The Journal of Productivity Analysis, 2, 127–142 (1991)

Technical Efficiency of Norwegian Banks: The Non-Parametric Approach to Efficiency Measurement*

SIGBJØRN ATLE BERG
Research Department, Bank of Norway, P.O. Box 1179, Sentrum, N-0107, Oslo, Norway

FINN R. FØRSUND
Department of Economics, University of Oslo, P.O. Box 1095, Blindern N-0317, Oslo, Norway

EILEV S. JANSEN
Research Department, Bank of Norway, P.O. Box 1179, Sentrum, N-0107, Oslo, Norway

Abstract

The nonparametric frontier methodology is applied to a sample of banks, where output levels are measured either by the number of accounts and their average size, or by the total balances of the accounts. The efficiency rankings of individual banks are found to depend substantially on our choice of output metric, whereas the estimated size of potential productivity improvements in the banking sector are less affected. The results on economies of scale are also largely unchanged.

Keywords: Bank outputs, data envelopment analysis, efficiency gains, economies of scale, efficiency ranking.

1. Introduction

The main contribution of this paper is empirical in nature. We apply a nonparametric frontier methodology to a sample of banks, and measure output levels in two alternative ways. We seek to determine whether our choice of output metric is important with respect to three applications of frontier analysis: We analyze the efficiency distributions of banks by considering how much total productivity in the banking sector could be improved, and we investigate the existence of economies of scale in banking. Both issues present important information for the design of public policy. We also consider the efficiency ranking of individual banks, which is an important micro application of frontier analysis.

The large number of average cost studies of banking firms have focused on scale and scope economies assuming all banks to be efficient. These studies have adopted different ways of measuring bank output. The main alternatives have been to measure output either by the number of accounts and their average size, or by the total balances of assets and liabilities. The choice between these alternatives does not seem to be essential for the qualitative characteristics of the cost functions estimated. See, for instance, the surveys by Kolari and Zardkoohi [1987] and by Clark [1988].

We are aware of few previous studies of bank efficiency frontiers; Sherman and Gold [1985] and Parkan [1987] have analyzed bank branches, whereas Rangan et al. [1988], Charnes et al. [1989, 1990], Ferrier and Lovell [1990], and Berger and Humphrey [1990a, 1990b] have analyzed banking firms.

*The refereeing process of this paper was handled through S. Grosskopf.

It is interesting to note that bank output is measured in different ways in most of these previous papers. For instance, Sherman and Gold used the number of transactions performed, Rangan et al. used total balances of deposits and loans, and Ferrier and Lovell used the number of accounts and their average size. We shall use the two latter alternatives, which are also the main alternatives from the average cost literature, and compare the results generated.

We adopt the same method of analysis as most previous bank efficiency studies: data envelopment analysis (DEA). DEA is a linear programming technique for constructing a nonparametric, piecewise linear envelope to the observed data. See e.g., Charnes and Cooper [1985] for an excellent presentation of the DEA method as developed since 1978 by these authors and associates. We briefly present and discuss DEA in Section 2 of the paper. In Section 3 we discuss the modelling of bank production, with special emphasis on alternative ways of measuring bank output. Section 4 presents the data, while Section 5 characterizes the efficiency frontiers under the two alternative ways of measuring output. We offer a few concluding remarks in Section 6.

2. Methodology

The general problem when measuring efficiency of micro units is to establish a benchmark. The seminal article by Farrell [1957] used a convex hull in the input coefficient space, assuming constant returns to scale, as the nonparametric best practice reference technology. Data envelopment analysis (DEA) represents a generalization of the Farrell approach because other assumptions than constant returns to scale can be accommodated within a convex piecewise linear best practice technology (see e.g., Grosskopf [1986]).

When making a choice between parametric frontiers (pioneered by Aigner and Chu [1968], and applied to banking cost functions in Ferrier and Lovell [1990], Berger and Humphrey [1990a, 1990b]) and nonparametric frontiers, one should bear in mind that the functional forms used for parametric frontiers are at best approximations to underlying production functions. As remarked by Afriat [1972], the properties of the functions "are not deliberate empirical hypotheses, but are accidental to technical convenience of the functions." The Farrell or DEA approach of fitting facets as close as possible to the observations seems more appropriate when our knowledge of underlying technologies is weak. It is also a technique that can be extended quite straightforwardly to multiple outputs, and that does not rely on price information for dual cost functions as in the parametric frontier cases referred to above.

Measures of efficiency are based on the distance of an observation from the frontier. This distance can in principle be measured in a number of ways, but are conventionally restricted to either the horizontal or the vertical direction in the output-input space. Measuring horizontally is here understood to mean that observed input usage is compared to the input bundle, with observed input ratios, needed with frontier technology at observed output levels. Measuring vertically means that observed outputs are compared with potential outputs at the frontier for observed inputs, keeping the relative composition of outputs as observed. For constant returns to scale these measures are of course identical. When the reference technology exhibits variable returns to scale, a need for a measure of scale efficiency arises. The benchmark is input coefficients at optimal scale. These can either be compared with observed input coefficients or coefficients adjusted for technical inefficiency in the two ways pointed out above (see Førsund and Hjalmarsson [1979]). This point is further elaborated below.

