INVESTIGACIONES ECONÓMICAS. vol. XXVII (1), 2003, 71-96

BANK COST EFFICIENCY AS DISTRIBUTION DYNAMICS: CONTROLLING FOR SPECIALIZATION IS IMPORTANT

EMILI TORTOSA-AUSINA
Universitat Jaume I

This paper analyzes the dynamics of cost efficiency scores in the Spanish banking industry during the 1985–1995 period and the way in which such dynamics are influenced by specialization. Efficiency is estimated by a non-parametric approach, and a model of distribution dynamics has been applied to assess the evolution of efficiency scores over time. Results show that efficiency at the industry level (mean efficiency) has increased importantly and that, considering how the entire distribution evolves, there are some important patterns that the mean and the standard deviation fail to uncover, such as multi-modality. In addition, more light is shed on efficiency dynamics when each firm's output mix is controlled for, as efficiency scores tend to become closer much faster.

Keywords: Banking, distribution dynamics, efficiency, product mix.

(*JEL* C14, C30, C61, G21, L5)

1. Introduction

The Spanish banking system has gone through a period of profound change over the last fifteen years, mainly as a result of deregulation which, along with other important features such as technological advances or an increasing culture of finance are contributing to significantly reshape the industry. These puzzling trends, together with the traditional importance of the banking system for the economy as a

I am indebted to my advisor, Francisco Pérez, for his help and advice, as well as two anonymous referees for very helpful comments which have greatly contributed to the improvement of the original manuscript. I also wish to thank the Instituto Valenciano de Investigaciones Económicas (IVIE) for kindly providing data and publishing a preliminary version of this article as a working paper (WP-Serie EC 99-15). Financial support from Fundación Caja Madrid, Generalitat Valenciana (GV01-129) and the CICYT (SEC2002-03651) is gratefully acknowledged.

whole, lead us to consider that change to be an outstanding topic of policy relevance and research interest.

One of the most intensively studied topics has been the efficiency of both commercial and savings banks. The aforementioned reasons have led many researchers to analyze how banks' efficiency is affected by this new competitive environment. More precisely, it is widely accepted that the removal of entry barriers forces banking firms to become more efficient, in order to raise industry competitiveness.

However, bank efficiency research studies vary widely in their aims and results, as there are differences regarding the techniques used, the definitions of outputs and inputs, the firms being analyzed (savings banks and/or commercial banks), the sample period uncovered, etc. With regard to the technique, some of them consider a nonparametric approach, while others choose econometric methods. In some cases only savings banks are focussed on, essentially because of their greater homogeneity and reliability of data, while others consider the whole industry, with commercial banks also analyzed. In addition, the period uncovered by research studies varies widely, which constitutes a further source of variation.

But even if a long period is considered, the conclusions drawn on efficiency dynamics are usually based on only two moments of the distribution of the efficiency scores, namely, the mean and the standard deviation. These statistics help a great deal to acquire a general picture of the overall efficiency trends in the industry, but they do not fully inform about the differences at the firm level. In particular, a time-invariant dispersion index may hide very different data structures, which may have important economic implications. Features such as multi-modality, non-normality or asymmetry of the data are impossible to uncover by any dispersion indicator.

In this study, a model of explicit distribution dynamics will enable us to fully characterize the dynamics of cost efficiency scores. And, in particular, due to the tighter competitive conditions currently facing firms, an upward trend should emerge. Yet if, for instance, a pattern such as bi-modality persisted (i.e., some firms being more efficient or inefficient than the average) and relatively *cost* inefficient firms did not abandon the industry, appropriate explanations should be sought out.

Among these, we will consider one which has not been sufficiently stressed in the literature: the role of each firm's output mix. Concerning the Spanish case, very few studies have explicitly considered that different specializations may entail different cost efficiency levels, as some activities are more costly than others (Prior and Salas, 1994). In particular, if different product mixes are not controlled for, an upward bias towards inefficiency could exist. Another way to consider the relevance of the output mix consists of defining different banking output specifications, as it involves assigning greater importance to different specializations. However, despite it being generally accepted that different output measures bias efficiency scores, this has been done only in Grifell-Tatjé *et al.* (1992). Our approach to control for specialization when analyzing efficiency and efficiency dynamics will not be the same, as the model considered to assess the dynamics of efficiency scores requires a different way of controlling for some variables.

The study is structured as follows. Section 2 briefly reviews the sources of debate and controversy when measuring efficiency issues and estimates the cost efficiency for banking firms through a DEA approach. Section 3 introduces the methodology to analyze efficiency dynamics, while Section 4 assesses how banks' output mixes may bias efficiency results and efficiency dynamics. Finally Section 5 concludes.

2. Measuring bank cost efficiency

Research studies on bank efficiency[1] yield a remarkable dispersion in the results achieved, even when applied to the same database and with roughly similar goals. Such a dispersion in results has a twofold source: the technique employed to measure efficiency and what is considered to be produced by banks.

With regard to the former, both parametric and nonparametric methods are the two broad categories to which different efficiency measuring techniques are affiliated. In particular, econometric methods are overwhelmingly used in the parametric case, while linear programming techniques are mostly employed within the nonparametric case. A clear trade-off exists between both methodologies: while econometric models specify a functional form which may entail less flexibility or error specification, linear programming techniques fail to decompose noise and

[1] See Berger and Humphrey (1997) for an excellent survey.

inefficiency.[2] Thus, no methodology dominates which makes the choice somewhat arbitrary and more dependent on the aims pursued.[3]

In this study a nonparametric approach has been considered, partly because we are primarily interested in all features possibly hidden by data, and linear programming techniques tend to envelope data more closely. We will consider the DEA (Data Envelopment Analysis) approach to efficiency analysis, initially developed to compute technical efficiency. However, if prices are considered, the methodology is not exactly DEA but ADEA (Allocative Data Envelopment Analysis), which enables cost efficiency scores to be computed. Such a methodology considers the specification of the following linear program:

$$\begin{array}{lll} Min_{x_{js}} & \sum_{j=1}^{n} \omega_{js} x_{js} & \\ s.t. & y_{is} \leq \sum_{s=1}^{S} \lambda_s y_{is}, & i = 1, \ldots, m, \\ & x_{js} \geq \sum_{s=1}^{S} \lambda_s x_{js}, & j = 1, \ldots, n, \\ & \lambda_s \geq 0, & s = 1, \ldots, S, \\ & \sum_{s=1}^{S} \lambda_s = 1 & \end{array} \qquad [1]$$

where the s firm uses an input vector $x = (x_1, \ldots, x_j, \ldots, x_n) \in \mathbb{R}_+^n$ available at $\omega = (\omega_1, \ldots, \omega_n) \in \mathbb{R}_+^n$ prices in order to produce $y = (y_1, \ldots, y_i, \ldots, y_m) \in \mathbb{R}_+^m$ outputs.

