ELSEVIER

Long- and short-run non-parametric cost frontier efficiency: An application to Spanish savings banks ☆

Diego Prior

Dpt. Economía de la Empresa, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Abstract

The utilisation rate of installed capacity is a popular concept in both the performance appraisal literature and in publications on industrial organisation. However, a common consensus has yet to be reached concerning the most appropriate way to measure the capacity utilisation of physical inputs and its final effect on company results. On the one hand, there are approaches that establish capacity utilisation with reference to the maximum level of production that can be achieved. In contrast, there are other approaches more strictly related to economic analysis of operating costs. In this paper, our main objective is to define an analytical process that uses non-parametric frontier methodology to provide the distance between the total costs of a given unit and the short-run frontier costs. As a natural extension of this proposal, it is possible to compute the short-run inefficiency caused by a non-optimal dimension of the fixed inputs: we define this as *capacity efficiency*.

The proposed evaluation process is applied to Spanish savings banks covering the period between 1986 and 1995. Throughout the period analysed, the greater part of cost inefficiency is due to capacity efficiency.
© 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

Studies focusing on the evaluation of managerial performance, whatever formal method they may use, usually conclude by stating what the potential reduction in costs or increase in profits would be, if managers were able to successfully rectify erroneous decisions from the past.

It is often assumed, without questioning real possibilities, that it is possible within each period to decide on the required technology, to adjust the inputs in accordance with their prices and to change the output-mix on the basis of the expected profits. However, from the point of view of operational decisions, it is commonly found that management face short-run rigidities limiting the decision making that would maximise profits. Here we refer to situations such as the presence of adjustment costs, administrative control or the intervention of external regulation. All these situations may limit the possibility to adjust inputs and may incorporate inefficiencies.

The purpose of this paper is to present a method of estimating the inefficiency related to existent capacity and the restrictions to adjust the levels of fixed inputs. The utilisation of installed capacity is a common concept in the literature on performance evaluation as well as in industrial organisation. Unfortunately, there is still no complete consensus on the most appropriate way to measure the capacity utilisation rate [1] and its effects on companies' profit and loss accounts.

In general, there are two main approaches to the concept of capacity utilisation:

(a) *Capacity as the maximum level of production in physical terms* (potential level of production that entirely uses the existent capacity).
(b) *Capacity as the desirable level of production in economic terms* (optimum amount of production at the lowest point of average total cost).

As discussed below, these two notions of capacity coincide when the reference technology exhibits increasing returns to scale. A problem arises when returns to scale have only a local significance and average cost curves are shaped in the well-known 'U' form. In this situation, the average total cost of the first approach is always higher than that corresponding to the second. This is the reason why, from on economic point of view, we advocate the second approach as a way of guaranteeing, under any circumstances, the optimal cost minimisation reference. More specifically, we propose a model that gives the degree of efficiency with respect to the short-run optimal total costs which need not necessarily coincide with the maximum level of production.

In the remainder of this paper, we first begin with a description of the most common notions of capacity utilisation. Then, Section 2 presents the frontier proposed model. Next, Sections 3 and 4 include the data and comment on the results obtained

---

[1] For instance, Cremeans (1978) explains how six industrial organisations in the USA, three public and three private, calculate variations in capacity utilisation. These indicators have rarely coincided, and occasionally even indicate clearly diverging trends.

by applying the proposed model to Spanish savings banks sector. Finally, we present a summary of our major findings in Section 5.

2. Concepts of capacity: Engineering and economic approaches

As indicated, there are two main notions of capacity output: a maximum level of output and an output level that minimises average costs.

The first notion was initially proposed by Johansen (1968):

> "Capacity is the maximum level of production per unit of time with the existing equipment and plant, assuming that the utilisation of the variable inputs is not restricted."

Färe et al. (1989) use this definition of maximum capacity to formulate frontier models establishing the capacity utilisation rate for multi-output firms. This notion makes economic sense when average cost has a negative slope due to increasing returns to scale.

However, it is unlikely that firms using their maximum capacity are operating with minimum production costs. When total costs are non-linear, the full utilisation of the existing capacity can contradict the general economic objectives of the firm. Indeed, cases arise in which a growth in activity leads to a fall in profits. This is obvious to Cremeans (1978), who states:

> "if physical measures are taken to establish economic objectives, such measures must contain economic concepts, otherwise they may be deceptive."

We may also consider cases described in Sunderland and Kane (1996) who refer to situations where increases in production generate disproportionate increases in costs, or increases in the volume of finished product stocks. In other cases, they refer to situations in which there is a fall in profits caused by outsourcing certain processes and thereby changing the size of the firm. In summary, there is sufficient evidence to suggest that improvements of physical outputs alone does not necessarily lead to improvement of economic capacity indicators.