It does not matter for the definition of the efficiency measures whether the reference technology is parametric or piecewise linear. In the DEA approach the efficiency scores are obtained directly, together with information on the reference technology. We have chosen to concentrate on the input saving efficiency measure due to the expressed interest in the banking sector of reducing costs. The DEA technique finds the input saving efficiency measure of a production unit by minimizing the distance from the observed point to the linear combination of best practice units along the factor ray of observed input proportions keeping outputs constant. Thus a linear programming problem is solved for each unit. Assuming k inputs, m outputs and a sample of n units, the formal problem can conveniently be stated in the following way:

$$\text{Min } E_{1j} \tag{1}$$

s.t.

$$Yz_j \geq y_j \tag{1a}$$

$$Xz_j \leq E_{1j}x_j \tag{1b}$$

$$z_j \geq 0 \tag{1c}$$

where

E_{1j} is the input saving efficiency measure for unit j,
Y is the mxn matrix of outputs from all units,
y_j is the $mx1$ vector of outputs from unit j,
X is the kxn matrix of inputs for all units,
x_j is the $kx1$ vector of inputs for unit j,
z_j is the $1xn$ vector of intensity weights defining the linear combination of best practice units to be compared with unit j.

The inequality (1a) states that the observed outputs must be less or equal to a linear combination of outputs of the best practice reference units. The next set of inequalities state that the use of inputs at the linear combination of reference units must be less or equal to the use of inputs of unit j adjusted to efficient operation.

As stated, problem (1) implies that the reference technology with output set $P(y) = \{y: Yz_j \geq y_j, Xz_j \leq x_j, z_j \geq 0, j = 1, \ldots, n\}$ is restricted to constant returns to scale. By introducing restrictions on the sum of intensity weights, DEA can accommodate either nonincreasing returns to scale or variable unrestricted returns to scale (see Grosskopf [1986]):

$$z_jI \leq 1 \text{ (nonincreasing returns to scale)} \tag{2}$$

$$z_j I = 1 \text{ (variable returns to scale)} \quad (3)$$

where I is the sum vector. The type of reference technology assumed is of course important for the efficiency distributions obtained.

When allowing for variable returns to scale, measures of scale efficiency may be calculated as described above. A convenient procedure is to establish the technical efficiency measures both with constant and with variable returns as the reference technology, i.e., solving problem (1) both without and with the constraint (3). With respect to input saving efficiency, the gross scale efficiency measure for the VRS technology is simply the efficiency measure for the CRS technology. Scale efficiency corrected for technical efficiency is the ratio between the two measures. Further clarification and interpretation follow below in connection with Figure 1.

Scale inefficiency is due to either decreasing or increasing returns to scale. To determine which case occurs, one can simply sum the weights z_j from the CRS technology problem (assuming unique solutions). Total weights exceeding one indicate decreasing returns, whereas weights totalling to less than one indicate increasing returns to scale.[1]

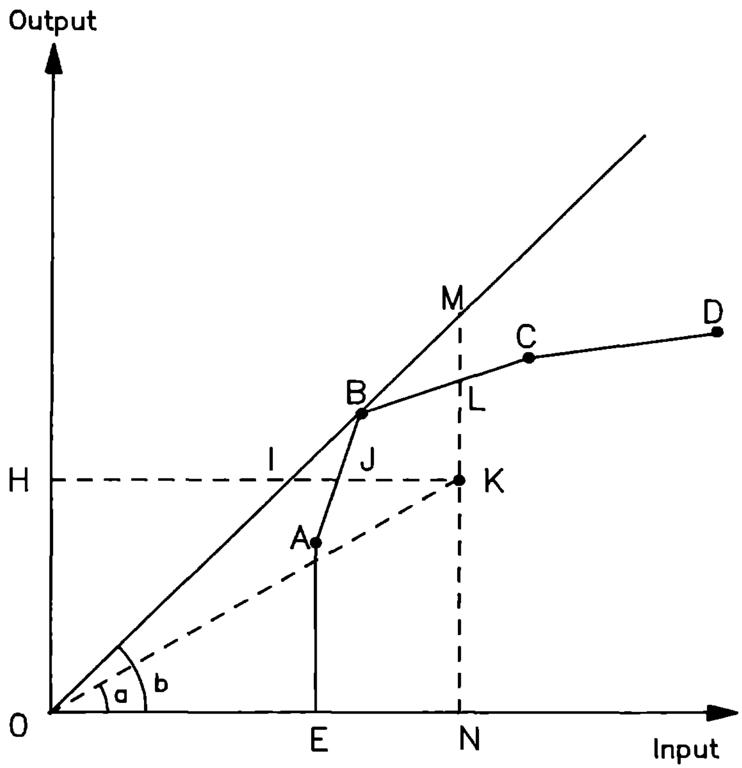

Figure 1. DEA-frontiers and efficiency measures.

To illustrate how the various measures are calculated, we adapt Figure 1 from Førsund and Hjalmarsson [1979] (parametric case), Färe et al. [1983] and Rangan et al. [1988] (non-parametric case). The best practice frontier with variable returns to scale [constraint (3) applies] is EABCD, the frontier with non-increasing returns [constraint (2) applies] is OBCD, while the constant returns to scale frontier is the ray through B. Input saving efficiency E_1 for the unit at K is measured as HJ/HK when the reference technology allows for variable returns to scale, and as HI/HK when constant returns are imposed.