The program [1] must be solved for each firm in each time period. The solution yields the cost minimizing vector x_s^*, given the price vector ω_s and the output vector y_s. Thereby, the cost efficiency score for each s firm is $ES_s = \frac{\omega_s' x_s^*}{\omega_s' x_s}$, which is bounded between 0 and 1. The unity corresponds to the efficient firms (also "best-practice" firms) which make up the efficient frontier.

Frontier techniques constitute the common framework within which bank efficiency is evaluated. However, this type of technique may be

[2] Some authors, such as De la Fuente (1998), argue this inability itself constitutes an assumption, as it implies assuming there is no error. On the other hand, Maudos *et al.* (1998) indicate these are but two distinct problems, namely, one decision regarding whether an error term exist (deterministic vs. stochastic approach), and another one dealing with the probability distribution for the inefficiency component. This opposed views reflect that the topic remains unsettled.

[3] Some research studies comparing both techniques have come to the conclusion that results do not vary dramatically when applied to the same database (Resti, 1997), and when this occurs it can be explained by the intrinsic features of each model.

inadequate for other purposes. For instance, De la Fuente (1998) claims it may be appropriate for a relatively homogeneous industry with a common technology in all firms, but not for a sample of regional aggregates which may not necessarily share the same technology. Similarly, if banks do not share the same specializations, it may be misleading to consider a common frontier for all firms. These arguments partly underlie our paper, and we will account for them, although in a different way, by entering each firm's distinctive product mix in a further stage of the analysis.

Bank output measurement constitutes the second source of debate and controversy regarding the measurement of bank efficiency. Our approach is mostly identified with the intermediation and, more closely, the asset approach, as it only deals with earning assets as bank output. However, it is a common belief that most banks raise a substantial portion of their funds through produced deposits and provide liquidity, payments, and safekeeping services to depositors to obtain these funds. If only the asset approach is considered, these activities will not be precisely captured by the analysis. This might be performed either by including deposits as an additional output or, in order to capture each distinct specialization more explicitly, by comparing firms only with those focusing on the same range of products and services, as adopting different output definitions implies emphasizing different specializations. This will be our approach, which is explained in more detail in Section 4. Also, this strategy allows us to get round the question of whether deposits should be treated as inputs, outputs, or both.

All variables are described in Table 1, which also reports some basic sample information provided by the Spanish commercial banks association (AEB, Asociación Española de Banca) and the Spanish savings banks association (CECA, Confederación Española de Cajas de Ahorro). In addition, in order to attain a homogeneous database, and regarding the important mergers and acquisitions (M&As) process undergone by the Spanish banking industry throughout the sample period, some modifications to the database are required. M&As are dealt with in the literature in different ways. One approach consists of dropping those firms involved in such a process; this, however, would involve ignoring most of the largest banks. This has led us to consider a different approach, namely, to backward sum the merged firms, des-

TABLE 1
Definition of the relevant variables

Variable	Variable name	Definition	1985-1988, mean (Std. dev)	1989-1992, mean (Std. dev)	1993-1995, mean (Std. dev)
Outputs					
y_1	Loans*	All forms of loans to customers	167,391	212,612	300,031
			(376,986)	(485,002)	(716,016)
y_2	Other earning assets*	Securities and loans to financial institutions	182,662	243,411	269,149
			(406,702)	(505,422)	(538,148)
Inputs					
x_1	Labour*	Total labour expenses	7,756	8,966	9,510
			(17,526)	(18,934)	(19,296)
x_2	Funding*	Savings deposits, other deposits, and interbank deposits	385,998	465,214	575,608
			(836,895)	(1,000,162)	(1,276,573)
x_3	Capital*	Physical capital	11,746	15,469	17,905
			(26,372)	(36,545)	(38,104)
Input prices					
w_1	Price of labour	labour expenses/number of employees	4.610	4.804	5.063
			(3.253)	(1.706)	(2.100)
w_2	Price of funds	financial costs/x_2	0.072	0.081	0.068
			(0.064)	(0.029)	(0.022)
w_3	Price of physical capital	(amortizations+other non-interest expenses)/x_3	0.481	0.636	0.639
			(0.671)	(0.985)	(1.034)

*In millions of 1990 pesetas.

pite this being considered a somewhat controversial approach, as fictious firms are created. However, it is the only method which enables us to consider the most extensive part of the system (more than 90 % of gross total assets in all years), ruling out only those firms which were not in continuous existence over the sample period. However, the basic reason behind this approach is that it permits the same firms to be considered at each period, which is important if dynamics are to be characterized. Most other studies with explicit dynamic attempts use the same approach (Fuentelsaz *et al.,* 2002).

The application of program [1] to this homogeneous database yields the results in Table 2. A common frontier has been estimated for each firm and each year. Results show a steady increase in mean efficiency, especially for savings banks, and more intense for unweighted mean efficiency. In particular, savings banks are, on average, much more inefficient than commercial banks at the beginning of the period under analysis (43.93 % vs 59.63 %), but they end up being more efficient (80.25 % vs 77.29 %). In addition, the standard deviation—computed annually for all commercial banks, savings banks, and total banking firms—shows a decreasing trend, and its magnitude is lower for savings banks. These patterns are the result of an ongoing process, with no remarkable ups and downs. Thus, it seems that when considering banking output made up only of earning assets, savings banks are becoming as efficient as commercial banks. This result differs substantially from previous research studies of cost efficiency which, in general, do not find a clear pattern towards an efficiency increase or decrease. However, the definition of output followed in our study is different from those considered in such studies.

TABLE 2
Evolution of efficiency, banking firms (1985-1995)

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
	Unweighted mean	59.63	59.81	61.46	59.87	70.60	70.89	73.34	75.51	73.66	69.49	77.29
Commercial banks	Weighted mean	82.83	81.87	82.46	84.47	87.80	88.29	85.60	89.95	81.34	81.74	88.28
	Standard deviation	21.90	21.70	24.78	24.28	19.70	20.86	19.88	18.53	19.97	22.17	17.73
	Unweighted mean	43.93	43.44	43.73	43.26	58.31	59.42	70.54	76.52	77.18	75.63	80.25
Savings banks	Weighted mean	57.19	58.85	60.44	62.55	75.67	77.26	75.34	84.90	82.09	84.67	85.79
	Standard deviation	17.80	14.44	17.15	16.79	13.99	13.84	9.85	10.56	10.40	10.92	11.16
	Unweighted mean	53.20	52.99	54.19	53.01	65.48	65.99	72.19	75.92	75.11	72.00	78.62
Total	Weighted mean	74.77	74.00	74.93	76.74	83.35	84.25	81.86	88.09	81.61	82.88	87.31
	Standard deviation	21.68	20.61	23.58	22.94	18.51	19.00	16.53	15.71	16.73	18.60	15.13
	# of commercial banks	71	69	71	70	69	66	71	71	70	71	71
	# of savings banks	50	50	50	50	50	50	50	50	50	50	50
	# of banking firms	121	119	121	120	119	116	121	121	120	121	121

3. An alternative methodology for the study of efficiency dynamics

In this section we attempt to explore the dynamics of efficiency scores explicitly. Previous research studies have done this by considering only two moments of their distribution: the mean and the standard deviation (or variance). However, dynamics may be too rich to be precisely captured by such indicators. Specifically, it is often the case that when both the time and cross-section dimensions of the distribution of efficiencies are considered, trends are difficult to identify.

