Full Length Research Paper

Long- and short-run non-parametric cost frontier efficiency of Tunisian commercial banks

Kamel Helali 1 and Maha Kalai 2

1 Department of Applied Quantitative Methods, Faculty of Economics and Management of Sfax, Aerodrome Road Km 4 BP 1088. 3018 Sfax, Tunisia.
2 Faculty of Economics and Management of Sfax. Aerodrome Road Km 4 B.P. 1088. 3018 Sfax, Tunisia.

Accepted 28 May, 2012

The capacity utilisation rate is a popular concept in both the performance assessment literature and in the publications on industrial organisations. However, a common consensus has yet to be reached concerning the most appropriate way to measure the capacity utilisation of the physical inputs and its final effect on the company results. On the one hand, there are approaches that establish the capacity utilisation with reference to the maximum level of production that can be achieved. In contrast, there are other approaches more strictly related to the economic analysis of the operating costs. In this paper, our main objective is to define an analytical process that uses non-parametric cost frontier methodology to show the gap 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 calculate the short-run inefficiency caused by a non-optimal dimension of the fixed inputs: we define this as the capacity efficiency. The proposed assessment process is applied to a set of Tunisian commercial banks covering the period between 1990 and 2009. Throughout the analysis period, the largest part of the cost inefficiency may be due to the capacity efficiency. JEL Classification: C61, D24, L6.

Keywords: Data Envelopment Analysis; short- and long-run cost efficiency; capacity utilisation; commercial banks.

INTRODUCTION

Studies on the evaluation of managerial performance, regardless of the formal method they can use, typically concluded by identifying what would reduce the potential costs or increase the profits if managers were able to correct the decision errors of the past. It is often assumed that, without questioning the real possibilities, it is possible in each period to decide on the technology necessary to adjust the inputs according to their prices and change the output on the basis of expected benefits. However, in terms of operational decisions, it is often found that the management deals with short-run rigidities limiting the decision-making that would maximise profits. We refer here to situations such as the presence of adjustment costs, administrative supervision or intervention of external regulation. All these situations may limit the possibility of adjusting the inputs and incorporate inefficiencies.

The aim of this work is to present an estimation method of the inefficiency associated with the existing capacity and restrictions to adjust the levels of the fixed input. The use of the installed capacity is a common concept in the literature on performance evaluation as well as in industrial organisations. Unfortunately, there is still no complete consensus on the most appropriate way to measure the rate of capacity utilisation (CU) and its effect on the profit and loss of firms.

In general, there are two main approaches to the concept of the capacity utilisation:
(a) In physical terms, capacity is the maximum level of production (potential level of output that fully employs existing capacity);
(b) In economic terms, capacity is the desirable level of production (optimal amount of production on the minimum point of the average total cost curve).

As we will see below, these two notions of capacity
coincide when the reference technology exhibits increasing returns to scale. A problem arises when these returns have only local significance and the average cost curves take the "U" form. In this situation, the average total cost of the first approach is always higher than that of the second. From an economic viewpoint, we favour the second approach as a way to guarantee, in all circumstances, the reference optimal cost minimisation. Specifically, we propose a model that provides the degree of efficiency with respect to the optimal total costs in the short run that does not necessarily coincide with the maximum level of production.

Concepts of capacity: technical and economic approaches

As mentioned before, there are two main concepts of production capacity: a maximum level of production and an output level that minimises the average costs. Johansen (1968) first proposed the first concept:

"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 the maximum capacity to formulate frontier models establishing the rate of capacity utilisation for firms with multiple outputs. This concept makes an economic sense when the average cost has a negative slope due to the assumption of increasing returns to scale.

However, it is unlikely that firms using their maximum capacity work with minimum production costs. When the total costs are non-linear, the full utilisation of the existing capacity may contradict the general economic objectives of the firm. Indeed, there are some cases in which a business growth leads to lower 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 can also examine the cases described by Sunderland and Kane (1996) who refer to situations where an increase in production generates a disproportional increase of the costs, or an increase of the volume of the stocks of finished products. In other cases, these authors refer to situations where there is a decline in profits caused by the outsourcing of certain processes and thus changing the size of the industry. In summary, there is enough evidence to suggest that the improvement of the physical relations alone does not necessarily lead to the improvement of the indicators of the economic capacity.

