INEFFICIENT BANKS OR INEFFICIENT ASSETS?

Emili Tortosa-Ausina*

Universitat Jaume I

Departament d'Economia

Campus del Riu Sec • 12071 Castelló (SPAIN)

Tel.: +34 964728606 / Fax: +34 964728591

E-mail: tortosa@uji.es

*I would like to express my gratitude to Fundación Caja Madrid, Fundació Caixa Castelló and Generalitat Valenciana for financial support, and to the Instituto Valenciano de Investigaciones Económicas, for kindly providing data.

1 Introduction

There exists a notable concern among academics, practitioners, and policy-makers about the discrepancies in efficiency performance across banking firms. Such a concern exists now, once the deregulation of geographic restrictions and the harmonization of regulatory and supervisory environments in most Western European countries have occurred, and existed before. But it remains an exciting quest, which is not yet over.

Focusing on the Spanish banking system, one faces a scenario where widespread changes—such as deregulation, disintermediation, technological change, the introduction of a single currency and the increased emphasis on the pursuit of noninterest income, among others—has taken place. However, despite all firms face the same competitive environment, there are some efficient firms but also pockets of poorer performance. We argue that if inefficiencies persist once deregulation has occurred, appropriate explanations must be pursued. Specifically, there is a common belief that productive efficiency is a survival condition in a competitive environment,[1] and that if no competitive pressures exist there might subsist pockets of inefficiency. But once competitive constraints are removed this should not be the case any longer, and firms' efficiency levels should be forced to converge.[2] In fact, the European Commission (1988) argued that "... the new competitive pressures brought about by the competition of the internal market can be expected to ... produce appreciable gains in internal efficiency ... [which will] constitute much of what can be call the dynamic effects of the internal market ...".

The suitable explanations for inefficiency persistence could be manifold. In this paper we address this issue in two main ways. First, although one estimates the cost efficiency scores of Spanish banking firms via standard approaches—in particular, by means of nonparametric techniques—, efficiency dispersion is assessed by others, not so standard, methods. With particular concern, we estimate nonparametrically the density functions of the efficiency scores at different points in time, in an attempt—among other interesting pursues—to draw some conclusions upon dynamics. Such conclusions are reinforced by computing transition probability matrices, as they encode information on both changes in firms' relative positions and ergodic distributions. These techniques could help explaining whether dispersion is really as high as traditional dispersion indi-

[1] See Leibenstein (1966, 1978). Such a widespread idea, though, has been partially contested recently in Stennek (2000).

[2] However, the forces for eliminating inefficiency may be weaker than suggested, and many firms might be just doing the best that is necessary to survive comfortably (Mayes et al., 1994; Simon, 1959).

cators (variance, standard deviation) show and if dynamics show any clear tendency at all.

The main contribution of the paper, though, consists of considering explicitly the structure of the Spanish banking industry, which differs much from that in the U.S., where the long-standing restrictions have shaped an extremely fragmented industry, with thousands of banks and bank holding companies. Conversely, in Spain a few very large institutions dominate. We argue that this intrinsic feature should be accounted for when analyzing banks, as it might help explain some tendencies in the industry.[3]

Industry's structure in Spain, though, has not stood still, and the distribution of banking institution sizes has undergone deep changes. In particular, there has been a flow of mergers and acquisitions (M&As) from 1985 which affected differently firms of differing sizes. However, in recent years this process has affected more markedly the largest commercial banks, in a way such that there remain only three banks from the formerly deemed as "large eight",[4] and one of them should not even be labelled as large anymore.[5] The M&As process affected also, and largely, savings banks, decreasing from 77 by 1985 to 50 few years after. As a result (and it will be corroborated throughout the paper) the degree of concentration among the largest firms in the industry was high in 1985, but remains high today despite the removal of entry barriers.