However, exactly what the economic optimum output level is remains to be defined. In the economic literature we find two possible interpretations. The first, suggested by Klein (1960) and more recently by Seguerson and Squires (1990), postulates that the optimum production level is precisely at the tangency point between short- and long-run average cost curves. The second, found in Cassels (1937) and Hickman (1964), takes as a reference the output level corresponding to the minimum level of long-run average total costs. [2] The difference between these

---

[2] Klein criticised this notion, since it is of limited practical value given that, as several studies conclude, the average cost curve are L-shaped, and it is difficult to precisely distinguish the minimum point of the long-run average cost curve.
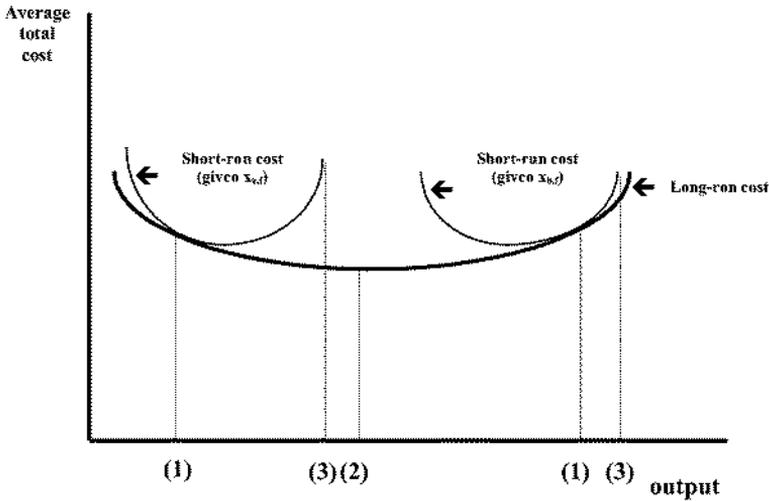
Fig. 1. Average cost curves in the short- and long-run: (1) tangency between short- and long-run average costs, (2) minimum long-run average costs, (3) maximum physical output given the fixed capacity.

two definitions, in the most general case of U-shaped average cost curves, is shown in Fig. 1.

The difference between the two definitions becomes clearer insofar as, in the long-run, the average total cost curve is U-shaped. However, in empirical applications, the decision to choose one definition or the other is of very little relevance if the coefficient of correlation between both is high. This is exactly the conclusion obtained by Nelson (1989) in his research on the capacity utilisation level in North American electricity generating plants. [3]

3. Short- and long-run non-parametric frontier models

As indicated in the previous section, when increasing returns to scale prevail and the average cost is diminishing, then the production level that minimises costs is also the one which maximises production. Hence the maximum production and minimum cost points coincide. However, increasing returns to scale cannot be taken for granted. For this reason, if the method to evaluate cost efficiency is to be general, it needs to be independent of a specific returns to scale assumption. This is the reason why the model of Färe et al. (1989) is not valid in general. Our proposal, by contrast,

---

[3] Nelson also finds that the majority of plants operate to the left of the short-run cost minimisation point and that the excess capacity increases production costs by between 8% and 10%. Furthermore, comparing this measurement with a technical capacity notion, he concludes that in all cases the latter approach overestimates the most reasonable levels of utilisation in economic terms.

does not require any specific return to scale case and works properly for both "L-shaped" and "U-shaped" average cost curves.

To specify the model we first describe the variables needed. Assume that for each of the k units of production to be evaluated, we know both the vector x_k of inputs consumed $[x_k = (x_{k,1}, \ldots, x_{k,J}) \in \Re_+^J]$ and the vector y_k of outputs $[y_k = (y_{k,1}, \ldots, y_{k,I}) \in \Re_+^I]$. It is also assumed that the production technology describing the transformation process of inputs into outputs is known and can be summarised by means of the following input requirement set:

$$L(y_k) = \{x_k : (y_k, x_k) \text{ is feasible}\}. \tag{1}$$

The input set $L(y_k)$ denotes the collection of all input vectors $x_k \in \Re_+^I$ that yield at least output vector $y_k \in \Re_+^I$. It provides a general representation of the technology in terms of input quantities and output quantities. No prices are involved, and no behavioural assumption is required. When input prices are available, and cost minimisation is a reasonable behavioural assumption, then it is possible to develop a price-dependent characterisation of technology.

Assume that prices (p_k) are known, and that inputs can be classified according to whether they are fixed and impossible to modify in the short-run $[x_{k,f} = (x_{k,1f}, \ldots, x_{k,Jf}) \in \Re_+^{Jf}]$, or variable and under the control of the company $[x_{k,v} = (x_{k,1v}, \ldots, x_{k,Jv}) \in \Re_+^{Jv}]$. The variable costs frontier provides a price-dependent characterisation of technology:

$$\mathrm{VC}(y_k, p_{k,v}, x_{k,f}) = \min_{x_v^*}\{p_{k,v} \cdot x_v^* | (x_v^*, x_{k,f}) \in L(y_k)\} \tag{2}$$

where $p_{k,v} = (p_{k,1v}, \ldots, p_{k,Jv}) \in \Re_+^{Jv}$ is the price vector of variable inputs for unit k and $x_v^* = (x_{1v}^*, \ldots, x_{Jv}^*) \in \Re_+^{Jv}$ is the input vector minimising variable costs. $\mathrm{VC}(y_k, p_{k,v}, x_{k,f})$ shows the minimum variable expenditures required to produce output vector y_k at variable input prices $p_{k,v}$ and with given fixed inputs level $x_{k,f}$.

