CONVERGENCE IN EFFICIENCY OF THE SPANISH BANKING FIRMS AS DISTRIBUTION DYNAMICS*

Emili Tortosa-Ausina†

WP-EC 99–14

Correspondencia: Emili Tortosa-Ausina, Universitat Jaume I
Facultat de Ciències Jurídiques i Econòmiques, Departament d'Economia
Campus del Riu Sec
12080 Castelló (Spain). Tel.: +34 964 728 606, fax.: +34 964 728 591, e-mail: tortosa@uji.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
First Edition: November 1999
Depósito Legal: V–4233–1999

IVIE working-papers offer in advance results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

*I must show my gratitude to Fundación Caja Madrid and Fundació Caixa Castelló for financial support, and to the Instituto Valenciano de Investigaciones Económicas, for kindly providing data. The usual disclaimers apply.

†Universitat Jaume I

CONVERGENCE IN EFFICIENCY OF THE SPANISH BANKING FIRMS AS DISTRIBUTION DYNAMICS

Emili Tortosa-Ausina

A B S T R A C T

During the last fifteen years the competitive conditions under which Spanish banking firms operate have become much tighter. Deregulation has affected both banks and savings banks, allowing them to expand geographically and to choose a less regulation-conditioned output mix.This paper analyzes how, in these circumstances, banking efficiency has been affected considering two specifications of output and analyzing the dynamics of the entire distribution, not only mean and standard deviation. Results differ depending on the output definition, but they show that, in general, efficiency scores exhibit dynamic patterns that only two moments of the distribution hardly capture. In particular, regardless of the output definition considered, efficiency scores were more dispersed in 1985 and more concentrated in 1995. In addition, the multi-modality of the distributions has almost disappeared at the end of the period.

Key words: Banking, distribution dynamics, cost efficiency, nonparametric density estimation, transition probability matrix.

JEL: C14, C30, C61, G21, L5

R E S U M E N

Durante los últimos quince años las condiciones competitivas bajo las que operan las empresas bancarias españolas se han intensificado considerablemente. La desregulación ha afectado tanto a bancos como a cajas de ahorro, permitiendo la expansión geográfica de estas últimas así como la elección de una determinada especialización menos condicionada por la regulación. Este trabajo analiza cómo, en estas circunstancias, la eficiencia de las empresas bancarias se ha visto afectada, a través de dos aproximaciones del output bancario y analizando la dinámica de la *totalidad* de la distribución, no únicamente media y desviación típica. Los resultados difieren de acuerdo con las distintas aproximaciones al output pero muestran que, en general, los índices de eficiencia muestran patrones dinámicos que sólo dos momentos de la distribución no captan. En particular, e independientemente de la definición del output, los índices de eficiencia estaban más dispersos en 1985 y más concentrados en 1995. Asimismo, la multi-modalidad de las distribuciones prácticamente ha desaparecido al final del periodo analizado.

Palabras clave: Bancos, dinámica de la distribución, eficiencia en costes, estimación no paramética de la densidad, matriz de probabilidad de transición.

1 Introduction

During the last decade, the Spanish banking system has experienced many changes, specially due to deregulation and removal of entry barriers, along with an increasing financial culture or technological advances. Although the pace of liberalization has increased only in the last few years, it is possible to assert that a new competitive environment has emerged, as some of its traditional competitive features –tight regulation, little innovation and barriers to foreign competition– have almost disappeared.

In these circumstances, it is appealing to study how banking firms might react or might be affected by the new competition. One of the most studied issues during the last years has been the efficiency of the banking firms, as the study of scale and scope economies has been proved not to be so important as a source of savings.[1] The Spanish banking system has not been the exception, and several studies have assessed the productivity, technical change, and efficiency of banking firms.

Such studies have considered different approaches in assessing efficiency issues, regarding the technique used (econometric versus linear programming techniques), definition of banking inputs and outputs or source of savings. Although no survey exists in the Spanish case,[2] and thus no comparison of the different results attained has been made, some conclusions are common in such studies, the most important one being that, despite the deregulation and increased competition in the industry, *mean* efficiency does not seem to have experienced major improvements.

In what follows, an analysis of *cost* efficiency of the Spanish savings and commercial banks will be carried out through a Data Envelopment Analysis approach, specifying two definitions of banking output. The rationale for this is clear-cut: although there has been controversy regarding the technique used, only Grifell-Tatjé, Prior and Salas (1992) have considered different output definitions. In a changing competitive environment, where more competition exists, banks might be choosing less regulation-conditioned strategies to face new competition; this might drive firms to experience strong variations in their output mixes, as these are one of the most important features of bank strategies.[3] Thus, as the multi-product nature of the banking firm influences the estimation of the efficiency scores, changes in the different banks' product mixes will bias such an estimation. Taking into account different output measures will allow to get more insights into the role of specialization when estimating efficiency, specially in a context of changing specialization.

But even if two output measures are considered, the analysis of efficiency scores' dynamics requires a further look than only two moments of the distribution. Other studies conclude that substantial efficiency gains have not been attained during the last fifteen years, but such conclusions do not consider that the evolution of the efficiency distribution

[1] See Berger and Humphrey (1991).
[2] Like the international survey by Berger and Humphrey (1997).
[3] This point was forcefully made in Pérez and Tortosa-Ausina (1998).

might hide important patterns, like bi-modality. If that were the case, with a group of firms being increasingly inefficient than average, it could entail important consequences for the industry, as such firms should abandon it. But mean and standard deviation do not inform on such phenomena. Bearing in mind such a limitation, a model of distribution dynamics which considers the *entire* distribution is proposed and devoted to the analysis of efficiency scores' evolution over time.

The paper is organized as follows: section 2 describes some features of the Spanish banking industry. Section 3 is devoted to the estimation of the cost efficiency scores, with a brief exposition of the technique and the model, and specifying the different output definitions considered. Section 4 illustrates the shortcomings of drawing conclusions from only two moments of the distribution, while section 5 shows how to overcome such shortcomings through a different econometric strategy, applying it to the scores of section 3. Section 6 concludes.

2 The Spanish banking industry: from regulation to competition

The 1980s and 1990s have been a remarkable period for the Spanish banking industry from the viewpoint of industrial organization. The strong liberalization process over the last two decades has resulted in record levels of entry and exit in the industry in terms of mergers, failures, office openings, and office closings. Probably, one of the features which makes more attractive the analysis of the banking organizations from an industrial organization viewpoint is the removal of restrictions on geographic expansion. As a result of these dynamics the overall industry structure has changed considerably.

Although it is a commonplace to label as a transition "from regulation to competition" the path followed by the Spanish financial system over the last twenty years or so, it is clearly a justified assertion.[4] The liberalization process has affected the majority of the European financial systems; however, as the tightness of regulation was very different in different countries, not all of them have assumed it equally. The Spanish banking industry was one of the most heavily regulated, thus undergoing one of the most intense deregulation, which we briefly summarize.

One of the most important changes in regulation was that affecting entry (de novo and by merger, but also potential entry) and exit. Although foreign banks are allowed to operate since 1978, under three settlement conditions (representation branches, divisions and branches) it was not until recently (1992) when they were allowed to freely operate as long as the bank satisfied the established conditions in the Second Coordination Directive.[5]

[4]See, for instance, Caminal, Gual and Vives (1988), Canals (1993) or Vives (1990, 1991a, 1991b).
[5]Clearly, this refers only to banks which home state is a member state of the European Union.

The other important issue in the liberalization process we must stress is the abolition of the limits on branching. They have been gradually liberalized since 1974, both for commercial banks and savings banks. However, the reshaping of the industry has come primarily from the removal of legal restrictions against geographic expansion by savings banks, which has allowed them not only to open branches in other regions but to define less regulation conditioned product mix strategies.

This issue is quite linked to the segmentation of institutions. Although regulation used to affect more tightly savings banks than commercial banks, now both types of firms operate under the same legal requirements and thus are allowed to perform the same operations; however, albeit firms are restructuring their product lines,[6] their output mixes still exhibit strong differences.

Other important regulatory modifications or removals have been those regarding investment requirements, interest rates (completely liberalized in 1987), reserve requirements (which fell from 19% at the beginning of 1990s to 2% nowadays) and capital requirements (which are still one of the highest in the world, as a consequence of the financial crisis).

There are other features which have contributed to the mutation of the system, like the growing role of technology, benefiting many average retail bank customers –the number of ATMs was almost negligible in the early 1980s, while at the late 1990s few people do not use them– or the increasing financial culture, which makes people choose their bank depending on variables other than simple physical proximity.

Although some features as the high concentration still persist,[7] the picture emerging is an industry where the overall level of competition has increased and where banking organizations are reacting differently. An illustration of the raise in the level of competition is the downfall for both commercial banks and savings banks in the mark-ups making up the profit and loss account, although such a pattern is not so clear when considering the traditional indicators of banking profitability (ROE and ROA).

