HUB RESEARCH PAPER
Faculteit Economie & Management

Optimal Capacity Utilization and Reallocation in a German Bank Branch Network: Exploring Some Strategic Scenarios

Kristiaan Kerstens, Bouye Ahmed Moulaye Hachem, Ignace Van de Woestyne and Niels Vestergaar

HUB RESEARCH PAPER 2008/59
DECEMBER 2008

KATHOLIEKE UNIVERSITEIT BRUSSEL
EHSAL
HUB
partners in
Hogeschool-Universiteit Brussel
ASSOCIATIE K.U.LEUVEN

Optimal Capacity Utilization and Reallocation in a German Bank Branch Network: Exploring Some Strategic Scenarios

Kristiaan Kerstens[1]

Bouye Ahmed Moulaye Hachem[1]

Ignace Van de Woestyne[2]

Niels Vestergaard[3]

December 2008

Abstract:
Quite a few studies have considered efficiency at the bank branch level by comparing mostly a single branch network, while an abundance of studies have focused on comparing banking institutions. However, to the best of our knowledge no study has ever assessed performance at the level of the branch bank network by looking for ways to reallocate resources such that overall performance improves. Here, we introduce the Johansen-Färe measure of plant capacity of the firm into a multi-output, frontier-based version of the short-run Johansen industry model. The first stage capacity model carefully checks for the impact of the convexity assumption on the estimated capacity utilization results. Policy scenarios considered for the short-run Johansen industry model vary in terms of their tolerance with respect to existing bank branch inefficiencies, the formulation of closure policies, the reallocation of labor in terms of integer units, etc. The application to a network of 142 bank branches of a German savings bank in the year 1998 measures their efficiency and capacity utilization and demonstrate that by this industry model approach one can improve the performance of the whole branch network.

Keywords: Bank Branch Network, Efficiency, Capacity, Reallocation

JEL classification: G21, M11.

[1] CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France. k.kerstens@ieseg.fr and a.bouye@ieseg.fr.

[2] HUB Hogeschool Universiteit Brussel, Stormstraat 2, B-1000 Brussel, Belgium, ignace.vandewoestyne@hubrussel.be.

[3] Department of Environmental and Business Economics, Centre of Fisheries and Aquaculture Management and Economics, University of Southern Denmark, Niels Bohrs Vej 9, DK-6700 Esbjerg, Denmark. nv@sam.sdu.dk.

1. INTRODUCTION

In today's integrated financial markets, banks face increasing competition for market share. The rapid changes in market conditions (e.g., disintermediation and deregulation trends, successive merger waves, new competition from the non-financial sector) raise a number of important questions from a regulatory perspective about the structure of the banking industry. But, equally important are the strategic issues related to the management of these financial service providers offering a wide range of increasingly complex products. Against this background, the issue of bank efficiency has become rather prominent, since inefficient banks may not survive these continuous challenges, especially when the sector implements massive investments in IT to foster productivity growth (improved information management, new delivery channels, etc.). While the literature on the efficiencies of banking institutions has been summarized from various perspectives (see, among others, Berger (2007), Berger and Humphrey (1997), Goddard, Molyneux and Wilson (2001), and the focused surveys on consolidation of Amel et al. (2004) and Berger, Demsetz and Strahan (1999)), the literature analyzing the drivers of performance in financial services delivery remains rather limited (see Harker and Zenios (2001)) as does the literature on the management of bank branch networks (see Paradi, Vela and Yang (2004) for a survey).

An abundant amount of studies has focused on comparing banking institutions, while fewer studies have studied efficiency at the bank branch level by comparing mostly a single branch network. However, to the best of our knowledge no study has ever assessed the performance at the level of the branch bank network by looking for ways to reallocate resources such that overall performance of the network improves. To put this topic in perspective, we first briefly summarize the efficiency literature on banking institutions and bank branch networks. Then, we expand on the reasons why the management of a branch network requires new models and how the short-run Johansen industry model shows some promise in this respect.

In view of the dual role of financial institutions as providers of transactions and as intermediates transferring funds from savers to investors, in the efficiency literature one finds mainly two types of models to measure the flow of services in a given period (see Berger and Humphrey (1997)):

- Production approach: Banks are considered as service providers to account holders that perform transactions and process documents for depositors (e.g., checks, loan applications, credit reports, etc.). Outputs are defined in terms of numbers of transactions or documents processed. Only current expenses related to physical inputs like labor and capital and their associated costs are considered, while interest payments are ignored. As a consequence, only input prices for physical inputs are considered.

- Intermediation approach: Banks are intermediating funds between savers and investors. The flow of services is seen as proportional to the stock of financial value in the accounts (e.g., value of loans, deposits, etc.). Outputs are defined in terms of financial value terms. In addition to the physical inputs, also the input of funds is considered. Costs therefore contain current expenses and interest payments and input prices for physical inputs and financial inputs are taken into account.

Both approaches have their relative advantages (see again Berger and Humphrey (1997)). The intermediation approach is more appropriate for evaluating entire banking institutions, since interest expenses are an important part of total costs and need to be minimized to guarantee overall cost minimization or profit maximization. The production approach is most suitable for bank branches, since intermediation is organized at a higher level. Certain studies employ both approaches.

Since the seminal article of Berger, Leusner and Mingo (1997), some progress had been made in analyzing bank branch efficiency. Some key results from this limited literature can be summarized as follows. (i) There are scale economies at the branch level. But, the excess costs of over-branching are rather low due to the relative flatness of average cost curves. Furthermore, additional revenues gained from the convenience offered to the customers at the network level probably compensate these additional costs due to scale inefficiency. (ii) The large dispersion of technical inefficiencies at the branch level implies that technical inefficiencies at the bank level are understated, since even efficient banks are likely to have some inefficient branches. (iii) Bank management only imperfectly controls the costs of branch offices through its procedures, incentives and supervision. The quality of local management remains a crucial determinant of branch performance. Further conclusions on bank branch efficiency are found in the surveys of Berger and Humphrey (1997) and Paradi, Vela and Yang (2004). International comparative network studies are still extremely rare (see Athanassopoulos, Soteriou and Zenios (2001) or McEachern and Paradi (2007) for exceptions).

Bank management has always monitored the operational efficiency of its branch network by a variety of tools to measure its performance. Traditional tools to measure efficiency are based on financial ratios (such as Return on Assets, Return on Equity, or similar ratios). While ratios provide a great deal of information about financial performance in comparisons across time or relative to other banks' performance, these tools have well-known limitations. An alternative approach is the use of deterministic or econometric frontier efficiency analysis using a production approach or eventually using accounting information (as it turns out that financial and production performance tends to be rather correlated: see, e.g., Elyasiani, Mehdian and Rezvanian (1994) or Feroz, Kim and Raab (2003)). Some success stories of using frontier benchmarking in evaluating branch networks have been well-documented (see, e.g., Sherman and Ladino (1995) or Athanassopoulos and Giokas (2000)). Straightforward uses of frontier benchmarking for managing branch networks have equally been testified in a variety of written sources. In particular, efficiency scores, rankings and frontier projections have, among others, been used as an instrument to reformulate budgetary and revenue targets; to identify branches needing a thorough internal audit; to rewrite internal procedures and test the implications of these reforms on performance; to induce a learning process for current personnel by assembling both weak and good performers and eventually move best-practice managers to poor performing branches; to train new employees at best practice branches, etc.

However, the rapid technological changes have led to the introduction of new delivery systems (Automatic Teller Machines (ATM), electronic fund transfer of point of sale (EFTPOS), phone and internet banking, e-money, centralized call centers, etc) that risk to erode away the earlier dominance of the brick-and-mortar bank branch. This increasing competition of distribution channels goes hand in hand with an increasing number of bank branches in the USA (Thirtle (2007)), even though these branches are becoming more concentrated in the networks of just a few institutions (due to industry consolidation). Though Thirtle (2007) finds no systematic relationship between branch network size and overall institutional profitability, which seems to suggest that banks somehow optimize the size of their branch network as part of an overall strategy, her findings do suggest that banks with mid-sized branch networks (101–500 branches) may be at a competitive disadvantage in branching activities relative to banks with larger branch networks. Together with the common knowledge that there remain unexploited scale economies at the branch level whereby the additional cost of “overbranching” seems to be compensated by the gains in additional revenues from providing extra customer convenience (see above), these findings point to the conclusion that the management of branch networks is going to remain a major challenge for the years to come.

