation of product-specific economies of scale and economies of scope, global and product-specific. The latter require the assumption of a zero level for at least one of the outputs. In this case, the translog takes the value zero.\textsuperscript{15} To avoid the problem, the authors substitute an arbitrary small value.\textsuperscript{16} This procedure is widely used, but remains questionable given the arbitrary choice of this value, which has the further inconvenience of being most often far outside the observations in the sample. Berger et al. (1987: 508) remark:

The results are highly dependent on the closeness of the approximation – any finding of scope economies could be reversed by making the approximation to zero more precise. Finally, these scope economy estimates are subject to unknown extrapolation error, since they require evaluation of a cost function at (or near) zero outputs which are generally far outside the sample over which the cost function is estimated.

The \textit{generalised translog form} is a means to overcome this problem. The output variables are replaced by the Box–Cox metric transformation of the actual output levels: $Y_i$ is replaced by $(Y_i^{1-1})^{\lambda}$ if $\lambda \neq 0$ and $\ln Y_i$ if $\lambda = 0$.\textsuperscript{17} This form allows for zero output, and is thus a good solution to circumvent the shortcomings of the translog when one wants to compute indicators of product-specific economies of scale and global and product-specific economies of scope. These indicators have however lost most of their appeal with the development of new measures. They have the disadvantage of giving information for non-observed situations (zero output level). They fail to capture reality.

Improvements in the multiproduct firm analysis have led to proposals of more satisfactory tools for examining cost structures (Baumol et al., 1982). Berger et al. (1987) have proposed new measures of multiproduct economies which are computed from existing observations of bank costs. They are termed expansion path scale economies and expansion path subadditivity, and

\textsuperscript{15} Certain authors assert that the translog is not defined for zero. This is wrong. To see why, take the translog cost function in simplest case, $ln C = a + b ln Y$, then $C = e^{a + b Y}$, which equals zero when $Y$ is zero.

\textsuperscript{16} Generally 0.001 or 0.01, or, as in the case of the Italian studies surveyed in the appendix, the smallest observed value.

\textsuperscript{17} For a detailed explanation see Maddala (1988: 316–7). In some studies, this functional form is also termed hybrid (translog) cost function when $\lambda = 1$. The function is a generalisation of the translog in the sense the expression for output approaches the natural logarithm of output as $\lambda$ approaches zero. This functional form was proposed by Caves et al. (1980), and was used in several recent researches on bank costs (Lawrence, 1989; Cebenoyan, 1990; Buono and Eakin, 1990; Glass and McKillop, 1992). Clark (1984), Kilbride et al. (1986), and Lanciotti and Raganelli (1988) also transformed the output variable by the Box–Cox metric, but within a Cobb–Douglas specification.
give better insights on the viability of multiproduct banks. They do not need extrapolation of near zero values and can then be calculated without reserve from the translog cost function. The recourse to the generalised translog form (with Box–Cox transformation) is not necessary for these more relevant indicators.\(^{18}\)

I will consider the following translog cost function:

\[
\ln TC = a + \sum_i b_i \ln Y_i + \sum_j c_j \ln P_j + \frac{1}{2} \sum_j \sum_h d_{jh} \ln Y_i \ln Y_h + \frac{1}{2} \sum_j \sum_i e_{ji} \ln P_j \ln P_i + \sum_j f_{ji} \ln Y_i \ln P_j + g \ln B + \sum_j h_j \ln Y_i \ln B + \sum_j k_j \ln P_j \ln B + \frac{1}{2} (\ln B)^2 + u 
\]  

(3)

where TC is the total cost, \(Y_i (i,k = 1...m)\) are the outputs, \(P_j (j,h = 1 ...n)\) are the input prices, \(B\) is the number of branches, \(u\) the error term, and \(a, b_j, c_j, d_{jh}, e_{ji}, f_{ji}, g, h_j, k_j,\) and \(l\) are the usual parameters to be estimated with a translog specification. Note that, since total cost and the regressors are in natural logarithms and have been normalised, the first order coefficients, \(b_j, c_j\) and \(g,\) can be interpreted as cost elasticities evaluated at the sample mean.

The share equations are included in the model. Since they do not involve any new coefficients, their inclusion increases the efficiency of the estimations. The input share equations are obtained using Shephard's lemma \((X_j = \partial T / \partial P_j\) for all inputs \(X_j):\)

\[
S_j = \frac{C_j}{TC} = \frac{(P_j X_j)/TC}{P_j / TC} (\partial T / \partial P_j) = \partial \ln TC / \partial \ln P_j 
\]  

(4)

where \(C_j\) is the expense in input \(j.\)

In the case of the above translog cost function, the input share equations are

\(^{18}\) Furthermore, while defined in zero, the generalised translog form, as all translog forms, behaves very badly in the neighbourhood of zero. Martin et al. (1992) show how the translog "degenerates" as any output approaches zero. This defect is of no consequence when the indicators are computed at relevant actual points, not close to zero in banking, where the translog has proved to be behaving properly. Very recently, other functional forms have been tried in the empirical literature of bank studies. Barnett and Lee (1985) show that the minflex Laurent translog form gives better approximations than the translog and the generalised Leontief flexible forms. It was used in recent bank cost studies: LeCompte et al. (1990) and Hunter et al. (1990). The separable quadratic cost function proposed by Röller (1990) has received a first application to banking in Martin and Sassenou (1992). Pulley and Braunstein (1992) introduced the composite cost function which they have applied to banks.
\[ S_j = c_j + \sum_{n} e_{kj} \ln P_j + \sum_{i} f_{ij} \ln Y_i + k_j \ln B \]  

(5)

Because of the restrictions of linear homogeneity in input prices, the input share equations must add up to one,

\[ \sum_{j} S_j = 1 \]  

(6)

One of the share equations is excluded in the estimation in order to avoid singularity problems. The parameters are invariant independently of which share equation is omitted. The share equation corresponding to the capital has then been excluded.

In addition, the system of equations (3) and (5) must satisfy certain regularity restrictions such as symmetry and homogeneity, in order to insure that a unique correspondence exists between the cost function and the underlying production function:  

- \[ d_{ik} = d_{ki} \] for all \( i, k = 1, \ldots, m \), \( e_{ih} = e_{hi} \) for all \( j, h = 1, \ldots, n \) (symmetry)
- \[ \sum_{i} c_{ij} = 1; \sum_{j} e_{kj} = 0; \sum_{j} f_{ij} = 0; \sum_{j} k_j = 0 \] (homogeneity)  

(7)

Finally, the system of equations (3) and (5), along with the restrictions (7), has to be estimated.  