Of course, the relationship between bank costs and size has been studied elsewhere. It ranks highly as an extremely appealing topic, mostly because of its policy implications regarding the optimal structure of the banking industry. But the topic remains unsettled, despite the vast literature related. To be specific, many authors have approached this topic analyzing the cost savings that can be achieved by increasing size—i.e., scale economies studies—.[6] Their general conclusion is that for the very smallest banks there are scale economies that allowed average costs to fall with increases in bank size; however, scale effects generally account for less than 5% of costs, and for the largest banks, constant average costs or even slight diseconomies prevail. In contrast, X-efficiency—consisting of managerial ability to control costs—is much more important than scale

[3]In fact, one of the directions for future research in the survey by Berger and Humphrey (1997) suggested as important for research and policy purposes to see if the U.S. results carry over into other nations with banking markets that are more national in scope with much tighter levels of concentration.

[4]Banco de Bilbao, Banco de Vizcaya, Banco Central, Banco Hispano Americano, Banco Español de Crédito, Banco de Santander, Banco Popular, and Banco Exterior.

[5]Banco Popular.

[6]See Berger et al. (1987), Gilligan and Smirlock (1984) or Clark (1988).

economies.[7]

More to the point of this paper, Berger and Humphrey (1992), Bauer et al. (1993) or Elyasiani and Mehdian (1995) investigate bank efficiency by class size. Conclusions of such studies are undoubtedly attractive, but subject to the differing size classes considered which, in addition, may change over time. We account for this issue from a *continuous* point of view, considering the differing sizes of firms, but without classifying them into mutually exclusive groups. In other words, we avoid the bias involved by choosing the limits among size classes. We do this by *weighting* both the density functions of efficiency scores and the transition probability matrices. With such an approach we may assess not only the differing efficiency levels across firms but also *across* assets, which provides a different, perhaps more interesting—considering the specific features of the Spanish banking system—view of inefficiency in industry.

The remainder of the paper is structured as follows. Section 2 presents the model to estimate cost efficiency scores for Spanish banking firms. Results are reported in section 3, along with the description of inputs and outputs. Sections 4 and 5 assess the dynamic patterns of the efficiency scores by estimating nonparametrically their density functions and computing transition probability matrices, respectively. Section 6 questions whether density functions differ when we focus on the efficiency of assets, rather than firms—i.e., taking into account explicitly industry's structure—. Finally, section 7 concludes.

2 Nonparametric estimates of efficiency scores

The survey by Berger and Humphrey (1997) provides an excellent view of the studies which have assessed the efficiency and productivity of financial firms.[8] More than 130 academic works that applied frontier efficiency analysis[9] were reviewed, applying both parametric and nonparametric techniques to the study of the efficiency of different types of financial institutions and different samples in different countries. In the case of Spanish banking firms, 11 studies were reviewed. Despite their differing attempts—according to the firms analyzed, the approaches to measure efficiency and even the choice of inputs and outputs—the emerging picture is not clear-cut in the sense of efficiency gains or

[7] See Berger and Humphrey (1991).

[8] See also Berger et al. (1993) and, more recently Berger et al. (1999).

[9] Measurement of X-inefficiency is the framework where measurement of bank efficiency is generally performed. Accordingly, assuming that there is a common efficiency frontier, the deviation of a bank from that frontier is a measure of X-inefficiency (Altunbas and Chakravarty, 1998).

losses throughout the post-deregulation period. It must be pointed out, though, that attempts to assess dynamics—at least carefully—were not pursued.

We examine here the <u>cost</u> efficiency of all Spanish banks—both commercial and savings—for each year and each firm in our sample (1985–1997). When this issue is explored, two chief controversies arise. One refers to the technique; the other one deals with our beliefs on what banks produce. The purpose of this study, though, is not to make a comparison between different techniques, or how results vary according to different outputs' choices.[10]

The nonparametric ADEA (Allocative Data Envelopment Analysis) technique[11] to measure cost efficiency has been selected because of its ability to closely envelope data—and data structure is a feature in which we are particularly concerned with—despite its inability to disentangle inefficiency from random error. On the contrary, parametric methods do this but, in turn, they must impose a functional form on the distribution of inefficiency which, in principle, involves less flexibility.[12]

According to ADEA methodology, the efficiency of an individual unit is measured *relative* to the other units in the sample, by constructing a linear piecewise frontier using the most efficient firms in the sample. Thus, the efficiency of each unit is determined by comparison to this "best-practice" frontier. Specifically, efficiency scores are computed by solving the following program for each firm in each time period:

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

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

Computing the individual cost efficiency scores requires solving program (1) for each s firm and year in our sample. The solution will be given by the x_s^* cost minimizing

[10] If that were the intended, the papers by Ferrier and Lovell (1990) or Resti (1997) provide excellent comparisons.