While measuring the efficiency of bank branch networks is fairly standard, few if any managerial tools are available to optimize existing bank branch networks while correcting for existing inefficiencies and accounting for targets of various kinds. A burgeoning literature exists that starts from efficiency measurements at the individual firm (plant or subunit) level to come up with some reallocation of resources at the level of the industry (firm). Early examples of such articles include Athanassopoulos (1995), Färe, Grosskopf and Li (1992), Golany and Tamir

(1995), Li and Ng (1995), among others. Meanwhile, a series of additional publications have appeared, including, for instance, Asmild, Paradi and Pastor (2009), Giménez-García, Martínez-Parra and Buffa (2007), Korhonen and Syrjänen (2004), and Lozano and Villa (2004). However, it is difficult to see a common structure in this large variety of research proposals. Furthermore, since few empirical applications exist and experience with practical implementations seems absent (at least it is not reported in publications), it is difficult to assess the relative advantages of these models from a managerial viewpoint. To the best of our knowledge, none of these reallocation models has ever been applied to the banking sector.

We have therefore opted to stick to a short-run industry model initially proposed in Johansen (1972) which received at least a minimum of discussion in the economics literature (see, e.g., Førsund and Hjalmarsson (1983) or Hildenbrand (1981)). Furthermore, it has been linked to the frontier-literature in Dervaux, Kerstens and Leleu (2000) who introduce frontier-based estimates of plant capacity (see Johansen (1968)) as a foundation for this short-run industry model, thereby distinguishing between variations in technical efficiency and capacity utilization. This methodological refined model has been applied in analyzing excess capacities in fisheries and further extended in Kerstens, Vestergaard and Squires (2006). Starting from the ex-post fixity of investments in production capacities, this short-run Johansen (1972) model allows for some substitution possibilities by reallocating inputs and outputs among the units composing the industry while eliminating technical inefficiencies and major variations in capacity utilization among units. Furthermore, over time substitution and technical change can be traced via shifts in successive short-run industry models. None of the other above mentioned models accounts for the notion of production capacity or distinguishes clearly between technical inefficiency and variations in capacity utilization. As far as we know, this short-run industry model has never been applied to banking.

Since the goal of performance benchmarking in this case is prospective (i.e., providing management with strategic information to actually improve performance), there are strong reasons to believe that many people object to unobservable projection points implied by the traditional convexity hypothesis. This is evidenced in remarks, scattered in the literature, on the problems encountered in communicating the results of efficiency measurement to decision makers. We offer three examples. In a study applying convex nonparametric frontier methods to measure bank branch efficiency, Parkan (1987: 242) notes: “The comparison of a branch which was declared relatively efficient, to a hypothetical composite branch, did not allow for convincing practical arguments as to where the inefficiencies lay.” In a similar vein, Bouhnik et al. (2001: 243), apart from criticizing extreme low scaling, also state: “… it is our experience that managers often question the meaning of convex combinations that involve what they perceive to be irrelevant DMUs." Finally, Epstein and Henderson (1989: 105) report similar experiences in that managers simply question the feasibility of the hypothetical projection points resulting from convex nonparametric frontiers. Thus, avoiding convexity may facilitate the implementation of frontier-based decision support models.[4] Therefore, in this contribution a lot of attention is devoted to testing for the impact of the convexity assumption in estimating capacity and in the results of the short-run industry model.

This contribution is structured as follows. We introduce in Section 2 the Johansen-Färe measure of plant capacity of the firm into a multi-output, frontier-based version of the short-run Johansen industry model. The first stage capacity model carefully checks for the impact of the convexity assumption on the estimated capacity utilization results. Policy scenarios considered for the short-run Johansen industry model vary in terms of their tolerance with respect to existing bank branch inefficiencies, the formulation of closure policies, the reallocation of labor in terms of integer units, etc. The data set of 142 bank branches of a German savings bank in the year 1998 is introduced in Section 3. The application to this German network of bank branches in Section 4 measures their efficiency and capacity utilization and demonstrate that by this industry model approach one can improve the performance of the whole branch network. A final section concludes and tries to outline some promising avenues for further research.

2. Methodology

2.1. Introduction

The theory of production is based on efficient technologies (production frontiers) and their value duals (such as minimal cost functions and maximum profit functions) and on envelope properties yielding cost-minimizing input demand functions and revenue maximizing output supply functions. In theory, emphasis is placed on efficient production and its consequences, and the evocative term "frontier" is applied to functions characterizing these boundaries. Using either parametric or nonparametric approaches, the standard cost structure is typically generated by imposing a specific functional form on the data and by obtaining the best fit by minimizing the deviations from the estimated structure. Efficiency measurement implies comparison between actual and optimal performance positioned on the relevant frontier. This frontier is called "best-practice", since it is an empirical approximation of the true but unknown

[4] We thereby ignore the theoretical arguments against convexity based upon, for instance, the indivisibilities in production. See, e.g., Scarf (1994).

frontier. The parametric approach is stochastically attempting to distinguish noise from inefficiency which requires strong assumptions, while the nonparametric approach does not run the risk of misspecification of the functional form but noise is not taken into account.[5]

We first offer several definitions to understand the mechanism of efficiency measurement. In general, efficiency analysis can be carried out at many levels of aggregation (i.e., at the plant, firm, industry or economy–wide level). The choice of level of aggregation is determined by – among other things – availability of data and the purpose of the study. Here, we focus on the linkages between the efficiency both at the firm (branch) level and the industry (branch network) level. Economic efficiency has both a technical and allocative component. Technical efficiency is generally about avoiding waste, i.e., reducing the use of inputs given output levels or increasing outputs given input levels (see Koopmans (1951) for a formal definition). Allocative efficiency is referring to optimal proportions in outputs and inputs connected to prevailing relative prices.

When it comes to measurement of technical efficiency, the so-called Debreu (1951)-Farrell (1957) measure is used. In an output-augmenting orientation, the Debreu-Farrell measure is defined as the maximum radial expansion in all outputs that is feasible with given technology. From an engineering capacity concept, Johansen (1968) defined plant capacity as the maximal amount of output that can be produced per unit of time with an existing plant and its equipment without any restrictions on the availability of variable inputs. Capacity arises due to fixity of one or more inputs, and is thereby inherently a short-run concept. Färe (1984) formally showed the existence of plant capacity for certain types of production functions, while Färe, Grosskopf and Kokkelenberg (1989) made the concept operational by using the Debreu-Farrell measure to calculate firm level capacity levels using nonparametric frontier approximations of technology. Their approach assumes that firms cannot exceed their use of fixed factors, but that their use of variable factors is unconstrained. A best-practice technology or frontier is constructed and the current output of each firm is evaluated against the maximum potential output at full capacity utilization, called “capacity output”.

Summing these firm-level capacity outputs across firms offers an estimate of the aggregate industry capacity output. Comparing this aggregate industry capacity output to current industry output provides a measure of overcapacity at the industry level. However, neither firm-level technical measures nor firm-level capacity levels allow for reallocation of inputs and outputs

[5] This is of course a simply presentation, but it presents the two essential differences between both approaches. For example, in recent years there has been a lot of work on the statistical foundation of the nonparametric approach: see Simar and Wilson (2008) for an overview.

across firms, precluding insight into the optimal restructuring and configuration of the industry. For example, the plant capacity measure implicitly assumes that production of capacity output is feasible and that the necessary variable inputs are available. In many other situations, relevant questions at the industry level are: What is the optimal firm-structure given the current aggregate output? How should the reallocation of inputs and outputs be performed between the firms? How does the reallocation look like if certain policy issues are taken into account? And what are the costs of pursuing these policy issues in terms of allocating more inputs than necessary?

To answer these questions, we combine the plant capacity notion (Johansen (1968)) at the individual and industry levels using a multiple-output and frontier-based version of the short-run Johansen (1972) sector model, a methodological refinement developed in Dervaux, Kerstens and Leleu (2000) and applied in, e.g., Kerstens, Vestergaard and Squires (2006). The short-run Johansen (1972) sector model analyses the industry structure resulting from underlying *ex post* firm-level production structures. Investment decisions imply a putty-clay production structure: while firms may eventually choose *ex ante* from a catalogue of production options exhibiting smooth substitution possibilities, most firms face fixed coefficients *ex post* and have a capacity that is entirely conditioned by the investment decision made. The short-run industry model nevertheless exhibits substitution possibilities when inputs and outputs can be reallocated across the units composing the industry. Over time, substitution and technical change can be traced via shifts in successive short-run industry models.

The revised short-run Johansen (1972) model proceeds in two phases. In a first step, the Johansen-Färe capacity measure determines capacity production for each individual firm at the production frontier. Second, this firm-level capacity information is employed in the industry model by a planning agency to select the level of activity at which individual firm capacities are utilized with the objective of minimizing fixed industry inputs given total outputs and capacities and the current state of technology. Following Dervaux, Kerstens and Leleu (2000) and Kerstens, Vestergaard and Squires (2006), the optimal industry or branch network configuration can be found by minimizing the total use of fixed inputs given that each firm cannot increase its use of fixed inputs and the production of the industry is at least at the current level.[6] The output

[6] Remark that, when appropriate price information is available, the technical optimization (in terms of primal or quantity based aspects) in both stages of the short-run Johansen industry model can be replaced by alternative economic capacity notions in the first stage and economic objective functions (e.g., industry cost functions as in Førsund and Hjalmarsson (1983), or industry revenue or profit functions) in the second stage. In the first stage, economic capacity notions based on, e.g, the cost function can be employed (e.g., Prior (2003)).

level of each firm in this type of model is the capacity output estimated from the firm-level capacity model.