---

19 Since the translog specification is a second-order logarithmic approximation to an arbitrary twice-differentiable cost function, the cross partial derivatives of the cost function must be equal and this implies the symmetry restrictions. The restrictions of homogeneity can easily be imposed with the translog form, but have to be accepted by a proper test on an unrestricted estimation of the translog form. Other regularity conditions, such as monotonicity, concavity and non-negative marginal costs (see footnote 12), cannot be imposed without a great loss in functional form flexibility. Instead, these conditions are checked at each data point. Zaardkooihi et al. (1986) showed that these restrictions must be met and pointed out the weakness of the numerous studies that omit them. This misspecification occurs in the following Italian studies: Baldini and Andrea (1990), Landi (1990), and Cardani et al. (1991). See Appendix.

20 It was not possible to include the revenue share equations in the model as proposed by Kim (1985) for permitting endogeneity of the outputs. The revenue share equation of output \( Y_i \) is \( R_i = p_i Y_i / TC \), where \( p_i \) is the price of output \( Y_i \). The cost function implies the problematic assumption of exogeneity of the output. Kim (1985) provides a theoretical solution allowing for the endogeneity of certain outputs. By exploiting the results of the profit maximisation with endogenous outputs, he demonstrates that the revenue share equation of output \( Y_i \) becomes

\[ R_i = \varepsilon_i (1 + 1/i0)^i \]
4. Measures of economies of scale and scope using the translog cost function

Following the development of the multiproduct firm analysis, several measures of cost economies have been proposed for the study of the cost structure.

4.1. Overall (or ray) economies of scale (ESC)

Brown et al. (1979: 266) have proposed to measure overall (or ray) economies of scale in multiproduct firms by computing the sum of the output cost elasticities of individual products. This gives the percentage change in costs that results from a proportional change in the output of all products:

\[ \text{ESC} = \sum_i \epsilon_i = \sum_i \left[ \frac{\partial TC}{\partial Y_i} \right] Y_i/TC = \sum_i MC_i \left( \frac{Y_i}{TC} \right) = \sum_i \left[ \frac{\partial lnTC}{\partial lnY_i} \right] \]

where \( MC_i \) is the marginal cost specific for the ith output and \( \epsilon_i \) its product cost elasticity.

There are overall economies of scale when a proportionate increase in all outputs produces a less than proportional increase in total cost, i.e., if ESC < 1. Inversely, ESC > 1 indicates ray diseconomies of scale.

In the case of the translog cost function presented in (3),

\[ \text{ESC} = \sum_i \epsilon_i = \sum_i \left( b_i + \sum_a d_{ab} \ln Y_a + \sum_j f_{ij} \ln P_j + h_i \ln B \right) \]

ESC can be estimated for any point. Generally, it is computed at particular points, and notably at the expansion point (the "typical" firm). At this point, \( Y_i = 1 \), \( P_j = 1 \) and \( B = 1 \), and thus, ESC in (9) is simplified by \( \sum_i b_i \).  

---

where \( \epsilon_i \) is the product cost elasticity and \( \theta_j \) is the own price elasticity of the ith endogenous output. Kim (1985) shows that the joint estimation of the model along with this equation is the solution for obtaining consistent estimates with endogenous outputs. Given that the price elasticities of the demand are generally unknown, this solution was never applied. The shortcoming which consists of assuming that the banking industry is competitive adjusting output so that marginal cost equals price is not possible neither.

21 Note that all variables are normalised by the sample means in the estimation of the cost function.
4.2. Augmented (or firm) economies of scale.

The indicator described above assumes a proportional increase in outputs with a fixed banking structure. This implicit assumption does not take into account that bank output can grow essentially through a change of the banking structure, that is, through an extension of the branching network. One can compute another indicator of overall economies of scale allowing for the change in the number of branches B. It is termed augmented economies of scale.\(^22\)

\[
ESCA = \sum_i \left[ (\partial \ln TC / \partial \ln Y_i) + (\partial \ln TC / \partial \ln B) \right] (\partial \ln B / \partial \ln Y_i) + \sum_i (\partial \ln B / \partial \ln Y_i) = ESC + (\partial \ln TC / \partial \ln B) \sum_i Z_i \tag{10}
\]

where \(Z_i\) is estimated separately by the following auxiliary regression:

\[
\ln B = \text{CONST} + \sum_i Z_i \ln Y_i \tag{11}
\]

Since the total cost increases as the number of branches rises, the cost elasticity of B, \(\partial \ln TC / \partial \ln B\), is expected to be positive. \(Z_i\) are also expected to be positive since the number of branches B is positively correlated to the outputs \(Y_i\) (equation 11). Thus, note from equation (10) that ESCA is expected to be superior to ESC.

In the case of the translog specification (3), ESCA becomes

\[
ESCA = ESC + (g + \sum_i h_i \ln Y_i + \sum_j k_j \ln P_j + 1 \ln B) \sum_i Z_i \tag{12}
\]

4.3. Product–specific economies of scale

Given the multiproduct nature of the firm, some measures have been proposed for evaluating the impact of the increase of one output on cost. These product–specific economies of scale make use of the theoretical relationship between marginal cost, average cost, and economies of scale. In a single product firm, where the marginal cost of producing a product is less than the average cost at a given level of output, average cost is declining in that range of output, implying economies of scale (and inversely). Baumol et al. (1982: 67–71) have proposed an approximation of this relationship in a multiproduct model. The concept labelled "average incremental cost" (AIC)

\(^{22}\) It is also referred to as "firm economies of scale" in opposition to "plant economies of scale". Others distinguish "returns to scale" (ESCA) and "returns to (network) density" (ESC).
is used instead of the average cost of the single product firm. The average incremental cost is the addition to total cost of producing a specific level of a product as opposed to not producing it at all (incremental cost IC), divided by the level of the output. The average incremental cost of producing \( Y_k \) is

\[
AIC_k = \frac{IC_k}{Y_k} = \left[ TC - TC(Y_1, ..., Y_{k-1}, 0, Y_{k+1}, ..., Y_m) \right] / Y_k
\]  

(13)

A measure of the product-specific economies of scale is given by \( PSSC_k = AIC_k / MC_k \), where \( MC_k \) is the marginal cost of producing the level of output \( Y_k \).

\[
PSSC_k = (AIC_k/MC_k) = \left[ \frac{(IC_k/Y_k)}{(\partial TC/\partial Y_k)} = \left[ \frac{(IC_k/TC)}{(\partial ln Y_k/\partial ln TC)}\right] \right] = \left[ TC - TC(Y_1, ..., Y_{k-1}, 0, Y_{k+1}, ..., Y_m) \right] / TC
\]  

(14)

In the case of the translog form presented in (3), \( PSSC_k \) is

\[
PSSC_k = \left[ TC - TC(Y_1, ..., Y_{k-1}, 0, Y_{k+1}, ..., Y_m)/TC \right] \times \left( b_k + \sum d_k ln Y_i + \sum f_{ij} ln P_j + h_k ln B \right)
\]  