3 The study of X-efficiency in banking

When analyzing the X-efficiency of banking firms, results tend to differ. The sources of dispersion are twofold: the technique used and the definition of inputs and outputs. There is a wide range of approaches available to measure these issues, and no generalized opinion of what is the best method exists. The only consensus we may find consists of the dominance of X-inefficiencies over scale and product mix ones.[8]

[6] See Pérez and Tortosa-Ausina (1998).

[7] And still growing, as recently two of the largest commercial banks have been involved in a merger process.

[8] See Berger and Humphrey (1991).

3.1 Methods to measure efficiency

The approaches to measure efficiency can be divided in two broad categories: parametric and nonparametric. Although both of them contain several sub-categories, the most used methods have been the econometric ones, in the parametric case, and the linear programming ones, in the nonparametric case.

Both parametric and nonparametric approaches have advantages and disadvantages, which use to constitute a source of dispersion. Following Resti (1997), results do not vary dramatically when applying both techniques to the same database, and when this happens it can be explained by the intrinsic features of the models. This is also argued in the most famous comparison of both econometric and linear programming techniques by Ferrier and Lovell (1990): while the latter fails in decomposing noise and inefficiency, the imposition of a parametric structure on technology or distribution of inefficiency of the former does not allow to distinguish specification error from inefficiency.

This trade-off makes the choice of technique somewhat arbitrary, depending on the aims pursued. Indeed, in a recent international survey by Berger and Humphrey (1997) of research studies on financial institutions' efficiency, 69 out of 130 surveyed studies were applications of nonparametric techniques. In this study, a linear programming model has been chosen, as these models tend to envelope data more closely. This possibility turns out to be quite interesting for us, as we are particularly interested (as it will be shown later on) in the structure of the data.

3.2 Two different approaches to measure output

The second source of controversy when studying efficiency in banking, contributing also to increase the dispersion of results, deals with the definition of inputs and outputs. Two main approaches exist to measure banking output: the *production approach* and the *intermediation approach*. While the former considers output composed by the number and type of transactions or documents processed during a certain period, and inputs being only labour and capital (physical inputs), the latter takes as outputs the money value of deposits, loans and securities, while inputs take also into account the financial costs involved by liabilities.

The production approach takes the financial institution as a services' producer for the depositors, and the intermediation approach considers it as a way to channel funds from agents with financial capacity to investors. Therefore, a new trade-off emerges: while the first approach ignores the intermediary nature of the bank, the intermediation approach captures more difficultly how banks produce services. The joint application of both definitions of output would be the desirable choice, but a sufficient database related to the production approach is not available. In addition, there are situations where the application of either definition turns out to be more appropriate: the production approach fits better when measuring efficiency at the plant (office) level, while the intermediation approach is more appropriate at the firm level.

These arguments have driven us to choose the intermediation approach. But the controversy remains even within this approach, due to the role of deposits. Although there is a generalized consensus on considering as output the overwhelming part of earning assets, in the liability side deposits generate an important field of disagreement, as they have both input and output nature. On one hand, they have input nature, as they are necessary in the intermediation task (loanable funds); on the other, they are output, if we consider them as a proxy for the volume of services and payment means offered by a bank (payment and safekeeping services).

3.3 Application to the Spanish banking firms

The linear programming approach to efficiency analysis, known as DEA (Data Envelopment Analysis), was initially developed to compute technical efficiency,[9] which does not requires the prices of inputs. However, if they are considered, the methodology is not exactly DEA but ADEA (Allocative Data Envelopment Analysis).[10] Such a methodology considers the specification of a linear programming problem like the following:

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

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.

In order to compute the efficiency scores, the program (1) must be solved for each s firm in the industry. The solution of such a program will be given by the cost minimizer vector x_s^*, given the price vector ω_s and the output vector y_s. Thus, the efficiency score for each s firm is:

$$ES_s = \frac{\omega_s' x_s^*}{\omega_s' x_s} \tag{2}$$

In a similar way, the inefficiency scores are:

$$IS_s = \frac{1}{ES_s} - 1 \tag{3}$$

The expression (3) show the amount in which firm s costs would be increased for oper-

[9] See Banker, Charnes and Cooper (1984).
[10] See Aly, Grabowski, Pasurka and Rangan (1990).

ating off the efficient frontier.

Considering the arguments of section 3.2, and in order to be comprehensive, two definitions of bank output within the intermediation approach have been considered. The first one (approach 1) considers only the intermediary nature of the banking firm, taking into account in the definition of output the majority of earning assets; the second one (approach 2) considers also the output nature of deposits and other magnitudes which can be considered as proxies of the services' rendering. Therefore, the second definition of output involves somewhat measuring efficiency from the production approach.[11]

Outputs (approach 1):

y_1 = fixed income securities+other securities+interbank loans

y_2 = credits to firms and households

Outputs (approach 2):

y_1 = cash and Bank of Spain+fixed income securities+other securities+interbank loans

y_2 = credits to firms and households

y_3 = savings deposits

Inputs (approaches 1 and 2):

x_1 = labor costs

x_2 = savings deposits+other deposits+interbank deposits

x_3 = physical capital

Input prices (approaches 1 and 2):

ω_1 = labor costs/number of workers

ω_2 = interest costs/x_2

ω_3 = cost of capital/x_3

Two reasons suggest the non convenience of including the item "cash and Bank of Spain" in the first approach: it has partly a compulsory nature, as it includes the deposits of banking firms in the Bank of Spain; in addition, it is more related to the services' rendering, as the firms more specialized in retail banking show higher values of this item.

3.4 Empirical results

In order to have a homogeneous database, and regarding the important mergers and acquisitions' process undergone by the banking industry, some modifications have been required.

[11] However, the options chosen by the different studies differ much. Some of them consider only the input nature of deposits (Sealey and Lindley, 1977; Berger and Mester, 1997a, 1997b; Mester, 1997; etc.); others consider only their output nature (Aly, Grabowski, Pasurka and Rangan, 1990; Berger and Humphrey, 1991; Berg, Førsund and Jansen, 1992; Ferrier and Lovell, 1990; Pastor, 1995, 1996; etc.), while others consider both their input and output nature (Aly, Grabowski, Pasurka and Rangan, 1990; Bauer, Berger and Humphrey, 1993; Maudos, 1996; Maudos, Pastor and Pérez, 1997; etc.).

Mergers and acquisitions are faced by the literature in different ways. One approach consists of dropping those firms involved in such process; this, however, would entail ignoring some of the largest banks. This has led us to consider a different approach, namely, to backward sum the merged firms, despite this is considered a somewhat controversial approach, as fictitious firms are created. However, it is the only method which allows considering the overwhelming part of the system (more than 90% of gross total assets), ruling out only those firms starting or ending up their operations throughout the sample period.

Tables 1 and 2 report the time evolution of both simple and weighted mean efficiency for commercial banks, savings banks and all banking firms, according to the results of the application of the program (1). Some interesting results emerge from their observation.

Firstly, results differ depending on the approach considered. Taking into account only the intermediary nature of the banking firm (approach 1, table 1), a steady increase in the efficiency can be appreciated, specially in the savings banks case, and more intense in the case of the simple mean than the weighted mean. Indeed, table 1 show that, starting from a much lesser mean efficiency (43.93%) than commercial banks (59.63%), savings banks end up the period being, in average, more efficient (80.25% versus 77.29%). This is the result of a very continuous process, with no strong fluctuations in the time evolution.

On the other hand, the efficiency scores estimated through the second approach (table 2) show a different pattern: although mean efficiency has increased comparing only the initial and final years, the observed tendency shows a much more restrained increase than the first approach. In this case, the commercial banks are the institutions that have experienced larger inefficiency decreases, while savings banks show similar values both at the beginning and at the end of the period.

Thus, it seems that savings banks are still more efficient than commercial banks in the field where they have been traditionally more specialized, i.e., in the services' rendering, more reflected in the second approach variables. But, in addition, savings banks have become also the most efficient firms when considering the "pure intermediation" approach.

When considering weighted means, thus considering firms' size, the emerging picture is similar: according to the first approach, savings banks end up the period being as efficient as commercial banks, while these are unable to overtake the first firms considering the second approach. Size biases the estimation of the weighted mean, thus any outlier observation in the largest firms influences it dramatically. It seems, therefore, that giving more importance to the largest firms, the efficiency of commercial banks and savings banks, specially according to the second approach, is much more similar.[12]

[12] We have considered relevant performing the analysis both on a weighted and an unweighted basis. Although many research studies consider only the former, an unweighted basis allows extracting meaningful amounts of information from banks of all sizes, rather than having the analysis dominated by the largest banks. See Berger and Mester (1999).