2.2. Definitions of Efficiency, Plant Capacity, and the Short-Run Industry Model

To develop these production models formally, the production technology *S* transforms inputs $x = (x_1, ..., x_N) \in \mathbb{R}_+^N$ into outputs $u = (u_1, ..., u_M) \in \mathbb{R}_+^M$ and summarizes the set of all feasible input and output vectors: $S = \{(x,u) \in \mathbb{R}_+^{N+M} : \ x \text{ can produce } u\}$. Let *J* be the number of firms/units ($j \in \{1, ..., J\}$). The *N*-dimensional input vector *x* is partitioned into fixed factors (indexed by *f*) and variable factors (indexed by *v*): $x = (x_v, x_f)$. To determine the capacity output or technical efficiency, a radial output-oriented efficiency measure $E^0(x,u) = \max\{\theta : (x, \theta u) \in S\}$ is computed relative to a frontier technology providing the potential output given the current use of inputs, where restrictions on input use determine the precise nature of the measure.

Nonparametric inner-bound approximations of the true technology can be presented by the following set of production possibilities, assuming strong disposal of inputs and outputs and variable returns to scale (*VRS*):

$$\begin{aligned} S^{\Lambda,VRS} = \Big\{ &(x,u) \in \mathbb{R}_+^{N+M} : u_m \le \sum_{j=1}^{J} z_j u_{jm}, \quad m = 1, ..., M; \\ &\sum_{j=1}^{J} z_j x_{jn} \le x_n, \quad n = 1, ..., N; \quad \sum_{j=1}^{J} z_j = 1, \quad z_j \in \Lambda, \quad j = 1, ..., J \Big\}, \end{aligned} \tag{1}$$

where $\Lambda \in \{C, NC\}$, with $C = \{z_j \in \mathbb{R}_+^J\}$ and $NC = \{z_j \in \mathbb{R}_+^J : z_j \in \{0,1\}\}$. $S^{\Lambda,VRS}$ assumes strong disposability of input and outputs, variable returns to scale, and it imposes either the traditional convexity (*C*) assumption or an alternative non-convexity (*NC*) hypothesis. From activity analysis, *z* is the vector of intensity or activity variables that indicates the intensity at which a particular activity is employed in constructing the reference technology by forming convex combinations of observations constituting the best practice-frontier.

From this same technology, a plant capacity version is defined by dropping the constraints on the variable input factors. This leads to Johansen's model definition of plant capacity whereby the availability of variable factors is unrestricted:

$$\hat{S}^{\Lambda,VRS} = \Big\{ (x,u) \in \mathbb{R}_{+}^{N+M} : u_m \le \sum_{j=1}^{J} z_j u_{jm}, \quad m = 1,\dots,M; \\ \sum_{j=1}^{J} z_j x_{jf} \le x_f, \quad f = 1,\dots,F; \quad \sum_{j=1}^{J} z_j = 1, \quad z_j \in \Lambda, \quad j = 1,\dots,J \Big\}, \tag{2}$$

where Λ is again defined as above. To remain consistent with the plant capacity definition, in which only the fixed inputs are bounded at their observed level, the variable inputs in the production model (2) are allowed to vary at will to exploit the full capacity of outputs conditioned by the fixed inputs.

The efficiency measure θ_1 is found by solving the linear programming problem for each firm $j = 1, 2, \dots, J$ relative to the production possibilities set with unrestricted variable inputs given by (2):

$$\max_{\theta_1^j, z_j} \left\{ \theta_1^j : (x, \theta_1^j u) \in \hat{S}^{\Lambda,VRS} \right\}. \tag{3}$$

The scalar θ_1 informs us by how much the production of each output of firm *j* can be increased. In particular, capacity output for firm *k* of the *m*th output is θ_1^{*k} multiplied by the actual production u_{km}. Hence, capacity utilization based on observed output (subscript 'oo') is:

$$CU_{oo}^{k} = \frac{1}{\theta_1^{*k}}. \tag{4}$$

Färe et al. (1994) note that this ray *CU* measure may be biased downwards, because there is no guarantee that the observed outputs are produced in a technically efficient way. The technical efficiency measure can be obtained by evaluating each firm $j = 1, 2, \dots, J$ relative to the production possibility set $S^{\Lambda,VRS}$. The outcome (θ_2) shows by how much production can be increased using the given vector of inputs:

$$\max_{\theta_2^j, z_j} \left\{ \theta_2^j : (x, \theta_2^j u) \in S^{\Lambda,VRS} \right\}. \tag{5}$$

The technically efficient output vector is θ_2^{*k} multiplied by observed production for each output. Total industry output can be found by aggregating the firm-level technically efficient output $\theta_2^{*k} u_k$ of each firm. Likewise, the aggregate industry capacity output can be found as the sum of firm-level capacity outputs ($\theta_1^{*k} u_k$). The unbiased ray measure of capacity utilization given technically efficient output (subscript 'eo') is then:

$$CU_{eo}^{k} = \frac{\theta_2^{*k}}{\theta_1^{*k}}. \tag{6}$$

The focus here is on reallocation of resources between branches in a network by explicitly allowing improvements in technical efficiency and capacity utilization rates. The model is developed in two steps as follows. In the first step, from model (3), an optimal activity vector z^{*k} is provided for firm k and hence capacity output and its optimal use of fixed and variable inputs can be computed:

$$u_{km}^{*} = \sum_{j=1}^{J} z_j^{*k} u_{jm}; \quad x_{kf}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jf}; \quad x_{kv}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jv} . \qquad (7)$$

In a second step, these "optimal" frontier figures (capacity output and capacity variable and fixed inputs) at the branch level are used as parameters in the industry model. In particular, the industry model minimizes the industry use of fixed inputs in a radial way such that the total production is at least at the current total level, or at a desired target level in the model extension developed below, by a reallocation of resources between firms or branches. Reallocation is allowed based on frontier production outputs and inputs used in each branch. In the short-run, it is assumed that current capacities cannot be exceeded either at the branch or industry level. Define U_m as the industry output level of output m and X_f (X_v) as the aggregate fixed (variable) inputs available to the sector of factor f (v):

$$U_m = \sum_{j=1}^{J} u_{jm}, \quad X_f = \sum_{j=1}^{J} x_{fj} \text{ and } X_v = \sum_{j=1}^{J} x_{vj} . \qquad (8)$$

The formulation of the multi-output and frontier-based short-run Johansen (1972) industry model can then be specified as:

$$\begin{aligned} & \min_{\theta, w, X_v} \theta \\ \text{s.t.} \quad & \sum_{j=1}^{J} u_{jm}^{*} w_j \geq U_m, \qquad m = 1,..,M, \\ & \sum_{j=1}^{J} x_{fj}^{*} w_j \leq \theta\, X_f, \qquad f = 1,...,F, \\ & \sum_{j=1}^{J} x_{vj}^{*} w_j \leq X_v, \qquad v = 1,...,V, \\ & 0 \leq w_j \leq 1, \quad \theta \geq 0, \quad j = 1,...,J. \end{aligned} \qquad (9)$$

Rather than reflecting a returns-to-scale hypothesis, the variables w now indicate which firms' capacity is utilized and by how much. The components of the activity vector w are bounded above at unity, such that current capacities can never be exceeded. The first constraint prevents total production by a combination of firm capacities from falling below the current output levels. The second constraint means that the total use of fixed inputs (right-hand side) cannot be less than the use by a combination of firms. The third constraint calculates the resulting total use of variable inputs. Note that the total amount of variable inputs is a decision variable. The objective function is a radial input efficiency measure focusing on the fixed inputs solely. This input efficiency measure has a fixed-cost interpretation at the industry level. The activity vector w

indicates which portions of the line segments representing the firm capacities are effectively used to produce outputs from given inputs.

To sum up, the optimal solution to this simple LP gives the combination of firms or branches that can produce the same or more outputs with less or the same use of fixed inputs in aggregate.[7] It measures the combined impact of the removal of any inefficiency, the exploitation of existing plant capacities, and the reallocation of inputs and outputs. Notice that an alternative could be to have an efficiency measure focusing on the expansion of industry outputs that has a revenue interpretation.

From a managerial point of view, the optimal solution of this short-run industry model provides information at two levels. First, at the level of the network it indicates the aggregate amount of variable inputs that is needed to realize the multiple aggregate outputs from given fixed aggregate inputs. If the optimal value of the aggregate variable inputs decision variable is larger than the current amount of aggregate variable inputs, then this implies additional recruitments are needed. Otherwise, a reduction in staff levels is required.