Adding the cost of fixed inputs to Eq. (2) yields the short-run total cost frontier:

$$\mathrm{SRTC}(y_k, p_{k,v}, x_{k,f}) = \mathrm{VC}(y_k, p_{k,v}, x_{k,f}) + p_{k,f} \cdot x_{k,f}. \tag{3}$$

Thus we can define an indicator of short-run frontier efficiency, $\mathrm{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})$ as the ratio between the minimum short-run total cost $[\mathrm{SRTC}(y_k, p_{k,v}, x_{k,f})]$ and the observed total cost $(p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f})$ of the firm under evaluation:

$$\begin{aligned} \mathrm{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) &= \frac{\mathrm{SRTC}(y_k, p_{k,v}, x_{k,f})}{p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f}} \\ &= \frac{\mathrm{VC}(y_k, p_{k,v}, x_{k,f}) + p_{k,f} \cdot x_{k,f}}{p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f}} \leqslant 1. \end{aligned} \tag{4}$$

If $\mathrm{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) = 1$, then evaluated firm is operating at the best practice costs, given the existent level of fixed inputs. However, when $\mathrm{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) < 1$, then the firm is not part of the short-run cost frontier.

$[1 - \text{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})]$ indicates the proportional reduction in costs that could be obtained if it would operate on the cost-efficient frontier.

The minimum variable cost frontier for unit k $[\text{VC}(y_k, p_{k,v}, x_{k,f})]$ is obtained from the following program: [4]

$$\text{VC}(y_{k,i} p_{k,jv}, x_{k,jf}) = \min \sum_{jv=1}^{Jv} p_{k,jv} \cdot x_{jv}^*$$

subject to:

$$x_{jv}^* - \sum_{s=1}^{K} z_s \cdot x_{s,jv} \geqslant 0, \quad jv = 1v, \ldots, Jv,$$

$$x_{k,jf} - \sum_{s=1}^{K} z_s \cdot x_{s,jf} = 0, \quad jf = 1f, \ldots, Jf,$$

$$-y_{k,i} + \sum_{s=1}^{K} z_s \cdot y_{s,i} \geqslant 0, \quad i = 1, \ldots, I, \tag{5}$$

$$\sum_{s=1}^{K} z_s = 1,$$

$$z_s \geqslant 0, \quad s = 1, \ldots, K.$$

Program (5) assumes a variable returns to scale technology. It is straightforward to develop a constant returns to scale version of the above short-run cost frontier, but the constant returns to scale assumption is inevitable linked to a long-run perspective.

We now focus on the determination of the long-run frontier cost-efficiency ratio $\text{LREFF}(y_k, p_{k,v}, p_{k,f})$, which, unlike $\text{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})$, compares the long-run efficiency costs (adjusting the level of fixed inputs) and the observed cost of the unit under evaluation:

$$\text{LREFF}(y_k, p_{k,v}, p_{k,f}) = \frac{\text{LRTC}(y_k, p_{k,v}, p_{k,f})}{p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f}} \leqslant 1. \tag{6}$$

This ratio is computed by taking information from the optimum of the following cost minimisation program: [5]

---

[4] This program has some connections with the non-parametric cost minimisation program presented in and in Färe et al. (1994). The differences are, basically, three: the technology exhibits variable returns to scale, we are only interested in minimising variable costs rather total costs and, finally, we impose strict equality in fixed inputs restrictions. To our knowledge, there are other papers presenting a similar restriction in the fixed inputs, Färe et al. (1990)—in the short-run expenditure constrained profit maximisation—and Primont (1993)—in the short-run cost minimization—but their formulation is less restrictive since their objectives are different to ours.

[5] This expression is an extension, for the variable returns to scale case, of Färe et al. (1994) standard program of total costs minimisation.

$$\text{LRTC}\big(y_{k,i}p_{k,jv}, p_{k,jf}\big) = \min \left(\sum_{jv=1}^{Jv} p_{k,jv} \cdot x_{jv}^{*} + \sum_{jf=1}^{Jf} p_{k,jf} \cdot x_{jf}^{*} \right)$$

subject to:

$$
\begin{aligned}
& x_{jv}^{*} - \sum_{s=1}^{K} z_s \cdot x_{s,jv} \geqslant 0, \quad jv = 1v, \dots, Jv, \\
& x_{jf}^{*} - \sum_{s=1}^{K} z_s \cdot x_{s,jf} \geqslant 0, \quad jf = 1f, \dots, Jf, \\
& -y_{k,i} + \sum_{s=1}^{K} z_s \cdot y_{s,i} \geqslant 0, \quad i = 1, \dots, I, \\
& \sum_{s=1}^{K} z_s = 1, \\
& z_s \geqslant 0, \quad s = 1, \dots, K.
\end{aligned}
\tag{7}
$$

Having quantified the short and long-run levels of frontier efficiency, we can now determine capacity inefficiency (excess in costs as a result of inappropriate level in fixed inputs): [6]

$$
\begin{aligned}
\text{capacity efficiency} &= \frac{\text{LREFF}(y_k, p_{k,v}, p_{k,f})}{\text{SREFF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})} \\
&= \frac{\text{LRTC}(y_k, p_{k,v}, p_{k,f})}{\text{SRTC}(y_k, p_{k,v}, x_{k,f})} \\
&= \frac{\text{LRTC}(y_k, p_{k,v}, p_{k,f})}{\text{VC}(y_k, p_{k,v}, x_{k,f}) + \sum_{jf=1}^{Jf} p_{k,jf} \cdot x_{k,jf}} \leqslant 1.
\end{aligned}
\tag{8}
$$

Using these computational results, the short-run fixed inputs utilisation rate can be defined as the ratio between optimal and observed fixed inputs:

$$\frac{x_f^{*}}{x_{k,f}}.
\tag{9}$$

This coefficient indicates whether the level of fixed inputs is correct in the long-run ($=1$), whether an excess of under-utilised fixed inputs exists (< 1) or whether the over-utilised fixed inputs are below the level minimising long-run total costs (> 1).