Bearing this in mind, we suggest that a more elaborate technique could be considered. Although the information provided by the aforementioned moments of the distribution provides insights into the evolution of the industry's efficiency, it fails to identify some important underlying changes. In particular, if the industry is undergoing rapid change, maybe not all firms react in the same way. If some specializations are changing, or if some firms are failing in the task of reducing costs, the efficiency scores of such firms could differ widely from the industry average. The standard deviation indicator fails to record these facts as it is unable to capture, for instance, whether or not some modes are present in the distribution.

The limitations of considering only two moments of the distribution have been underlined by different lines of research. This is the case of the literature on stochastic dominance, which states that distributions of stock price changes are inconsistent with the assumption of normal probability functions. It also stresses that the incorporation of higher moments of a distribution and the adoption of alternative approaches to portfolio selection have largely been ignored in favour of the more familiar mean-variance approaches.

However, even if we incorporate higher moments, a full view of efficiency density is not achieved. For this, a direct estimation of the density functions must be performed, which may be done both via parametric and nonparametric methods. As usual, there is a trade-off between both methods: the parametric approach assumes a family of parametric density functions $f(\cdot|\theta)$ such as the normal $\underset{\sim}{N}(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$, and attempts to obtain the best estimator $\widehat{\theta}$ of θ. However, in the nonparametric case the aim is to obtain a good estimator $\widehat{f}(\cdot)$ of all the density function $f(\cdot)$. This approach is more suitable for our goals: since we are trying to get insights on dynamics, along with the usual descriptive statistics, we may also consider how the densities of efficiency scores vary over time, without making any assumption on their shapes. This requires them to be estimated via nonparametric methods.

Dynamic analysis is further improved if one evaluates not only how densities evolve over time, but also whether firms change in their relative efficiency scores. This is quite plausible in an industry undergoing rapid change, where firms may be more, or less, efficient than others at any particular moment, but this ranking may not stand still. Hence, it is of interest to capture how this ranking varies over time, and what the long term distribution of efficiency scores might be.

Therefore, our analysis will be based on two elements: the analysis of the cross-section distribution of the efficiency scores, along with an explicit attempt to model distribution dynamics and to characterize their long term behaviour. This type of analysis follows from Quash (1993a,b), who also pointed out the limitations of confining the analysis of the time evolution of one variable—in his case, per capita income—to summary statistics.

Before proceeding, we normalize efficiency scores. This enables the effect on each firm of the variable's behaviour to be corrected for all firms in the industry, which may exhibit generalized fluctuations. But it is also useful because it allows the distorting effects of outlying observations, to which nonparametric methods to measure efficiency are particularly sensitive, to be offset. In addition, it is a necessary modification if results in this section and Section 4 are to be compared. Accordingly, the new efficiency scores will be $NES_s = ES_s / \frac{1}{S}\sum_{s=1}^{S} ES_s$, where NES_s are the normalized efficiency scores for all s firms in the sample, $s = 1, ..., S$. The interpretation is straightforward: if $NES_s = 2$ it would indicate that firm s is twice as efficient as the average, while a value of $NES_s = 0{,}5$ means that it is half as efficient.

3.1. Univariate nonparametric density estimation

Hence, the first step to evaluate how the *entire* distribution of efficiency scores evolves over time is to estimate nonparametrically their corresponding density functions in each sample year. The dynamic implications of this analysis are clear: if probability mass tends to be more markedly concentrated around a certain value, convergence is achieved, namely, (normalized) efficiency scores tend to equalize. If such a value were unity, the outcome would be a convergence process to the average. Although the opposite outcome (divergence) would imply probability mass being increasingly spread across a wider range, there is a broad spectrum of additional outcomes, such as different modes emerging or vanishing, phenomena with strong economic implications.

A variety of techniques to estimate density functions nonparametrically exists. The most simple one is the histogram, but its shortcomings, such as having to choose an origin and the bin width, not always providing an accurate view of data, lead to it being used only as a starting point. In contrast, kernel smoothing[4] provides a way of uncovering data structure much more accurately, while maintaining the advantages of the histogram—i.e., without imposing any parametric structure on the data. It is not the only method for uncovering data structure and, according to Silverman (1986), it is not always the best. But it is the most widely applicable in many situations, its attributes are easily understood and its discussion permits a better understanding of other methods of density estimation.[5]