Table 1: Evolution of efficiency, banking firms (1985–1995) (*approach 1*)

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
	Simple 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
	Simple 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
	Simple 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

Table 2: Evolution of efficiency, banking firms (1985-1995) (*approach 2*)

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
	Simple mean	71.58	71.61	74.80	69.32	78.94	79.27	76.32	78.12	76.50	73.92	81.39
Commercial banks	Weighted mean	86.95	86.34	86.79	74.11	89.63	87.94	89.81	90.45	82.83	89.20	89.98
	Standard deviation	17.27	17.26	17.34	16.16	17.16	18.81	19.61	17.94	19.53	21.92	15.89
	Simple mean	83.48	79.81	80.92	75.80	81.53	82.77	81.68	84.31	83.16	82.44	87.78
Savings banks	Weighted mean	87.84	85.67	86.58	81.47	88.62	90.32	85.74	90.01	88.63	89.45	92.21
	Standard deviation	9.71	10.19	11.01	12.12	10.36	10.39	11.40	10.03	9.93	10.29	8.68
	Simple mean	76.46	74.94	77.31	72.00	80.02	80.76	78.51	80.66	79.25	77.41	84.27
Total	Average mean	87.23	86.11	86.72	77.62	89.26	88.81	88.41	90.29	84.89	89.30	90.85
	Standard deviation	15.74	15.30	15.32	14.92	14.71	15.81	16.89	15.46	16.54	18.51	13.47

4 Evolution of the efficiency: are two moments of the distribution enough to draw conclusions?

Up to now, the conclusions on the evolution of efficiency have been drawn according to the behavior of one moment of the efficiency distribution, namely, mean (and weighted mean). However, in an environment of increasing competition it is not obvious that tendencies were generalized, and a high level of dispersion could be underlying. Consequently, standard deviation can help in acquiring some more insights into efficiency dynamics. According to the first output definition it seems to have been falling steadily since 1987. But the same conclusion cannot be drawn from approach 2 to output measurement, as in this case no clear tendency exists.

However, a time-invariant standard deviation of efficiency might be hiding a distribution with, for instance, high multi-modality. In the banking industry such a feature turns out to be of a paramount importance. In a strongly regulated industry, which precludes the entry of new competitors and does not permit the geographical expansion of some of its institutions (savings banks), inefficiencies might be persistent. If deregulation occurs, such inefficiencies should disappear, as less inefficient firms would be unable to face the new competitive pressures.

Mean and standard deviation do not inform about such phenomena. Data might be strongly non-normal or multi-modal, something involving important implications. In particular, a unimodal distribution of efficiency which turns into a bimodal distribution over time could be reflecting a group of inefficient firms becoming increasingly inefficient. Such firms should abandon the industry. If this does not occur, different explanations can be explored.

Particularly, as mentioned above, we might consider that efficiency scores depend on what we are thinking banks produce. Different output definitions involve considering different specializations. In such a case, a firm might show very different efficiency scores, according to various definitions of what bank output is. Thus, the persistence of a very low efficiency score would be possible in a strongly competitive industry if such a firm has a high efficiency score according to other output definition which gives more importance to other specializations.

5 A new approach to study the dynamics of the efficiency scores

In order to analyze the time evolution of an economic variable, and its tendency towards convergence or divergence, the literature on inequality and economic convergence provides appropriate instruments. Quah (1993a, 1993b, 1996b, 1997) has tried to connect such areas, pointing out their shortcomings and introducing a new econometric approach.[13] Such an approach is used to study what in probability terms is known as random fields, which is based in considering the entire distribution of the analyzed variable, both in its time and cross-section dimensions.

5.1 Defining a new efficiency score

The alternative econometric strategy we are considering requires the normalization relative to the mean of the variable to be analyzed, i.e., the efficiency scores (according to both output definitions). This exercise permits isolating shocks that could bias the analysis, which turns out to be particularly interesting, as *mean* cost efficiency has been growing (specially according to the first definition of output).

The new efficiency scores are:

$$NES_s = \frac{ES_s}{\frac{1}{S}\sum_{s=1}^{S} ES_s} \tag{4}$$

where ES_s are efficiency scores and NES_s are the normalized efficiency scores for all s firms in the sample, $s = 1, ..., S$. This normalization will be carried out for both output definitions, and the interpretation of the "new" efficiency scores is straightforward: if $NES_s = 2$ it would indicate that firm s is twice efficient than the average, while a value of $NES_s = 0.5$ means that it is half efficient.[14]

Once the variables to be analyzed have been determined, a three dimensional plot (figure 1) to see how they evolve both in the cross-section (firms) and in time (years) dimensions is drawn. From this plot it is clear than the dynamics of the efficiency scores are hardly captured by only two moments of the distribution like mean and standard deviation.

The three-stage methodology to be presented captures much more precisely the dynamics of figure 1, which might be very rich. It allows to draw conclusions on the behavior of the relevant variables (efficiency for two output definitions) during the analyzed period, along with its long-run features. Bearing this in mind, the study is structured in three stages:

1. The analysis of the cross-section distribution of the variable at different points in time through the nonparametric estimation of density functions.

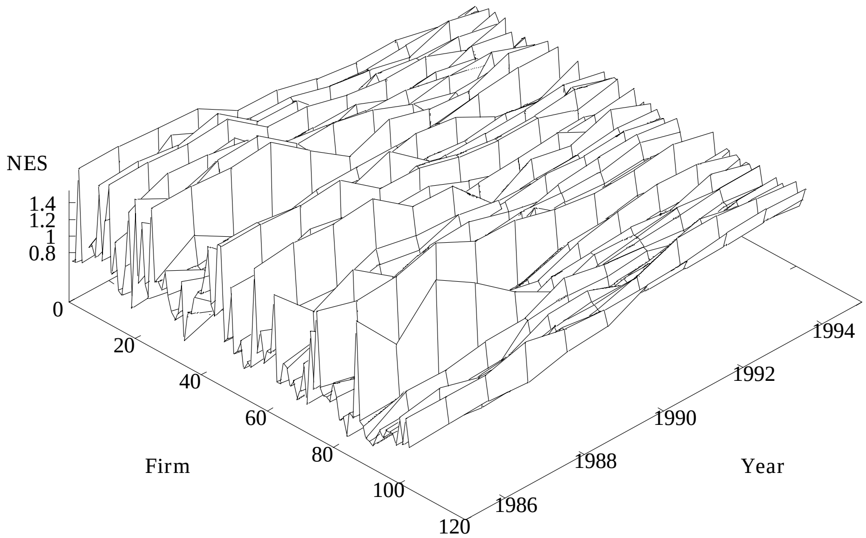
2. Modelling the *law of motion* of such a distribution, i.e., how it evolves over time.

3. Identifying its long-run characterization (ergodic distribution).

[13] See also Andrés and Lamo (1995) and Koopmans and Lamo (1995).

[14] Normalization permits the existence of firms with efficiency scores larger than unity. If such a normalization had not taken place, the maximum value would be (in unitary terms) precisely the unity.

Figure 1: Efficiency evolution, banking firms (*approach 1*)



5.2 Nonparametric estimation of the univariate density functions

The estimation of the cross-section distribution of efficiency at each point in time permits to uncover any particular pattern in its time evolution. As we are primarily interested in the underlying structure of the data, nonparametric techniques are the most suitable. In this context, convergence would exist if the probability mass tended to be gradually more concentrated around a certain value, and if such a value were the unity, then it would be convergence "to the mean".

The analysis to be done is mainly visual and nonparametric. Although the parametric analysis is the most powerful, data might be strongly non-normal, asymmetric or multimodal. In this sense, one of the most important challenges of data analysis consists of uncovering all complexities they could hide and, with such attempts, the parametric approach turns out to be clearly unsatisfactory.

However, relying too much in the visual aspect of data has suffered strong criticisms from a historical perspective, although some of its defendants were rather famous[15]. The first objection the sceptic may argue is natural: this type of analysis is nonsense if graphical representation permits to uncover any intrinsic feature in the data. However, in most situations, as the number of observations increases ($s \rightarrow \infty$) we can see *nothing*.[16] In order

[15] K. Pearson, for example.

[16] Scott (1992) even argues that such an exercise leads to a problem of *too much ink*.

to solve such a problem, data must be smoothed, the histogram being the most simple example of smoothing. Indeed, this is the second objection against the nonparametric approach to estimate density functions: why not simply using the histogram to uncover data structure? Although it is not a bad starting point,[17] it has well-known shortcomings[18] that lead us to choose another way to smooth data.