However, it remains to define exactly the optimum level of economic output. In economic literature, we find two possible interpretations. The first, which was proposed by Klein (1960) and Seguerson and Squires (1990), postulates that the optimum level of production is precisely the point of tangency between the average total cost curves in the short- and long-run. The second, which was developed by Cassels (1937) and Hickman (1964), takes, as reference, the level of the output corresponding to the minimum level of the average total costs in the long-run. In fact, Klein has criticised this concept as it is of a limited value in practice because the average cost curve can be L-shaped, and it is difficult to distinguish the precise minimum point of the average cost curve in the long-run. The difference between these two definitions, in the most general case of the average cost curves in U-shapes, is shown in Figure 1.

The difference between the two definitions is precise to the extent that, in the long-run, the average total cost curve is U-shaped. However, in the empirical applications, the decision to choose either definition is of a very little importance if the correlation coefficient between both of them is high.

Short- and long-run non-parametric frontier models

As mentioned in the previous section, when increasing returns to scale prevail and the average cost declines, then the production level that minimises the costs is one that maximises production. Hence, the point at which the production is at its maximum and the costs at its minimum coincide. However, the increasing returns to scale cannot be taken for granted. For this reason, if the method of estimating the cost efficiency has to be general, it must be independent of the hypothesis of a specific return to scale. In general, this is why the model of Färe and al. (1989) is not valid. In contrast, our proposal does not require a specific level of return to scale but works properly for both "L" and "U" shaped average cost curves.

To specify the model, we first describe the necessary variables. Suppose that for each k units of production to be evaluated \( k \in \{1, \ldots, K\} \), we know both vector \( x_k \) of the consumed inputs \( \left[ x_k = (x_{k,1}, \ldots, x_{k,J}) \in IR^J_+ \right] \) and \( y_k \) the vector of outputs \( \left[ y_k = (y_{k,1}, \ldots, y_{k,J}) \in IR^J_+ \right] \). It is also assumed that the production technology describing the process of transforming inputs into outputs is known and can be summarised by the overall condition of the following input:

\[
L(y_k) = \{ x_k : (y_k, x_k) \text{ is feasible} \}
\]

The input set \( L(y_k) \) is the collection of all the input vectors \( x_k \in IR^J_+ \) that yield at least the output vector \( y_k \in IR^J_+ \). It provides a general technology representation in terms of the amounts of input and output. No price is involved, and no assumption of behaviour is required. When the prices of raw materials are available, and the cost minimisation is a reasonable
behavioural assumption, it is possible to develop a price characterisation dependent on technology.

Suppose that the prices \((p_k)\) are known, and that the inputs can be classified as fixed and therefore impossible to change in the short run \([x_{k.f} = (x_{k,1.f}, \ldots, x_{k,K}) \in IR_{+}^{k.f}]\), or variable, and under the control of the firm \([x_{k,v} = (x_{k,1.v}, \ldots, x_{k,J}) \in IR_{+}^{k.v}]\). The variable frontier cost provides a characterisation of prices dependent on technology:

\[
\text{VC}(y_k, p_{k,v}, x_{k,f}) = \min_{x_v} \left\{ p_{k,v} \cdot x_v \left| \left( x_v, x_{k,f} \right) \in L(y_k) \right. \right\}
\]

where \(p_{k,v} = (p_{k,v_1}, \ldots, p_{k,v_J})\) is the vector of the variable input prices for unity \(k\) and \(x_v = (x_{v_1}, \ldots, x_{v_J})\) is the vector of the input minimising the variable costs. \(\text{VC}(y_k, p_{k,v}, x_{k,f})\) indicates the minimum variable costs required to produce the output vector \(y_k\) at the prices of the variable raw materials \(p_{k,v}\) according to the level of fixed input level \(x_{k,f}\).