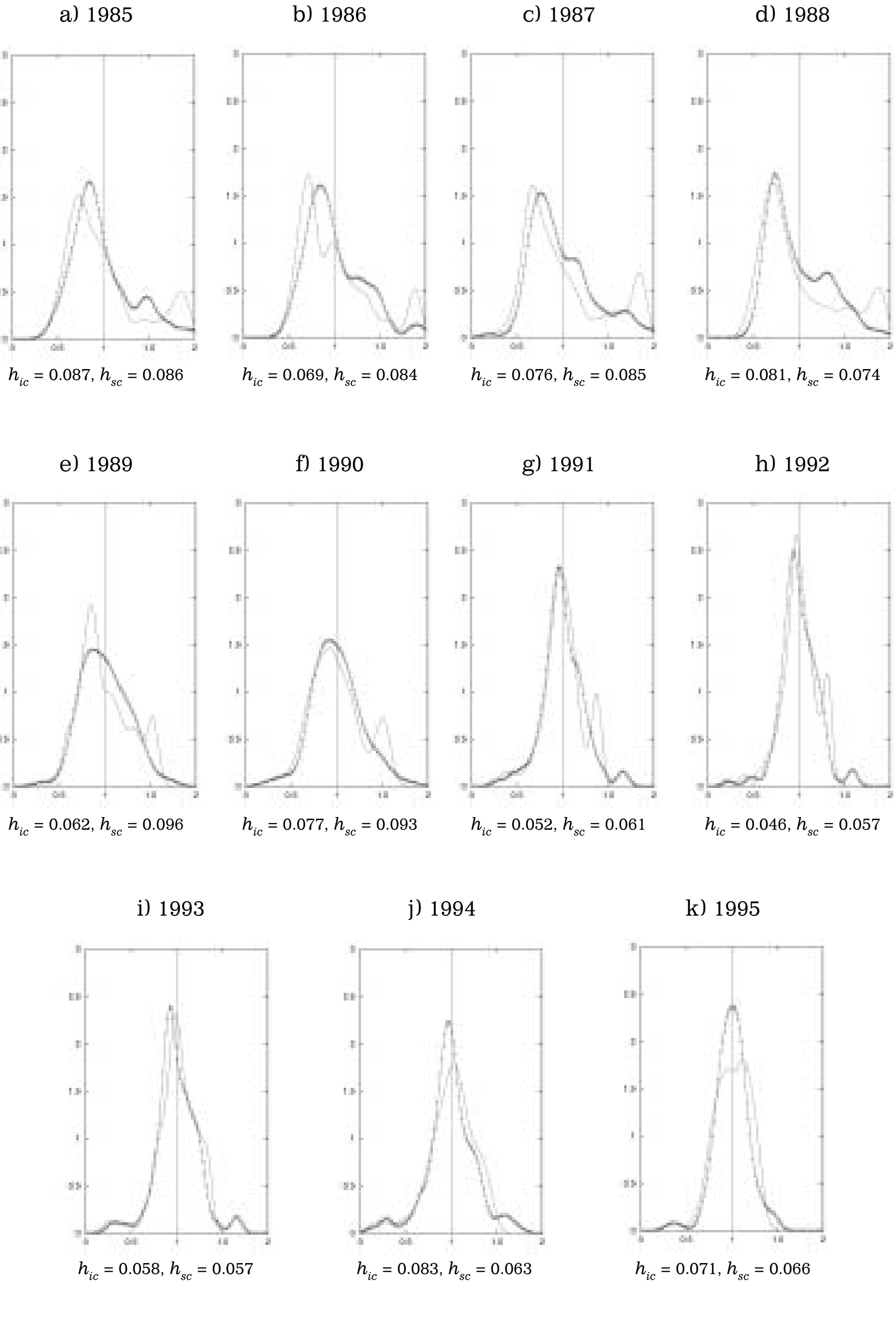
Kernel smoothing has been applied to the efficiency scores computed in Section 2.[6] Results appear in Figure 1 (solid lines), showing the time evolution of the distribution of the efficiency scores for all sample years. The most important feature is a steady trend towards convergence in efficiency, as probability mass tends to be increasingly more concentrated around unity, despite an apparent deceleration in the process since 1993. The interpretation is straightforward: banking firms' efficiency scores are increasingly closer to the industry average, although the features dominating the initial situation (Figure 1.a) are not only a much higher dispersion level (the standard deviation was 0.407 in 1985 and 0.192 in 1995) but also a very different shape from the final situation (Figure 1.k). Hence, while in 1985 multi-modality is clear, it has almost vanished by 1995. Thus, the initial scenario reveals the existence of a group of firms which were much more efficient than the industry average which, in time, approaches it.

[4] Some of the most interesting monographs related to this topic are those by Silverman (1986) and Wand and Jones (1995).

[5] Such as the naive estimator, the nearest neighbour method, the variable kernel method, the orthogonal series estimators, the maximum penalized likelihood estimators, etc.

[6] For this, we chose the Gaussian kernel. The bandwidths, or smoothing parameters, were selected in accordance to the solve-the-equation plug-in approach suggested by Sheather and Jones (1991).

FIGURE 1
Normalized efficiency densities (industry- and specialization-conditioned)



a) 1985: h_{ic} = 0.087, h_{sc} = 0.086

b) 1986: h_{ic} = 0.069, h_{sc} = 0.084

c) 1987: h_{ic} = 0.076, h_{sc} = 0.085

d) 1988: h_{ic} = 0.081, h_{sc} = 0.074

e) 1989: h_{ic} = 0.062, h_{sc} = 0.096

f) 1990: h_{ic} = 0.077, h_{sc} = 0.093

g) 1991: h_{ic} = 0.052, h_{sc} = 0.061

h) 1992: h_{ic} = 0.046, h_{sc} = 0.057

i) 1993: h_{ic} = 0.058, h_{sc} = 0.057

j) 1994: h_{ic} = 0.083, h_{sc} = 0.063

k) 1995: h_{ic} = 0.071, h_{sc} = 0.066

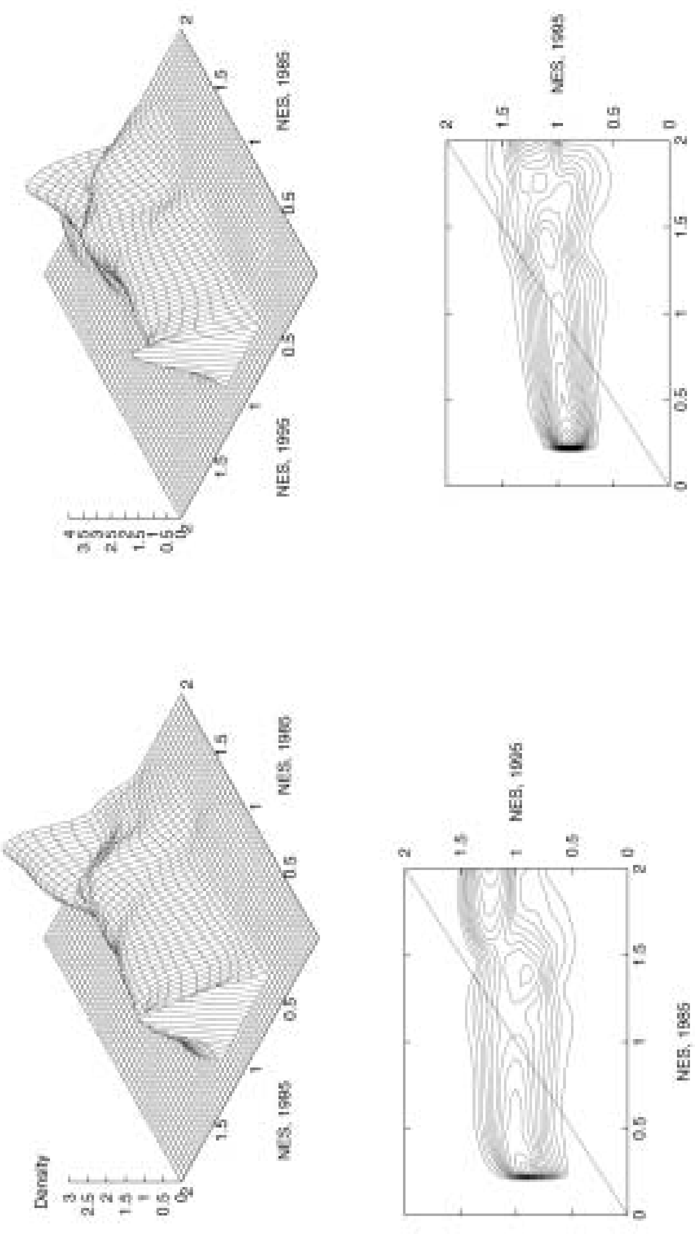
If the analysis is extended to the remaining periods, one notices that the 1995 scenario is the result of a continuous process, with some very efficient firms coming increasingly closer to the industry average and many others below it which, in time, approach it. In sum, initially less efficient firms are increasingly efficient and vice versa, although the initially more efficient firms approach the average more steadily. Thus, there has been a transition from relatively high dispersion and multi-modality to another one where the probability mass tends to be more markedly concentrated about 1. Consequently, it is clear that the behaviour of the cross-section of the distribution cannot be captured only by the mean and the standard deviation, but requires a more detailed analysis of the distributions and their mobility through time.