The gross scale efficiency measure, E_3, for unit K compares the observed input coefficients with those obtained at the optimal scale (scale elasticity equal to one). In Førsund and Hjalmarsson [1979] it is shown that this can be calculated as a/b. By simple geometry we see from Figure 1 that this measure is the same as HI/HK, i.e., the input saving measure E_1 with reference to the CRS technology. This property generalizes to multiple outputs and inputs. Correcting for technical inefficiency in the input direction, observation K is moved to point J on the frontier. The corrected scale measure, E_4, is then HI/HJ. Clearly, $E_4 = E_3/E_1$, where both measures now refer to the VRS technology.

Note the interpretation of the scale efficiency measures.[2] An E_3 or E_4 efficiency score of say 0.8 means that the input coefficients at optimal scale and with frontier technology are 80 percent of the observed input coefficients (E_3) or of the coefficients corrected for technical inefficiency (E_4), when retaining the observed output and input ratios.[3]

When solving problem (1) for the unit K in Figure 1, unit B will clearly be the only one with a positive weight z in the reference set. The reference point is I and the weight attributed to unit B is thus OI/OB, which is less than one. This implies that the elasticity of scale for unit K, both evaluated at point J and at point K, must be greater than one. This interpretation of the sum of weights z in the CRS case generalizes to multiple outputs and inputs.

By the nature of placing the linear faceted convex *lid* over the observations, increasing returns can only be experienced from the start of the output range. There must be at least one unit at the other end of the size range that experiences constant returns. Very efficient large units could lead to *most* of the reference set exhibiting constant returns, while very efficient medium sized units divide the reference set into small units exhibiting increasing returns and large units exhibiting decreasing returns to scale.

When the reference technology allows for variable returns to scale, units operating with input and output quantities sufficiently far from other units at both ends of the size distribution can be identified as fully efficient simply for lack of other comparable units. Also, increasing the number of units in the sample should lead to decreasing average levels of efficiency due to the positive probability of including more efficient outliers. One should keep in mind that efficiency estimates are relative measures, conditional on the sample actually at hand.

3. Modelling Bank Production

The banking literature is divided on whether deposits should be treated as inputs or outputs. See, for instance, the thorough discussion by Berger and Humphrey [1990b]. The DEA approach implicitly views banking as a number of (interdependent) activities, each producing a *technical output* in the sense of Sealey and Lindley [1977]. Within that analytical framework, deposits should be treated as outputs simply because they represent a resource-consuming activity. This corresponds to what is sometimes called the *value added approach* to bank modelling, see, e.g., Fixler [1988] and Berger and Humphrey [1990b].

In accordance with this view, we specify five banking activities, namely supplying demand and time deposit services, short and long term loan services and other services (mainly brokerage services, property management and the provision of safe deposit boxes). These activities entail operating costs in terms of four inputs, namely labor, machines, materials and buildings.

The costs incurred depend on the activity levels. The banking literature does not provide a clear answer to how these levels should be measured. Although treating deposits as outputs is often associated with measuring activity levels by the number of accounts (see, e.g., Berger et al. [1987]), total balances seem equally reasonable (see, e.g., Kim [1989]). We have already noted that the previous studies of bank efficiency have used a number of different output measures.

We shall investigate two of these alternatives, by carrying out parallel analyses, measuring deposit and loan activities both by the number of accounts and their average size, and by the total balances. These are also the two main alternatives used in the average cost literature. Prominent examples are Benston et al. [1983] who use the number of accounts and their average size, and Gilligan and Smirlock [1984] who use total balances.

In both cases the efficiency concept will be the technical efficiency measure, but the exact meaning of efficiency will differ in the two cases. If activity levels are measured by the number of accounts and their average size, the efficiency concept will be close to pure operating efficiency. If we measure by total balances, the efficiency concept will be broader and include the bank's ability to influence the sizes of deposit and loan accounts.

The levels of other services will in both cases be measured by the income generated from these services. The implicit assumptions are that income differences represent differences in volume, i.e., equal service prices are charged by all banks, and that prices are appropriate weights for aggregation of different services, i.e., the relative prices of different services reflect their relative costs. The first of these assumptions is probably closer to reality than the second one. An imperfect measure of the activity level still seems preferable to ignoring this activity.

The quantity of inputs will also be measured in the same way in both cases. We use expenditure data from the annual accounts, see below.

4. Data

Our data are the primary data from the official Norwegian Bank Statistics for 1985. That year has been chosen because it is the only year for which data on the number of accounts are available. Such data were collected from 121 of the total sample of 218 Norwegian banks, subsidiaries of foreign banks not included. Among these 121 banks, 14 had given information we found to be obviously incorrect. Data on the number of accounts are thus available from 107 banks, whereas data on total balances are available from all 218 banks. Notice that bank branching is permitted in Norway, and that most banks operate several branches.

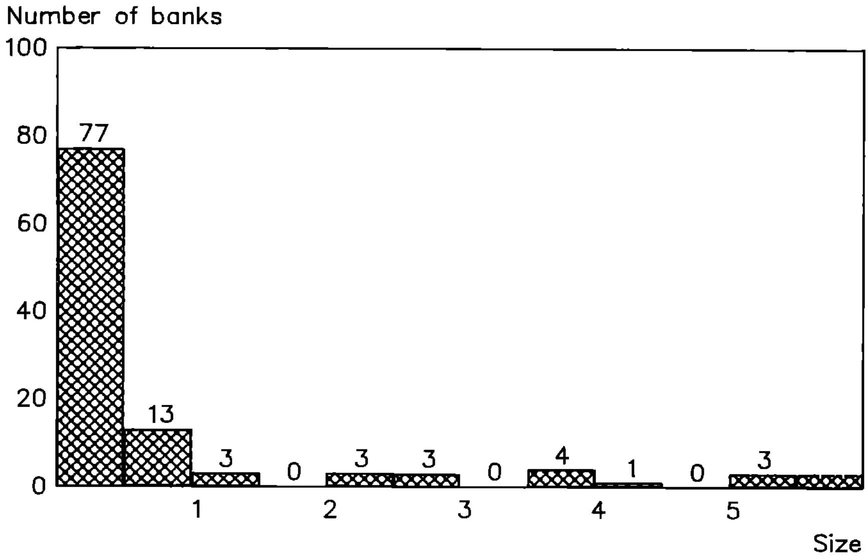

Figure 2. Banks by size (total loans in billions NOK).