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

\[
\text{SRTC}(y_k, p_{k,v}, x_{k,f}) = \text{VC}(y_k, p_{k,v}, x_{k,f}) + p_{k,f} \cdot x_{k,f}
\]

Thus, we can define an indicator of the efficiency frontier in the short run, \(\text{SREF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})\) as the ratio between the minimum short run total cost \(\text{SRTC}(y_k, p_{k,v}, x_{k,f})\) and the observed total cost \(p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f}\) of the firm being evaluated:

\[
\text{SREF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) = \frac{\text{SRTC}(y_k, p_{k,v}, x_{k,f})}{p_{k,v} \cdot x_{k,v} + p_{k,f} \cdot x_{k,f}} \leq 1
\]

If \(\text{SREF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) = 1\), then the evaluated firm is operating as the best practice costs, given the existing level of fixed input \(2\) otherwise, if \(\text{SREF}(y_k, p_{k,v}, p_{k,f}, x_{k,f}) < 1\), then the firm is not a part of the short run frontier cost. \(1 - \text{SREF}(y_k, p_{k,v}, p_{k,f}, x_{k,f})\) indicates the proportional reduction in costs that can be obtained if it would operate on cost-efficient frontier.

The minimum variable cost frontier for unit \(k\) \([\text{VC}(y_k, p_{k,v}, x_{k,f})]\) is obtained from program (5). This program is related to the nonparametric cost minimisation program presented by Färe et al. (1994). There are mainly three differences: technology exhibits variable returns to scale, interest is paid only the minimisation of the variable costs instead of the total costs and, finally, imposing a strict equality on the restrictions of the fixed inputs. According to Prior (2003), (3) are other documents of a similar limitation on the fixed inputs, Färe et al. (1990), in the constraint of expenditure in the short run profit maximisation and Primont (1993) in minimising the costs in the short run. However, their formulation is less restrictive because their objectives are different from ours.

Program (5) assumes a variable returns to scale
technology. It is easy to determine constant returns to scale for the above short run cost frontier, but the assumption of constant returns to scale is inevitably linked to a long run perspective.

\[ V(J_{ij}, P_{v,f} x_{ij}) = \min_{j=1}^{J_f} \sum_{i=1}^{I} p_{ij} x_{ij} \]

subject to

\[ \sum_{i=1}^{I} j_{ij} x_{ij} \geq 0 \quad j=1, \ldots, J_f \]
\[ \sum_{i=1}^{I} y_{ij} = 0 \quad jf=1, \ldots, J_f \]
\[ -y_{ij} + \sum_{i=1}^{I} j_{ij} y_{ij} \geq 0 \quad i=1, \ldots, I \]
\[ \sum_{i=1}^{I} z_{ij} = 0 \quad \forall s = 1, \ldots, K \]

Now we will focus on determining the long-run frontier cost-efficiency ratio \( LREF(y_k, P_{k, v}, P_{k, f}) \), which, unlike \( SREF(y_k, P_{k, v}, P_{k, f}, x_{k, f}) \), compares the long run efficiency costs (adjusting the level of fixed input) and the observed cost of the unit being evaluated: (See Coelli and Battese (1988) and Dogramaci and Färe (1988)).

\[ LREF(y_k, P_{k, v}, P_{k, f}) = \frac{LRTC(y_k, P_{k, v}, P_{k, f})}{p_{k, v} x_{k, v} + p_{k, f} x_{k, f}} \leq 1 \]

This expression is an extension, of the variable returns to scale case, of the standard program of the minimisation of the total costs of Färe et al. (1994). This ratio is calculated by taking information from the optimum of the cost minimisation program \( \gamma \).

\[ LRTC_{k, v, k, f}(P_{k, v}, P_{k, f}) = \min_{\beta} \left( \sum_{i=1}^{I} p_{ij} x_{ij} + \sum_{j=1}^{J_f} p_{j} x_{j} \right) \]

subject to

\[ \sum_{i=1}^{I} j_{ij} x_{ij} \geq 0 \quad j=1, \ldots, J_f \]
\[ \sum_{i=1}^{I} y_{ij} = 0 \quad jf=1, \ldots, J_f \]
\[ -y_{ij} + \sum_{i=1}^{I} j_{ij} y_{ij} \geq 0 \quad i=1, \ldots, I \]
\[ \sum_{i=1}^{I} z_{ij} = 0 \quad \forall s = 1, \ldots, K \]