3.2. Intra-distribution dynamics and long-run trends

Section 3.1 fails to uncover some patterns. Specifically, despite the fact that the dynamic behaviour of a distribution may not offer a clear pattern towards either convergence or divergence, intra-distribution mobility may be present. In other words, changes in firms' relative positions may be taking place without being reflected in the shape of the density functions. In addition, no long-run behaviour of the cross-section distribution can be inferred from the results achieved so far.

In order to approach such questions, a *law of motion* of the cross-section distribution of efficiency scores within a formal structure (i.e., modelling its dynamics) must be obtained. Therefore, one is interested in finding the mechanism or, more precisely, the operator that maps distributions to distributions, i.e., the distribution of efficiency scores at period t to the distribution of efficiency scores at period $t + l$. This mathematical operator comes from traditional Markov process theory, and it is known as the stochastic kernel (Stokey and Lucas, 1989). If that operator M is the identity map, distributions at periods t and $t+l$ are identical, and no intra-distribution movements occur. But even if distributions at both periods are identical, the stochastic kernel may not be the identity because it measures, precisely, intra-distribution dynamics.

FIGURE 2
Stochastic Kernels, nomalized efficiency



Stochastic kernels *inform* us about the different distribution dynamics of the analyzed variables and they will be obtained through nonparametric estimation of bivariate density functions. Thus, assuming that observations refer to each firm's position and correspond to a year, changes in firms' relative positions will be analyzed for two years. More concretely, 11-year transitions will be analyzed, from 1985 to 1995.[7]

Stochastic kernels are depicted in Figure 2.[8] The top relevance of this analysis would emerge if Figure 1 were time-invariant. Such a situation would be fully compatible with changes in firms' relative positions, which could only be detected by means of stochastic kernels.[9] This is not our case, as Figure 1 exhibits strongly dynamic patterns.

The left-hand side in Figure 2 shows firms' mobility (in relative efficiency terms) between periods 1985 and 1995. Through its analysis, especially through the contour plots, it is noticeable that inter-annual mobility is high, as probability overwhelmingly concentrates off the positively-sloped diagonal, with a narrow range of dispersion in 1995 and a much wider one in 1985. Thus, initially more efficient (inefficient) firms than the industry average may end up being as efficient as the initially more inefficient (efficient). Analogously, if probability were concentrated along the positive-sloped diagonal, we would have persistence in firms' relative efficiency positions.

The long run hypothetic distribution —*ergodic distribution*— poses a question which remains unsolved. In order to characterize it, the efficiency scores' observations space must be properly discretized. In such a case, the stochastic kernel M becomes a matrix Q. In other words, M and Q both refer to the stochastic kernels, but in the continuous and discrete case, respectively, and $Q_{r\times r}$ is a transition probability matrix from one state to another, assuming a finite space of states $E = \{e1, e2, \ldots, e_r\}$.

The discretization of the observations' space that the analyzed variable (efficiency) may take into r states e_i, $i = 1, \ldots, r$, allows intra-distribution mobility to be clearly interpreted in such a way that state $e_i = (0.5, 3)$ would include those firms whose relative efficiency ranks

[7] For a similar application, see Lamo (2000).
[8] In this case, we chose the Epanechnikov kernel, and the plug-in bandwidths suggested by Wand and Jones (1994).
[9] Those authors familiar with the concepts of σ and β-convergence may find a sort of parallelism here, as the absence of σ-convergence does not preclude the existence of β-convergence.

between half and three times the industry average. In addition, cell p_{ij} in matrix $Q_{r\times r}$ would indicate the probability of a firm initially in state i of relative efficiency transiting to state j throughout the period or periods (l) being analyzed. In the same way, each row is a probability transitions' vector, which helps us to further understand the analogy to the continuous counterpart: they are equivalent to each probability density for each point in E, when cutting stochastic kernels' figures by a plane parallel to the $t+l$ axis.

Results on transition probability matrices are shown in Table 3. The limits of the states have been chosen to give the same probability in each state (20 %) in the initial period (1985), which enables us to assess whether the ergodic distribution differs substantially from that initial (uniform) distribution. If this transition probability matrix were the identity matrix, distributions would be invariant and, in addition, no intra-distribution movements would occur. In contrast, if probability tended to be more strongly concentrated off the main diagonal then high intra-distribution mobility would exist. Reality confirms this. For instance, in Table 3, the top left-hand entry shows that the most inefficient 20 % of banking firms—those being less efficient than 67 % of the industry average—remained with relative efficiency scores in that range with probability 0.00: all firms (100 %) transited to other states of higher relative efficiency (13 %, 26 % and 61 % transited to states 2, 3 and 4 respectively). The same holds for the following 20 % of most inefficient firms, which abandoned the second state of relative efficiency and moved to the upper-efficiency states (second row in Table 3). Thus, 40 % of the most inefficient firms, those being 77 % less efficient than the industry average, have totally moved to states of higher relative efficiency. On the other hand, the initially most efficient banks show persistence, although at poor levels (less than 50 %).

The probability of a firm ending up in a certain state is indicated by the ergodic distribution, the last row in the table. It shows that the highest probability mass tends to concentrate in the fourth state of relative efficiency (38 %), i.e., ranging between 0.968 and 1.226 times the average efficiency of the industry. This ergodic distribution differs greatly from the initial uniform distribution, suggesting that, in the long run, firms' relative positions will be much closer.