(15)

There are product-specific economies of scale for the product \( Y_k \) if \( PSSC_k > 1 \). This concept can be easily extended to a subset of products. It must be pointed out that this measure implies a knowledge of the cost when one level of output is zero. With a translog form a questionable approximation has to be made.\(^{23}\)

**4.4. Global economies of scope**

The analysis of the multiproduct firm introduced the concept of economies of scope. "There are economies of scope where it is less costly to

\(^{23}\) In order to avoid these inconveniences, Gilligan et al. (1984) have considered marginal cost. "Whether product-specific returns to scale are increasing or decreasing can be evaluated by the change in marginal costs. Marginal costs that are decreasing for the ith output should be indicative of product-specific returns to scale" (Gilligan et al., 1984: 397). The indicator is then

\[
PSSC_2 = \frac{\partial MC_i/\partial Y_i}{(TC/Y_i)} = \frac{\partial \left( \frac{TC}{\partial ln Y_i} \right)}{\left( \frac{TC}{\partial ln Y_i} \right) - 1}
\]

If \( PSSC_2 < 0 \), then marginal cost of product \( Y_i \) is declining, and this would imply product-specific economies of scale. This, however, constitutes only a very broad approximation.
combine two or more product lines in one firm than to produce them separately" (Panzar and Willig, 1981: 268). The source of economies of scope are sharable inputs. These are particularly important in the banking activity (safety system, equipment, computers, branches, labour, information).

Global economies of scope are measured by calculating the cost differential that would arise between independent and joint production of specific output levels of all products:

\[
ESCO = [TC(Y_1, 0, ..., 0) + ... + TC(0, ..., 0, Y_m) - TC]/TC \quad (16)
\]

If \(ESCO > 0\), then there are global economies of scope.

4.5. Product–specific economies of scope

A usual measure is to compute the cost increase or decrease that arises from producing a specific product both independently from, and jointly with, the remaining product mix.

\[
PSSO_i = [TC(Y_1, ..., Y_{i-1}, 0, Y_{i+1}, ..., Y_m) + TQ(0, ..., 0, Y_p, 0, ..., 0) - TC]/TC \quad (17)
\]

\(PSSO_i > 0\) implies product–specific economies of scope for product \(Y_i\) (conversely for \(PSSO_i < 0\)).

4.6. Cost complementarity

Because of the difficulty of obtaining a reliable estimate of TC with any zero level of output, it is possible to use a stronger condition for the existence of global or product–specific economies of scope. The presence of cost complementarity between two products is a sufficient condition for product–specific economies of scope between those two products. Existence of cost complementarities amongst all pairs of products is a sufficient condition for global economies of scope.

Cost complementarity occurs when the marginal cost of producing one product declines with an increase in the level of production of another one:

\[
CC_{ik} = \frac{\partial MC}{\partial Y_i} = \frac{\partial^2 TC}{\partial Y_i \partial Y_k} = \frac{TC(Y, Y_i)}{((\partial lnTC/\partial nY_i) \partial nY_k) + (\partial nTC/\partial nY_i) (\partial nTC/\partial nY_k)} \quad (18)
\]

\(CC_{ik} < 0\) implies product–specific economies of scope between products \(Y_i\) and \(Y_k\). In the case of the translog form (3), \(CC_{ik}\) becomes
\[ CC_{ik} = \left[ T/(Y_i Y_k) \right] (d_{ik} + \epsilon_i \epsilon_k) \]
with \( \epsilon_i = b_i + \sum_{k} d_{ik} \ln Y_k + \sum_{j} f_{ij} \ln P_j + \ln \ln \)

(19)

Since TC/ \( Y_i Y_k > 0 \) for any non zero level of production \( Y_i \) and \( Y_k \), a sufficient condition for cost complementarity is that

\[ d_{ik} + \epsilon_i \epsilon_k < 0 \]

(20)

4.7. Subadditivity property of the cost function

The various indicators presented above give an analysis of the cost structure, but focus on different aspects of the costs. The measures of economies of scale and economies of scope can give considerable insight into the behaviour and structure of multiproduct industries. They are however not sufficient, as shown by Baumol et al. (1982). For instance, as shown by the aforementioned authors, the presence of both the overall economies of scale and global economies of scope is neither sufficient nor necessary for determining a situation of natural monopoly, as could be commonly believed. In a single product industry, presence of economies of scale over the whole range of production is a sufficient condition for a tendency towards a natural monopoly. In a multiproduct industry, a necessary and sufficient condition is the subadditivity of the cost function. "A cost function is subadditive for a particular output vector \( Y \) when \( Y \) can be produced more cheaply by a single firm than by any combination of smaller firms" (Baumol et al., 1982: 170). Formally, "a cost function \( C(Y) \) is strictly and globally subadditive in the set of commodities \( N = 1,...,n \), if for any \( m \) output vectors \( Y_1,...,Y_m \) of the goods in \( N \) we have \( C(Y_1 + \ldots + Y_m) < C(Y_1) + \ldots + C(Y_m) \)" (Baumol, 1977: 810).

Unfortunately, there apparently exists no straightforward means by which to test whether a particular function is subadditive or not (Baumol et al., 1982: 170). However, it is possible to establish quantifiable sufficient conditions for subadditivity.\(^{24}\) For instance, if the cost function is strictly ray concave and strictly trans–ray convex (Baumol, 1977: 811).\(^{25}\) The ray concavity of the cost function and the transray convexity of the cost function can be considered as a generalisation to the multiproduct firm of respectively economies of scale and economies of scope.

\(^{24}\) The main relationships found between cost related concepts in the multiproduct framework are summarised in Baumol (1977: 812).

\(^{25}\) There are other possibilities of sufficient conditions for subadditivity. For instance, product–specific economies of scale for each product in conjunction with economies of scope. See Baumol (1977) and Baumol et al. (1982).
A cost function is strictly (output) ray concave if

\[ C[\nu_1 Y + (1-k)\nu_2 Y] > k C[\nu_1 Y] + (1-k)C[\nu_2 Y] \]

for any \( \nu_1, \nu_2 \geq 0, \nu_1 \neq \nu_2, \) and \( 0 < k < 1 \) \hspace{1cm} (21)

A cost function \( C(Y) \) is said to be transray convex through \( Y^* = (Y_1^*, ..., Y_n^*) \) if there exists any set of positive constants \( w_1, ..., w_n \) such that for every two output vectors \( Y^a = (Y_1^a, ..., Y_n^a), Y^b = (Y_1^b, ..., Y_n^b) \) lying on the same hyperplane \( \sum w_i Y_i \) through \( Y^*, \) that is, for which \( \sum w_i Y_i^a = \sum w_i Y_i^b = \sum w_i Y_i^* \), we have

\[ C[kY^a + (1-k)Y^b] \leq kC(Y^a) + (1-k)C(Y^b) \] for any \( k, 0 < k < 1 \) \hspace{1cm} (22)

However, checking the properties defined above is difficult in empirical studies because knowledge of the cost function at non-observable points, notably the origin, is required. Berger et al. (1987) have proposed new measures of multiproduct economies based on the subadditivity concept, which have the advantage of being computed from existing observations.