After quantifying the short and long run levels of the frontier efficiency, we can now determine the capacity inefficiency (excess in costs as a result of an inappropriate level in the fixed inputs):

\[ \text{Capacity Inefficiency} = \frac{LREF_{k, v, k, f}(P_{k, v}, P_{k, f})}{SREF_{k, v, k, f}(P_{k, v}, P_{k, f}, x_{k, f})} = \frac{LRTC_{k, v, k, f}(P_{k, v}, P_{k, f})}{V(J_{ij}, P_{k, v, k, f} x_{ij}) + \sum_{j=1}^{J_f} p_{j} x_{j, f}} \leq 1 \]

Berndt and Fuss (1986) developed this measure of economic capacity utilisation in a seminal article dealing with the problem of measurement of the multifactor productivity with quasi-fixed inputs. Berndt-Fuss’s approach can be adapted using either the traditional practices of growth accounting or the parametric estimation using econometric techniques (See also Berndt and Hesse (1986) for a discussion of problems with the capacity measures). Here, we adapt this economic measure of capacity utilisation in a non-parametric approach to assess efficiency.

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

\[ CU_k = \frac{x^*_{k, f}}{x_{k, f}} \]

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

The estimation process proposed, if we assume a single output and one fixed input, is (6) in Figure 2. Unit \( k \) produces outputs \( y_k \) and incurs total cost \( TC_k \). Applying program (5), we see that the short run total cost \( SRTC_k \) appears to maintain the observed fixed input \( x_{k, f} \). Program (7) reports the unrestricted long-run frontier total cost \( LRTC_k \), but requires the adjustment of the fixed input from \( x_{k, f} \) to \( x^*_{k, f} \). For unit \( k \), the frontier coefficients are:

\[ LREF = \frac{LRTC_k}{TC_k} < 1 \]
\[ SREF = \frac{LRTC_k}{TC_k} < 1 \]

The capacity utilisation rate of the fixed input which is equal to \( x^*_{k, f} / x_{k, f} < 1 \), means the under utilisation of this factor.

In the example shown in Figure 2, the under-utilisation of the fixed input can be illustrated in two ways. Firstly, by producing \( y_k \), the optimal fixed input \( x^*_{k, f} \) minimises the total costs \((x^*_{k, f})\) is less than the actual \( (x_{k, f}) \). Second, we notice that the real production level \( (y_k) \) is lower than the one required to reach the point of tangency between the short- and long-run total cost frontiers \((y_k)\). In general, the second comparison is applicable only if a single output is produced. However, the proposed model works for multi-output technologies, and the short run fixed inputs utilisation can be deduced by comparing optimal \( (x^*_{k, f}) \) and observed \( (x_{k, f}) \) fixed inputs levels.

Figure 2 also illustrates the weakness of the technical capacity notion mentioned above. The maximum output concept of capacity \( (y^*_{k, max}) \) propose \( (8) \) are et al. (1989) presents a higher average cost than both \( SRTC \) and \( LRTC \) cost frontier. Therefore, the coherence of the maximum output notion with profit maximisation or cost
minimisation behaviour is not granted in advance. In summary, care should be taken in evaluating the managerial performance to avoid penalising the rational behaviour.

Tests on the Tunisian commercial banks: data, variables in the sample and results

The proposed method is applied on a sample of Tunisian commercial banks in the period between 1990 and 2009. To avoid Tunisian heterogeneous banking system, we can use variable return to scale model, which ensures that a firm is compared only with firms of a similar size. Thus, the sample the covers of the following units: Agricultural National Bank (BNA) Tunisian Banking Company (STB), Tunisia Arab International Bank (BIAT), Banking International Union (UIB), Housing BanK (BH), Attijari Bank [ex South Bank], Tunisian Bank (BT), Banking Union for commerce and Industry (UBCI), Arab Tunisian Bank (ATB) and Amen Bank (AB).