[11] See Aly et al. (1990).

[12] McAllister and McManus (1993), Mitchell and Onvural (1996), and Wheelock and Wilson (2000) test and reject the translog specification of bank cost functions, and suggest semi-nonparametric and nonparametric methods for estimating bank costs.

vector, given the price vector ω_s and outputs vector y_s.

Accordingly, the efficiency scores are given by:

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

Similarly, the inefficiency estimates will be given by:

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

which reveals the amount to which firms s costs are increased for performing off the efficient frontier made up of those "best-practice" banks.

3 Data and results

For our definition of inputs and outputs we have initially chosen the intermediation approach as suggested by Sealey and Lindley (1977), which contemplates banks as financial intermediaries between liability holders and those who receive bank funds, treating inputs and outputs in a mutually exclusive way. Complementarily, we may consider 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. Accordingly, we will treat savings deposits both as inputs and outputs, in line with other research studies applying the value-added method (Berger et al., 1987), under whose views liabilities may have simultaneously input and output characteristics.[13]

All variables are described in table 1, which reports also some descriptive statistics.[14] Information is provided by the Spanish association of commercial banks (AEB, Asociación Española de Banca) and the Spanish association of savings banks (CECA, Confederación Española de Cajas de Ahorro).[15] Some firms were dropped—those which

[13]The different approaches to bank output measurement are accurately explained in Berger and Humphrey (1992).

[14]An additional controversial issue regards whether firms face input price differentials or not. In line with most previous work applied to Spanish banking firms, we use firm-specific input prices. Although some recent contributions (Mountain and Thomas, 1999) argue that banks face competitive factor markets, and thus equal input prices, which may lead to mis-measurement of efficiency and economies of scale, others demonstrate significant differences in input prices across firms exist (Fukuyama et al., 1999).

[15]Although things have changed much, virtually all current differences between commercial and savings banks are attributable to their different type of ownership: commercial banks are privately-owned, whereas savings banks are foundations. Whether the type of ownership might bias efficiency has been studied in Altunbas et al. (2000).

were not in continuous existence over the sample period 1985–97—and banks were backward merged in order to have the same number of firms at every year. Although this could seem an important loss of data, our sample involved always around 90% of total assets in industry.

Efficiency scores have been computed for each year, considering jointly commercial banks and savings banks, hence assuming all firms are faced with the same opportunities to combine labour, physical capital, and financial capital to produce outputs that are virtually identical.[16] Results are reported in table 2. They reveal efficiency enhancement has not taken place at firm level, despite deregulation and unrestrained entry. To be more specific, although some fluctuations have taken place, initial (1985) and final (1997) simple mean values are very similar, regardless of the analyzed group of firms—commercial, savings or total banking firms—. However, some further conclusions diverge upon the type of firm considered, or firm's size. Specifically, savings banks' efficiency is always higher—at least on a simple mean basis—than commercial banks', a result common to most previous research studies.[17] Regarding firms' size, an increase in savings banks' weighted mean efficiency is appreciated (from 86.24% by 1985 to 93.24% by 1997). This tendency could suggest that the lift of regulations and, specially, the possibility of this type of firms to expand geographically has allowed them or, at least, some of them, to engage in different activities to the (expensive) ones they could have in their home region; in fact, the largest savings banks are those which have been expanding geographically in the last sample years at very high rates. In the case of commercial banks, weighted values show also that firms' size is an issue to account for, as they are always higher, or much higher, than their simple counterparts, despite a time invariant pattern. These trends further contribute to make us appreciate that conclusions drawn at firm level might mask important tendencies at industry level, and vice versa.