The proposed evaluation process, assuming one output and one fixed input, is shown in Fig. 2. Unit k produces y_k outputs and incurs a total cost TC_k. Applying program (5), the short-run total cost (SRTC_k) appears which maintain the observed

---

[6] This economic measure of capacity utilisation, was developed by Berndt and Fuss (1986) in a seminal article treating the problem of multi-factor productivity measurement with quasi-fixed inputs. This Berndt–Fuss approach can be implemented using either traditional growth accounting practices or by parametric estimation using econometric techniques (see also Berndt and Hesse (1986) for a discussion of problems with capacity measures). Here we adapt this economic measure of capacity utilisation to the non-parametric approach to efficiency evaluation.
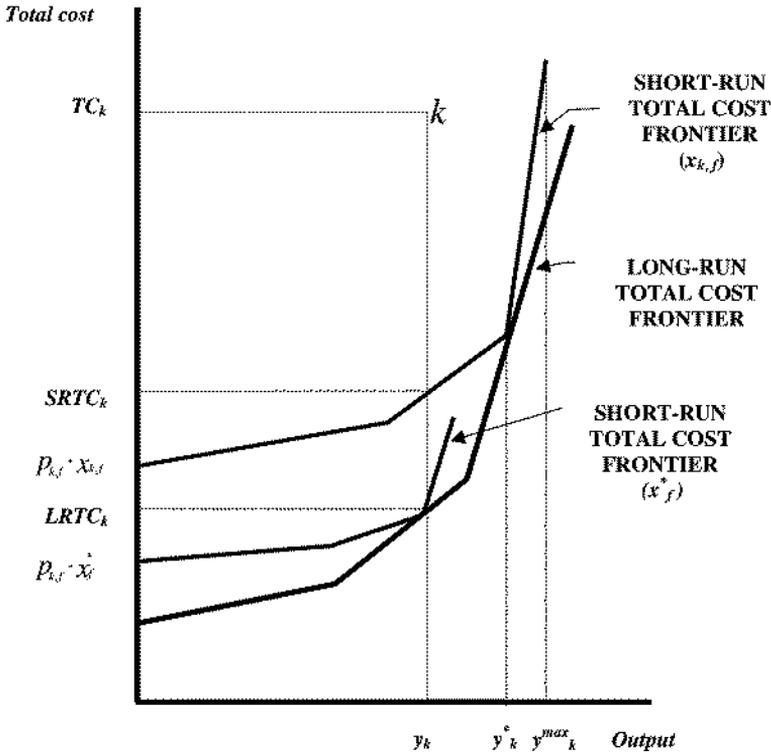
Fig. 2. Long- and short-run total cost efficiency.

fixed input $x_{k,f}$. Program (7) yields the unrestricted long-run frontier total cost LRTC_k, but requires the adjustment of the fixed input from $x_{k,f}$ to x_f^*. For unit k, the aforementioned frontier coefficients are:

$$\text{LREFF} = \frac{\text{LRTC}_k}{\text{TC}_k} < 1,$$
$$\text{SREFF} = \frac{\text{SRTC}_k}{\text{TC}_k} < 1,$$
$$\text{capacity efficiency} = \frac{\text{LRTC}_k}{\text{SRTC}_k} < 1,$$
$$\text{fixed input utilisation} = \frac{x_f^*}{x_{k,f}} < 1 (\text{under-utilisation}).$$

In the example presented in Fig. 2, the fixed input under-utilisation can be illustrated in two ways. First, by verifying that, when producing y_k, the optimal fixed input that minimises total costs (x_f^*) is lower than the real fixed input ($x_{k,f}$). Second, by observing that the real output level (y_k) is lower than that required to reach the tangency point between short- and long-run total cost frontiers (y_k^*). In general, the second comparison is applicable only when one output is produced. However, the proposed model works for multi-output technologies, and the short-run fixed

inputs utilisation rate can be deducted comparing optimal (x_f^*) and observed $(x_{k,f})$ fixed input levels.

Fig. 2 also illustrates the weakness of the technical capacity notion mentioned in Section 1. The maximum output notion of capacity (y_k^{max}) proposed in Färe et al. (1989) presents a higher average cost than both SRTC and LRTC cost frontier points. Therefore, the consistency of the maximum output notion with profit maximization or cost minimization behaviour is not granted in advance. In summary, care should be taken when evaluating managerial performance not to penalize rational behaviour.