4.8. Expansion path scale economies

Overall or ray scale economies (ESC) only provide information about the economies of scale in the particular case when all outputs are changed proportionally. However, they provide limited insight into cost efficiency questions in the more general case when the product–mix is allowed to vary. Expansion path scale economies (EPSCE) is a concept that enables the comparison of the efficiency of banks of different sizes and different mixes. One considers the expansion in the \( m \)-dimensional output space from a "small" bank \( A \) characterised by its output bundle \( Y^A \) to a larger bank \( B \) with an output bundle \( Y^B \), with a different output mix. EPSCE is defined as the elasticity of incremental cost with respect to incremental output along the segment \( AB \). Figure 1 gives an illustration of EPSCE in a two output space. Note that EPSCE generalises the conventional ESC. ESC is a special case, where the smaller firm is at the origin.

\[
EPSCE(A,B) = \sum_i \left( \frac{Y_i^B - Y_i^A}{Y_i^B} \frac{\partial \ln TC(B)}{\partial \ln Y_i^B} \right) \left( \frac{TC(B) - TC(A)}{TC(B)} \right) \hspace{1cm} (23)
\]
For evaluating EPSCE at the firm level, that is, when the number of branches is allowed to vary, I propose an augmented indicator EPSCA, derived in a similar way as ESCA is derived from ESC.

\[
EPSCA(A,B) = \sum_i \left( \frac{Y_i^b - Y_i^d}{Y_i^b} \right) \left( \frac{\partial \ln TC(B)}{\partial \ln Y_i} + \frac{\partial \ln TC(B)}{\partial \ln B} \right) \left( \frac{\partial \ln B}{\partial \ln Y_i} \right)
\]

(24)

\[
= EPSCE(A,B) + \frac{\partial \ln TC(B)}{\partial \ln B} \sum_i \left( \frac{Y_i^b - Y_i^d}{Y_i^b} \right) \left( \frac{\partial \ln B}{\partial \ln Y_i} \right)
\]

with the translog form (3), EPSCE(A,B) becomes

\[
EPSCE(A,B) = \frac{TC(B)}{TC(B) - TC(A)} \times \sum_i \left( (Y_i^b - Y_i^d)/Y_i^b \right) \left( b_i + \sum_k d_{ik} \ln Y_k + \sum_f f_{if} \ln P_f + h_i \ln B \right)
\]

(25)

and EPSCA(A,B)

\[
EPSCA(A,B) = EPSCE(A,B) + \left( g + \sum_i h_i \ln Y_i + \sum_k k_i \ln P_k + \ln B \right) \sum_i \left( (Y_i^b - Y_i^d)/Y_i^b \right) \frac{TC(B)}{TC(B) - TC(A)} Z_i
\]

(26)

where \( Z_i \) are obtained with the same auxiliary regression used for ESCA.

EPSCE(A,B) < 1 indicates scale economies along the expansion path spanning A and B, and conversely for EPSCE(A,B) > 1. EPSCA(A,B) gives the same information, but in the more realistic case in which the number of branches is allowed to change.

4.9. Expansion path subadditivity

Estimates of economies of scope (ESCO and PSSO) in banking necessitate the extrapolation of estimated cost function well beyond the range of observed data. The expansion path subadditivity is a measure associated with plausible observations. Expansion path subadditivity, EPSUB, is computed between two points of the m-dimensional output space, a "small"
one A and a "larger" B (see Figure 1 for an illustration in a two–output space).

Considering these two banks, whose respective output bundle are $Y^A$ and $Y^B$, and such that $Y_i^B \geq Y_i^A \geq 0$ for all outputs i, a bank $\hat{A}$ with an output bundle $Y^A$ is introduced such that $Y^A + Y^A = Y^B$.

Expansion path subadditivity gives the proportional cost increase from two–firm instead of one firm production of $Y^B$, using the smaller firm A on the expansion path. Formally, it is

$$EPSUB(A,B) = \frac{[TC(A) + TC(\hat{A}) - TC(B)]}{TC(B)}$$  \hspace{1cm} (27)

If $EPSUB(A,B) > 0$, the cost function is said to be expansion path subadditive between A and B, since it implies a cost advantage for the "large" firm B. Conversely, if $EPSUB(A,B) < 0$ then firm B may be considered as competitively not viable, since firms A and $\hat{A}$ could drive B from a competitive market. Note that EPSUB generalises the conventional scope measure ESCO, since ESCO is a special case where the smaller firm specialises at point C or D.
All the measures proposed for examining the cost structure of an industry can, theoretically, be computed at any point, once the cost function is estimated. In this work, they will be computed at the average points, that is for the average bank of each class, and for the average of the whole sample.

5. The data

The estimation is made with 1991 balance sheet data collected from 'Il Mondo' (supplemento Novembre 1992) and from several 'Bolletino Statistico' and 'Supplementi' published by the Bank of Italy. The sample is representative of the Italian banking market, 231 banks amongst a total of 255, once special credit institutions and foreign banks are discarded.\(^{26}\)

The independent variables of the translog cost function defined in (3) are three outputs, two inputs, and a control variable (number of branches).

The first output, customer deposits ('raccolta da clientela'), \(Y_1\), consists of demand deposits, time deposits and deposit certificates hold by ordinary customers. The second output, bank deposits ('raccolta interbancaria'), \(Y_2\), groups all types of funds from credit institutions. The third output, customer loans ('impieghi economici'), \(Y_3\), represents all the assets to the ordinary customers. The total cost of labour, \(CL\), consists of all salaries and related charges. Other variables had to be calculated following usual procedures used in the literature.

A proxy for the unit cost of labour, \(WL\), is obtained for each bank by dividing the total cost of labour (salaries and related charges) by the number of employees.

Obtaining a proxy for the unit cost of physical capital is a problem which pervades the whole literature. It is very rare to have a totally satisfying variable, and this delicate issue needs some comment. In most studies, the price is not specific to each bank.\(^{27}\) On the grounds that banks would face

\(^{26}\) Following previous studies and for the sake of a minimal homogeneity of the institutions in the sample, only banks having retail banking activities are considered, and the credit institutions which operate only medium and long term activities have been discarded. The categories of banks included in the sample are Public Law Institutes (Istituti di diritto pubblico), National Interest Banks (Banche di interesse nazionale), Ordinary Credit Banks (Banche di credito ordinario), Popular Banks (Banche popolari), and Savings Banks (Casse di risparmio). Foreign Banks (Banche estere), Country Banks (Casse rurali), and Special Credit Institutes (Istituti di credito speciale) have been excluded. The banks for which one or more variables were missing have not been included in the sample.