For different reasons, the banking sector and the period covered in the sample are relevant. First, despite the significant merger activity that took place, banks have very heterogeneous sizes. In addition, some of them faced a significant and substantial reduction in the financial results. Finally, following the deregulation of the banking sector, the Tunisian banks followed a strategy of growth both in the product range and in the number of the open agencies. This strategy created a problem of an excess capacity. In brief, we have a sector where the growth strategies related to the size led to a situation of capacity inefficiency caused by an excessive investment in the physical capital, as shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Average evolution of branch size in the period analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unit 1990</td>
</tr>
<tr>
<td>Number of agency</td>
<td>unit</td>
</tr>
<tr>
<td>Total Assets</td>
<td>MD</td>
</tr>
<tr>
<td>Inter bank lending</td>
<td>MD</td>
</tr>
<tr>
<td>Customer credit</td>
<td>MD</td>
</tr>
<tr>
<td>PortfolioTitle</td>
<td>MD</td>
</tr>
<tr>
<td>Physical capital</td>
<td>MD</td>
</tr>
</tbody>
</table>

Table 1 Table 1 summarises the average change of the total number of agencies, the total assets, the physical capital and the aggregate output, which will be measured according to three outputs: loans to customers (CC), inter
bank loans \((P_l)\) and title portfolio \((PT)\). The increase of the number of agencies that implies an increase in the number of ATMs is obvious. The increase of the total assets and aggregate outputs is also clear. In short, this trend implies an increase in the quality of the provided services (being closer to customers and expanding the amount of the services provided) and also illustrates how the connection has evolved over time.

There are many recent articles, such as that of Domenech (1992), Pastor (1995), Grifell-Tatjé and Lovell (1996), Kumbhakar and Losano-Vivas (2001), Prior (2003) and Athanasoglou et al. (2008), which apply the nonparametric frontiers to assess the efficiency of banks. Although it shares the methodology, this paper is different. In comparison with the work of Pastor (1995) and Grifell-Tatjé and Lovell (1996), we will refer to Prior (2003) to try to determine not only the level of the total cost efficiency but also the technical efficiency. Domenech (1992) also evaluates the cost efficiency but with no separation between the fixed and the variable inputs. Furthermore, this author follows the so-called \textit{intermediation approach}, which differs from ours.

Here, we follow the so-called \textit{production approach}, because it almost complies with our main objective, which is to investigate the causes that explain the differences in the operating costs. The analysis focuses basically on the study of the service production as evidenced by claims on banks and financial institutions and loans administered by each organisation. This service production requires the consumption of physical inputs where the cost is recorded as an operating expense in the profit and loss accounts.

After defining the methodological approach, we focus on the structure of the fixed and variable inputs in the banking sector. The literature has not devoted excessive attention in this regard. This can reduce the reliability of the results (As stated previously, we can mention Noulass and al., 1990; Berger and al., 1993; Hunter and Timme, 1995; Maudos and al., 2002, and Isik and Hassan, 2002). However, there are important differences between these contributions and the paper mentioned. First, these authors apply the econometric cost or profit functions, which require a priori specification of a particular function that reflects the technological relationships of which the existence is not guaranteed. In addition, they have not examined whether the existence of fixed inputs introduces rigidities that separate the levels of the short-run efficient costs from the long-run ones (However, Hunter and Time (1995) explicitly recognise that physical capital investment in the banking agencies and other equipment show 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 emphasise different institutional rigidities (rigid structures of the organisation, staff, or the irreversibility of the

### Outputs \((y)\)

- \(y_1\): Inter-bank loans \((P_l)\) which represents cash and balances with the BCT, CCP and TGT + claims against banks and financial institutions;
- \(y_2\): Loans to customers \((CC)\), which represent the receivables from the customers;
- \(y_3\): Portfolio Title \((PT)\), which represents the commercial title portfolio + title portfolio of investment.

### Variable Inputs \((x_{var})\)

- \(x_{var}\): Labour input \((L)\) which represents all the employees;
- \(x_{var}\): The financial capital factor \((F)\) is the set of the borrowed capital. The adoption of all the resources that produce financial charges is not in contradiction with the existing literature (See Hirtle (2007)). Problems \((5)\) and \((7)\) are calculated using GAMS.). The financial capital will be measured as follows: Financial Capital = Interest and similar charges incurred + Commissions + losses on business portfolio and financial operations.

### Fixed Input \((x_{fix})\)

- \(x_{fix}\): Physical capital input \((K)\) which represents the net immobilisation.