4. Spanish savings banks: Data and variables in the sample

The proposed method is applied to a sample of Spanish savings banks in the period between 1986 and 1995. There are three types of banking institutions in Spain: private banks, savings banks and credit co-operatives. Savings banks concentrate on retail banking, providing savings and loan services to customers. Historically, Spanish savings banks were regulated by a number of constraints on their development. The main components of this regulation were the geographic restrictions on their operations, a more limited allowance to offer financial services than private banks and, finally, additional reserve requirements.

The deregulation of the Spanish savings bank industry increased in 1989 when the constraint on the territorial expansion was abolished. This process continued until savings banks were allowed to offer the same financial services as private banks. Nowadays, private banks and savings banks compete in a global market of financial services.

For different reasons, both the savings bank sector and the time period covered in the sample are relevant. First, in spite of the considerable merger activity that occurred, savings banks have very heterogeneous sizes. [7] Furthermore, they faced a considerable and significant reduction in financial results. [8] Finally, as a result of the deregulation of the sector, Spanish savings banks have followed a strategy of growth both with respect to product range and branch office network. [9] This strategy created a problem of excess capacity (in the period analysed the number of branches increased by 33%). In brief, we have a sector where the growth strategies with regard to size have led to a situation of capacity inefficiency, caused by an excess investment in physical capital, as illustrated in Table 1.

---

[7] In 1995, of a total of 50 savings banks, 9 could be classified as very small (with less than 100 branch offices), while 8 would be included in the subgroup of large savings banks (with a network of at least 600 branch offices, one savings bank having double this figure and another four times this figure).

[8] Common indicators are return on assets (ROA) and also a financial version of the coverage ratio (Intermediation Margin/Total Operating Costs). For the period analysed, both indicators show a fall.

[9] Grifell-Tatjé and Lovell (1996), in a dynamic analysis of the efficiency of savings banks between 1986 and 1991, confirm that the aggressive policy of opening new branch offices increased competition despite generating some excess capacity.

Table 1
Evolution of branch size in the period analysed

	Year		
	1986	1990	1995
Number of savings bank accounts	44,803,094	46,481,526	42,466,601
Number of branch offices	11,306	13,683	15,008
Number of ATMs	3059	9440	15,292
Accounts per branch office	3962.77	3397.02	2829.59
ATMs per branch office	0.27	0.68	1.01

Table 1 summarises the evolution of the total number of savings bank accounts, number of branch offices and number of ATMs. The important increase in the number of branches and in the number of ATMs is evident. The drop in the number of accounts by branch and the increase in the number of ATMs by branch is also clear. In short, this trend implies an increase in the quality of the service provided (being closer to customers and expanding the amount of services provided) and also illustrates how over-branching has evolved over time. [10]

There is a large number of recent articles—Domenech (1992), Pastor (1995) and Grifell-Tatjé and Lovell (1996), Kumbhakar and Lozano-Vivas (2001)—that apply non-parametric frontiers to evaluate the efficiency of Spanish savings banks. Although it shares the methodology, this paper is different. Compared with the work of Pastor (1995) and Grifell-Tatjé and Lovell (1996), we seek to determine the level of total cost efficiency and not only technical efficiency. Domenech (1992) also evaluates cost efficiency but no separation is made between fixed and variable inputs. Furthermore, this author follows the so-called *intermediation approach* that differs from ours.

Here we follow the so-called *production approach*, [11] since it complies closely to our essential objective: to investigate the causes that explain the variations in operating costs. More specifically, the analysis is focused on the study of service production as reflected in the total number of accounts and loans administered by each organisation. This service production requires the consumption of physical inputs, whose cost is recorded as operating costs in the profits and loss accounts.

Having defined the methodological approach followed, we focus attention on the structure of fixed and variable inputs in banking. The literature has not devoted an excessive amount of attention to this topic. This may reduce the reliability of the obtained results. As known precedents, one can mention Noulas et al. (1990), Berger et al. (1993) and Hunter and Timme (1995). However, there are important differences between these mentioned contributions and this paper. Firstly, these authors apply econometric cost or profit functions, and this demands the a priori specification of a particular function reflecting technological relationships whose existence is not

---

[10] As pointed out by a referee, Spain has the largest per capita number of branch offices in Europe.

[11] Survey of the various conceptual approaches can be found in Berger and Humphrey (1992).

guaranteed. Furthermore, they do not examine whether the existence of fixed inputs introduces inflexibilities separating short- and long-run efficient cost levels. [12]

Given these considerations, the variables selected to conduct the efficiency analysis are the following:

Outputs (y)
   y_1: Number of loans (physical variable)
   y_2: Number of current and savings accounts (physical variable)
   y_3: Number of fixed term savings accounts (physical variable)
   y_4: Service charges applied (monetary variable)

Variable inputs (x_{var})
   x_{1var}: Material consumption (monetary variable)
   x_{2var}: Staff (full time equivalent)

Variable costs (VC):
   VC_1: Material consumption ($VC_1 = x_{1var}$)
   VC_2: Labour cost ($x_{2var} \cdot p_{2var}$)

Fixed input (x_{fix})
   x_{1fix}: Number of branches [13,14]

Fixed cost (FC)
   FC_1: Depreciation and other operating expenses ($x_{1fix} \cdot p_{1fix}$)

These variables provide the real total cost of each firm:

$$\text{Total operating cost(TC)} = (+)\, VC_1 \, (\text{material consumption})$$
$$(+)\, VC_2 \, (\text{labour costs})$$
$$(+)\, FC_1 \, (\text{depreciation and other operating expenses}).$$

---

[12] However, Hunter and Timme (1995) explicitly acknowledge that physical capital investments in branch offices and other types of equipment show very little variability in the short-run and require high levels of adjustment costs and time to change. These authors also reaffirm that there are other factors unconnected with adjustment costs that may explain the *quasi-fixed* nature of certain inputs, and they point out various institutional inflexibilities (inflexible organisational structures, personnel, irreversibility or immobility of certain inputs, regulatory restrictions...) as causing situations of immobility in the short run.