Figure 2 shows that most of the 107 banks are quite small; in fact the median bank has only approximately 200 million NOK (30 million USD) in total loans. Only 17 banks have total loans in excess of 1 billion NOK (150 million USD). This means that data envelopment analysis based on a variable returns to scale technology, is likely to find most large banks efficient (see the discussion in Section 3). The sample of 107 banks is unbiased in the sense that the relative size distribution remains about the same when all 218 banks are included. The latter distribution is therefore not shown.

The definitions of our output and input measures are given in Table 1. In principle we should measure inputs in physical units, but that has not been possible. We have to make do with expenditures, implicitly assuming equal input prices for all banks. We believe that this assumption is quite close to the realities of Norwegian banking. Labor remunerations are regulated by national wage agreements, and machines and materials are easily transportable

Table 1. The production activities of banks. Bank statistics codes in parentheses, see Central Bureau of Statistics of Norway [1987].

A. Activities:	1. Demand deposits. (2011+2012+2111)
	2. Time deposits. (2112+2114)
	3. Short-term loans. (161+162+163+164+165)
	4. Long-term loans. (166)
	5. Other services. (431+433)
B. Inputs:	1. Labor. (3211+3212+3213)
	2. Machines. (3311+3312+3330)
	3. Materials. (334+335+336)
	4. Buildings. (3321+3322+3411+3412)

Table 2. The correlation between alternative measures of activity levels.

	No. of accounts vs Average size	No. of accounts vs NOK volume	NOK volume vs Average size
Demand deposits	0.02	0.97	0.10
Time deposits	−0.12	0.99	−0.01
Short-term loans	0.33	0.94	0.42
Long-term loans	0.20	0.99	0.26

goods where the law of one price should approximately apply. However, building services are measured by maintenance costs plus depreciation allowances or rental costs. This user cost is the best measure available, even though it is not very likely to be a correct measure of building services. We have still chosen to retain buildings as a specified input.

Table 2 is based on the sample of 107 banks. It shows for each of the deposit and loan activities the correlation between the proposed measures of activity levels. Coefficients above 0.16 are statistically significant at the 5% test level. Notice that the average size of accounts is only weakly correlated with the number of accounts or the total balances of accounts, whereas the number of accounts and total balances are highly correlated. The latter finding is consistent with the standard result that average cost functions do not depend much on which set of output measures we adopt.

5. Technical Efficiency Scores

In this section we solve for each bank in the sample of 107 the linear programming problem (1) with and without the additional constraint (3) from Section 3, for each of the two alternative ways of measuring activity levels. Under the first alternative a bank is found to be inefficient if we are able to construct a reference bank as a linear combination of other banks, such that the reference bank has at least as many accounts and at least the same average size of accounts, while using fewer inputs than the real bank. Under the second alternative a bank is found to be inefficient if the reference bank has at least the same total balances of each type of deposit and loan accounts, while using fewer inputs. In both cases the reference bank must supply at least the same amount of other services as the real bank.

As an example consider two banks from the sample, with the activity levels and input quantities listed in Table 3. The input vectors of the banks are approximately equal. Consider first the case where activity levels are measured in value. Bank A's deposit and loan activities are greater than those of bank B, whereas bank B produces more of other services. It is still evident that bank A is the more efficient. The VRS efficiency scores generated by data envelopment analysis are 1.00 and 0.78, respectively. Bank A is one of six banks, whose linear combination constitutes the reference bank relative to bank B.

However, consider also the case where activity levels are measured by the number of accounts, and their average size. The table shows that bank B has more and smaller accounts of all four types than bank A. Bank A is no longer close to dominating bank B. In fact both banks now obtain a VRS efficiency score equal to 1.00.

Table 3. Data from two savings banks.

Inputs in 1000 NOK.

	Labor	Machines	Materials	Buildings
Bank A	2472	1617	946	767
Bank B	2266	1664	1288	676

Activity levels measured in 1000 NOK.

	Demand deposits	Time deposits	Short-term loans	Long-term loans	Other services
Bank A	235927	66104	54687	126213	1
Bank B	109147	24397	33724	98795	27

Activity levels measured by the number of accounts, and the average size of accounts (NOK).

	Demand deposits	Time deposits	Short-term loans	Long-term loans
Bank A				
- number	9434	1890	1080	1998
- size	25008	34976	50636	63169
Bank B				
- number	10836	2197	1354	2070
- size	10073	11104	24907	47727

The example illustrates that the efficiency scores obtained may be sensitive to our measurement of activity levels. We proceed to explore the differences between the two cases.