\(^{27}\) Amongst about a hundred recent studies (from 1982), very few use a specific price for each bank (surveyed in Forgeserti, 1992, and Math, 1992). One can cite Hardwick (1990), Noula et al. (1990), Glass and McKillop (1992), and Muldor and Sassenou (1991). The other studies calculating a unit cost specific to each bank are American ones employing the
Table 1

<table>
<thead>
<tr>
<th></th>
<th>Total assets (%)</th>
<th>Total deposits (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banks in the sample</td>
<td>82.9</td>
<td>84.12</td>
</tr>
<tr>
<td>Other institutions</td>
<td>17.1</td>
<td>16.87</td>
</tr>
<tr>
<td>(Foreign banks)</td>
<td>(0.75)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

almost no difference in input prices, and that costs are examined on a cross-sectional basis, many studies have ignored the input prices in the specification of the cost function.\(^{28}\) In this case, the remaining parameters of the cost function are assumed to be a function of those unspecified input prices. The goal of the studies is rarely to look at the relationship between total costs and input prices. Some US studies consider regional differences in input prices, with a division of the USA in nine distinct regions. The physical capital prices are approached by the average office rental rates of the region (rent per square meter). However, all banks located in one of these nine very large regions are assumed to face the same capital price. This very questionable assumption has led authors to totally ignore the input price. For instance, Kolari \textit{et al.} (1992: 404) in their study of Finnish savings institutions have argued that "the relatively small size of Finland suggests that assuming constant prices is comparable to standard practice in US studies".

The approximation chosen here for the capital cost is also not devoid of strong flaws. Following Parigi (1989), I have assumed that the variable provision for depreciation ('accantonamenti per ammortamenti') represented an estimation of the expenditure in capital, \(\text{CK}\). This measure added to the cost of labour, \(\text{CL}\), gives the total operating cost of the bank \(\text{TC}\), since in the production approach only labour and physical capital are considered as inputs. This procedure has to be interpreted with great caution since the expenditures in capital may not be adequately reflected by the variable 'accantonamenti per ammortamenti', which is also influenced by fiscal considerations. The number of branches ('sportelli') has been used as a measure of the factor capital, following an argument usually confirmed by the fact that actual physical capital per branch is constant between banks. The detailed Functional Cost Analysis data (FCA). The samples with such data are however not representative, being biased, and thus irrelevant for useful conclusions. For the inconveniences of FCA data, see Muldur (1992), and Math (1992).

\(^{28}\) This is the case in the following recent Italian studies: Lanciotti \textit{et al.} (1988), Cossutta \textit{et al.} (1988), Baldini and Andrea (1990), and Landi (1990). See the Appendix.
Table 2
Size Distribution of Banks

<table>
<thead>
<tr>
<th>Class/ total assets in billions liras (1991)</th>
<th>Number of banks (before exclusion for missing variables)</th>
<th>Average assets in million liras</th>
<th>Average number of branches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0–300</td>
<td>17 (30)</td>
<td>159512</td>
<td>4.53</td>
</tr>
<tr>
<td>2 300–600</td>
<td>38 (41)</td>
<td>441005</td>
<td>9.97</td>
</tr>
<tr>
<td>3 600–1200</td>
<td>46 (47)</td>
<td>863809</td>
<td>18.11</td>
</tr>
<tr>
<td>4 1200–2500</td>
<td>47 (50)</td>
<td>1803645</td>
<td>34.17</td>
</tr>
<tr>
<td>5 2500–5000</td>
<td>33 (34)</td>
<td>3383209</td>
<td>48.88</td>
</tr>
<tr>
<td>6 5000–20000</td>
<td>34 (36)</td>
<td>9212847</td>
<td>118.85</td>
</tr>
<tr>
<td>7 &gt; 20000</td>
<td>16 (17)</td>
<td>67049331</td>
<td>428.31</td>
</tr>
<tr>
<td>Total</td>
<td>231 (255)</td>
<td>7106703</td>
<td>66.67</td>
</tr>
</tbody>
</table>

Minimum asset value = 69.5
Maximum asset value = 138012.6

The unit cost of capital, WK, is then obtained by dividing the estimated cost of capital by the number of branches.29

The sample has been divided in 7 classes in the analysis of the cost function (the same classes as Landi (1990), Cossutta et al., (1988), and Baldini and Andrea (1990)). These classes by asset size are presented, along with some statistics, in Table 2.

6. Results

6.1. Estimation procedure

The system of equations (3) and (5), along with restrictions (7) is estimated using Zellner’s seemingly unrelated regressions (SUR) procedure.30

---

29 For the estimation of (3) and (5), the notations of the outputs Y, are respectively Y1, Y2, and Y3 for i = 1, 2, 3. Pj are the unit cost of labour and the unit cost of capital for j = 1, 2. TC is the notation for total operating costs.

30 This technique uses estimates of the covariance of the residuals across equations to improve the efficiency of the estimates. The system estimates are preferred to the single-equation ordinary least squares estimates because of their greater efficiency, as is shown by the decrease of standard errors for all parameters. OLS results are not added for the sake of brevity. The estimation was carried out with the Syslin procedure in SAS.
6.3. Economies of scale

Ray scale economies, ESC, and augmented (or firm) economies of scale, ESCA, were calculated at the eight mean points, using equations (9) and (12) respectively. The results reported in Table 4 are in line with the previous measures for the Italian banking market. Ray scale economies are found over the whole sample. This conventional measure assumes that the number of branches remains unchanged as a bank expands. This hypothesis is not realistic. The augmented scale economies formula, ESCA, gives a more appropriate measure, since it includes both the direct effect on total cost of a (proportional) change in outputs and the indirect effect of the induced change in the number of branches. As expected, ESCA is found superior to ESC. The conclusion is nevertheless not modified, and returns to scale are reported at all sizes.

These measures give information on cost output elasticity when all outputs vary proportionally. This represents, however, only a special case, which may be inadequate to determine the competitive viability of banks in the real market. Actually, the product mix varies considerably when the size increases in the Italian banking industry, as shown in Table 5.