[13] As noted by the referees, other fixed factors (home office, ATMs and other capital equipment) have been left out. We acknowledge this but consider that there are several reasons behind these choices. First, ATMs are not included given the impossibility of obtaining reliable information about the annual cost of maintaining the ATM network. Second, we could have taken fixed assets (in accounting terms) as the aggregate value of the physical capital invested, but the specific accounting regulation of the Spanish financial firms and the historical-cost accounting principle bias these calculations. To sum up, we have chose the most reasonable fixed input, expressed in physical units, in retail banking for the Spanish case.

[14] Branches have also elements of an output since they provide services to customers, but we ignore this possibility. To reinforce the role of branches as the variable representing the physical capital, we introduce the output variable 'service charges' as a proxy for these services. We believe that in the end this is a better output variable than the branch offices.

The output variables can also be presented in another way, closer to management accounting literature:

Intermediation margin $= (+)$ interest and returns on loans $(y_1 \cdot p_{y1})$

$(+)$ other interest income

$(+)$ service charges (y_4)

$(-)$ interest and charges paid to current and

savings accounts $(y_2 \cdot p_{y2})$

$(-)$ interest and charges paid to fixed term

savings accounts $(y_3 \cdot p_{y3})$

$(-)$ other interest expenses.

Having defined operating costs and the intermediation margin, it is possible to formulate the coverage ratio, a common metric in the banking literature:

$$\text{Coverage ratio} = \frac{\text{Intermediation margin}}{\text{Total operating cost}}.$$

The coverage ratio provides evidence for the ability to meet total operating cost from the normal intermediation margin. It helps to determine the company's ability to survive in the long run provided that the coverage ratio assumes values above unity.

Table 2 shows the descriptive statistics for these variables. As mentioned, the specification of variables follows the well-known production approach. The number of branches have also been previously considered fixed input by Hunter and Timme (1995). The consideration of staff as a variable input could be criticised, especially in 1986 when the Spanish labour market was still strongly regulated. To analyse the effect of changing the specification of the fixed inputs, we also ran program (5) with staff as a fixed input. We do not report these results for reasons of space, but it is fair to state that the impact of including staff does not affect the basic conclusions.

To compute program (5) one needs variable input prices. Given the limited information available, we assume a unitary price $(p_{1v} = 1)$ for material consumption. Thus, material consumption (x_{1v}) is assumed to equal current expenditures (VC_1). The price of labour input is deduced from labour cost (VC_2) and the number of staff (x_{2v}). Therefore, program (5) is solved by taking into account the specific average wage for each firm.

Program (7) requires the fixed input price (p_{1f}). This price is computed by dividing depreciation and other operating expenses (FC_1) by the number of branches (x_{1f}). This calculation provided the specific average cost per branch for each savings bank.

Table 2
Descriptive statistics of the variables

	Mean	Standard deviation	Maximum value	Minimum value
Year 1986				
Number of loans (y_1)	51612	73385	414713	1530
Number of current accounts (y_2)	508663	739670	5054716	7047
Number of fixed term saving accounts (y_3)	73195	87267	429019	332
Service charges applied (y_4)	341	795	6099	5
Variable costs (VC)	1119	1589	10655	26
Fixed costs (FC)	3705	5479	38065	93
Inter-mediation margin	8402	12992	81030	205
Coverage ratio	1.684	0.225	2.662	1.242
Year 1990				
Number of loans (y_1)	83109	120446	635050	1530
Number of current accounts (y_2)	638035	1013036	7220492	14725
Number of fixed term saving accounts (y_3)	88239	117409	798288	987
Service charges applied (y_4)	960	2019	13912	29
Variable costs (VC)	2439	4171	29900	77
Fixed costs (FC)	7026	12783	93757	178
Inter-mediation margin	15237	25481	161234	378
Coverage ratio	1.568	0.242	2.220	0.943
Year 1995				
Number of loans (y_1)	192415	208332	978985	3956
Number of current accounts (y_2)	722340	870828	4669976	17973
Number of fixed term saving accounts (y_3)	126992	141905	804622	6525
Service charges applied (y_4)	4272	8952	56603	72
Variable costs (VC)	5370	8083	48903	162
Fixed costs (FC)	14130	20836	124921	317
Inter-mediation margin	30436	43372	235741	669
Coverage ratio	1.551	0.196	2.190	1.165

5. Short- and long-run cost frontiers: Empirical results

Using the specified variables, we ran three times (for each year in the sample) the programs (5) and (7). [15] Table 3 presents a summary of the results. Observe that, in the years analysed, two different trends co-exist. Between 1986 and 1990, both the short- and long-run cost efficiency coefficients show a steady fall. Specifically, in 1989 the deregulation process was concluded and savings banks were increasing their branch network. In 1995, a clear growth in efficiency is observed. This is explained by the improvement in the relationship between the number of accounts and the level of

---

[15] Problems (5) and (7) are computed using GAMS. A copy of the models is available upon request.