Dispersion indicators reveal that savings banks' are much closer in their efficiency levels than commercial banks, and get even closer over time, as suggested by lower standard deviation values. No clear time pattern emerges, though, for all banking firms. But this statistic does not encode important information such as the asymmetry of the

[16]In contrast, Mester (1989, 1993) and Cebenoyan et al. (1993) consider it may be inappropriate to compare efficiency scores for different ownership categories if firms in one category utilize a different technology. In order to enable comparisons between ownership types relative to the industry "best practice" cost frontier, and bearing all firms currently face the same regulation, we consider a common frontier for all banks.

[17]See, for instance, Maudos (1996) or Pastor (1995).

distribution of efficiency scores, their shape compared to the normal distribution, or the existence of multiple modes, which might carry an important economic interpretation. In our particular instance, they miss the "pockets" of inefficiency that could exist in the industry. Accordingly, it could occur that a cluster of very efficient and very inefficient firms existed at the same time and were approaching, or coming apart over time. These invisible features for variance and/or standard deviation could carry strong industry implications.

4 A first attempt to detect pockets of inefficiency

More specifically, all descriptive statistics we might use help in gaining insights on the cross section probability density function of the efficiency scores at different points in time. But a more fully description is given if we know the exact *shape* of the densities, which may be estimated either through parametric or nonparametric methods. The former consists of specifying a functional form for the distribution, such as the normal distribution, the t-Student distribution or the log-normal distribution (Aitchison and Brown, 1954), which depend on some parameters. Hence, the problem is confined to accurately estimate those parameters which completely characterize the density of the efficiency scores.

But these methods depend on the (possibly) arbitrary choice of the functional form, as there is no generally-accepted criterion to infer the most apposite function. Conversely, the nonparametric approaches do not assume any particular form for the density. However, simple nonparametric density estimators—such as the histogram, the frequency polygon or varying the bin width—, although free from the "parametric straitjacket" of rigid distributional assumptions, suffer from other drawbacks, as they are not smooth and they are not sensitive enough to local properties.

Kernel smoothing overcomes these serious drawbacks, but keeps their advantages.[18] Thus, it constitutes an important data analytic tool which provides a very effective way of showing structure in a set of data. In particular if, as suggested, multi-modality exists, the kernel method would uncover it, whereas it would be completely missed by imposing a uni-modal parametric model such as the log-normal. In sum, with this method we

[18]See Silverman (1986), Scott (1992), Wand and Jones (1995), Simonoff (1996), Devroye and Györfi (1985) or Nadaraya (1989). The literature on this topic is vast and still grows at high rates.

may precisely capture how the *entire* cross section of efficiency scores evolves over time.

This method consists of—after normalizing[19] efficiency scores (just for convenience)—estimating the following density function for each output specification and year (or period):

$$\widehat{f}(x) = \frac{1}{Sh}\sum_{s=1}^{S} K(\frac{x - NES_s}{h}) \tag{4}$$

where S is the number of firms in our sample, NES_s is the normalized efficiency score for firm s, and h is the bandwidth, window width or smoothing parameter, which determines the amount to which data will be smoothed.

K is a kernel function satisfying:

$$\int_{-\infty}^{+\infty} K(t)dt = 1 \tag{5}$$

Kernel's choice consists of several alternatives.[20] In our case we have selected the Gaussian kernel which, in the univariate case we are dealing with is expressed by:

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

The relevant choice, though, is not the kernel's but, by large, the h's or bandwidth's. While the kernel determines the shape of the bumps when plotting function (4), the smoothing parameter has a different effect, conditioning bumps' width. If h is too small, an excessive number of bumps is generated and data structure is difficult to appreciate; in other words, data are *undersmoothed*. On the other hand, if h is too large *oversmoothing* occurs, and some data features are hidden. What we find under these graphic facts is the traditional trade-off between bias and variance which, indeed, depends on the smoothing parameter: the larger is h, less variance and more bias, and vice versa.

The relevance of this decision has led us to take some cautions on this topic and, finally, to choose the smoothing parameter suggested by (Sheather and Jones, 1991),

[19] Or dividing by the mean. Consequently, if the normalized efficiency score of certain firm had a value of 2, it would indicate such a firm is twice efficient than industry average. On the other hand, if such a value were 0.5, it would indicate its efficiency is half of industry average. An advantage of this strategy consists of offsetting the distorting effects outliers might cause—to which nonparametric techniques to measure efficiency are particularly sensitive—.