### Variable Costs \((VC)\)

- \(VC_1\): Cost of labour \((CL)\) equal to the total personnel costs;
- \(VC_2\): Cost of financial capital \((FC)\) equal to the charges on the treasury and inter-bank operations + Interest on the customer’s deposits + charges on bonds.

### Fixed Cost \((FIXC)\)

- \(FIXC\): Costs of physical capital factor \((CK)\) equal to the charges on various operations + overall Operating expenses + Depreciation and provision.

These variables provide the real total cost \((TC)\) of each bank that covers the operating and the financial costs, let
Table 2. Descriptive statistics of variables

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Moyenne</th>
<th>Écart-Type</th>
<th>r %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>142777</td>
<td>4796044</td>
<td>1401851.35</td>
<td>970507.982</td>
<td>8.46</td>
</tr>
<tr>
<td>$y_2$</td>
<td>38479</td>
<td>1733488</td>
<td>236032.39</td>
<td>230759.025</td>
<td>9.05</td>
</tr>
<tr>
<td>$y_3$</td>
<td>353</td>
<td>991880</td>
<td>191381.915</td>
<td>181040.596</td>
<td>10.26</td>
</tr>
<tr>
<td>$CV_1$</td>
<td>3939</td>
<td>52190</td>
<td>21312.515</td>
<td>10715.767</td>
<td>9.54</td>
</tr>
<tr>
<td>$CV_2$</td>
<td>3762</td>
<td>109226</td>
<td>29707.565</td>
<td>21627.302</td>
<td>4.92</td>
</tr>
<tr>
<td>$CFIX$</td>
<td>11581</td>
<td>185719</td>
<td>62830.03</td>
<td>39095.896</td>
<td>4.10</td>
</tr>
<tr>
<td>$MI$</td>
<td>3052.46</td>
<td>61710.49</td>
<td>20119.265</td>
<td>13070.849</td>
<td>8.94</td>
</tr>
<tr>
<td>$RC$</td>
<td>0.267</td>
<td>2.576</td>
<td>1.008</td>
<td>0.542</td>
<td>4.67</td>
</tr>
<tr>
<td>$AG$</td>
<td>17</td>
<td>158</td>
<td>85.6</td>
<td>34.4</td>
<td>4.02</td>
</tr>
</tbody>
</table>

Table 3. Average Scores of efficiency and capacity utilisation rates of fixed factor

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-run Efficiency</td>
<td>86.2%</td>
<td>80.4%</td>
<td>73.7%</td>
<td>75.1%</td>
<td>72.1%</td>
</tr>
<tr>
<td>Long-run Efficiency</td>
<td>78.2%</td>
<td>69.0%</td>
<td>67.1%</td>
<td>67.5%</td>
<td>64.8%</td>
</tr>
<tr>
<td>Capacity Efficiency</td>
<td>90.8%</td>
<td>85.9%</td>
<td>91.1%</td>
<td>90.1%</td>
<td>89.9%</td>
</tr>
<tr>
<td>CU of fixed Capital</td>
<td>82.5%</td>
<td>72.5%</td>
<td>65.3%</td>
<td>69.7%</td>
<td>66.9%</td>
</tr>
</tbody>
</table>

$TC = VC + FIXC = CL + CF + CK$. The output variables can also be presented in another way, closer to the literature of management accounting: The intermediation margin = Total assets. After defining the operating costs and the intermediation margin ($MI$), it is possible to formulate the coverage ratio ($CR$), a common metric in the banking literature:

Coverage Ratio = \frac{\text{Intermediation Margin}}{\text{Total Operating Costs}}

The coverage ratio provides evidence for the ability to meet the total operating cost of the normal intermediation margin. It helps to determine the bank’s ability to survive in the long-run, provided that the coverage ratio has values greater than the unit. Moreover, and in order to follow the number of branches to reduce the cost inefficiencies, we present the average growth rate of the number of agencies ($AG$).

Table 2 shows the descriptive statistics for these variables. As mentioned before, the specification of variables follows the procedure of the production approach.

To run program (5), we need the input price variables. Given the lack of information, and in accordance with the development in literature on this subject, the unit price of each input variable is measured by the ratio between its cost and its stock, let: price of labour is $p_L = CL/L$; price of financial capital factor is $p_F = CF/F$. Therefore, program (5) is solved by taking into account the average salary for each specific bank.