Table 3
Average efficiency scores, optimal branch levels and coverage ratios

	1986	1990	1995
Average efficiency scores			
Long-run efficiency (LREFF)	76.55%	66.18%	81.01%
Short-run efficiency (SREFF)	93.11%	89.22%	94.25%
Capacity efficiency	81.84%	73.57%	85.47%
Capacity with respect to optimum value			
Branches$_k$ < Branches*[a]	9 (12%)	7 (11%)	3 (6%)
Branches$_k$ = Branches*[b]	23 (30%)	23 (37%)	15 (30%)
Branches$_k$ > Branches*[c]	45 (58%)	33 (52%)	32 (64%)
Coverage ratios			
Coverage ratio based on observed costs[d]	1.684	1.568	1.551
Coverage ratio adjusted to short-run efficiency	1.807	1.757	1.648
Coverage ratio adjusted to long-run efficiency	2.204	2.374	1.924

[a] Number of banks whose network of branches is below the optimum value.
[b] Number of banks whose network of branches coincides with the optimum.
[c] Number of banks whose network of branches is above the optimum value.
[d] (Intermediation margin/total operating cost.)

physical capital invested. It is also evident that, throughout the period analysed, the major part of cost inefficiency is explained by capacity efficiency (according to non-parametric Sign and Wilcoxon tests, all differences in the cost frontier coefficients are statistically significant). By way of example, we observe that in 1995 the possibilities for reducing cost inefficiency without altering the number of branches (SREFF) are certainly limited (less than 6%, 100–94.25). Since the most important cost inefficiencies are always related to capacity inefficiency, this situation significantly reduces the possibilities to improve efficiency in the short-run.

Another way of presenting the results is to re-compute the coverage ratio replacing observed costs with short- and long-run efficient costs. Its temporal variation clearly shows a tendency towards a smaller difference between intermediation margin and operating costs. There are two reasons for this reduction: (i) a fall in the intermediation margins (due to the growth of competition in interest rates for both asset side and debit side transactions) and (ii) a growth in operating costs. In summary, the average coverage ratio, based on observed costs, moves from 1.68 to 1.55 between 1986 and 1995. Adjusting for the inefficiencies, we observe substantial increases in the coverage ratio. [16] Table 3 also shows that the most common situation is the under-utilisation of branches. Otherwise stated, if total cost minimisation were the firms' main objective, the Spanish savings banks maintain a higher number of branches than required. By contrast, the number of cases with over-utilisation of branches reduce from nine in 1986 to only three in 1995.

---

[16] Non-parametric Sign and Wilcoxon tests reveal that the differences between the average coverage ratios and the adjusted coverage ratios are statistically significant.

These results present another interesting perspective. As previously mentioned, between 1989 and 1995 an important merger process has been observed (in total 17 mergers). Comparing the capacity efficiency coefficients before (1986) and after the mergers (1995), we compute that in only five cases capacity efficiency was worse in 1995 than in 1989. This situation clarifies the real possibilities for merger operations in the Spanish financial sector due to the adjustment of the branch network and the reallocation of physical capital invested.

To summarise, the cost frontier efficiency level of Spanish savings banks depends almost entirely on the adequate network of branches, and little can be done by solely adjusting the other variable inputs. [17] On the one hand, this situation provides interesting strengths to the companies analysed. On the other hand, it can be the source of important weakness in the future because of the potential future proliferation of the use of communication technology in banking operations. These results reflect the real importance of strategy and long-run efficiency improvements. Once the branch network has been adjusted, the efficiency in total costs can be maintained over the years. Although e-banking, at present, is not very well developed in Spain, there is a clear opportunity to compete with savings banks by offering more attractive interest rates to consumers, provided that e-banks could better control total operating costs without depending on branches.

6. Concluding remarks

The present study addresses both theoretical and applied objectives. From a theoretical point of view, the main interest has been to point out that the traditional formulation of cost minimisation models mainly quantifies long-run frontier efficiency levels. This is due to the general assumption that all inputs are variable (adjustable) in the short-run. When this is not the case, the practitioner should adapt a short-run perspective.

In practice, the most common technological situation is the presence of both variable and fixed inputs. In such a cases it is worth adapting the frontier methodology to determine the different inefficiencies found: first, the adjustable inefficiency in the short-run, second, long-run inefficiency and, finally, the capacity inefficiency caused by the presence of non-optimal levels of fixed inputs. This is the specific theoretical objective in the first part of the paper. Our proposal completes the traditional formulation of non-parametric production and cost frontier efficiency models by distinguishing two components: short-run and capacity efficiency.

---

[17] However, as pointed out by one referee, any approach to determine cost or branch efficiency is necessarily overstated since we only have information on bank costs and thereby neglect costs incurred by the bank customer. This overestimation could be corrected in two ways. First, combining bank and customer expenses and using this combined cost instead of the bank production cost. Second, defining a profit rather than a cost frontier to observe the trade-off between operating costs, financial costs and financial revenues.