Second, at the level of the individual production units (bank branches) the model yields a complete planning for service production. Per unit, one obtains optimal fixed ($x^*_{fj} w^*_j$) and variable ($x^*_{vj} w^*_j$) inputs as well as optimal outputs ($u^*_{jm} w^*_j$). This may imply reallocations of inputs: fixed and variable inputs may be redistributed among units. Obviously, adjusting fixed inputs may be costly (e.g., renegotiating an existing office rental contract) and may furthermore require time to implement (e.g., legal terms of notification prevent immediate changes). Equally so, adjusting variable inputs may be subject to a series of constraints (especially labor is under legal protection). This plan may also imply reallocations of outputs: this simply means that one adjusts the output targets within the planning horizon so as to better exploit the existing capacity of the whole network. Obviously, this may imply accompanying policy measures that are not necessarily part of the model (e.g., marginal changes in global and local marketing campaigns in an effort to gear consumer demand towards these targets).

2.3. Short-Run Industry Model: Additional Scenarios

Now, we turn to a discussion of some additional scenarios that extend the frontier-based short-run industry model to adapt to managerial concerns.

[7] In fact, this short-run industry model is geometrically speaking a set consisting of a finite sum of line segments known as a zonotope (see Hildenbrand (1981: 1096).

1. *Restriction on number of branches:*

Assume the number of branches should be restricted to *N*. Since the variable w_j represents the utilization of the corresponding branch, this restriction can be modeled with the following constraints:

$$\begin{aligned}&w_j \leq b_j \ (j=1,...,J);\\&\sum_{j=1}^{J} b_j \leq N; \qquad\qquad (10)\\&b_j \in \{0,1\} \ (j=1,...,J).\end{aligned}$$

By adding these constraints to model (9), it becomes a mixed integer program. The binary variable b_j indicates whether the corresponding branch is used in the optimal solution or not. The amount by which it is used can then be read from variable w_j.

2. *Allow for existing inefficiency*

The capacity outputs and the corresponding optimal fixed and variable inputs as computed in (7) presuppose that all eventually existing technical inefficiency can be eliminated in an effort to exploit the existing capacity of production. However, starting from the optimal activity vector $z^{*k} = (z_1^{*k},...,z_J^{*k})$ obtained from solving model (3), it is also possible to define capacity outputs and the corresponding optimal fixed and variable inputs while maintaining the existing levels of technical inefficiency by computing:

$$\overline{u}_{km}^{*} = \frac{1}{\theta_2^k}\sum_{j=1}^{J} z_j^{*k} u_{jm}; \quad x_{kf}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jf}; \quad x_{kv}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jv} . \qquad (11)$$

Hence, while the optimal fixed and variable inputs remain the same, the capacity outputs are maintained or scaled down by the measured amount of technical inefficiency (θ_2). Referring to the capacity output in (7) as the fully efficient one, the adjustment in (11) is called the fully inefficient capacity output. Both these capacity outputs can be considered special cases of the 100α % inefficient capacity output and the corresponding optimal fixed and variable inputs that can be defined as:

$$\overline{u}_{km}^{*}(\alpha) = \frac{1}{1+\alpha(\theta_2^k - 1)}\sum_{j=1}^{J} z_j^{*k} u_{jm}; \quad x_{kf}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jf}; \quad x_{kv}^{*} = \sum_{j=1}^{J} z_j^{*k} x_{jv}, \qquad (12)$$

with $0 \leq \alpha \leq 1$. Clearly, the 0% inefficient capacity output corresponds with the fully efficient capacity output, while the 100% inefficient capacity output coincides with the fully inefficient capacity output. When fully inefficient capacity output are used in the short-run industry model, this implies that one measures the impact of reallocation only.

3. Restrictions on the personnel transfer

Assuming the number of employees is a variable input, personnel transfer for a given branch with respect to the current situation is then measured by the difference between the optimal variable input resulting from the industry model and the observed fixed input (i.e., $x_{vj}^{*}w_j - x_{vj}$). It could be meaningful to allow personnel transfer only in integer multiples of some unit β. For instance, $\beta = 0.5$ would mean that the number of employees must change in multiples of one half (e.g., because the basic unit of a labor contract in some countries is either a part-time of a full-time contract). Since this change can be either positive (reflecting an increase in number of employees) or negative (referring to a decrease), this condition can be modeled by the constraint:

$$x_{vj}^{*}w_j - x_{vj} = \beta(i_1 - i_2)\,, \qquad (13)$$

with i_1 and i_2 integer variables. The difference of both integer variables measures exactly the change in personnel expressed in units of β (e.g., $\beta = 0.5$ means this difference of integer variables measures personnel change in half units). Note that adding this type of constraint transforms model (9) to a mixed integer problem.

4. Imposing alternative aggregate output targets

If it is possible to impose alternative target values on the outputs, then the first set of constraints in model (9) needs to be changed to:

$$\sum_{j=1}^{J} u_{jm}^{*}w_j \geq (1+\gamma_m)U_m\ , \qquad (14)$$

with $\gamma_m \geq -1$. A value of $\gamma_m \geq 0$ (implying $1+\gamma_m \geq 1$) means that the aggregate output *m* of the industry model must be at least $100\gamma_m$% larger than the current industry level of output *m*. Obviously, positive values correspond with increases, while negative values reflect decreases with respect to the current industry level of output *m*. If all $\gamma_m = 0$, then no alternative target values are proposed and the original model (9) is obtained based upon observed aggregate outputs.

Remark that, in general, imposing a positive target value (i.e., above the output aggregate) additionally restricts the constraints. This lead to worse objective function values in the case of a minimization problem. Put differently, a positive target value leads to a higher efficiency measure θ. Ultimately, too large positive target values may result in infeasibilities. By contrast, negative target values (i.e., below the output aggregate) relax the corresponding constraint, which results in a lower or equal efficiency measure value. Whether this phenomenon actually occurs, however, depends on the status of the corresponding constraint and on its relation with other constraints. For instance, adding a negative target value to a nonbinding output constraint has no influence on the optimal solution. Even if an output constraint is binding, other binding output constraints could prevent a reduction of the efficiency measure θ when adding a negative target value.

Additional scenarios that could eventually be envisioned are: (i) limiting the range of plant capacity utilization for the units in the optimal solution (see, e.g., Kerstens, Vestergaard and Squires (2006)), and (ii) aggregating some of the outputs to reduce the number of dimensions (at the risk that the required more spectacular changes are more difficult to implement).

3. DATA: BANK BRANCHES OF A GERMAN SAVINGS BANK

Data are obtained from the article by Porembski, Breitenstein and Alpar (2005). These authors analyze a sample of 142 German bank branches in the year 1998. In this work, we measure the efficiency of these branches of a German savings bank and demonstrate that by a different industry model approach one can improve the efficiency over the whole network.

German thrift institutions are owned by communities or counties. Today, these institutions participate in all types of banking activities, either directly or through a central institution that is commonly owned. These banks are independent of each other, but share a number of resources. An important characteristic of these banks is that the goal of profit maximization is conditioned by the requirement of providing services to their stakeholders (e.g., community or county, to small businesses, and the middle-class). For example, nobody who wants to open an account can be rejected. These special characteristics cause some serious problems, since, for instance, it is not allowed to restrict branches to regions with profitable customer bases only. Moreover, increased competition is faced due to the globalization of financial markets, the spread of internet banking, and the increasing operational cost of personnel, whereas interest rates and profits have been decreasing over the last few years. This explains why these banks are very keen on increasing their productivity.

The bank analyzed is among the ten largest of its type in Germany. Its total assets in 1998 were in the tens of billions US $. To develop the bank branch industry model, we follow Porembski, Breitenstein and Alpar (2005) and basically adopt a so-called, production approach to defining the transformation of banking inputs into financial services. Bank branches are considered as service providers to account holders performing transactions and processing documents. Outputs are therefore normally defined in terms of the numbers of transactions or documents processed. The outputs chosen cover most of the products offered by a branch and the level of disaggregation is high (e.g., one distinguishes between demand deposits for business and for households). However, very often, and also in this case, detailed transaction flow data are unavailable, whence the stock of the number of accounts of various types is employed instead. Furthermore, only physical inputs like labor and capital and their associated costs are taken into account. Actually, around 60% of the operating costs are due to personnel. Hence, the labor input is one of the most important at the branch level. A major part of the remaining operating costs are building and equipment costs. Since these costs are very difficult to determine (e.g., the corresponding book value is often biased), the input office space serves as a surrogate input measure.