Specifically, the grounds of our analysis will be the *kernel smoothing*, which is becoming increasingly popular.[19] It provides a way to uncover data structure without imposing any parametric model. This allows us to prevent from features like a bimodal structure (which might have an important economic meaning), impossible to uncover through a parametric uni-modal model.[20]

Kernel smoothing consists primarily of estimating, for both efficiency scores computed, the following density function:

$$\widehat{f}(x) = \frac{1}{Sh}\sum_{s=1}^{S} K(\frac{x - NES_s}{h}) \tag{5}$$

where S is the number of firms being analyzed, NES_s are the efficiency scores (computed according to both output measures), K is a kernel function and h is the bandwidth, window width or smoothing parameter.

There exist multiple options for the kernel selection.[21] In our case the Gaussian kernel has been the choice; its expression in the univariate case is the following:

$$K(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2} \tag{6}$$

Thus, equation (5) becomes:

$$\widehat{f}(x) = \frac{1}{Sh}\sum_{s=1}^{S} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-NES_s}{h})^2} \tag{7}$$

While kernel selection determines the form of the bumps when graphically representing

[17] In addition, it was the only nonparametric density estimator until 1950s.

[18] See Silverman (1986) for an illustration of some of them.

[19] It is not the only available method to approach our problem and, following Silverman (1986), we can conclude that it is not the best in all circumstances. However, it is the most widely used in many situations, their properties are easily understood and their discussion allows to better understand other methods, like the naive estimator, the orthogonal series estimator, or the penalized maximum likelihood estimator. Together with the study of Silverman, other interesting monographs in this field are Scott (1992), Wand and Jones (1995) and Simonoff (1996). In order to approach the topic with more insights, see Devroye and Györfi (1985) or Nadaraya (1989).

[20] The parametric approach and nonparametric approach differ widely. The former, starting from a family of parametric density functions $f(\cdot|\theta)$ like the normal $N(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$, focuses primarily in obtaining the best estimator $\widehat{\theta}$ for θ. In the nonparametric case the focus relies more heavily in obtaining a good estimator $\widehat{f}(\cdot)$ for all the density function $f(\cdot)$.

[21] E.g., Epanechnikov, triangular, Gaussian, rectangular, etc. Given their efficiency levels use to be around 90%, the choice must be based on other considerations, like computing straightforwardness. Anyway, the relevant choice is the bandwidth selection, as it will be shown later on.

function (7), the smoothing parameter h influences it differently, determining the width of such bumps. However, bandwidth selection is much more important than kernel's. If h is too small, an excessive number of bumps is generated, thus being difficult to clearly distinguish data structure. This phenomenon is known as undersmoothing. On the other hand, when h is too large we have oversmoothing, in such a way that some features present in the data (like multi-modal structures) are hidden. Underlying these concepts lies the traditional trade-off between bias and variance which, indeed, depends on the smoothing parameter: as h increases, variance decreases and bias increases, and vice versa.

Prior research studies applying the nonparametric estimation of density functions to the analysis of convergence or time evolution of inequalities hardly emphasize the h chosen. In most of them no mention exists, while others simply indicate that the smoothing parameter has been chosen *automatically*.[22] However, as it has been pointed out, choosing different h's influences significantly the results, which forces us to look for a more suitable bandwidth.

Jones, Marron and Sheather (1996) compare different h's, coming to conclusions stating the importance of this topic. Among them, they state that some first generation methods do not sufficiently smooth data in many circumstances (undersmoothing), while the opposite occurs for others (oversmoothing). Second generation methods offer a reasonable balance between these two extremes or, equivalently, between bias and variance. The higher performance of the second generation method is being increasingly reported in the literature on kernel smoothing.[23]

These arguments lead us to choose the bandwidth proposed by Sheather and Jones (1991) according to the study of Park and Marron (1990). It is based on the second generation method solve-the-equation plug-in-approach, and its superior performance relative to the first generation methods has been further verified.[24] Its known as h_{SJPI}, which has its origins in the authors' names and in the approach followed.[25]

The first stage of the new methodology has been applied to the problem being analyzed, i.e., the time evolution of cost efficiency of the Spanish banking firms. The density function (7) has been estimated for both series of (normalized) efficiency scores, for some years and several-years periods (trying to cover all the period). To be exact, density functions have been estimated for:

1985 and 1995: which allows to visually compare the shape of the cross-section distribution both at the beginning and the end of the considered period.

[22]In such a way Silverman (1986) refers to h_{LSCV} (*least squares cross validation*), which leads us to consider that it is the chosen bandwidth.

[23]See, for example, the simulation studies by Cao, Cuevas and González-Manteiga (1994) or Park and Turlach (1992).

[24]See Jones, Marron and Sheather (1996).

[25]Details on its estimation are not reported, as it involves considering too many technicalities; these can be found in Sheather and Jones (1991) and Park and Marron (1990). In addition, the `Matlab` routine which permits its computation is available through Steve Marron *web*'s page (`http://www.stat.unc.edu/faculty/marron.html`).

1990: as it is the year dividing the considered period into two equally-length subperiods. Besides, it was a year when some important changes in the Spanish banking sector took place.

1986–89 and 1991–94: unavoidable, as we are trying to cover the whole time dimension of data.

1985–95: this allows to make comparisons of the results obtained according to both output measures.

Results are plotted in figures 2 and 3. The most outstanding feature according to figure 2 is a steady tendency towards convergence in efficiency scores, as probability mass tends to be gradually more concentrated around unity, even though it seems that the process has undergone a deceleration in 1995 (figure 2.e). The interpretation is straightforward: banking firms' efficiency scores are approaching industry average (figure 2.e), starting from a situation (figure 2.a) with higher dispersion (standard deviation in 1995 is 0.192, while in 1985 it was 0.407) and with very different features from which the final distribution shows. In this way, multi-modality is present in 1985, but such a phenomenon is quite lessened in 1995. The initial situation reflects the existence of a group of institutions much more efficient than industry average which, as time goes by, get closer to it.

Patterns are not the same in accordance with the second approach to output measurement (figure 3). In such a case multi-modality is still present in 1985 (figure 3.a), but in a lesser dispersion context (standard deviation in 1985 was 0.206, while in 1995 it is 0.160). Even though transitions to situations where mean efficiency is more concentrated around unity is a common feature to both approaches, the existence of two strong clubs or clusters in 1991–94 period must be stressed in this case (figure 3.d).

Limiting the analysis of distribution dynamics to only two points in time would be a mistake, as it involves losing a lot of relevant information. If the remaining considered periods are analyzed, we might check that the situation in 1995 is the result of a continuous process, with some very efficient banks gradually less separated from the average and some others which approach it as time goes by (output measures 1 and 2). Summing up, less efficient firms are increasingly efficient and vice versa, although the initially most efficient reach the average faster.

Thus, there has been a transition from a situation with high dispersion and multi-modality to other where the probability mass tends to concentrate around a certain value. Differences have lessened, both in terms of lower dispersion (in accordance with output measures 1 and 2) and disappearance of high multi-modality. Obviously, mean and standard deviation alone are not enough to reflect the rich dynamics of the entire distribution being considered.

Figure 2: Evolution of normalized efficiency density (banking firms) (*approach 1*)