[20] Epanechnikov, triangular, Gaussian, rectangular, etc.

based on (Park and Marron, 1990). It relies on the second generation method *solve-the-equation plug-in*, and its higher performance relative to first generation methods has been verified in further research studies.[21]

Nonparametric density estimates are reported in figure 1. Splitting them into sub-figures has been done in an attempt to better capture the trends at different points in time. The features such figures reveal are manifold but, probably, the most interesting ones—for our purposes—lie in that distributions are not, by large, time invariant, nor uni-modal. Specifically, multi-modality exists at all periods, whose shape has been changing over time, and seems more marked comparing initial and final years. This could suggest that there exist differing speeds of adjustment to changes in the financial system: multi-modality has increased after 1985—the "twin-peaks" seem more marked in 1986–90—, decreases in 1992–96, and appears to arise again in 1997. Overall, these tendencies—which are absolutely invisible for any dispersion indicator—reveal us that the distance between the "best practice" institutions and those others off the efficient frontier has been varying over time, and that the number of efficient firms is large enough to form a perceptible mode.

However, we ignore whether efficient firms are always the same or, conversely, there are changes in firms' relative efficiency. Thus, a new—and, perhaps, more profound—issue is addressed: is inefficiency just a temporary phenomenon, due to restrained competition and random shocks? Or, somehow surprisingly, inefficiencies might be persisting over time? Just considering a snapshot of efficiency in an industry might lead us to understate some firms' efficiency; but if by not producing at maximum capacity such firms are more flexible and adapt faster to shifts in the competitive conditions then, over time, they will be more efficient. In other words, the ranking of efficiency might vary over time, and firms' relative positions might be changing. This is, precisely, the basic attempt of section 5.

[21] See, for instance, Jones et al. (1996) or the simulation studies by Park and Turlach (1992) or Cao et al. (1994). More details on our bandwidth are available from Sheather and Jones (1991) and Park and Marron (1990) papers. In addition, Steve Marron's *web*'s page provides the Matlab routine which enables its obtaining (URL: http://www.stat.unc.edu/faculty/marron.html).

5 Additional dynamic patterns: changes in firms' relative positions and long run behaviour

Thus, if we deem inefficiency as a dynamic rather than as a static concept, maybe a more accurate assessment of the dynamics of banks' efficiency scores is required.[22] Although the analysis performed in section 4 has provided a powerful starting point, intra-distribution mobility has still not been analyzed. And, precisely, a large number of changes in firms' relative positions could be undergoing despite time-invariant shapes of the density functions.

Computing correlation ranks offers a first prior into such changes. However, despite carrying a lot of meaningful information, they are summary statistics which cannot detect where firms lie at different points in time. Accordingly, we are unable to corroborate whether inefficiency is occurring only because of differing speeds of adjustment to changes in concurrence (dynamic inefficiency) or it is a persistent phenomenon, and initially inefficient firms remain inefficient over time.

The technique applied here—and elsewhere—does exactly that.[23] In particular, transition probability matrices Q are estimated, in order to identify firms' mobility over time. Then, if their positions were invariant relative to the mean—as we previously normalized data—these transition probability matrices should be the identity matrix: the distributions are invariant and, in addition, firms' mobility does not exist. On the other hand, if entries off the diagonal were different from zero, then intra-distribution mobility over time would be undergoing.

Previously, the space of efficiency scores observations' must be properly discretized in $r = 5$ states. In our case, for the annual transitions the criterion has been to divide all observations from the 1985–97 period in order to have the same probability mass (20%) in each case, although this obviously yields intervals with different widths. For the 13-year transitions, the upper endpoints refer to 1985 data, and form intervals with the same amount of probability.

Such a discretization into r states s_i, $i = 1, \ldots, r$ allows for clear interpretation of distribution mobility. For instance, state $s_i = (0.5, 3)$ would include all those companies with efficiency ranging between half and three times the industry average. In addition,

[22]See Yoo (1992).