Note that the estimated cost function satisfies all the regulatory conditions when evaluated at the means of the sample (see footnotes 12 and 19). As a Chow type test indicates, the null hypothesis for the linear homogeneity in input prices is accepted (at the 5% level), i.e., the imposed restrictions of homogeneity do not distort the estimates given by the unrestricted specification. Since the model has already imposed and checked symmetry and homogeneity (equations 7), the regularity conditions are satisfied if the cost function is monotonically increasing and concave in input prices.
### Table 3
**Translog Cost Function Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Translog cost equation estimates</th>
<th>t-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.037631 *</td>
<td>1.894</td>
</tr>
<tr>
<td>b1</td>
<td>0.349132 ***</td>
<td>3.960</td>
</tr>
<tr>
<td>b2</td>
<td>0.042175 **</td>
<td>2.450</td>
</tr>
<tr>
<td>b3</td>
<td>0.156340 *</td>
<td>1.830</td>
</tr>
<tr>
<td>c1</td>
<td>0.776040 ***</td>
<td>121.445</td>
</tr>
<tr>
<td>c2</td>
<td>0.223960 ***</td>
<td>35.048</td>
</tr>
<tr>
<td>¼ d11</td>
<td>0.021268</td>
<td>-1.358</td>
</tr>
<tr>
<td>¼ d22</td>
<td>0.004080</td>
<td>-0.874</td>
</tr>
<tr>
<td>½ d33</td>
<td>-0.165920</td>
<td>-0.921</td>
</tr>
<tr>
<td>d12</td>
<td>0.024622</td>
<td>0.428</td>
</tr>
<tr>
<td>d13</td>
<td>0.294492</td>
<td>0.968</td>
</tr>
<tr>
<td>d23</td>
<td>0.035971</td>
<td>0.613</td>
</tr>
<tr>
<td>½ e11</td>
<td>0.050886 ***</td>
<td>25.899</td>
</tr>
<tr>
<td>½ e22</td>
<td>0.050886 ***</td>
<td>25.899</td>
</tr>
<tr>
<td>e12</td>
<td>-0.101772 ***</td>
<td>-25.899</td>
</tr>
<tr>
<td>f11</td>
<td>0.011670</td>
<td>0.482</td>
</tr>
<tr>
<td>f12</td>
<td>-0.011670</td>
<td>-0.482</td>
</tr>
<tr>
<td>f21</td>
<td>0.010106 **</td>
<td>2.275</td>
</tr>
<tr>
<td>f22</td>
<td>-0.010106 **</td>
<td>-2.275</td>
</tr>
<tr>
<td>f31</td>
<td>0.037427</td>
<td>1.590</td>
</tr>
<tr>
<td>f32</td>
<td>-0.037427</td>
<td>-1.590</td>
</tr>
<tr>
<td>g</td>
<td>0.422045 ***</td>
<td>10.963</td>
</tr>
<tr>
<td>h1</td>
<td>0.084473</td>
<td>0.751</td>
</tr>
<tr>
<td>h2</td>
<td>-0.053264 **</td>
<td>-2.252</td>
</tr>
<tr>
<td>h3</td>
<td>-0.006693</td>
<td>-0.063</td>
</tr>
<tr>
<td>k1</td>
<td>-0.054948 ***</td>
<td>-5.133</td>
</tr>
<tr>
<td>k2</td>
<td>0.054948 ***</td>
<td>5.133</td>
</tr>
<tr>
<td>½ l</td>
<td>0.005109</td>
<td>0.163</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Labour share equation estimates</th>
<th>t-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>0.776040 ***</td>
<td>121.445</td>
</tr>
<tr>
<td>c11</td>
<td>0.101772 ***</td>
<td>25.899</td>
</tr>
<tr>
<td>c12</td>
<td>-0.101772 ***</td>
<td>-25.899</td>
</tr>
<tr>
<td>f11</td>
<td>0.011670</td>
<td>0.482</td>
</tr>
<tr>
<td>f12</td>
<td>0.001016 **</td>
<td>2.275</td>
</tr>
<tr>
<td>f21</td>
<td>0.010106 **</td>
<td>1.590</td>
</tr>
<tr>
<td>f31</td>
<td>-0.054948 ***</td>
<td>-5.133</td>
</tr>
<tr>
<td>k1</td>
<td>-0.054948 ***</td>
<td>-5.133</td>
</tr>
</tbody>
</table>

Adjusted $R^2$ : 99.81

d.f. : 441

* *** and ** mean statistically significant at the 10 %, 5 %, and 1 % level, respectively, for a two-tailed test.*
Table 4
Ray Scale Economies and Augmented Economies of Scale

<table>
<thead>
<tr>
<th>Size class (assets in billions lira)</th>
<th>ESC</th>
<th>ESCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 less than 300</td>
<td>0.52462</td>
<td>0.94721</td>
</tr>
<tr>
<td>2 300 – 600</td>
<td>0.55506</td>
<td>0.94692</td>
</tr>
<tr>
<td>3 600 – 1200</td>
<td>0.54078</td>
<td>0.94665</td>
</tr>
<tr>
<td>4 1200 – 2500</td>
<td>0.53497</td>
<td>0.94622</td>
</tr>
<tr>
<td>5 2500 – 5000</td>
<td>0.53645</td>
<td>0.94187</td>
</tr>
<tr>
<td>6 5000 – 20000</td>
<td>0.50939</td>
<td>0.93968</td>
</tr>
<tr>
<td>7 more than 20000</td>
<td>0.49804</td>
<td>0.92796</td>
</tr>
<tr>
<td>whole sample (&quot;typical&quot; bank)</td>
<td>0.54765</td>
<td>0.93388</td>
</tr>
</tbody>
</table>

Auxiliary regression branch-output

| Intercept                          | 0.117839 *** |
| in Y1                              | 0.821151 *** |
| in Y2                              | -0.117221 ***|
| in Y3                              | 0.211208 *** |
| Adjusted R²                         | 0.9027      |
| F 712.45                           |             |

Note: ESCA have also been computed for the seven classes by doing auxiliary regression (11) on each class. The results for ESCA do not differ significantly from those ones presented here and using an auxiliary regression for the whole sample.

*** statistically significant at the 1 % level.

6.4. Expansion path scale economies

Expansion path scale economies, EPSCE, is a means for evaluating scale economies of banks that vary in scale and product mix simultaneously. An augmented indicator, EPSCA, is also developed in order to take account of the induced change in the number of branches as outputs expand. The two measures are calculated with equations (25) and (26) respectively, and are reported in Table 6.32

---

32 EPSCE (A,B) and EPSCA (A,B) are indicated for the mean bank B, since they measure the elasticity of incremental cost with respect to incremental output along the segment AB.
Table 5  
Comparison of the Product Mix at the Eight Mean Points  

<table>
<thead>
<tr>
<th>Class</th>
<th>Output Y1 (base = 100)</th>
<th>Output Y2 (Y1 = 100)</th>
<th>Output Y3 (Y1 = 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 less that 300</td>
<td>100</td>
<td>3</td>
<td>53</td>
</tr>
<tr>
<td>2 300-600</td>
<td>100</td>
<td>11</td>
<td>60</td>
</tr>
<tr>
<td>3 600-1200</td>
<td>100</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>4 1200-2500</td>
<td>100</td>
<td>11</td>
<td>66</td>
</tr>
<tr>
<td>5 2500-5000</td>
<td>100</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>6 5000-20000</td>
<td>100</td>
<td>25</td>
<td>80</td>
</tr>
<tr>
<td>7 more than 20000</td>
<td>100</td>
<td>58</td>
<td>89</td>
</tr>
<tr>
<td>All</td>
<td>100</td>
<td>42</td>
<td>83</td>
</tr>
</tbody>
</table>

As already noted by Berger et al. (1987), EPSCE results are "much more variable and erratic" than ray scale economies. The same observation applies to EPSCA. EPSCA is more relevant than EPSCE for the same reasons ESCA is preferred to ESC. With this more relevant indicator, the results are then very different since EPSCA measures indicate that economies of scale are exhausted at a small size, and that the cost function exhibits diseconomies of scale for the five larger classes.