Listing the inputs and outputs constituting the production technology in detail, the following inputs are available:

- Employees (number);
- Office space (square meters);

whereby the units of measurement are put in between braces. Notice that it is common to consider office space as a fixed input that cannot be modified in the short-run. Hence, employees are the sole variable inputs. In addition, there is information on the following 11 output dimensions:

- Private demand deposits (accounts);
- Business demand deposits (accounts);
- Time deposits (accounts);
- Saving deposits (accounts);
- Credits (accounts);
- Bearer securities (accounts);
- Recourse guarantees (accounts);
- Bonds (accounts);
- Investment deposits (accounts);
- Insurances (contracts);
- Contributions to a building society (contracts).

Descriptive statistics, including mean, variance, skewness, the minimum and the maximum, for these input and output dimensions are reported in Table 1. We can make the following observations. First, there is a lot of variation among these bank branches as witnessed by the standard deviation. Furthermore, the positive skewness of the distribution reveals the dominance of certain large units, mainly reflecting substantial differences in size. Second, notice that some branches do not seem to produce time deposits, recourse guarantees, or insurance since these outputs are zero at the minimum. This may reveal a variety of patterns of specialization among this sample bank branches. In addition, the last row contains the sum of all inputs and outputs at the level of the branch network. This serves as a benchmark to assess the impact the various scenarios in the industry models.

< Table 1 about here >

4 EMPIRICAL RESULTS

First, we report extensively on the estimation results of the plant capacity measure and its underlying efficiency measures. We thereby focus on the impact of the convexity hypothesis and the impact of correcting the capacity definition for the presence of technical inefficiency or not. Thereafter, we turn to the basic results from the short-run industry model and also investigate the implied reallocations at the level of the individual branches. We thereby report on a series of different scenarios.

4.1. Estimation of Plant Capacity: Testing for Convexity

Descriptive statistics for the capacity-related efficiency measure (θ_1), the ordinary technical efficiency measure (θ_2), and the plant capacity measure (CU_{eo}) are reported in Table 2 for both the convex and non-convex case. Four key observations can be made: (i) the output-oriented inefficiency measures are on average much higher in the convex case than in the non-convex case; (ii) in the non-convex case all bank branches except three are technically efficient in contrast to just about 40% of observations in the convex case; (iii) two thirds of all branches (97) operate at full capacity in the non-convex case compared to about one fifth (33) in the convex case; and (iv) these phenomena result in rather low average measures of capacity utilization in the convex case compared to the non-convex case.

< Table 2 about here >

The difference between the densities of the output efficiency measures obtained with the convex and non-convex models as well as the resulting ray CU measure can be tested with a statistic developed by Li (1996) and later refined by Fan and Ullah (1999). This test statistic has the critical advantage to be valid for dependent and independent variables, the former dependency being typical for frontier estimators. The null hypothesis states the equality of both distributions.

Table 3 summarizes the obtained results. In total, three efficiency measures (θ_1, θ_2 and CU_{eo}), both in the convex and non-convex case, are compared two by two. Notice that the symmetry of the table immediately follows from the symmetry of the test itself. The values of these test statistics must be compared with the reference value for the target significance level. A value higher than the reference value leads to a rejection of the null hypothesis (implying that both density distributions can be considered statistically different). Table 3 also shows the conclusion depicted with symbols when tested for a significance level of 1%: an asterisk (*) is used when the null hypothesis is rejected (different densities) and an equality sign (=) flags that the null hypothesis cannot be rejected (equal densities). We notice that all density distributions can be considered different, except for θ_1 and CU_{eo} in the non-convex case. The latter exception is explained by the fact that only three observations are technically inefficient ($\theta_2 > 1$) in the non-convex case (hence, the ratio CU_{eo} is inevitably very close related to θ_1). In conclusion, statistical tests indicate that these efficiency measures follow different distributions. Put differently, adding the traditional convexity hypothesis is not as innocuous as it is traditionally assumed.

< Table 3 about here >

Table 4 reports descriptive statistics of plant capacity inputs and outputs for two variations: (i) convex vs. non-convex; and (ii) full efficiency vs. full inefficiency. These results need to be contrasted with the descriptive statistics on the inputs and outputs of the original data in Table 1. Comparing Tables 4 and 1, one immediately observes that: (i) the capacity inputs remain on average close to the observed inputs, while the choice for the output orientation of efficiency measurement implies that capacity outputs are quite above observed outputs; (ii) this divergence between capacity and observed outputs is more substantial for the convex case than for the non-convex case; and (iii) the difference between capacity outputs without and with technical inefficiency is again largest in the convex case. This analysis serves to underscore the importance of the convexity axiom and, to some lesser extent, the impact of eliminating technical inefficiency or not.

< Table 4 about here >

4.2. Short-Run Industry Model: Basic Results and Additional Scenarios

Instead of using the fully efficient capacity output in the short-run Johansen industry model formulated in (9), the fully inefficient capacity output (11) as well as the 100α% inefficient capacity output for a given α (12) can be employed, leading to a series of variations of this basic model. By examining these different models, the impact of allowing for inefficiency can be measured in combination with the difference between convex and non-convex estimates of capacity.

Table 5 summarizes exactly this impact of both convexity and inefficiency on several key decision variables. First, there is the influence on the optimal industry efficiency measure θ^*. In the next row, the influence on the number of branches is reported for which full capacity is used in realizing at least the aggregate outputs with only a fraction of the fixed aggregate inputs. Similarly, the next rows indicate the number of branches that are only partially used or not used at all to realize the set of constraints in model (9).

< Table 5 about here >

In the convex case, the effect of allowing for inefficiency is noticeable. We observe, for instance, an increase of the efficiency measure with 0.1 when allowing for all existing technical inefficiency (this is a relative increase of 17%). Since capacity outputs are lower when one allows for inefficiency, it is harder to economize on fixed inputs and an increase of its optimal value can indeed be expected. Furthermore, notice that the full efficiency case only utilizes 106 of the 142 branches. Since the number of branches only partially used is limited to only three, this means that 33 branches are not used at all to implement the optimal solutions obtained in the Johansen industry model. This is quite a substantial amount (23.2% of the total number of branches), making one doubt whether such solution is implementable in practice. When inefficiency is allowed for, then the number of unused branches is reduced to 28 (19.7%), which remains considerable.

Remark that, contrary to what one may expect, the branches that are no longer used in the optimal solution remain not necessarily the same when moving from the fully efficient to the fully inefficient case. Put differently, the 28 branches observed with zero capacity in the fully efficient scenario are not necessarily contained in the 33 branches that are no longer utilized in the fully efficient scenario. Examining the individual branches, we detect 11 of the 28 branches that are used in the fully efficient case but not used at all in the fully inefficient scenario. Except for one, these are even used at full capacity.

We end by looking at the results in the non-convex case. With respect to the optimal efficiency value θ^*, we notice only a minor increase of 0.003 (this is a relative increase of only 0.4%) when moving from the fully efficient to the fully inefficient industry model. From the individual results per branch, it can be observed that there is no shift in the optimal solution. Thus, all branches used at full capacity in the fully efficient case are also maintained at full capacity in the fully inefficient scenario. The same holds true for the branches used at partial capacity and for those that are no longer used at all. Only a minor change can be detected in the capacity of two branches used at partial capacity. Consequently, the effect of allowing inefficiency in the non-convex case can be neglected. The same holds for the other decision variables reported in this case, since there is no difference at all. Intermediate inefficiency levels for the non-convex model are therefore of limited interest in this particular study.

Notice that the number of unused branches reduces to 24 (16.9%) which is substantially lower compared to the convex model (33 in the fully efficient scenario and 28 in the fully inefficient case). From additional examination of individual branch results, it can be noticed that the 24 branches that are no longer used following the non-convex methodology are not necessarily contained in the unused branches according to the convex methodology. Indeed, with respect to full efficiency, 11 branches are found with zero capacity in the non-convex case, but with full capacity in the convex case. In the fully efficient scenario, even 13 branches can be detected having zero capacity according to the non-convex methodology, but with full capacity following the convex methodology. This underscores that the fundamentally different nature of the convex and non-convex technologies may have far reaching managerial consequences.

To complement Table 5, Figures 1a and 1b trace the evolution of the industry efficiency measure as a function of a given α for the convex respectively the non-convex case. As could already be anticipated from considering the extreme cases in Table 5, the function for the convex case is much steeper because industry efficiency changes over a wider range. The relative flatness of this function in the non-convex case is related to the small amount of technical inefficiency that can be detected under this assumption in the first place.