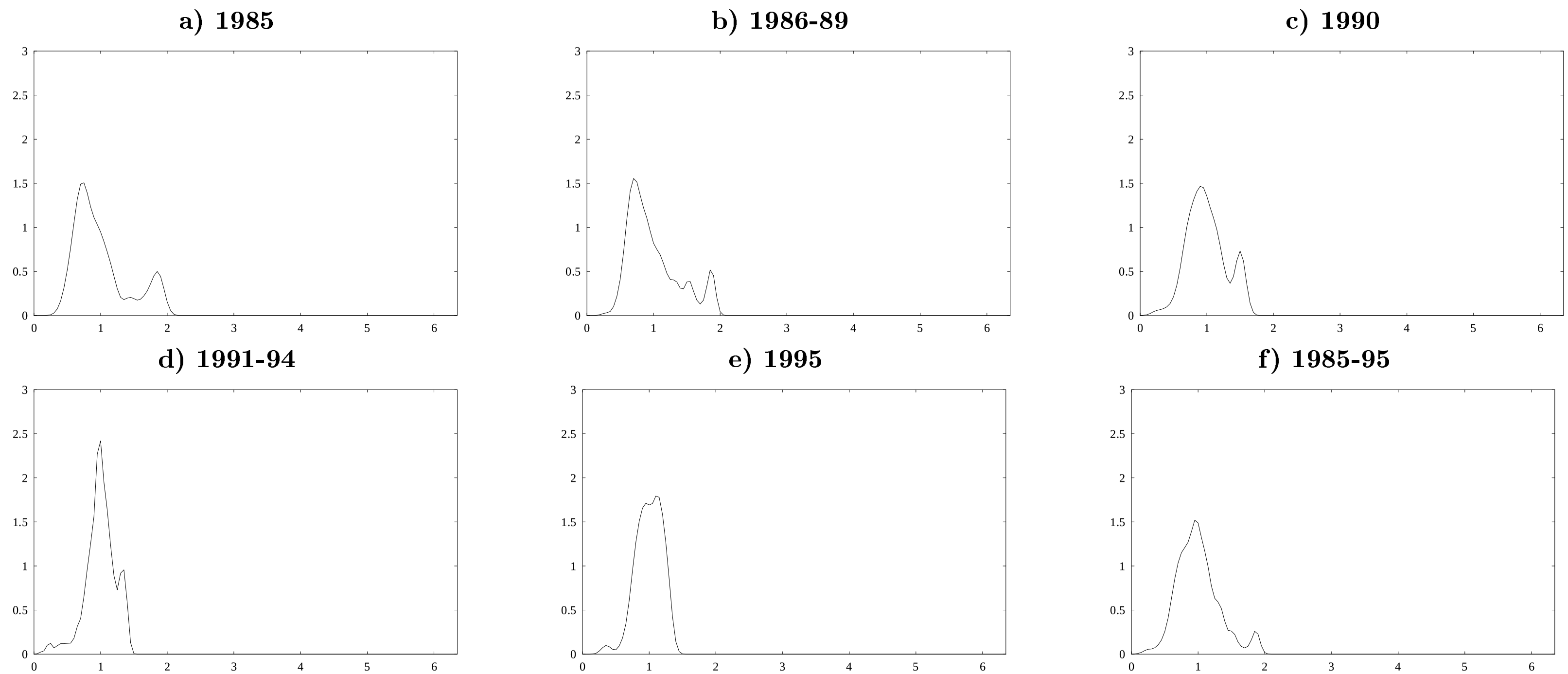
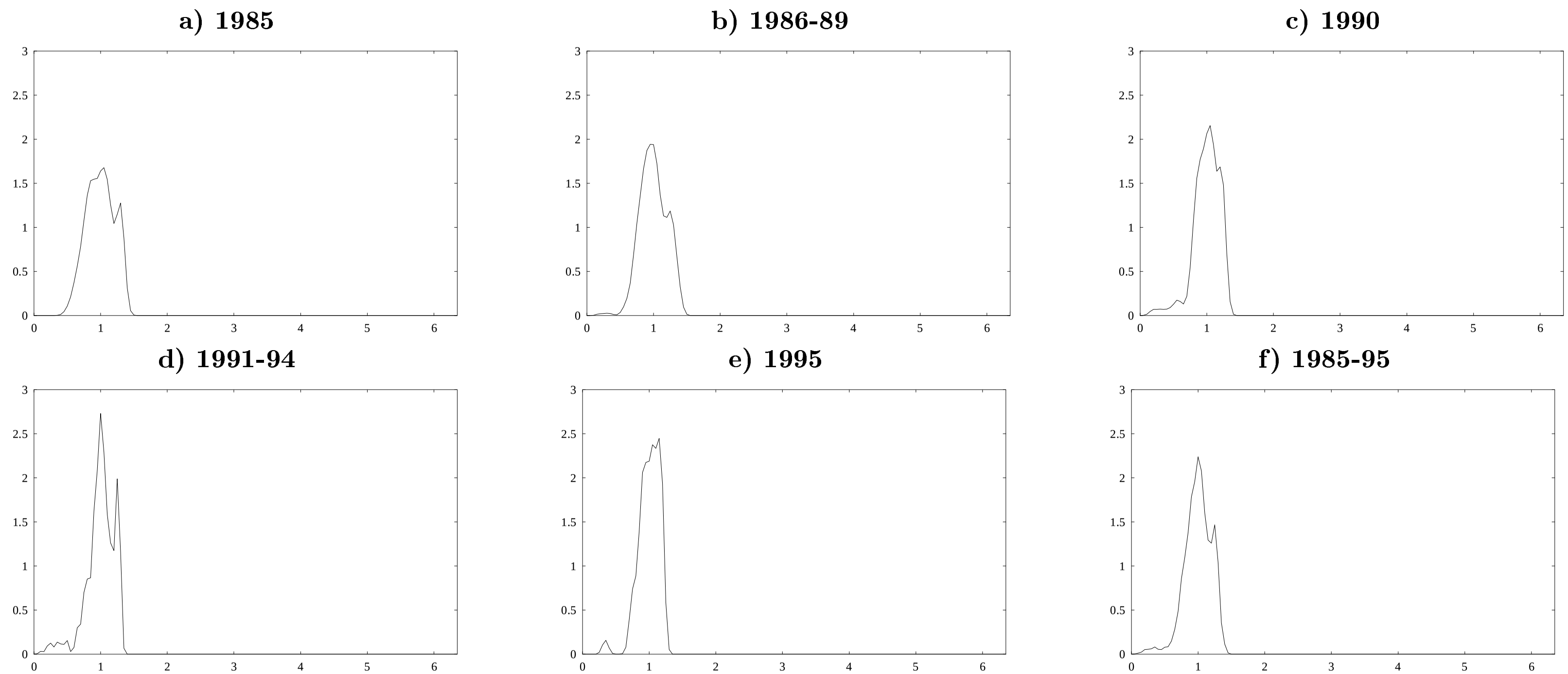


Figure 3: Evolution of normalized efficiency density (banking firms) (*approach 2*)



5.3 Intra-distribution mobility: estimation of the stochastic kernels

The exercise in section 5.2 informs about the dynamics of the distribution, but not completely. It has some limitations that makes desirable a second stage to overcome them. In particular, it can be argued that the dynamic evolution of a distribution might not offer a clear pattern towards convergence or divergence, in the sense described above, but important intra-distribution movements were taking place. In other words, the external shape of the density function might not be affected, but changes in firms' relative positions might be taking place.

In order to overcome such shortcomings, a law of motion of the cross-section distribution is required. Thus, the dynamics can be modelled. Knowing such a law and, therefore, drawing conclusions on the patterns of the variables' cross-section distribution dynamics, need to model of the stochastic process which takes values that are probability measures (λ_t) associated to the cross-section distribution at time t (F_t), where:

$$\forall\ y \in \mathbb{R}:\ \lambda_t((-\infty, y]) = F_t(y) \tag{8}$$

Such an aim enables us to build a formal statistical structure which captures the stated phenomena (intra-distribution mobility and long-run behavior). However, the standard econometric analysis does not provide suitable instruments to model the sequence of distributions' dynamics. Pursuing such aims, we can resort to Markov Processes Theory and establish a duality in order to approach the problem.

The same as transition probability functions describe the dynamics of a scalar process, the *stochastic kernels* describe the dynamics or law of motion of a sequence of distributions.[26] In other words, the equivalent for distributions of the dynamics of a scalar process is being considered.[27]

Let λ_t be the probability measure associated to the distribution of each output specification efficiency scores F_t (one for each output specification) at time t, then the stochastic kernel[28] describing the evolution from λ_t to λ_{t+1} is the mapping M_t to [0,1] of the Cartesian product of efficiency scores and Borel-measurable sets such that:[29]

$$\forall \text{ set } A \text{ Borel-measurable}:\ \lambda_{t+1}(A) = \int M_t(y, A)d\lambda_t(y) \tag{9}$$

[26] Stokey and Lucas (1989), secs. 8.1 and 8.3.

[27] Details on this has been intentionally omitted, due to its excessively technical nature. We have attempted only to provide the ideas such concepts entail. Anyway, we will follow the ideas by Quah (1996a, 1997), Andrés and Lamo (1995), Koopmans and Lamo (1995) and Stokey and Lucas (1989).

[28] It is hard to completely understand the links between the analysis of distribution dynamics and Markov Processes Theory, in general, and the stochastic kernels, in particular. The study by Durlauf and Quah (1998) is the one which more precisely captures such links which, as stated, turn out to be quite complex.

[29] See the technical appendix in Tortosa-Ausina (1999).

Notice that the values equation (9) takes are measures or distributions instead of scalars or finite dimensional vectors. Additionally, assuming M_t time-invariant, equation (9) could be re-written as:

$$\lambda_{t+1} = M * \lambda_t \tag{10}$$

where M is a representation of the stochastic kernel encoding information on how starting with a probability measure λ_t associated to the cross-section distribution F_t we end up in λ_{t+1} (associated to F_{t+1}), i.e., on the different firms' relative positions, which is equivalent to knowing partly the dynamics we attempt to model. Thus, estimation of M from the available data allows empirically quantifying distribution dynamics.

Additionally, considering equation (9) and iterating:

$$\lambda_{t+s} = (M * M * \cdots * M) * \lambda_t \tag{11}$$

This expression allows characterizing (when $s \to \infty$) the ergodic distribution, thus completely characterizing the efficiency scores' distribution dynamics.[30]

The estimation of the stochastic kernels will be based on the nonparametric estimation of bivariate density functions. Thus, assuming each variable's observations correspond to a year or period of years, changes in firms' relative positions between two years or periods of years will be analyzed. In particular, k-year transitions will be analyzed, being $k = 1, 11$.