TABLE 3
Industry-conditioned intra-distribution mobility (11-year transitions)

Normalized efficiency					
	Upper endpoint:				
	State 1 0.672	State 2 0.767	State 3 0.968	State 4 1.226	State 5 ∞
State 1	0.00	0.13	0.26	0.61	0.00
State 2	0.00	0.00	0.33	0.63	0.04
State 3	0.08	0.14	0.41	0.32	0.05
State 4	0.05	0.09	0.36	0.36	0.14
State 5	0.06	0.00	0.22	0.33	0.39
Ergodic distribution	0.06	0.09	0.35	0.38	0.12

4. The role of specialization and its influence on efficiency dynamics

Therefore, according to the above results, many banking firms are becoming increasingly efficient, and are approaching the industry average. Explanations for this have been given in Section 1. Yet substantial inefficiency in the industry still persists, a common observation of many other studies. To account for this, different explanations could be explored. We consider that the role of specialization has not been sufficiently stressed in the literature and, in a banking industry undergoing profound shifts, if specializations change, efficiency scores could be affected. This point has been forcefully argued by Maudos *et al.* (2001), who consider that different product mixes entail different efficiency levels, as different products and services require a more intense input use, so that some firms would be mislabelled as inefficient.

In order to assess whether banks' efficiency, and its dynamics, is influenced by specialization, we suggest an approach which first considers which firms focus on the same scope of products and services. Thus, a cluster analysis is performed as if no *a priori* classification of firms exists according to their output mix similarities, multivariate techniques must be applied in order to obtain unbiased groups or clusters. However, regardless of the variables used and the attempts pursued, studies with these attempts face serious problems. Specifically, there is no consensus on the number of clusters to consider across the industry or how to assess their stability over time.

Regarding to the first question, approached in several ways, there is no single criterion which determines the optimal number of groups. A quick glance at some of the research studies using statistical multivariate techniques to segment the banking industry into groups of firms with similarities would confirm this variety of criteria. If we additionally consider the multiplicity of techniques within the multivariate techniques, the variables chosen or the periods being analyzed, the number of groups in the industry may vary widely, as may their membership and stability over time.

Previous research into industries segmented into groups of heterogeneous firms and how this might affect their performance is closely related to the literature on strategic groups. When this type of analysis is performed with a focus on banks, groups tend to be formed through a cluster analysis based on the firms' portfolio composition. This composition is added to the analysis as the relevant variable upon which to perform industry segmentation. Both input and output related balance sheet items are usually considered—although the precise choice of variables varies from study to study—not only because they reveal their specializations or product mix strategies, but also because of their strong links to funding and investment strategies.

Accordingly, Amel and Rhoades (1988) consider 15 variables, including both asset and liability categories, in a similar way to Kolari and Zardkoohi (1987). Studies of Spanish banking firms followed similar approaches, including different asset and liability ratios as rough product mix indicators (Gual and Hernández, 1991; Sánchez and Sastre, 1995; Freixas, 1996).[10] We have followed these contributions, with the inclusion in the analysis of the major assets and liabilities categories (as a percentage of the balance sheet): i) cash and balances with the central bank, ii) fixed-income securities, iii) interbank loans, iv) credits to firms and households, v) savings deposits, vi) other deposits, vii) interbank deposits, and viii) issued securities. Together, these items account for more than 90 % of total assets.[11]

[10] However, the disaggregation available for the U.S. banking industry is much higher than that for the Spanish case.

[11] The *circular* 4/91 of the Bank of Spain (CBE 4/91) modified the way in which banking firms disclose information, forcing us to merge both types of public balance sheets. Our selected items, according to CBE 4/91, correspond to: i) *caja y depósitos en bancos centrales*, ii) *deudas del Estado, obligaciones y otros valores de renta fija*, iii) *entidades de crédito* (assets), iv) *créditos sobre clientes*, v) *depósitos de*

Other considerations[12] finally led us to select nine groups which, despite the problems faced when using these techniques, meet our initial requirements, as they exhibit both within group homogeneity and between group heterogeneity, according to firms' product mixes. In sum, firms in each group are more similar to each other with regard to their output mixes than when compared to firms in other groups.

On this point, our clusters are adequate not only from the methodological point of view, but also because they suit some features of the Spanish banking sector. Savings banks are affiliated almost exclusively to four clusters, hence those four clusters identified by Gual and Hernández (1991) for this type of banks emerge. As in Sánchez and Sastre (1995), these groups also include several regional and local commercial banks, *i.e.*, it seems that the geographical expansion restrictions on savings banks up to 1989 affected their competitors' membership. Another group includes virtually all large commercial banks, which have traditionally been regarded as tough competitors, with the exception of two of them (one of which, Banco Popular, has a much closer product mix to that of a savings bank). Two other groups contain the vast majority of Basque and Catalan banks, regardless of their type of ownership (commercial/savings). Two further clusters include the most important foreign banks, and those banks raising a substantial proportion of funds via interbank deposits, respectively. The rationale for our groups is in line with the ideas of Aldenderfer and Bashfield (1984), who suggest that it is often more important to achieve clusters with sound economic significance than to follow a roughly objective criterion.

The other controversial decision in cluster analysis regards the stability of groups over time. However, although our groups are stable according to the Amel and Rhoades (1988) rule, changes in group membership are unavoidable, as the analyzed period has undergone a remarkable process of changes in firms' product mix strategies. Some studies for other countries confirm this assertion, by finding several Stable Strategic Time Periods (SSTP) throughout the sample period.[13]

ahorro, vi) *otros débitos*, vii) *entidades de crédito* (liabilities), and viii) *créditos representados por valores negociables.*

[12] More technical issues regard the distance measure employed—in our case the Euclidean squared distance—or the method to form clusters—the Ward method, which minimizes the intra group variance.

[13] See Fiegenbaum and Thomas (1990) for a definition of a SSTP.

Once firms have been satisfactorily clustered into different groups, drawing conclusions on how specialization influences the efficiency convergence process requires a slight modification of the variable of analysis. In particular, while in Section 3 efficiency scores were divided by the industry average, in this case they are divided by each firm cluster's average, in the following way, i.e., $NES_{sk} = ES_{sk}/\frac{1}{S_k}\sum_{s=1}^{S_k} ES_{sk}$, where NES_{sk} is the normalized efficiency score of firm s affiliated to cluster k relative to its cluster, ES_{sk} the efficiency score of firm s affiliated to cluster k and S_k the number of firms in cluster k, $k = 1, \ldots, 9$. The interpretation of the new variables of study is different, as we now control for the cluster average instead of industry average. In this case, if convergence occurs in the way described above, we will have convergence to the cluster average. In other words, we are not considering *absolute* convergence but *conditional* convergence. This means that other factors exist apart from industry wide factors, which are more important in the convergence process, such as product mix factors.