< Figures 1a and 1b about here >

Notice that the industry efficiency measure has a fixed cost interpretation and denotes the potential budgetary gains from closing down the branches indicated by zero utilization in the industry model. However, one must realize that in practice a host of additional considerations may be necessary to choose among these in defining a coherent closure policy. As already pointed at previously, adjusting fixed inputs may be costly both when one is owner of the office space (e.g., should one rent out part of the excessive office space assuming this is technically feasible, or should one sell of the property and buy a smaller one somewhere nearby?) and when one is renting these (e.g., renegotiating an existing office rental contract may be costly). Furthermore, these changes require time to implement (e.g., legal terms in buying and selling contracts as well as in rental contracts prevent changes overnight). In addition, it may be necessary to include additional consideration into this decision making process. For instance, it makes a difference whether one closes down a branch in a town with two additional branches of the same bank or in a small village with no other branch around in the neighborhood. These decisions may thus need to be conditioned on a variety of geographical information that is currently ignored in the model.

We now restrict attention to the non-convex methodology. Furthermore, since the effect of allowing for inefficiency is negligible in the non-convex case, we also limit the analysis to the case of full efficiency. We discuss the following three scenarios of interest that have been formally introduced in subsection 2.3. Firstly, the impact of adding restrictions on the number of branches (10) in model (9) is considered. Secondly, we investigate the influence of adding restrictions on the personnel transfer (13) to the short-run industry model. Finally, we evaluate the effect of imposing some alternative aggregate output targets (see (14)). Results for all these scenarios are reported in Table 6.

< Table 6 about here >

Restrictions on the number of branches

The results of adding the constraints on the number of branches for some key reference values of *N* to the model are reported in the first five columns of Table 6. On one extreme, we notice that the problem becomes infeasible when limiting the number of branches to 95 or less. This means that we need at least 96 branches to deliver the current level of network outputs from given fixed inputs. On the other side of the range, we see that efficiency no longer improves when passing the limit of 118 branches. Furthermore, observe that in all cases, the number of branches used at full capacity is very close to the imposed limit *N*. Put differently, the number of branches used at partial capacity is very low (only one to two), meaning there seems to be little or no advantage of moving to scenarios that promote the use of partial capacities. Obviously, the value of the efficiency measure θ decreases as *N* increases. This observation corresponds with intuition since an increase in the number of branches implies using branches that are less efficient and/or that have less capacity.

Restrictions on the personnel transfer

Adding restrictions on the personnel transfer, the middle part of Table 6 reports the effect of adding such a restriction for two values of β. In particular, personnel transfer is only possible in integer multiples of either $\beta = 0.5$ (number of employees must change in multiples of one half) or $\beta = 1$ (number of employees must change in multiples of one). This scenario has two noticeable effects. First, the industry efficiency score increases substantially, implying that less fixed inputs can be economized. Second, there is a substantial move from branches working at full capacity to branches functioning at some partial capacity level. This actually turns out to be the only scenario producing such a result.

We add two remarks on potential implementation problems. First, the transfer of personnel can be difficult in view of geographical distances. For instance, it would make little sense to reallocate a person for say about 10% of his working time (about a half day per week in a five day working week) to a bank branch located at 500 km from his/her initial location. The current model ignores this issue basically because geographical information is lacking. However, in principle it is possible to extend the current model by restricting patterns of reallocation among units within a certain geographical radius (see, e.g., Giménez-García, Martínez-Parra and Buffa (2007) for an example).

Second, the empirical model only employs aggregate information on personnel. Disaggregating personnel may yield more detailed results that are easier to implement and that have positive additional results. For instance, in Sherman and Ladino (1995) the efficiency results have been used to look for reductions in the number of branch managers by looking for possibilities to share managers for specific nearby bank branches. This again necessitates detailed geographical information. In a similar vein, the efficiency and capacity results could be used to make sure reallocations of managers go from high performance to low performance branches such that these relatively more successful managers can induce best practice behavior throughout the branch network.

Imposing alternative aggregate output targets

The last part of Table 6 reports on some aggregate output target scenarios. In a first scenario, we impose a positive output target of 10% on the number of saving deposits only. As a result, the optimal efficiency measure increases substantially from its original value of 0.702 to 0.775. To achieve this target, the number of branches needed at full capacity must be increased from 116 to 120, reducing the number of branches at zero capacity by 4. Increasing the target beyond 30% of current aggregate output is infeasible. For instance, using a negative reduction of 20% on the number of saving deposits has no influence at all on the optimal solution. Clearly, the other output constraints prevent such a reduction. When systematically looking for output variables that do have an influence when imposing a, for instance, 20% negative target, we observe that only the number of bearer securities accounts and the number of insurance contracts do make a difference. This effect is valid under ceteris paribus conditions, i.e., assuming no targets are imposed for the other outputs. First, in the case of the bearer securities, the efficiency measure θ is further reduced to 0.675, hereby using only 108 branches at full capacity compared to 116 originally (resulting in an increase of the number of unused branches from 24 to 32). Second, with respect to the number of insurance contracts, a more modest effect is observed: the efficiency measure only drops with 0.001. This result is obtained by utilizing 115 branches at full capacity instead of 116 initially. The number of branches no longer used remains the same (24), but when looking at individual results, we notice a minor shift. One branch previously not used is now used partially, and simultaneously another branch previously used only partially is now no longer used at all.

5. CONCLUSIONS

Briefly summarizing the main contributions of this work, we focus shortly on the methodology employed as well as on the results. The efficiency literature analyzing the financial sector shows that even well performing banking institutions may have technical inefficiencies and some excess capacities at the level of their network of bank branches. Instead of relying on a burgeoning literature that starts from efficiency measurements at the individual level to come up with reallocations of resources at the firm level, we have opted to continue in the tradition of the revised short-run Johansen (1972) industry model, which is firmly grounded in the economics literature.

By way of example, we have analyzed the financial services supplied by a bank branch network of a rather large sized German savings bank (see Porembski, Breitenstein and Alpar (2005)) using a production approach. The ordinary technical efficiency measure, the capacity-related efficiency measure, and the plant capacity measure have been computed using both convex and non-convex technologies. The resulting difference between the densities of these output efficiency measures and the resulting ray capacity utilization measure have been tested: the Li (1996) test statistic reveals that the resulting densities are almost all different from one another. This provides strong support to opt for a non-convex production technology rather than the traditional convex one for frontier benchmarking purposes.

Empirical results of the short-run industry model reveal a potential for closing down part of the network while maintaining current service levels, even under the most conservative estimates of efficiency and capacity (i.e., the ones based on a non-convex technology). Three additional scenarios related to the impact of adding restrictions on the number of branches on the one hand and on personnel transfer on the other hand, and the fixing of alternative aggregate output targets have also been documented.

Obviously, these scenarios do not exhaust the possibilities to adjust this network model to managerial needs. We have mentioned on several occasions the usefulness of including geographical information. Additional policy considerations could include local and regional market share considerations (competition issues in general). Obviously, while including these additional parameters need not be impossible, one must be aware that the inclusion of additional constraints lowers the potential benefits of the short-run industry model and that some combinations of constraints may even lead to infeasibilities.

The implementation cost of efficiency and capacity analysis and the resulting short-run industry models is high for single shot exercises, but this cost becomes low once the needed data on inputs and outputs are integrated into the accounting system (e.g., eventually as part of an activity based costing (ABC) strategy: see Kantor and Maital (1999)). Furthermore, while the computation of efficiency measures and capacity measures is rather straightforward and meanwhile a host of software options are around (e.g., in GAMS: see Olesen and Petersen (1996); in the freeware R: see Wilson (2008); in SAS: see Emrouznejad (2005), etc.), it is clear that the utilization of the short-run industry model as a strategic planning tool would ideally require its integration into a DSS. We are unaware of written accounts reporting on the regular use of frontier benchmarking software in organizations.[8] This remains an important issue for future research.

Overall, we hope this contribution has shown convincingly that there is scope to employ efficiency-based models to manage bank branch networks both at a strategic and operational level. Obviously, more research is needed to come up with more detailed branch network models geared towards a more complete set of managerial needs.

REFERENCES

Amel, D., C. Barnes, F. Panetta, C. Salleo (2004) Consolidation and Efficiency in the Financial Sector: A Review of the International Evidence, *Journal of Banking & Finance*, 28(10), 2493-2519.

Asmild, M., J.C. Paradi, J.T. Pastor (2009) Centralized Resource Allocation BCC Models, *Omega*, 37(1), 40-49.

[8] Non-convex frontier technologies have been used for years to assess credit union performance by their trade association (Credit Union National Association (CUNA)): see, e.g., Fried, Lovell and Vanden Eeckaut (1995). CUNA recently launched “CU Benchmarker” as a web based, paid service for benchmarking to its member credit unions. See the CUNA webpage http://advice.cuna.org/cu_benchmarker.html (consulted December 7, 2008).

Athanassopoulos, A. (1995) Goal Programming & Data Envelopment Analysis (GoDEA) for Target-Based Multi-Level Planning: Allocating Central Grants to the Greek Local Authorities, *European Journal of Operational Research*, 87(3), 535-550.