In the bivariate case, nonparametric density estimation departs again from kernel method. Generalizing to the multivariate case, the function to be estimated is:

$$\widehat{f}(\mathbf{x}; \mathbf{H}) = S^{-1} \sum_{s=1}^{S} K_{\mathbf{H}}(\mathbf{x} - \mathbf{NES}_s) \tag{12}$$

where $\mathbf{H}$ is a $d \times d$ bandwidth matrix (2×2 in the bivariate case) and K is a kernel d-variate function.

In this bivariate case we have $\mathbf{x} = (x_1, x_2)$ and $\mathbf{H} = \mathbf{h} = (h_1, h_2)$, h_1 and h_2 being the bandwidths for each coordinate direction. Thus, the function to be estimated would turn into:

$$\widehat{f}(\mathbf{x}; \mathbf{h}) = (Sh_1h_2)^{-1} \sum_{s=1}^{S} K(\frac{x_1 - NES_{s1}}{h_1}, \frac{x_2 - NES_{s2}}{h_2}) \tag{13}$$

[30]The ergodic distribution should not be considered exactly as a prediction of the future, as future realizations of the variables could be influenced by a wide range of ways. This concept should be more properly considered a characterization of last years' tendencies.

Relative to the kernel, the Epanechnikov kernel has been the chosen:

$$K_e(\mathbf{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\mathbf{x}^T\mathbf{x}) & \text{if } \mathbf{x}^T\mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where c_d is the volume of the unitary d-dimensional sphere: $c_1 = 2$, $c_2 = \pi$, $c_3 = 4\pi/3$, etc.

In what the bandwidth selection is concerned, the state of the art is in a very preliminar stage, much more than in the univariate case. However, the problems associated to its estimation through first generation methods are similar in both bivariate and univariate cases.

In this study the solve-the-equation plug-in approach will be used. Particularly, we will depart from Wand and Jones (1994), where individual smoothing parameters for each coordinate direction are provided which, in general, perform better than least squares cross validation bandwidths and thus have been applied when estimating stochastic kernels.

Intra-distribution mobility of the (normalized) efficiency scores distributions according to both output measures is reported in figures 4 and 5, which graphically represent bivariate density functions estimates for both variables. Each coordinate direction represents a period, and the density functions try to reflect firms' transitions between these two periods.

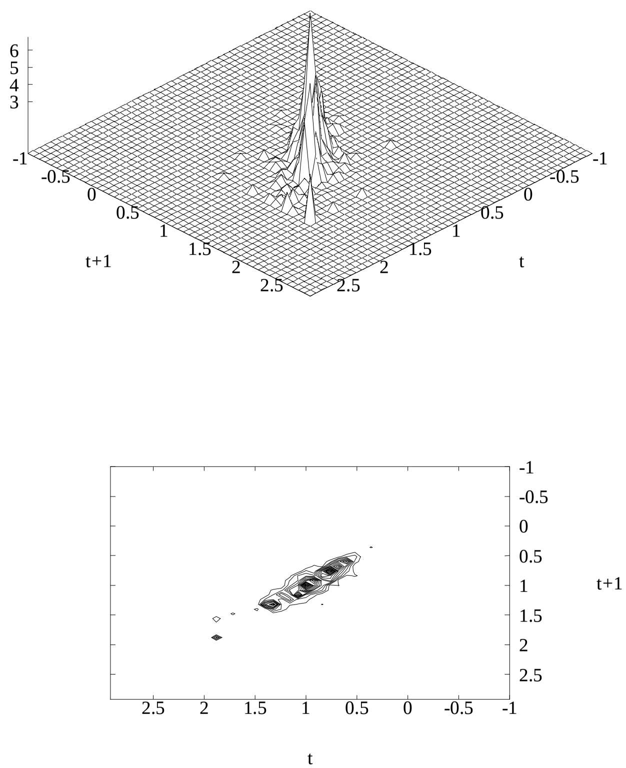
If figures 2 and 3 were time-invariant it would be perfectly compatible with a situation where changes in firms' relative positions were taking place; estimation of the stochastic kernels makes possible identifying such patterns. Although figures 2 and 3 do not show an invariant dynamic pattern, it is desirable to detect whether intra-distribution movements occur.

Figures 4.a and 5.a show intra-distribution mobility between periods t and $t+1$, for all sample periods; thus, they show changes in firms' relative positions between years 1985 and 1986, 1986 and 1987, 1987 and 1988, etc. Through their analysis, and specially considering the contour plots, we may come to the conclusion that inter-annual mobility is rather low for both output measures (even lower for the second output measure). Such a pattern is given by a probability mass concentrated along the positive sloped diagonal in the contour plots, which indicates persistence in the (normalized) efficiency relative positions. Similarly, the probability being concentrated along the negative sloped diagonal (or simply off the positive sloped diagonal) would imply evidence for strong intra-distribution mobility.

The conclusions are not the same when considering figures 4.b and 5.b, displaying transitions for the whole period (11-year transitions) or persistence in the firms' relative poistions. In these cases intra-distribution mobility is high. As contour plots show, probability does not overwhelmingly concentrates on the positive sloped diagonal. To be exact, it is not possible to assert whether it concentrates along either diagonal. Thus, initial relative positions are more disperse than final ones, resulting irrelevant for such final positions, much closer to each other. In sum, initially more efficient (inefficient) firms than average might end up

Figure 4: Stochastic kernels, efficiency (banking firms) (*approach 1*)

a) 1-year transitions



b) 11-year transitions

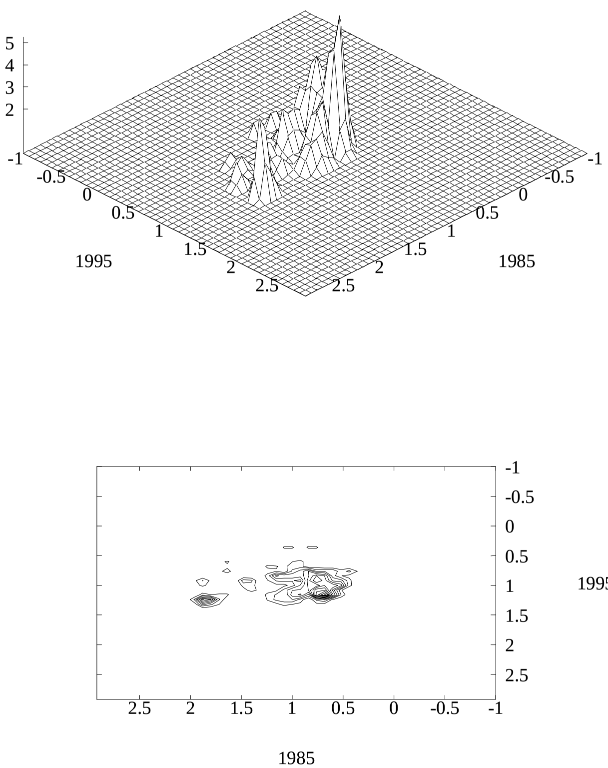
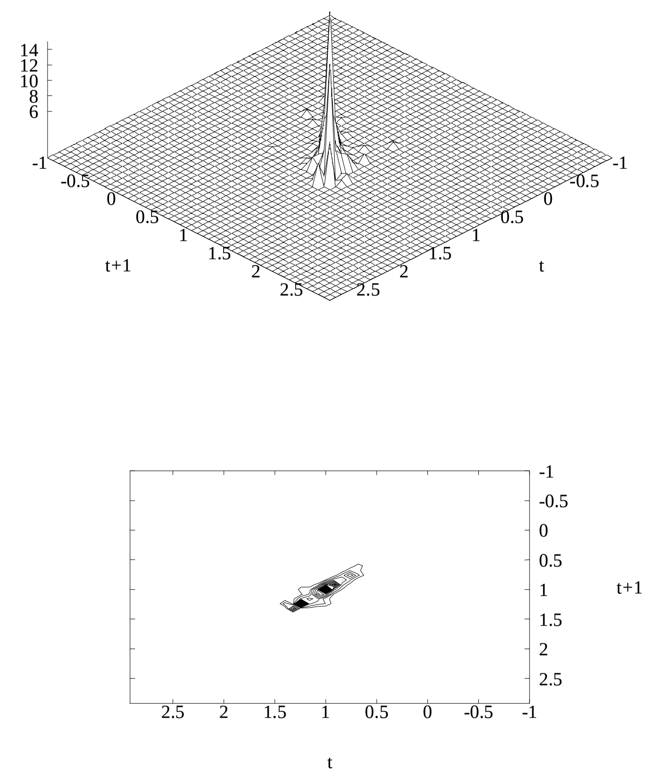
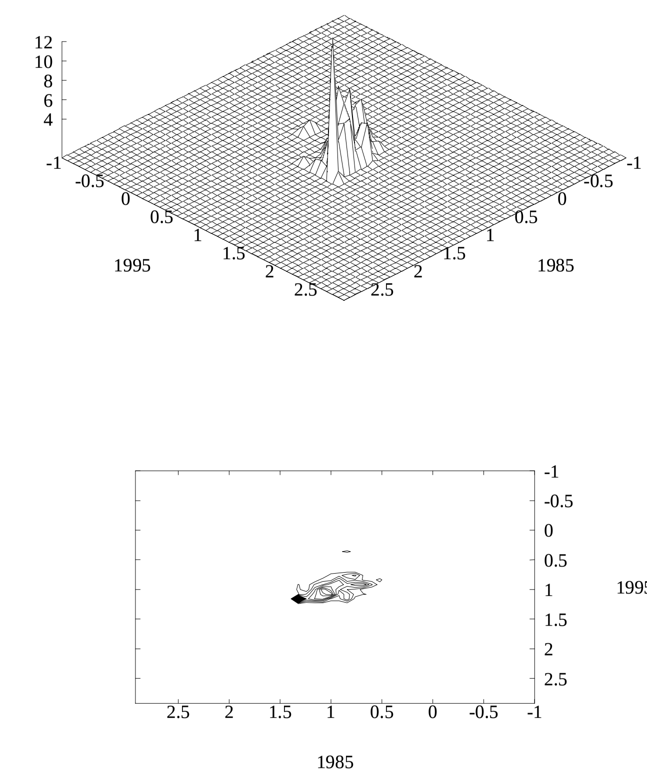