Program (7) requires the calculation of the price of fixed input ($p_{1\text{fix}}$). This price is calculated by dividing the depreciation and other operating costs ($K$) by the cost factor of the physical capital ($CK$) is $p_K = CK/K$. This calculation provides the average cost per bank for each specific period.

Using the specified variables, we ran programs (5) and (7) for each sample period (Problems (5) and (7) are calculated using GAMS). Table 3 summarises the empirical results. We observe, during the analysed period, the coexistence of two different trends. Between 1990 and 1999, the coefficients of the short- and long-run cost efficiency showed a steady decline. Specifically, in the late 90s, the process of deregulation and open competition was concluded and the dynamics of clusters in the innovation process took place. From 1999, we observed a stable net growth efficiency. This is explained by the stability of the use of quasi-fixed factors despite the improvement of the relationship between the number of agencies and the level of physical capital investment (see figure 3).

It is also clear that, throughout the analysed period, most of the cost inefficiency is due to the capacity efficiency. For example, we observe that, in 1999, opportunities to reduce costs inefficiencies without changing the number of branches ($SREFF$) were certainly limited (less than 10%, from 100 to 90.8%). Since the most important cost inefficiencies are always linked to capacity inefficiency, this situation reduces the possibilities to improve the short run efficiency.

Table 3 also shows that the most common situation is the under-utilisation of branches. In other words, if the
total cost minimisation was the main objective of the firms, the Tunisian banks must maintain a larger number of branches as needed. However, the number of agencies is experiencing a net average increase of 4.02%. Thus, the banking system shows a strong under-utilisation of the productive capacity. The capacity utilisation rate of the capital input fell from 82.5% to 66.9% between 1990 and 2009.

To summarise, we can say that the level of the cost frontier efficiency of the Tunisian banks is almost entirely dependent on the adequate network of branches, and we cannot do that only by the poor adjustment of the other variable inputs. On the one hand, this situation provides the strengths of interest to the analysed banks. On the other hand, it may be the source of the material weakness in the future because of the potential proliferation use of communication technologies in the banking operations. These results reflect the real importance of the strategy and the improvement of the long run efficiency. Although the technological innovation is not very developed in Tunisia at present, there is a clear opportunity to compete with banks by offering more interest rates that are attractive for consumers if banks could better control the total operating costs, without depending on the branches.

CONCLUSIONS

This study focuses on both the theoretical and empirical objectives. From a theoretical viewpoint, the main interest is to emphasise that the traditional formulation of models of cost minimisation mainly quantifies the long-run frontier efficiency levels. This is due to the general assumption that all the inputs are variable (adjustable) in the short-run. When this is not the case, the practitioner must adapt a short-run perspective.

In the empirical perspective, our interest was to analyse the frontier efficiency levels in the Tunisian banks. The sector interesting because of the special circumstances of the study period. As noted in Sections 2 and 3, between 1990 and 2009, the Tunisian financial sector has undergone a process of deregulation. This lead to a steady increase in the number of branches (especially in the early years of deregulation), which significantly reduced their level of capacity utilisation rates. In the last years of the sampling period, this process was interrupted and there was a substantial increase in the number of transactions per bank. The main conclusions of the application are that the cost inefficiencies are structural in nature and depend on the network size. For the adjustment of fixed inputs, other factors show that inefficiencies are almost low. In other words, the key factor explaining the inefficiencies in the Tunisian financial sector is the inefficient production capacity. Surprisingly, little attention has been devoted to this factor in the efficiency analysis and in the previous literature on the Tunisian banking sector.

Finally, it is interesting to note that we focused on quantifying the cost associated with an inappropriate level of the physical invested capital, ignoring the adjustment costs when banks are downsizing. Therefore, we overestimate the capacity inefficiency costs by assuming no adjustment costs whatsoever. This research can be extended by focusing on the profit rather than on the cost function. Maintaining a vast network of branches could provide a greater level of service to the customers. This may be a competitive strategy that protects banks from a deterioration of there intermediation margin, since situations of competition in the interest rates are more easily avoided.

REFERENCES