Athanassopoulos, A., D. Giokas (2000) The Use of Data Envelopment Analysis in Banking Institutions: Evidence from the Commercial Bank of Greece, *Interfaces*, 30(2), 81-95.

Athanassopoulos, A., A. Soteriou, S.A. Zenios (2001) Disentangling Within- and Between Country Efficiency Differences of Bank Branches, in: P.T. Harker, S.A. Zenios (eds.) *Performance of Financial Institutions: Efficiency, Innovation, Regulation*, Cambridge, Cambridge University Press, 336-363.

Berger, A.N. (2007) International Comparisons of Banking Efficiency, *Financial Markets, Institutions & Instruments*, 16(3), 119-144.

Berger, A.N., R.S. Demsetz, P.E. Strahan (1999) The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future, *Journal of Banking and Finance*, 23(2-4), 135-194.

Berger, A., D. Humphrey (1997) The Efficiency of Financial Institutions: International Survey and Directions for Future Research, *European Journal of Operational Research*, 98(2), 175-212.

Berger, A.N., J.H. Leusner, J.J. Mingo (1997) The Efficiency of Bank Branches, *Journal of Monetary Economics*, 40(1), 141-162.

Bouhnik, S., B. Golany, U. Passy, S.T. Hackman, D.A. Vlatsa (2001) Lower Bound Restrictions on Intensities in Data Envelopment Analysis, *Journal of Productivity Analysis*, 16(3), 241-261.

Debreu, G. (1951) The Coefficient of Resource Utilization, *Econometrica*, 19(3), 273-292.

Dervaux, B., K. Kerstens, H. Leleu (2000) Remedying Excess Capacities in French Surgery Units by Industry Reallocations: The Scope for Short and Long Term Improvements in Plant Capacity Utilization, in: J.L.T. Blank (ed) *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement,* Amsterdam, Elsevier, 121-146.

Elyasiani, E., S. Mehdian, R. Rezvanian (1994) An Empirical Test of Association between Production and Financial Performance: The Case of the Commercial Banking Industry, *Applied Financial Economics*, 4(1), 55-59.

Emrouznejad, A. (2005) Measurement Efficiency and Productivity in SAS/OR, *Computers and Operations Research*, 32(7), 1665-1683.

Epstein, M., J. Henderson (1989) Data Envelopment Analysis for Managerial Control and Diagnosis, *Decision Sciences*, 20(1), 90-119.

Fan, Y., A. Ullah (1999) On Goodness-of-fit Tests for Weakly Dependent Processes Using Kernel Method, *Journal of Nonparametric Statistics*, 11(1), 337–360.

Färe, R. (1984) The Existence of Plant Capacity, *International Economic Review*, 25(1), 209-213.

Färe, R., S. Grosskopf, E. Kokkelenberg (1989) Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach, *International Economic Review*, 30(3), 655-666.

Färe, R., S. Grosskopf, S-K. Li (1992) Linear Programming Models for Firm and Industry Performance, *Scandinavian Journal of Economics*, 94(4), 599-608.

Farrell, M. (1957) The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A: General*, 120(3), 253-281.

Feroz, E.H., S. Kim, R.L. Raab (2003) Financial Statement Analysis: A Data Envelopment Analysis Approach, *Journal of the Operational Research Society*, 54(1), 48-58.

Førsund, F., L. Hjalmarsson (1983) Technical Progress and Structural Change in the Swedish Cement Industry 1955-1979, *Econometrica*, 51(5):1449–67.

Fried, H.O., C.A.K. Lovell, P. Vanden Eeckaut (1995) Service Productivity in U.S Credit Unions, in: P.T. Harker (ed) *The Service Productivity and Quality Challenge*, Boston, Kluwer, 365-390.

Giménez-García, V.M., J.L. Martínez-Parra, F.P. Buffa (2007) Improving Resource Utilization in Multi-Unit Networked Organizations: The Case of a Spanish Restaurant Chain, *Tourism Management*, 28(1), 262-270**.**

Goddard, J.A., P. Molyneux, J.O.S. Wilson (2001) *European Banking: Efficiency, Technology and Growth*, New York, Wiley.

Golany, B., E. Tamir (1995) Evaluating Efficiency-Effectiveness-Equality Trade-offs: A Data Envelopment Analysis Approach, *Management Science*, 41(7), 1172-1184.

Harker, P.T., S.A. Zenios (2001) What Drives the Performance of Financial Institutions?, in: P.T. Harker, S.A. Zenios (eds.) *Performance of Financial Institutions: Efficiency, Innovation, Regulation*, Cambridge, Cambridge University Press, 3-31.

Hildenbrand, W. (1981) Short-Run Production Functions based on Microdata, *Econometrica*, 49(5), 1095–1125.

Hirtle, B. (2007) The Impact of Network Size on Bank Branch Performance, *Journal of Banking & Finance*, 31(12), 3782–3805

Johansen, L. (1968) Production Functions and the Concept of Capacity, Namur, Recherches Récentes sur la Fonction de Production (Collection "Economie Mathématique et Econometrie", no 2). Reprinted in: Førsund, F.R. (ed.) (1987) *Collected Works of Leif Johansen, Volume 1*, Amsterdam, North Holland, 359–82.

Johansen, L. (1972) *Production Functions: An Integration of Micro and Macro, Short Run and Long Run Aspects*, Amsterdam, North Holland.

Kantor, J., S. Maital (1999) Measuring Efficiency by Product Group: Integrating DEA with Activity-Based Accounting in a Large Mideast Bank, *Interfaces*, 29(3), 27-36.

Kerstens, K., N. Vestergaard, D. Squires (2006) A Short-Run Johansen Industry Model for Common-Pool Resources: Planning a Fishery's Industrial Capacity to Curb Overfishing, *European Review of Agricultural Economics*, 33(3), 361-389.

Koopmans, T. (1951) Analysis of Production as an Efficient Combination of Activities, in: T. C. Koopmans (ed) *Activity Analysis of Production and Allocation*, New Haven, Yale University Press, 33-97.

Korhonen, P., M. Syrjänen (2004) Resource Allocation Based on Efficiency Analysis, *Management Science*, 50(8), 1134-1144.

Li, S.K., Y.C. Ng (1995) Measuring the Productive Efficiency of a Group of Firms, *International Advances in Economic Research*, 1(4), 377-390.

Li, Q. (1996) Nonparametric Testing of Closeness between Two Unknown Distribution Functions, *Econometric Reviews*, 15(1), 261–274.

Lozano, S., G. Villa (2004) Centralized Resource Allocation Using Data Envelopment Analysis, *Journal of Productivity Analysis*, 22(1-2), 143-161.

McEachern, D., J.C. Paradi (2007) Intra- and Inter-Country Bank Branch Assessment Using DEA, *Journal of Productivity Analysis*, 27(2), 123-136.

Olesen, O., N. Petersen (1996) A Presentation of GAMS for DEA, *Computers and Operations Research*, 23(4), 323-339.

Paradi, J.C., S. Vela, Z. Yang (2004) Assessing Bank and Bank Branch Performance: Modeling Considerations and Approaches, in: W.W. Cooper, L.M. Seiford, J. Zhu (eds) *Handbook on Data Envelopment Analysis*, Kluwer, Boston, 349–400.

Parkan, C. (1987) Measuring the Efficiency of Service Operations: An Application to Bank Branches, *Engineering Cost and Production Economics*, 12(1-4), 237-242.

Porembski, M., K. Breitenstein, P. Alpar (2005) Visualizing Efficiency and Reference Relations in Data Envelopment Analysis with an Application to the Branches of a German Bank, *Journal of Productivity Analysis*, 23(2), 203-221.

Prior, D. (2003) Long- and Short-Run Non-Parametric Cost Frontier Efficiency: An Application to Spanish Savings Banks, *Journal of Banking & Finance*, 27(4), 655-671.

Scarf, H.E. (1994) The Allocation of Resources in the Presence of Indivisibilities, *Journal of Economic Perspectives*, 8(4), 111–128.

Sherman, D., G. Ladino (1995) Measuring Bank Productivity Using Data Envelopment Analysis (DEA), *Interfaces*, 25(2), 60-73.

Simar, L., P.W. Wilson (2008) Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives, in: H. Fried, C.A.K. Lovell, S. Schmidt (eds) *The Measurement of Productive Efficiency and Productivity Change*, New York, Oxford University Press, 421-521.

Wilson, P.W. (2008) FEAR: A Software Package for Frontier Efficiency Analysis with R, *Socio-Economic Planning Sciences*, 42(4), 247-254.

Reviewed by Diego Prior (Departament d'Economia de l'Empresa, Universitat Autònoma de Barcelona).