Figure 5: Stochastic kernels, efficiency (banking firms) (*approach 2*)

a) 1-year transitions



b) 11-year transitions



being as efficient as initially more inefficient (efficient).

Therefore, although figures 4.a and 5.a show persistence in firms' relative positions between two consecutive years, by means of probability concentrated along the positive sloped diagonal, when analyzing 11-year transitions (figures 4.b and 5.b) results tend to differ. In such cases, several firms abandon the positive sloped diagonal, ending up with a clear narrowing of the range of values in 1995.

5.4 Long-run tendencies: ergodic distribution

The developed analysis helps in overcoming one of the limitations of section 5.2, as it identifies firms' changes in their relative efficiency scores or intra distribution dynamics. However, it leaves still unsolved the long-run behaviour or ergodic distribution.

Computing the ergodic distribution and characterizing long-run impels us to discretize the efficiency scores' space (for both output choices). In such a case, measures λ_t are probability vectors and the stochastic kernel M becomes a transition probability matrix Q.[31] Thus, the discrete counterpart to equation (10) is:

$$F_{t+1} = Q_{r \times r} \cdot F_t \tag{15}$$

where $Q_{r \times r}$ is a transition probability matrix from one state of efficiency to another, assuming a countable state space:

$$E = \{e_1, e_2, \ldots, e_r\} \tag{16}$$

for each of the analyzed variables. The discretization of the observations' space in which the analyzed variables may take values in r states e_i, $i = 1, \ldots, r$ permits straightforwardly interpret intra-distribution mobility. For example, state $e_i = (0.5, 3)$ includes those firms with efficiency between half and three times average efficiency for the total sample. In addition, cell p_{ij} in $Q_{r \times r}$ matrix shows the probability that a firm initially belonging to state i transits during the period or periods (l) being considered to state j. Cell p_{ij} is defined as:

$$p_{ij} = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{N_{ij,t}}{N_{i,t}} \tag{17}$$

where T is the number of periods in the sample (11 years), $N_{ij,t}$ is the number of firms transiting during a period from state i to state j and $N_{i,t}$ is the total number of firms starting the period in state i. In addition, each row in the matrix represents a transition probability vector. Such vectors help in better understanding the analogy with the continuous case:

[31] Namely, M and Q refer both to the stochastic kernel in the continuous and discrete cases, respectively.

they are equivalent to the density probability defined for each point in E, when cutting the figure at that point by a plane parallel to $t+l$.[32]

When computing annual transitions (1-year transitions) through transition probability matrices, the available observations for the eleven years are divided into five states $E = \{e_1, e_2, \ldots, e_5\}$. The states' upper limits have been selected in a way such that the initial distribution (1985) is uniform.[33] Of course, this strategy gives different limits to the states according to the different measures of output.

The first column in each table (see tables 3 and 5) displays the total number of observations in that relative efficiency state at the beginning of the period. Thus, the first cell of row number four in table 3 would indicate that 402 observations (out of the total number of observations for the 11 years in the sample) were initially in such a state of relative efficiency (e_4), i.e., they were between 0.968 and 1.226 times more efficient than the average. In addition, 63% out of these 402 observations stayed in the same state of relative efficiency in the following period, while 13% moved to another state of higher efficiency (e_5); the remaining 24% moved to states of less relative efficiency (e_1, e_2 and e_3).

The values in tables 3 and 5 show persistence in the firms relative positions. The persistence is maximum when the values in the main diagonal are close to 100%. However, in our case the pattern is somewhat irregular. According to the first approach to output measurement, 62% of firms starting each period in the state e_1 of relative efficiency stayed in the same state in the next period, while 18% of firms moved to state e_2, 16% to e_3, 2% to e_4 and 2% to e_5. Persistence, though, is lower according to approach 2 to output measurement, as only 38% of firms stay, moving the remaining 62% to the other states of relative efficiency.

Which is the probability of a firm ending up in a certain state of relative efficiency? This would be given by the ergodic distribution, which shows that the probability mass, according to approach 1, would be more concentrated in the fourth state of relative efficiency (35%), i.e., in a state of relative efficiency between 0.968 and 1.226 times the average of the sector. However, according to output measure 2 (table 5), the ergodic distribution differs, as the probability mass tends to be more uniformly distributed. Although e_3 has higher probability, differences are not so important.

Patterns in table 4 differ strongly from those in table 3,[34] with the probability mass much more concentrated in states e_3 and e_4. All firms starting the period in state e_1 of relative efficiency (with efficiency under 0.672 times the average) move to states e_2, e_3 and e_4 (13%, 26% and 61%, respectively). Such a tendency is similar for firms in e_2 in 1985,

[32] Andrés and Lamo (1995).

[33] However, other authors select limits with different criteria. E.g., Quah (1993a) chooses states simply "reasonable to him".

[34] We must always bear in mind that transition probability matrices are just discretized versions of the stochastic kernels which allow us to compute the long run tendencies. Thus, tables 3 and 4 would be the discretized counterparts to figures 4.a and 4.b, respectively, and the same would occur to tables 5 and 6 relative to figures 5.a and 5.b.

all of them leaving such an state and transiting to e_3, e_4 and e_5. Thus, firms initially less efficient have moved completely to situations of more relative efficiency. On the other hand, initially more efficient firms show persistence, although to a somewhat lesser extent (less than 50%). Differences with 1–year transitions disappear when comparing long run tendencies: probability mass ends up being more concentrated across states e_3 and e_4. The same occurs when considering approach 2 to output measurement (table 6), being the probability spread across such states (36% in e_3, 32% in e_4).

Table 3: Convergence in efficiency, banking firms (1–year transitions) (*approach 1)*

Normalized efficiency					
	Upper limit				
	0.672	0.767	0.968	1.226	∞
(170)	0.62	0.18	0.16	0.02	0.02
(146)	0.16	0.38	0.39	0.04	0.03
(347)	0.03	0.12	0.55	0.27	0.03
(402)	0.01	0.01	0.22	0.63	0.13
(255)	0.01	0.01	0.05	0.22	0.71
Ergodic distribution	0.07	0.09	0.29	0.35	0.20

Table 4: Convergence in efficiency, banking firms (11–year transitions) (*approach 1*)

Normalized efficiency					
	Upper limit				
	0.672	0.767	0.968	1.226	∞
	0.00	0.13	0.26	0.61	0.00
	0.00	0.00	0.33	0.63	0.04
	0.08	0.14	0.41	0.32	0.05
	0.05	0.09	0.36	0.36	0.14
	0.06	0.00	0.22	0.33	0.39
Ergodic distribution	0.06	0.09	0.35	0.38	0.12

Table 5: Convergence in efficiency, banking firms (1–year transitions) (*approach 2)*

Normalized efficiency					
	Upper limit				
	0.813	0.942	1.060	1.206	∞
(205)	0.67	0.20	0.08	0.02	0.03
(269)	0.15	0.53	0.26	0.03	0.03
(343)	0.03	0.21	0.52	0.19	0.05
(271)	0.01	0.05	0.26	0.52	0.16
(232)	0.02	0.03	0.07	0.27	0.61
Ergodic distribution	0.14	0.22	0.28	0.21	0.15

Table 6: Convergence in efficiency, banking firms (11–year transitions) (*approach 2*)

Normalized efficiency					
	Upper limit				
	0.813	0.942	1.060	1.206	∞
	0.24	0.43	0.19	0.14	0.00
	0.20	0.30	0.25	0.25	0.00
	0.08	0.16	0.40	0.36	0.00
	0.00	0.17	0.44	0.39	0.00
	0.05	0.05	0.05	0.85	0.00
Ergodic distribution	0.10	0.22	0.36	0.32	0.00

6 Conclusions

This paper tries to assess the dynamics of cost efficiency in the Spanish banking industry over the last decade through a model of distribution dynamics. The (cost) efficiency scores are computed using a nonparametric method (DEA) and specifying two output measures, which enables us to further assess whether firms' specializations are varying over time. The model of distribution dynamics is a three-stage model which tries to identify how the distribution of the efficiency scores evolves (estimating nonparametrically density functions, via the kernel method), if there exist changes in firms' relative positions over time and which would be the likely long run (ergodic) distribution of such scores.