6.5. Economies of scope  

Conventional scope economies measures are rejected for methodological and theoretical reasons, as explained in Section 4. It was however straightforward to calculate a cost complementary indicator from the estimated cost function for the three pairs of product. The presence of cost complementarity between the three pairs of products is a sufficient condition for global economies of scope. The results are nevertheless not conclusive since no cost complementarity has been found for the three pairs of products and at the eight average points. The indicator \( CC_n \) obtained using equation (19) is found to be positive in all cases. At the typical point, the indicator \( CC_n \) is reduced to \( d_n + b_1 b_2 \) (equation 19), and thus \( CC_{12} = 0.0409, CC_{13} = 0.3625, \) and \( CC_{23} = 0.0442.^{33} \) No conclusion can be drawn in terms of economies of scope since cost complementarity is a sufficient condition, not a necessary one. Moreover, economies of scope may be misinterpreted as

---

33 Since all the cost complementarity indicators are positive, whatever the size, they are not reported for the sake of brevity.
Table 6
Expansion Path Scale Economies

<table>
<thead>
<tr>
<th>Class</th>
<th>EPSCE</th>
<th>EPSCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 less than 300</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2 300 – 600</td>
<td>0.55980</td>
<td>0.92385</td>
</tr>
<tr>
<td>3 600 – 1200</td>
<td>0.81748</td>
<td>1.31470</td>
</tr>
<tr>
<td>4 1200 – 2500</td>
<td>0.74953</td>
<td>1.21039</td>
</tr>
<tr>
<td>5 2500 – 5000</td>
<td>0.78489</td>
<td>1.21093</td>
</tr>
<tr>
<td>6 5000 – 20000</td>
<td>0.60846</td>
<td>1.05966</td>
</tr>
<tr>
<td>7 more than 20000</td>
<td>0.56374</td>
<td>1.02584</td>
</tr>
</tbody>
</table>

Economies of scale because of the use of aggregated outputs. Kim (1986) has demonstrated that the problem of the validity of a consistent output aggregate remains unsolved whatever the level of aggregation.

6.6. Expansion path subadditivity

The subadditivity based measures are more significant than conventional measures of scale economies. Expansion path subadditivity measure, EPSUB, is given by equation (27), and the results are reported in table 7. For banks in the three largest classes, dividing output amongst two smaller firms, one located at the mean of the previous size class, reduces costs slightly. These results are consistent with EPSCA, showing a cost disadvantage for large banks. The results for other measures of EPSUB, when associated with EPSCA, are not as easily interpretable. The cost function is found to be expansion path subadditive between representative banks of class 2 and class 3, and between representative banks of class 3 and 4, implying a cost advantage for the larger bank (B) on a combination of the smaller one (A) and (A). EPSUB could lead one to give conclusion opposed to that of EPSCA, which shows diseconomies of scale on segment AB. It should be remembered however that they measure different features of the cost function. The variation in scale and scope are not the same with these measures. This apparent contradiction in findings is found again for the representative bank

---

34 EPSUB (A,B) defined in (27) is indicated for average bank B. In the estimation of TC for bank A, the average unit cost of capital and labour were chosen, and the number of branches was estimated with (11). I also computed EPSUB by using the average for each size class unit cost of capital and labour. This has not led to fundamentally different results (same sign and similar magnitude).
of class 2, but with inverse conclusions (economies of scale with EPSCA, disadvantage of larger bank with EPSUB).

7. Conclusions

This paper is an attempt to explore cost characteristics in the Italian banking industry. A multiproduct translog cost function is estimated on a representative cross-sectional sample of 1991 bank data. Several cost measures are compared and confronted. Concerning these measures, two major crucial points are emphasised. First, cost measures must take into account the change in the organisational form as scale and/or scope of bank production is modified. Consequently, measures including the induced change of the number of branches are preferred. Second, the concept of cost subadditivity is of major importance in the analysis of a multiproduct industry. Two indicators, expansion path scale economies and expansion path subadditivity, are derived from this concept, and provide information on the competitive viability of banks that vary in scale and product mix simultaneously. They are therefore preferred to the conventional measures of scale and scope economies. They have the further advantage of being computed at observable points, and are thus expected to shed more light on the real banking market.

Considerable caution must be exercised in interpreting the results, notably because of the use of accounting data whose economic significance is always suspicious, and because of the arguable approximations of some variables. Acknowledging these shortcomings, I refrain from declaring definitive conclusions on Italian banking cost structure. The results can nevertheless be compared to those of previous studies since the latter used similar data and variables.

The conventional measures of scale economies are in line with the previous findings since presence of ray scale economies is found for all sizes. The use of more relevant tools leads however to different results. From expansion path scale and expansion path subadditivity measures, it appears that the banks in the three largest size classes would not benefit from cost advantages resulting from scale and scope. This result conflicts with the general implications of previous Italian studies, but is in line with most of other European ones.

This study has also shed some light on the limits of cost banking studies. Two extensions would need to be undertaken. One concerns more directly empirical work, the other the theory of the banking firm.

First, any bank cost study is subject to the following strong criticism: even with good data and a proper methodology, we would be unable to detect real scale and scope effects since these effects may be completely hidden by
Table 7
Expansion Path Subadditivity

<table>
<thead>
<tr>
<th>Class</th>
<th>EPSUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 less than 300</td>
<td>-</td>
</tr>
<tr>
<td>2 300-600</td>
<td>-0.05942</td>
</tr>
<tr>
<td>3 600-1200</td>
<td>+0.10838</td>
</tr>
<tr>
<td>4 1200-2500</td>
<td>+0.05174</td>
</tr>
<tr>
<td>5 2500-5000</td>
<td>-0.03116</td>
</tr>
<tr>
<td>6 5000-20000</td>
<td>-0.04475</td>
</tr>
<tr>
<td>7 more than 20000</td>
<td>-0.00002</td>
</tr>
</tbody>
</table>

inefficiencies. This means a rejection of the approach analysing scale and scope economies by employing cost functions where all banks are assumed to be approximately equally efficient. More fundamentally, the essential assumption underlying the whole literature, i.e., the cost minimisation behaviour of banks, is rejected. Actually, these inefficiencies are often reported in descriptions of a specific banking system, such as the Italian one, but are also often underlined as a more general feature of banking markets. Why are there so many inefficiencies in banking especially? Several answers are advanced, for instance, regulatory features specific to banking, or imperfections of markets inherent in the banking activity itself (Colwell and Davies, 1992). Whatever the reasons, any empirical research should use another method able to cope better with the presence of inefficiencies. One possibility is to determine a frontier production function, not in a stochastic way, but in a deterministic way by keeping only the most relatively efficient banks to draw this frontier.