The densities plotted with lines-points in Figure 1 are specialization-conditioned (or product mix-conditioned) counterparts to those plotted with solid lines. The first conclusion reached on studying them is that they are tighter and more concentrated than those of industry-relative efficiency scores. When 1985 standard deviations are compared, conditioning on specialization gives a reduction of 10.32 % over that of industry-relative series. Also, it can be observed that the strong bi-modality in the industry-relative series almost disappears for its product mix-relative counterparts, suggesting that efficiency differences tend to diminish when comparing firms with similar output mixes. In addition, although after controlling for specialization some differences still persist, we must bear in mind that the cluster analysis has been carried out on a 1995 basis, and that specializations have changed remarkably throughout the period. Specifically, Pérez and Tortosa-Ausina (2002) show that the nine groups considered are much more similar in their specializations in 1995 than in 1985. Hence, convergence in efficiency could be partly driven by convergence in specialization. Thus, if the cluster analysis were performed for each year separately, the conditioned analysis would probably yield much tighter densities for every period.

Shifts in estimated densities from industry-conditioned to specialization-conditioned efficiency series are presented in Figure 3 by subtracting industry-relative density estimates from the specialization-conditioned estimates. A Kolmogorov-Smirnov test of the equality of both distributions indicates that the shifts in the specialization-conditioned distributions are statistically significant at the $p = 0{,}01$ level.

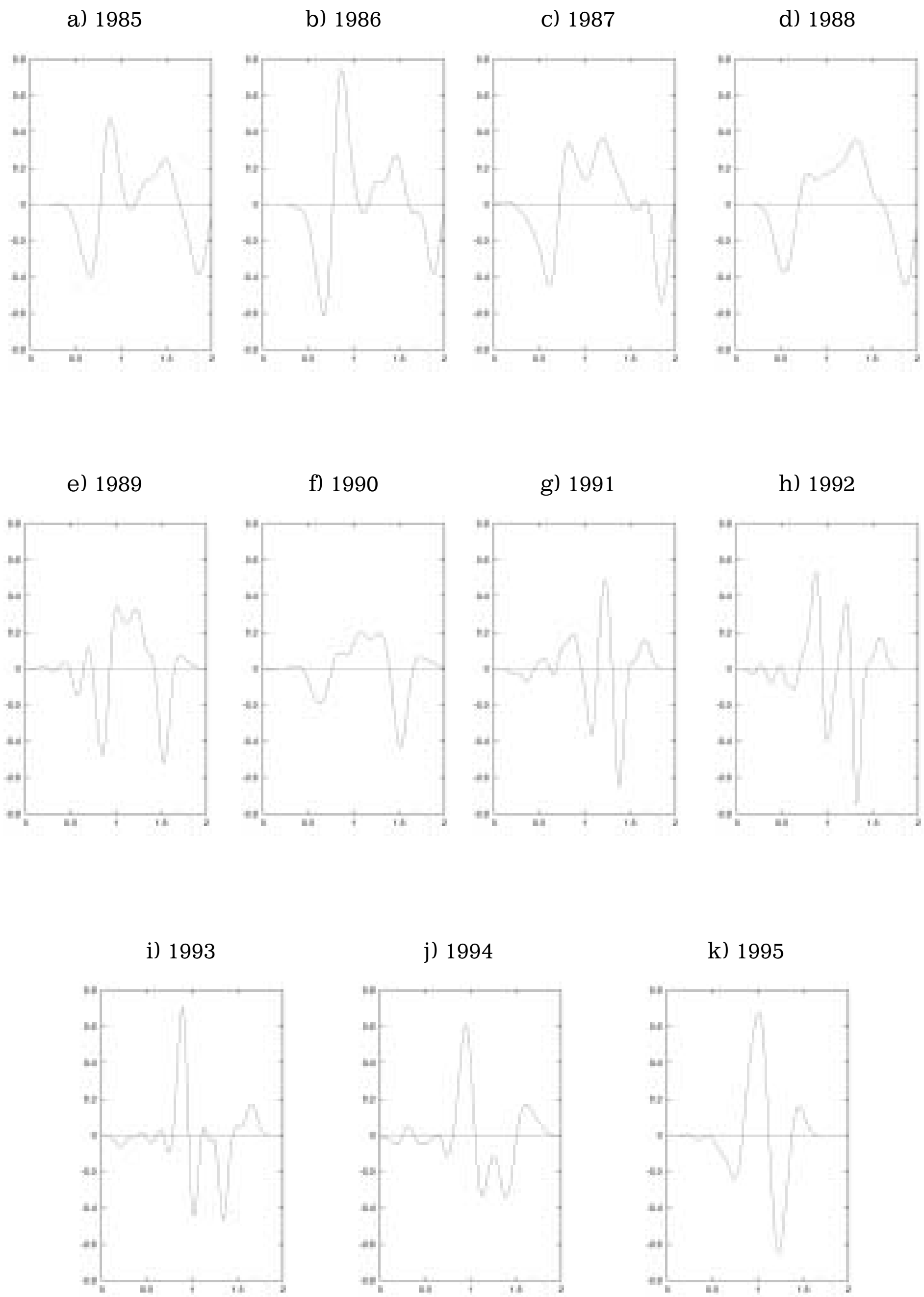
Figure 2.b is the product mix-conditioned counterpart to Figure 2.a. It also shows that persistence is quite low. Probability tends to abandon the positive sloped diagonal and to be more concentrated along the horizontal axis (1985) than the vertical axis (1995). Consequently, firms' relative positions are less dispersed in 1995, and this occurs to a greater extent than in the industry conditioned counterpart figure (see Figure 2.a). However, it is not clear that strong dynamic patterns exist, as in this product mix-conditioned case, firms' initial relative efficiency scores are less dispersed.

Table 4 is the discrete counterpart (transition probability matrix) to the stochastic kernels in Figure 2.b. It shows that probability tends to concentrate more markedly in a unique relative efficiency state. In particular, transitions show that 49 % of probability mass will tend to concentrate in state 4 (ergodic distribution). On the other hand, probability in state 1 of relative efficiency is only 6 %. This, again, turns out to be of paramount importance, as probability in 1985 was uniformly distributed (20 %) across the five considered states.

TABLE 4
Specialization-conditioned intra-distribution mobility
(11-year transitions)

Normalized efficiency					
	Upper endpoint:				
	State 1 0.727	State 2 0.846	State 3 0.964	State 4 1.236	State 5 ∞
State 1	0.05	0.09	0.48	0.38	0.00
State 2	0.04	0.13	0.33	0.50	0.00
State 3	0.00	0.05	0.25	0.65	0.05
State 4	0.09	0.26	0.13	0.43	0.09
State 5	0.05	0.00	0.19	0.52	0.24
Ergodic distribution	0.06	0.17	0.21	0.49	0.07

FIGURE 3
Shifts in estimated densities (industry-vs. specialization-conditioned)



5. Final remarks

The analysis of efficiency in the Spanish banking industry has been thoroughly studied during the last few years. Different reasons motivated this research, but they primarily stem from the reshaping of the industry caused by a dramatic deregulation process, important technological advances, significant financial market innovation, a growing financial culture or an increasing internationalization of banking firms.