5.1. Potential Efficiency Gains

In Figure 3 total loans are cumulated according to efficiency scores. Consider first the case with variable returns to scale. When we measure activity levels by the number of accounts, banks with 84% of total loans are found to be fully efficient, as compared to 86% when we measure by total balances. The potential efficiency gains seem small. The picture changes drastically, however, if we impose constant returns to scale. In that case only 4% or 11%, respectively, of total loans are given in fully efficient banks. The potential savings now seem very important.

The reason for the difference in conclusions is that large banks are all fully efficient in the variable returns to scale case. This in turn is due to the fact that there are so few large banks and that they differ widely in size. Reference banks of comparable size are therefore hard to construct. In the constant returns to scale case large banks are compared with proportional blow-ups of smaller banks, and very few large banks are then found to be efficient.

For our purpose, however, the important conclusion is that the size of the potential input saving does not depend much on how the activity levels are measured.

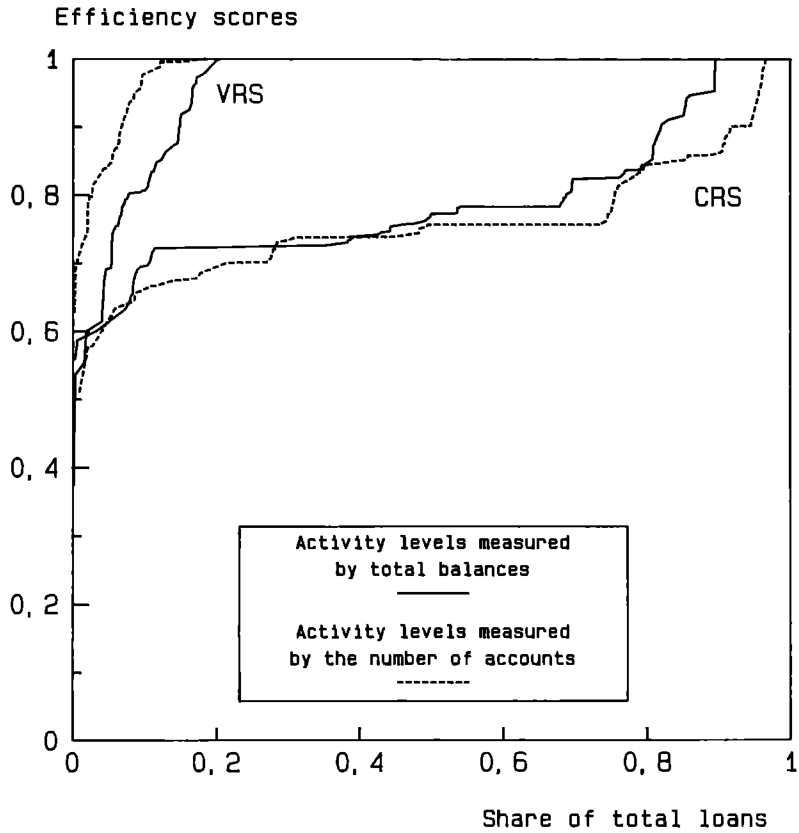

Figure 3. The distribution of efficiency scores against share of total loans.

We have not investigated in depth how the efficiency scores correlate with profitability. But as a piece of anecdotal evidence, consider two banks which met with serious economic problems in 1988. These banks both obtain efficiency scores of 1.00 however activity levels are measured. This indicates that inefficiency was not the main reason for their economic problems. We should notice, however, that giving loans without using resources to determine the creditworthiness of the customers is an easy way to become efficient within our analytical framework. We note also that the two banks are comparatively large, with total loans between 2 and 3 billion NOK (300–500 million USD). As mentioned above, very few banks of that size are found to be inefficient in our sample.

5.2. Economies of Scale

Economies of scale at individual banks can be identified in the way explained in Section 3 above. The results from the sample of 107 banks are summarized in Figure 4, for the