This new approach overcomes some limitations of prior research studies of cost efficiency in the Spanish banking industry. Firstly, it has not been sufficiently stressed how efficiency scores vary according to different output specifications[35] and specially how this might be linked to changes in firms' output mixes. One of the possible consequences of the liberalization process undergone by the Spanish banking industry is the re-definition of balance sheet strategies and output mixes; this has been shown in Pérez and Tortosa-Ausina (1998) and is reinforced by the results achieved in this study, as efficiency scores differ substantially depending on the output specification. While according to the first approach (which considers only earning assets as outputs) mean cost efficiency increases from 53.20% to 78.62%, and savings banks overtaking commercial banks, the second approach (which considers also deposits as outputs, thus considering also the service production nature of the banking firm) estimates a much more stable pattern (specially considering only 1985–94 period), and savings banks being always more efficient than commercial banks.

While these findings are undoubtedly interesting, the conclusions we may draw on the dynamics of efficiency scores improve dramatically when applying a model of distribution dynamics. Mean efficiency scores may not precisely depict the level of efficiency in the industry, as important differences might subsist. Dispersion measures help, but not fully,

[35]The exception being the paper by Grifell-Tatjé, Prior and Salas (1992), although it is a commonplace to simply point out the importance of such a decision, but without carrying out any further analysis.

as they are unable to identify multi-modality. Conclusions are highly reinforced when considering the entire distribution.

The three-stage model applied in the paper informs us precisely of the dynamics. The nonparametric estimation of the univariate density functions shows that, according to either output measure, efficiency scores are getting closer, but in different ways. The first approach to output measure shows that the multi-modality in 1985 has almost disappeared in 1995, i.e., the existence of firms much more efficient than the average is a feature which does not longer exist in 1995. However, according to approach 2 to output measure, the efficiency scores were already quite close in 1985, although in 1995 are even closer. Thus, the conclusion we might draw is that not only efficiency scores are approaching the average but also that specializations are changing over time. In addition, the estimation of bivariate density functions which attempt to identify changes in firms' relative positions show that, according to both approaches to measure banking output, probability mass is much more concentrated in 1995. Finally, the long run tendencies confirm that banks initially more inefficient are leaving the states of less relative efficiency, contributing to make probability mass less uniformly distributed across all the states of relative efficiency.

Thus, the importance of the conclusions is twofold. First, in a context of major changes, primarily due to deregulation, the estimation of efficiency depends heavily on the output specification, as changes in banks' product mix might be taking place. Secondly, it is important when analyzing the *dynamics* of the efficiency scores to consider the *entire* distribution and not only two of its moments. Such an approach gives much more robustness to the conclusions we might draw on banks' efficiency.

References

Aly, H.Y., Grabowski, R., Pasurka, C. and Rangan, N. (1990). Main Patterns of Economic Growth in OECD Countries. *The Review of Economics and Statistics*, 72:211–218.

Andrés, J. and Lamo, A. (1995). Dynamics of the income distribution across OECD countries. Discussion Paper 252, Centre for Economic Performance.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30:1078–1092.

Bauer, P.W., Berger, A.N. and Humphrey, D.B. (1993). Efficiency and productivity growth in u.s. banking. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency: Techniques and Applications*, pages 386–413. Oxford University Press, Oxford.

Berger, A.N. and Humphrey, D.B. (1991). The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics*, 28:117–148.

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

Berger, A.N. and Mester, L.J. (1997a). Efficiency and productivity change in the U.S. commercial banking industry: A comparison of the 1980s and 1990s. Working Paper 97–5, Federal Reserve Bank of Philadelphia.

Berger, A.N. and Mester, L.J. (1997b). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance*, 21:895–947.

Berger, A.N. and Mester, L.J. (1999). What explains the dramatic changes in cost and profit performance of the U.S. banking industry? Working Paper 99–1, Federal Reserve Bank of Philadelphia.

Berg, S., Førsund, F.R. and Jansen, E.S. (1992). Technical efficiency of Norwegian banks: The non-parametric approach to efficiency measurement. *The Journal of Productivity Analysis*, 2:127–142.

Caminal, R., Gual, J. and Vives, X. (1988). Competition in Spanish Banking. In Dermine, J., editor, *European Banking in the 1990s*, chapter 8. Blackwell, Oxford.

Canals, J. (1993). *Competitive strategies in European banking.* Oxford University Press, Oxford.

Cao, R., Cuevas, A. and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis*, 17:153–176.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Durlauf, S.N. and Quah, D.T. (1998). The new empirics of Economic Growth. Discussion Paper 384, Centre for Economic Performance.

Ferrier, G.D. and Lovell, C.A.K. (1990). Measuring cost efficiency in banking. *Journal of Econometrics*, 46:229–245.

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

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407.

Koopmans, R. and Lamo, A. (1995). Cross-sectional firm dynamics: Theory and empirical results from the chemical sector. Discussion Paper 229, Centre for Economic Performance.

Maudos, J. (1996). Eficiencia, cambio técnico y productividad en el sector bancario español: una aproximación de frontera estocástica. *Investigaciones Económicas*, 20(3):339–358.

Maudos, J., Pastor, J.M. and Pérez, F. (1997). Competencia y evolución de la eficiencia en el sector bancario español: la importancia de la especialización. Unpublished manuscript paper.

Mester, L. (1997). Efficiency in the savings and loan industry. *Journal of Banking and Finance*, 17(2–3):267–286.

Nadaraya, E.A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves.* Kluwer, Dordrecht.

Park, B.U. and Marron, J.S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409):66–72.

Park, B.U. and Turlach, B.A. (1992). Practical performance of several data-driven bandwidth selectors. *Computational Statistics*, 7:251–285.

Pastor, J.M. (1995). Eficiencia, cambio productivo y cambio técnico en los bancos and cajas de ahorro españolas: un análisis de la frontera no paramétrico. *Revista Española de Economía*, 12(1):35–73.

Pastor, J.M. (1996). Diferentes metodologías para el análisis de la eficiencia de los bancos y cajas españoles. Documentos de trabajo 123–1996, Fundación FIES.

Pérez, F. and Tortosa-Ausina, E. (1998). Product mix of the Spanish banking firms: Do competition clubs exist? Working Paper-Serie EC 98–02, IVIE.

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

Quah, D.T. (1993b). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95(4):427–443.

Quah, D.T. (1996a). Convergence empirics across economies with (some) capital mobility. *Journal of Economic Growth*, 1(1):95–124.

Quah, D.T. (1996b). Empirics for economic growth and convergence. *European Economic Review*, 40:1353–1375.

Quah, D.T. (1997). Empirics for growth and distribution: Stratification, polarization and convergence clubs. *Journal of Economic Growth*, 2(1):27–59.

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*, 21:221–250.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.

Sealey, Jr., C.W. and Lindley, J.T. (1977). Inputs, outputs and a theory of production and cost at depository financial institutions. *Journal of Finance*, 32(4):1251–1266.

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(3):683–690.

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

Simonoff, J.S. (1996). *Smoothing Methods in Statistics.* Springer, New York.

Stokey, N.L. and Lucas Jr, R.E. (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, Massachusetts.

Tortosa-Ausina, E. (1999). *Especialización productiva, eficiencia y convergencia de las empresas bancarias españolas.* PhD thesis, Universitat Jaume I.

Vives, X. (1990). Deregulation and competition in Spanish banking. *European Economic Review*, 34:403–411.

Vives, X. (1991a). Banking competition and European integration. In Giovannini, A. and Mayer, C., editors, *European Financial Integration*, chapter 2. Cambridge University Press, Cambridge.

Vives, X. (1991b). Regulatory reform in European banking. *European Economic Review*, 35:505–515.

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

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