Results vary from study to study, but an observation common to most of them is that inefficiencies continue to persist over time. However, studies have not explicitly considered how efficiency scores vary over time at firm level, and conclusions generally pertain to the whole industry wide. In addition, how the different specializations chosen by banks might bias both the efficiency scores and their dynamics has not been fully assessed.

In this study a nonparametric approach has been considered to estimate Spanish banking firms' cost efficiency throughout the period 1985–1995. Results show that mean efficiency, according to our definition of bank output, has grown considerably, and more steadily for savings banks. This conclusion is drawn from industry-wide data. But differences at firm level might be very important as, if significant differences persist, industry structure could be affected. However, if individual differences are to be assessed, a somewhat different view is required. In particular, we focus on how distributions of efficiency scores evolve over time and, according to this, what might be the (probable) long run distribution. This type of analysis permits the uncovering of some features of the distribution hidden if only means and standard deviations are analyzed, such as a strong bi-modality at the beginning of the period which has been declining over time, and the fact that the decrease of efficiency inequalities underwent a slowdown in 1995.

In order to assess how different specializations might bias efficiency dynamics, we have considered multivariate statistical techniques to cluster firms with similar output mixes. According to this, when conditioning each firm's efficiency score on its cluster's average, the convergence process is stronger, as probability tends to concentrate faster around unity. This constitutes an important explanation as to why differences in cost efficiency at firm level may persist, with the relatively inefficient firms not abandoning the industry: different output mixes entail some firms being more costly but, once specializations have been controlled for, not more inefficient.

The picture emerging is that of an industry where cost efficiency levels are increasing both at industry and firm level, but this pattern is significantly influenced by each firm's choice of what products and services it focuses on. Once this has been controlled for, most of the remaining differences in efficiency scores are removed.

References

Aldenderfer, M. S. and Blashfield, R. K. (1984), *Cluster Analysis.* Sage Publications, Newbury Park.

Amel, D. F. and Rhoades, S. A. (1988): "Strategic groups in banking", *The Review of Economics and Statistics* 70, pp. 685-689.

Berger, A. N. and Humphrey, D. B. (1997): "Efficiency of financial institutions: International survey and directions for future research", *European Journal of Operational Research* 98, pp. 175-212.

De la Fuente, A. (1998): "Fronteras de producción eficiente y convergencia regional: un comentario al trabajo de Maudos, Pastor y Serrano", *Revista Española de Economía* 15, pp. 265-271.

Fiegenbaum, A. and Thoma, H. (1990): "Strategic groups and performance; the U.S. insurance industry, 1970-84", *Strategic Management Journal* 11, pp. 197-217.

Freixas, X. (1996), *Los límites de la competencia en la banca española.* Fundación BBV, Bilbao.

Fuentelsaz, L., Gómez, J. and Polo, Y. (2002): "Followers' entry timing: Evidence from the Spanish banking sector after deregulation", *Strategic Management Journal* 23, pp. 245-264.

Grifell-Tatjé, E., Prior, D. and Salas, V. (1992): "Eficiencia frontera y productividad en las cajas de ahorro españolas", Documento de Trabajo 92-1992, Fundación FIES.

Gual, J. and Hernández, A. (1991): "Costes operativos, tamaño y especialización en las cajas de ahorro españolas", *Investigaciones Económicas* 15, pp. 701-726.

Kolari, J. and Zardkoohi, A. (1987), *Bank Costs, Structure, and Performance.* Lexington, Massachusetts/Toronto.

Lamo, A. (2000): "On convergence empirics: Some evidence for Spanish regions", *Investigaciones Económicas* 24, pp. 681-707.

Maudos, J., Pastor, J. M. and Pérez, F. (2002): "Competition and efficiency in Spanish banking sector: the importance of specialisation", *Applied Financial Economics* 12, pp. 505-516.

Maudos, J., Pastor, J. M. and Serrano, L. (1998): "Fronteras de producción eficiente y convergencia regional: réplica", *Revista Española de Economía* 15, pp. 273-279.

Pérez, F. and Tortosa-Ausina, E. (2002): "Product mix clubs, divergence and inequality of Spanish banking firms", *Applied Financial Economics* 12, pp. 431-445.

Prior, D. and Salas, V. (1994): "Eficiencia técnica de las cajas de ahorro españolas y sus factores determinantes", *Papeles de Economía Española* 58, pp. 141-161.

Quah, D. T. (1993a): "Empirical cross-section dynamics in economic growth", *European Economic Review* 37, pp. 426-434.

Quah, D. T. (1993b): "Galton's fallacy and test of the convergence hypothesis", *Scandinavian Journal of Economics* 95, pp. 427-443.

Resti, A. (1997): "Evaluating the cost-efficiency of the Italian banking system: What can be learned from the joint application of parametric and non-parametric techniques?", *Journal of Banking and Finance* 21, pp. 221-250.

Sánchez, J. M. and Sastre, M. T. (1995): "¿Es el tamaño un factor explicativo de las diferencias entre entidades bancarias?", Documento de trabajo 9512, Banco de España.

Sheather, S. J. and Jones, M. C. (1991): "A reliable data-based bandwidth selection method for kernel density estimation", *Journal of the Royal Statistical Society,* Ser. B 53, pp. 683-690.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Stokey, N. L. and Lucas, R. E. (1989), *Recursive methods in economic dynamics.* Harvard University Press, Cambridge, Massachusetts.

Wand, M. P. and Jones, M. C. (1994): "Multivariate plug-in bandwidth selection", *Computational Statistics* 9, pp. 97-116.

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing.* Chapman and Hall, London.

Resumen

El presente trabajo tiene por objeto el estudio de la dinámica de los índices de eficiencia en la industria bancaria española durante el periodo 1985–1995, así como la manera en que dicha dinámica se ve condicionada por la especialización. La eficiencia se estima a través de un enfoque no paramétrico, y se aplica un modelo de dinámica de las distribuciones para valorar la evolución de los índices de eficiencia en el tiempo. Los resultados indican que la eficiencia a nivel de industria (eficiencia media) se ha incrementado y que, considerando la evolución de la totalidad de la distribución, existen patrones importantes imposibles de detectar por la media y la desviación típica, como la multi-modalidad. Asimismo, condicionar por la especialización de cada empresa es relevante al analizar la dinámica de la eficiencia, pues los índices tienden a aproximarse entre sí más rápidamente.

Palabras clave: Empresa bancaria, dinámica de las distribuciones, eficiencias, especificación.

Recepción del original, octubre de 1999
Versión final, abril de 2002