The second necessary extension is related to the theory of the banking firm. How should one measure banking production? What does a bank

---

35 See Berger and Humphrey's article "The dominance of inefficiencies over scale and product mix economies in banking" (1991). Two kind of inefficiencies are usually distinguished, technical inefficiencies due to an overuse of all inputs, and allocative inefficiencies due to an improper mix of inputs.

36 Other variants exist. This recent work focuses on supposedly inefficient multiproduct firms, such as public services, hospitals, and financial institutions. For an overview of this growing literature using parametric or non-parametric frontier analysis in banking, see Colwell and Davies (1992). For a recent survey of the non parametric approach, see Holvad and Hougaard (1993).
produce? What are the inputs? As noted in section 2, these basic questions are still waiting for definitive answers. Any analysis based on profit maximisation requires prior answers to these questions. In their survey of existing rival approaches, Colwell and Davies (1992: 126) conclude that one approach would be to suggest that banking output and productivity are cases of "measurement without theory". This would suggest empirical work should be given a lower priority than development of the theory of financial intermediation and its application to output. Such a theory should coherently link the services of the financial sector and thus offer consistent measures of output and productivity.

This constitutes an ambitious program for further research.

Appendix
Survey of cost function analyses of the Italian banking sector

The previous scarce and less advanced Italian studies are not of great interest. For a survey of this older literature, see Landi (1990: 100-5), and Conigliani (1990: 99-106).

Authors: Lanciotti and Raganelli (1988)

Approach: Aggregate output index, 2 inputs, only the price of labour is specific to each bank.
Estimation: Cobb-Douglas function with Box-Cox transformation of dependent and independent variables. The dependent variable is not the total cost, but the average cost. Control variables: branches, concentration index.
Results: Average cost-output elasticity. Average cost-branch elasticity. Presence of economies of scale.

Authors: Cossutta et al. (1988)

Approach: Intermediation approach, 6 outputs, 3 inputs, but input prices for labour and physical capital are ignored on the grounds they are assumed identical for all banks. An input price for financial capital is calculated. Translog function without factor share equations.
Estimation: Translog not fully specified (missing variables). Control variable: branches. The cost function is decreasing in three outputs.
Results: 7 size classes. Ray scale economies. Product-specific economies of scale for 4 outputs and for 2 sets of 3 outputs (problem of approximation of 0-output level by the minimum observed value). Evidence of economies of scale for all sizes, especially when the number of branches is maintained constant. Product-specific economies of scope for 4 outputs and for 2 sets of 2 outputs (problem of 0 output level). Presence of economies of scope in most cases, particularly for the largest banks.

Author: Parigi (1989)

Approach: Production approach, 4 outputs, 2 inputs.
Estimation: Translog function with factor share equations. Control variable: branches. The cost function is decreasing in two outputs.

Results: 3 size classes. Presence of scale economies for the 2 smallest sizes (method not clearly specified). Tests on the null hypothesis of absence of economies of scope (cost complementarity). The hypothesis is rejected.

Authors: Baldini and Landi (1990)

Data: Balance sheet data, 294 (and a reduced sample of 288 for verification of the stability of results) credit institutions for 1987.
Approach: Production approach, 3 outputs. Input prices are ignored.
Estimation: Translog cost function without factor share equations. Necessary restrictions are omitted. Control variable: branches. The cost function is decreasing in one output.

Results: 7 size classes. Ray economies of scale and augmented (firm) economies of scale. Presence of economies of scale for all sizes, more particularly for the small ones. Cost complementarity for the three pairs of products: no clear evidence.

Author: Landi (1990)

Approach: Production approach, 3 outputs, Input prices are ignored.
Estimation: Translog cost function without, factor share equations. Translog function is not correctly specified (absence of certain variables). Necessary restrictions are omitted. Control variable: branches. The cost function is decreasing in one output.

Results: Same as above (Baldini and Landi), 2 indicators of product-specific economies of scope for 2 sets of 2 outputs. Presence of economies of scope if the separated production requires the double of branches, but not if the total number of branches is maintained constant, that is, in the hypothesis there would be a repartition of existing branches between the two sets of outputs. Calculation for each bank of a relative efficiency indicator (derived from the difference between
the actual operating cost and the estimated one).

**Authors: Conigliani et al. (1991)**

*Data:* Bank of Italy data, two samples from different sources of respectively 256 and 233 credit institutions for 1987.

*Approach:* Intermediation approach, 6 outputs, 3 inputs.

*Estimation:* Translog cost function with factor, share equations. Control variable: branches.

*Results:* 5 size classes. Ray scale economies and augmented scale economies. Presence of economies of scale for all sizes. Product-specific economies of scope for the six outputs (problem of the approximation of zero-output level by the minimum observed value). No evidence of global economies of scope, only some clear presence of economies of scope for few products and for specific size.

**Authors: Cardani et al. (1991)**

*Data:* Balance sheet data, 94 credit institutions for 1986.

*Approach:* Intermediation approach, but considering a single aggregate, output index, 3 inputs.

*Estimation:* Translog cost function without factor share equations. The necessary restrictions on the parameters are omitted. Control variable: branches (not specified correctly in the translog form).

*Results:* 5 size classes, Ray economies of scale. Presence of economies of scale for the smallest class. Economies of scale are totally excluded for the largest banks. Calculation of indices of global inefficiency.

**References**


Bell, F. W. and Murphy, N. B. (1968), *Costs in Commercial Banking: A Quantita-


CONIGLIANI, C. (1990), La concentrazione bancaria in Italia, Il Mulino.


METU STUDIES IN DEVELOPMENT

LANDI, A. (1990), Dimensioni, costi e profitti delle banche italiane, Il Mulino.


OECD (1992), Banks under Stress, Paris: OECD.


**Özet**

İtalyan bankacılık sektöründe ölçek ve kapsam ekonomileri

Bu çalışmada İtalyan bankacılık sektörünün maliyet özellikleri araştırılmıştır. 1991 yılında sektörde faaliyet gösteren bankaların büyük bir kısmına ilişkin veriler kullanılarak çok-ürünli translog maliyet fonksiyonu tahmin edilmiş ve değişik maliyet unsurları incelenmiştir.