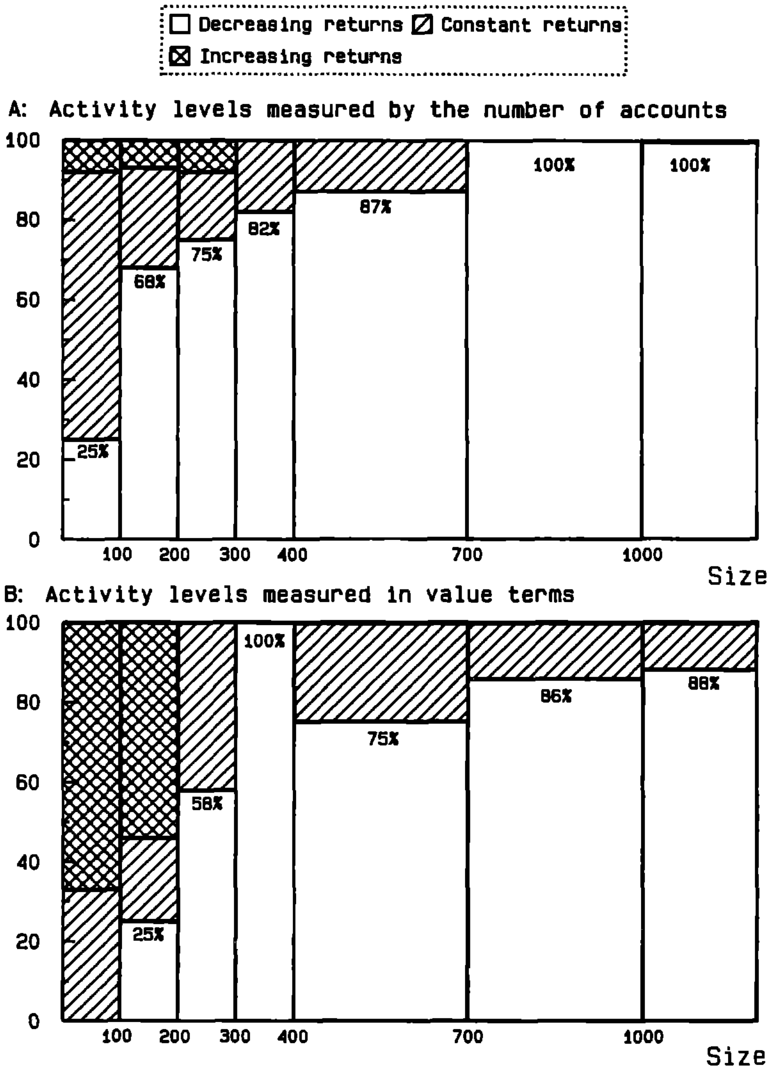

Figure 4. Returns to scale by size (total loans in million NOK).

two alternative measures of activity levels. When activity levels are measured by the number of accounts, only 5 small banks experience increasing returns to scale. 28 banks have constant returns, and the majority have decreasing returns. All banks with total loans in excess of 1 billion NOK (150 million USD) show decreasing returns. The largest bank with increasing returns to scale has total loans of 270 million NOK (40 million USD).

In the alternative case where activity levels are measured by total balances, the results are not as extreme. Half of all banks are still found to experience decreasing returns to scale. Increasing returns are only found among banks with less than 200 million NOK

(30 million USD) in total loans, whereas constant returns are found even among the largest banks.

As pointed out in Section 3, the convexity of the frontier ensures that increasing returns will be more frequent at smaller banks. That result can be inferred from Rangan et al. [1988]. Ferrier and Lovell [1990] found that 85–90% of all banks (except the smallest size class) exhibit increasing returns. But they did not find nonincreasing returns at all large banks, as we do.

Lack of scale economies beyond a certain size is also the standard result in conventional average cost studies. Clark [1988] states in his survey of that literature that "only two studies, however, find significant overall economies of scale above $100 million of deposits." Recent studies based on samples of large banks (see, e.g., Shaffer and David [1986], Shaffer [1988] and Noulas et al. [1990]) find economies of scale at much larger banks. The critical size found apparently depends on the size distribution of the sample. This could possibly be explained by the fact that larger banks are involved in more activities. The specified activities thus tend to be less representative of their true activities. In the present study we have tried to alleviate this problem by introducing other services than deposits and loans as a separate banking activity, however this did not alter the standard conclusions on economies of scale.

For our present purpose the important conclusion is that the two alternative ways of measuring output generate about the same conclusion.

5.3. Efficiency Rankings

So far we have found that the distribution of efficiency scores does not differ much in the two cases. But the picture changes when we turn to the efficiency rankings.

First, the two approaches do not identify the same number of efficient banks. When activity levels are measured by total balances 51 banks are found to be fully efficient, but that number increases to 66 when activity levels are measured by the number of accounts and their average size Further, 3 of the original 51 banks are no longer found to be efficient. Second, the two approaches do not identify the same banks as being the least efficient. Among the 20 banks with the lowest efficiency scores in the two cases, only 7 are the same.

The correlation between the rankings of the banks under the two alternatives is 0.69, which is certainly statistically significant. The correlation coefficient still confirms that the rankings do depend on the way we measure activity levels. If we impose constant returns to scale, the correlation between the two rankings becomes weaker, namely 0.59. The correlation between the two pairs of rankings is illustrated in Figures 5A and 5B.

5.4. Increasing the Sample Size

We proceed to consider how efficiency scores are affected by expanding the sample from 107 banks where data on the number of accounts are available, to 218 banks where data only permit measurement by total balances. In the VRS case the share of fully efficient