Table 1: Descriptive Statistics of Inputs and Outputs

	Inputs		Outputs (all in numbers)										
	Personnel (number)	**Office space (m^2)**	**Private demand deposits**	**Business demand deposits**	**Time deposits**	**Saving deposits**	**Credit**	**Bearer securities**	**Recourse guarantees**	**Bonds**	**Investment deposits**	**Insurance**	**Contributions to a building society**
Mean	5.42	297.34	1846.91	272.31	37.32	5155.47	124.14	284.68	46.53	95.89	365.73	25.74	47.46
St. Dev.	4.17	213.12	1455.95	265.39	39.15	4086.80	100.01	196.27	44.45	85.79	288.29	26.67	48.81
Skew	1.58	1.71	1.68	2.19	2.78	1.78	1.56	1.48	1.98	2.07	1.79	2.67	2.22
Min.	1.0	64.00	432.00	31.00	0.00	1257.00	6.00	33.00	0.00	7.00	74.00	0.00	3.00
Max.	20.89	1228.00	7851.00	1563.00	285.00	20523.00	499.00	1020.00	271.00	503.00	1673.00	185.00	293.00
Total	769.84	42222	262262	38668	5300	732077	17628	40424	6607	13616	51934	3655	6739

Table 4: Descriptive Statistics of Plant Capacity Inputs and Outputs: Convex vs. Non-Convex; Full Efficiency vs. Full Inefficiency

	Personnel	Office space	Private demand deposits	Business demand deposits	Time deposits	Saving deposits	Credit	Bearer securities	Recourse guarantees	Bonds	Investment deposits	Insurance	Contributions to a building society
Full Efficiency	**Convex**												
Mean	8,13	296,37	2840,26	456,36	70,46	7642,41	212,73	398,15	86,95	155,05	562,67	42,14	71,86
St. Dev.	4,43	210,57	1524,33	306,12	47,83	4370,98	107,50	189,55	57,81	88,98	317,85	25,97	49,99
Min	2,00	64,00	552,00	46,00	0,00	1335,00	6,00	67,00	0,00	15,00	74,00	3,00	14,00
Max	20,89	1228,00	7851,00	1563,00	285,00	20523,00	499,00	1020,00	271,00	503,00	1673,00	185,00	293,00
	Non-Convex												
Mean	6,71	282,08	2308,20	365,72	54,79	6290,96	172,71	333,47	68,27	126,29	453,43	33,35	59,31
St. Dev.	4,95	204,79	1646,01	342,33	56,05	4695,95	124,40	191,35	71,28	98,43	329,60	29,33	52,76
Min	1,00	64,00	471,00	31,00	0,00	1335,00	6,00	57,00	0,00	14,00	74,00	0,00	3,00
Max	20,89	1228,00	7851,00	1563,00	285,00	20523,00	499,00	1020,00	271,00	503,00	1673,00	185,00	293,00

Full InefficiencyConvex

Mean	2568,37	413,68	63,33	6922,23	191,76	358,87	77,81	140,60	509,14	38,37	65,72
St. Dev.	1542,24	305,14	46,58	4405,07	107,96	193,75	55,53	90,37	320,86	26,33	50,41
Min	552,00	46,00	0,00	1335,00	6,00	67,00	0,00	15,00	74,00	3,00	14,00
Max	7851,00	1563,00	285,00	20523,00	499,00	1020,00	271,00	503,00	1673,00	185,00	293,00
Non-Convex											
Mean	2303,87	365,13	54,68	6279,34	172,24	332,71	68,17	126,06	452,57	33,27	59,14
St. Dev.	1646,57	342,48	56,01	4696,89	124,14	191,27	71,31	98,45	329,67	29,32	52,67
Min	471,00	31,00	0,00	1335,00	6,00	57,00	0,00	14,00	74,00	0,00	3,00
Max	7851,00	1563,00	285,00	20523,00	499,00	1020,00	271,00	503,00	1673,00	185,00	293,00

Table 2: Descriptive Statistics for θ_1, θ_2 and CU_{eo}

	Convex			Non-Convex		
	θ_1	θ_2	CU_{eo}	θ_1	θ_2	CU_{eo}
Mean	1,533	1,147	0,801	1,086	1,002	0,939
St. Dev.	0,556	0,204	0,170	0,171	0,016	0,107
Min	1,000	1,000	0,343	1,000	1,000	0,565
Max	3,475	1,982	1,000	1,873	1,133	1,000
# Eff. Obs	33	57	32	97	139	97

Table 3: Li (1996) Test Statistic for Differences in Densities

		Convex			Non-Convex		
		θ_1	θ_2	CU_{eo}	θ_1	θ_2	CU_{eo}
Convex	θ_1	0.000 =	7.728 *	13.013 *	26.211 *	54.730 *	27.061 *
	θ_2	7.728 *	0.000 =	12.804 *	6.672 *	26.543 *	7.693 *
	CU_{eo}	13.013 *	12.804 *	0.000 =	27.074 *	53.955 *	24.841 *
Non-Convex	θ_1	26.211 *	6.672 *	27.074 *	0.000 =	6.205 *	0.506 =
	θ_2	54.730 *	26.543 *	53.955 *	6.205 *	0.000 =	6.215 *
	CU_{eo}	27.061 *	7.693 *	24.841 *	0.506 =	6.215 *	0.000 =

H_0: The two density distributions are equal. Conclusions: * : Reject H_0, = : Accept H_0. Reference values: 1.28 for 10% sign. level, 1.64 for 5% sign. level, 2.33 for 1% sign. level.

Table 5: Basic Short-Run Industry Model Results: Impact of Convexity and Technical (In)efficiency

	Decision Variables	Full efficient $(\alpha = 0)$	Full inefficient $(\alpha = 1)$
Convex	Industry efficiency θ^*	0.588	0.688
	# Full Capacity *w*	106	112
	# Partial Capacity *w*	3	2
	# Zero Capacity *w*	33	28
Non Convex	Industry efficiency θ^*	0.702	0.705
	# Full Capacity *w*	116	116
	# Partial Capacity *w*	2	2
	# Zero Capacity *w*	24	24

Figure 1a: Industry Efficiency measure in Relation to α in Convex Case

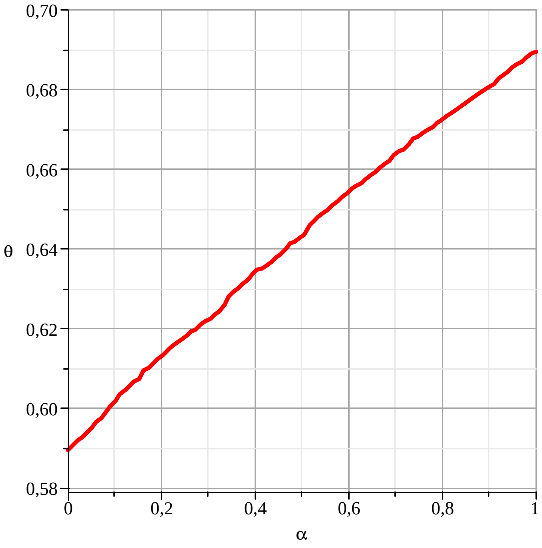

Figure 1b: Industry Efficiency measure in Relation to α in Non-convex Case

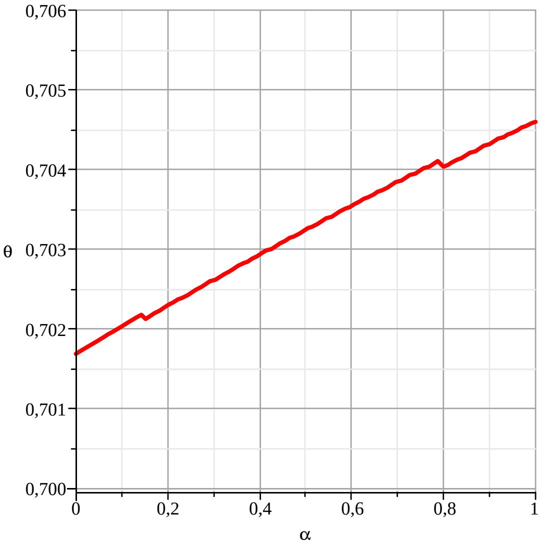

Table 6: Short-Run Industry Model Results: Additional Scenarios

	N					β		Aggregate output targets			
	≤ 95	96	100	117	≥ 118	0.5	1.0	S1*	S2	S3	S4
θ^*	–	0.766	0.722	0.702	0.702	0.711	0.723	0.775	0.702	0.675	0.701
# Full Cap.	–	95	99	115	116	84	78	120	116	108	115
# Partial Cap.	–	1	1	2	2	47	57	1	2	2	3
# Zero Cap.	–	46	42	25	24	11	7	21	24	32	24

* S1: Impose a target value of +10% on the number of saving deposits.
S2: Impose a target value of -20% on the number of saving deposits.
S3: Impose a target value of -20% on the number of bearer securities account.
S4: Impose a target value of -20% on the number of insurance contracts.