From the applied point of view, our interest has been to analyse frontier efficiency levels in Spanish savings banks. The sector is of interest due to the special circumstances during the time period studied. As observed in Sections 3 and 4, between 1986 and 1995 the Spanish financial sector experienced a process of deregulation. This lead to a steady increase in the number of branches (especially in the initial years of deregulation), significantly reducing their level of capacity utilisation. In the final years of the sample period this process was halted and one observes a substantial increase in the number of operations per branch. The main findings of the application are that most cost inefficiencies are structural and depend on the size of the branch networks. Adjusting the fixed inputs, the remaining factors showing inefficiency are almost insignificant. In other words, the essential factor explaining inefficiency in the Spanish financial sector is capacity inefficiency. Surprisingly, little attention has been devoted to this factor in the previous literature analysing efficiency in the Spanish banking industry.

Finally, it is worth noting that we have focused on the quantification of the cost related to the inappropriate level of invested physical capital, ignoring the adjustment costs when savings banks would be downsizing. Therefore, we are overestimating the capacity inefficiency costs by assuming that there are no adjustment costs whatsoever. This research may be extended by focusing on the profit rather than the cost function. The maintenance of an extensive network of branch offices could provide a greater level of service offered to the customers. This may be a competitive strategy that protects the savings banks from a deterioration in its intermediation margin, since situations of competition in interest rates are more easily avoided.

Acknowledgements

References

Berger, A.N., Hancock, D., Humphrey, D., 1993. Bank efficiency derived from the profit function. Journal of Banking and Finance 17 (2/3), 317–347.

Berger, A.N., Humphrey, D., 1992. Measurement and efficiency issues in commercial banking. In: Griliches, Z. (Ed.), Output Measurement in the Service Sectors, National Bureau of Economic Research, Studies in Income and Wealth, vol. 56. The University of Chicago Press, Chicago, IL.

Berndt, E., Fuss, M., 1986. Productivity measurement with adjustments for variations in capacity utilisation and other forms of temporary equilibrium. Journal of Econometrics 33 (1), 7–29.

Berndt, E., Hesse, D.M., 1986. Measuring and assessing capacity utilisation in the manufacturing sectors of nine OECD countries. European Economic Review 30, 961–989.

Cassels, J.M., 1937. Excess capacity and monopolistic competition. Quarterly Journal of Economics 51, 426–443.

Cremeans, J.E., 1978. Capacity utilization rates—what do they really mean. Business Economics XIII (3), 41–46.

Domenech, R., 1992. Medidas no paramétricas de eficiencia en el sector bancario español. Revista Española de Economía 9 (2), 171–196 (English summary).

Färe, R., Grosskopf, S., Kokkelenberg, E.C., 1989. Measuring plant capacity, utilization and technical change: a nonparametric approach. International Economic Review 30 (3), 655–666.

Färe, R., Grosskopf, S., Lee, H., 1990. A nonparametric approach to expenditure-constrained profit maximization. American Journal of Agricultural Economics 72 (3), 574–581.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1994. Production Frontiers. Cambridge University Press, Cambridge.

Grifell-Tatjé, E., Lovell, C.A.K., 1996. Deregulation and productivity decline: The case of Spanish savings banks. European Economic Review 40 (6), 1281–1303.

Hickman, B.G., 1964. On a new method for capacity estimation. Journal of the American Statistical Association 59, 529–549.

Hunter, W.C., Timme, S.G., 1995. Core deposits and physical capital: A re-examination of bank scale economies and efficiency with quasi-fixed inputs. Journal of Money, Credit, and Banking 27 (1), 165–185.

Johansen, L., 1968. Production functions and the concept of capacity. Collected Works of Leiv Johansen, (Chapter 19), pp. 359–382.

Klein, L.R., 1960. Some theoretical issues in the measurement of capacity. Econometrica 28 (2), 272–286.

Kumbhakar, S.C., Lozano-Vivas, A., 2001. The effects of deregulation on the performance of financial institutions the case of spanish savings banks. Journal of Money, Credit, and Banking 33 (1), 101–120.

Nelson, R.A., 1989. On the measurement of capacity utilization. Journal of Industrial Economics 37, 273–286.

Noulas, A.G., Ray, S.C., Miller, S.M., 1990. Returns to scale and input substitution for large U.S. banks. Journal of Money, Credit, and Banking 22 (1), 94–108.

Pastor, J.M., 1995. Eficiencia, cambio productivo y cambio técnico en los bancos y cajas de ahorros españoles: un análisis frontera no paramétrico. Revista Española de Economía 12 (1), 35–73 (English summary).

Primont, D., 1993. Efficiency measures and input aggregation. In: Diewert, W.E., Spremann, K., Stehling, F. (Eds.), Mathematical Modelling in Economics. Essays in Honor of Wolfgang Eichhorn. Springer, Berlin, pp. 288–294.

Seguerson, K., Squires, D., 1990. On the measurement of economic capacity utilization for multi-product industries. Journal of Econometrics 44 (3), 347–361.

Sunderland, F.O., Kane, M.E., 1996. Measuring productivity on a value basis. National Productivity Review 15 (4), 57–76.