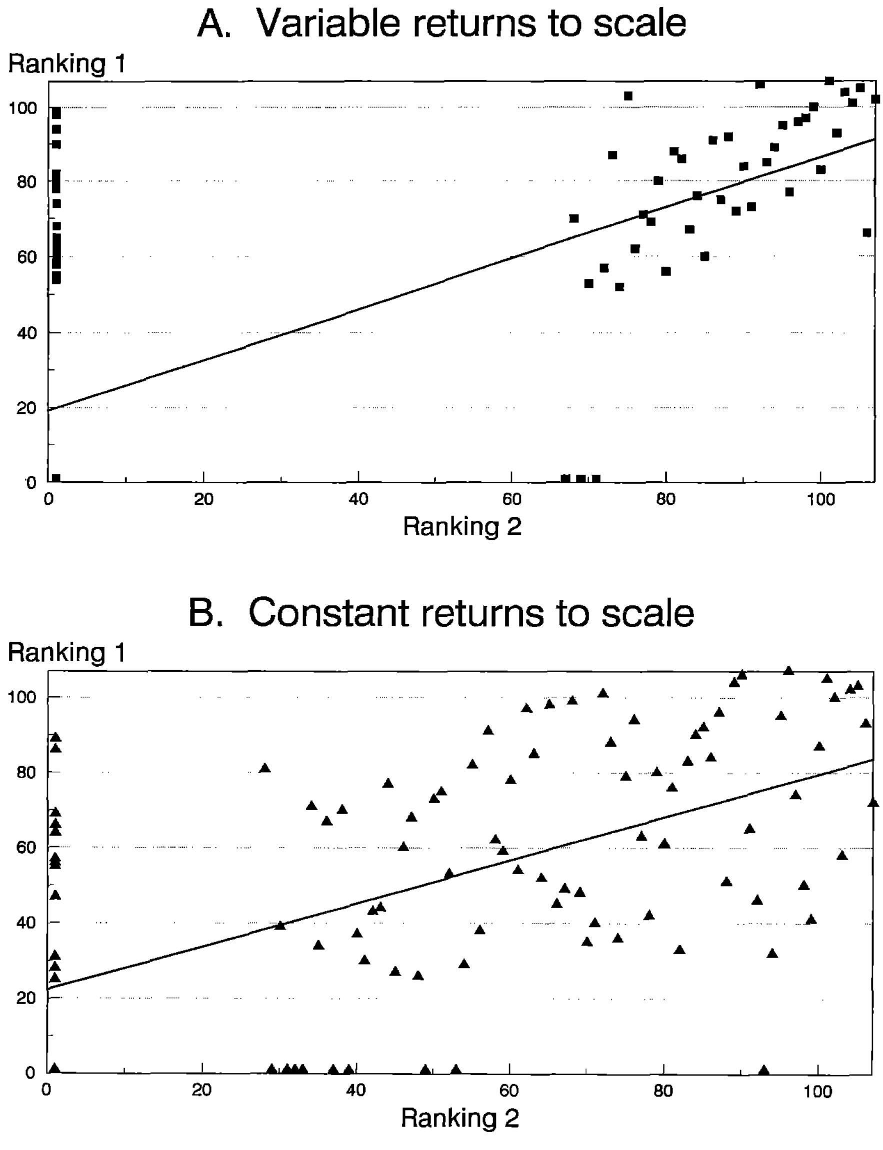

Figure 5. Efficiency rankings of individual banks when bank output is measured by: 1. the number of accounts and their average size; 2. total balances.

banks now decreases from 86% to 73% of total loans. This still implies that the potential efficiency gains are small.

The average efficiency score drops to 0.81 when the sample is expanded to 218 banks. This can be compared to the average score of 0.72 reported by Rangan et al. [1988] in their sample of 215 American banks (where output is also measured by total balances). The higher average in our sample indicates that the efficiency spread is lower. This might perhaps be taken as an indication that competition is a bit more severe in the Norwegian banking market than in the segment of the American market analyzed by Rangan et al. [1988].

The number of fully efficient banks increases modestly from 51 to 59, but only half of the efficient banks from the small sample remain efficient in the larger sample. These differences are as we should expect.

Increasing the sample size also changes the ranking of the original 107 banks. The rank correlation coefficient is 0.73. Considering the 20 banks with the lowest efficiency scores, we find that only 12 of these banks are the same on both rankings. From this we conclude that rankings within a subsample may change significantly when the size of the total sample is increased. The change in efficiency scores is most pronounced for the class of medium sized banks where the number of banks increases the most.

6. Concluding Remarks

In this study we have found that important characteristics of the efficiency frontier for Norwegian banking are about the same whether we choose to measure output by the number of accounts and their average size or by the total balances of the accounts. This applies to the size of potential efficiency gains as well as to our results on economies of scale.

The Norwegian banking market may be about to become more competitive, both because of deregulation of the national banks, and because of entry by foreign banks. Assuming constant returns to scale, we find that the efficiency gains from increased competition may be substantial. Assuming variable returns, the estimated gains are more modest. A realistic estimate is somewhere in between, which would still be substantial. The development of a more competitive market should therefore be encouraged.

However, the banking sector is also about to become more concentrated, through mergers and acquisitions. We do not find evidence of cost savings from increased bank size. There is thus a case for restrictions on the concentration process in order to retain a sufficient number of independent banks on the market.

While the efficiency frontier is not much affected, the efficiency rankings of individual banks depend heavily on how we choose to measure bank output. As mentioned in Section 4, the interpretation of the efficiency concept will be different for different sets of output measures. Before undertaking efficiency studies, one should therefore take great care to clarify what kind of efficiency one wants to measure, and define output measures accordingly.

The study also reveals the sensitivity of data envelopment analysis to sample size. The rankings within a subsample were shown to change significantly when the total sample size was expanded.

Our findings on large banks may indicate that even the larger sample of 218 banks is too thin in that end of the size distribution: Large banks are found to be very efficient and to exhibit decreasing returns to scale. These two findings may be closely connected: If with a variable returns to scale technology large banks have been found to be efficient merely for lack of a truly efficient reference set, their true inefficiency will in the DEA framework show up as decreasing returns to scale. This may very well be the case in our sample. It illustrates that assumptions about technology may be quite important for the results obtained from thin data sets.

Notes

Useful comments from Rolf Färe and from three anonymous referees are gratefully acknowledged.

1. Rangan et al. [1988] state that one has to solve a third problem, namely (1) with constraint (2), to determine the nature of the scale property. This approach yields consistent inferences even when solutions are not unique, but is unnecessary in case of uniqueness.
2. Rangan et al. [1988] wrongly state that $1 - E_4$ (E_4 being defined as above) shows the proportional reduction in input usage which could occur if the observation K were scale efficient. Their reference point I is in fact outside the feasible technology. When moving to the feasible point B, which exhibits constant returns to scale, both inputs and outputs have to change.
3. An output increasing efficiency measure E_2 is equal to NK/NL and NK/NM, respectively, for the two technologies, see Førsund and Hjalmarsson [1979]. When correcting for inefficiency in the output direction the measure $E_5 = E_3/E_2$ corresponds to E_4, with observed input and output ratios retained.

References

Afriat, S.N. (1972). "Efficiency Estimation of Production Functions." *International Economic Review* 13, 568–598.

Aigner, D.J. and S.F. Chu. (1968). "On Estimating the Industry Production Function." *American Economic Review* 58, 826–839.

Benston, G.J., A.N. Berger, G.A. Hanweck, and D.B. Humphrey. (1983). "Economies of Scale and Scope in Banking." In *Proceedings from a Conference on Bank Structure and Competition.* Chicago: Federal Reserve Bank of Chicago.

Berger, A.N., G.A. Hanweck, and D.B. Humphrey. (1987). "Competitive Viability in Banking," *Journal of Monetary Economics* 20, 501–520.

Berger, A.N. and D.B. Humphrey. (1990a). *The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking.* Finance and Economics in Discussion Series, No. 107. Washington D.C., Board of Governors of the Federal Reserve System.

Berger, A.N. and D.B. Humphrey. (1990b). *Measurement and Efficiency Issues in Commercial Banking.* Paper presented at a NBER Conference on Research on Income and Wealth, Charleston, South Carolina.

Central Bureau of Statistics of Norway (1987). *Credit Market Statistics. Private and Public Banks 1985.* Oslo.

Charnes, A. and W.W. Cooper. (1985). "Preface to Topics in Data Envelopment Analysis." *Annals of Operations Research* 2, 59–94.

Charnes, A., W.W. Cooper, Z.M. Huang, and D.B. Sun. (1989). *DEA Cone-Ratio for use in Developing Decision Support Systems to Monitor Performance in a Collection of Banks.* Paper presented at a Conference on New Uses of DEA in Management, Austin, Texas.

Charnes, A., W.W. Cooper, Z.M. Huang, and D.B. Sun. (1990). "Polyhedral Cone-Ratio DEA Models with an Illustrative Application to Large Commercial Banks." *Journal of Econometrics* 46, 73–91.

Clark, J.A. (1988). "Economies of Scale and Scope At Depository Financial Institutions: A Review of the Literature." *Federal Reserve Bank of Kansas City Economic Review,* 16–33.

Färe, R., S. Grosskopf, and C.A.K. Lovell. (1983). "The Structure of Technical Efficiency," *Scandinavian Journal of Economics* 85, 181–190.

Farrell, M.J. (1957). "The Measurement of Productive Efficiency." *Journal of the Statistical Society,* Series A, General 120, Part 3, 253–281.

Ferrier, G.D. and C.A.K. Lovell. (1990). "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence." *Journal of Econometrics* 46, 229–245.

Fixler, D.J. (1988). *A Commercial Bank Output Index.* Working paper No. 179. Washington D.C., Bureau of Labor Statistics.

Førsund, F.R. and L. Hjalmarsson. (1979). "Generalized Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants." *Economic Journal* 89, 294–315.

Førsund, F.R. and L. Hjalmarsson. (1987). *Analyses of Industrial Structure: A Putty-Clay Approach.* Stockholm: Industrial Institute for Economic and Social Research.

Gilligan, T.W. and M.L. Smirlock. (1984). "An Empirical Study of Joint Production and Scale Economies in Commercial Banking." *Journal of Banking and Finance* 8, 67–77.

Grosskopf, S. (1986). "The Role of the Reference Technology in Measuring Productive Efficiency." *Economic Journal* 96, 499–513.

Hancock, D. (1986). "A Model of the Financial Firm with Imperfect Asset and Deposit Elasticities." *Journal of Banking and Finance* 10, 37–54.

Kim, M. (1989). "Total Factor Productivity Growth in Banking: The Israeli Banking Sector 1979–1982." *Journal of Productivity Analysis* 1, 139–153.

Kolari, J. and A. Zardkoohi. (1987). *Bank Costs, Structure, and Performance*. Lexington, MA: Lexington Books.

Noulas, A.G., S.C. Ray, and S.M. Miller. (1990). "Returns to Scale and Input Substitution for Large U.S. Banks." *Journal of Money, Credit and Banking* 22, 94–108.

Parkan, C. (1987). "Measuring the Efficiency of Service Operations: An Application to Bank Branches." *Engineering Costs and Production Economics* 12, 237–242.

Rangan, N., R. Grabowski, H.Y. Aly, and C. Pasurka. (1988). "The Technical Efficiency of US Banks." *Economics Letters* 28, 169–175.

Schmidt, P. (1985). "Frontier Production Functions." *Econometric Reviews* 4, 289–328.

Sealey, C.W. and J.T. Lindley. (1977). "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions." *Journal of Finance* 32, 1251–1266.

Shaffer. S. (1988). *A Revenue-Restricted Cost Study of 100 Large Banks.* Research Paper No. 8806. Federal Reserve Bank of New York.

Shaffer, S. and E. David. (1986). *Economies of Superscale and Interstate Expansion.* Research Paper No. 8612. Federal Reserve Bank of New York.

Sherman, H.D. and F. Gold. (1985). "Bank Branch Operating Efficiency. Evaluation with Data Envelopment Analysis." *Journal of Banking and Finance* 9, 297–315.