[23]For applications similar to those we are dealing with see, among others, Andrés and Lamo (1995), Lamo (2000), or Quah (1996).

cell p_{ij} in the transition probability matrix $Q_{r \times r}$ would show the probability of a firm initially in state i to end up in state j throughout the period observed (l), and each row would constitute a transition probability vector, with its cells summing to one.

Table 3 displays these transition probabilities from approach, for annual and 13-year transitions. The overall view they provide suggests that, indeed, intra-distribution mobility occurs. Specifically, the top left-hand entry in table 3.a shows that the less efficient 20% of banking companies—with efficiency scores less than 80.2% of the average—remained in the following year with efficiencies in that range with probability 0.69. The remainder 0.31 moved overwhelmingly (0.21) to state 2—including the following 20% of less efficient firms, ranging from 80.2% to 88.5% of average efficiency—, to state 3—0.07, ranging between 88.5% and 96.4% of industry average—, state 4—0.01, in the 96.4–107.7% range—and, surprisingly, even some firms end up in state 5 of upper relative efficiency—0.03, above 107.7% of industry average—. Persistence is even lower in states 2, 3, and 4. On average, only 47.3% of firms remain in their initial state of relative efficiency, and transitions occur both to inferior and superior states. Persistence, though, is higher for the upper state of relative efficiency; as low right-hand entry in table 3.a shows, 74% of these firms remain in the same state of relative efficiency each year.

Overall, persistence in table 3.a averages to 57.0%. In the case of 13-year transitions (table 3.b), when only initial (1985) and final (1997) years are accounted for, this average value decreases to 31.8%, suggesting much deeper intra-distribution movements and, accordingly, that the persistence of inefficiencies—at least in terms of relative positions—is extremely low. These findings strongly corroborate our prior assumption regarding the *dynamic* nature of inefficiency: although density functions' shape does not seem to vary much over time, firms' efficiency ranking varies much. In addition, this feature seems to affect more markedly to the most inefficient firms, as they abandon their initial states more overwhelmingly, specially comparing initial and final years.

But more information is available. Last row in both tables 3.a and 3.b contains information on the ergodic or, in other words, hypothetic long-run distribution if those tendencies in 1985–97 period continued over time. Table 3.a shows that departing from an initial uniform distribution for all observations, probability ends up also very uniformly distributed, implying efficiency scores' will not approach over time. If 13-year transitions are considered, the trend is different, as bi-modality emerges: 29% of firms will concentrate in state 2, whereas 22% will remain the highest state of relative efficiency. However, the interpretation of the ergodic distribution in this case is less clear.

6 Firms vs assets

The second contribution—probably, the first in importance—of this paper consists in taking into account explicitly and accurately the differing sizes of firms. This is done in an attempt to capture the peculiar structure of the Spanish banking system, as differing results could make policy makers reassessing their judgments on the overall efficiency of the industry.

In order to get further insights on the probable discrepancies that might arise we may consider some standard measures of industry concentration, which constitute perhaps the most prominent aspect of market structure. The Herfindahl index takes into account all firms in an industry and their relative sizes. But if very few large firms dominate, concentration ratios provide more meaningful information, as they only consider the dominant firms in an industry.[24] The m-firm concentration ratio—the sum of the market shares of the m largest firms—constitutes the most common measure of concentration in empirical studies, expressed as:

$$\mathrm{CR}m = ms_1 + ms_2 + \ldots + ms_S \tag{7}$$

where ms_s is the market share of firm s, there are S firms in the industry, and firms are numbered so that firm 1 has the largest number share, firm 2 has the second-largest market share, and so on.

Such concentration ratios are provided by table 4 for assets, loans and savings deposits.[25] The CR1 value has fluctuated up and down mostly because of the mergers which affected the largest firms throughout the period.[26] More interesting patterns

[24] It is also debatable whether the inclusion of all firms is beneficial, as it can be argued that the entry and exit of firms at a small-scale have little effect on concentration (Hart, 1971).

[25] However, it could be argued that banking firms specialize differently, and concentration ratios could differ substantially for other balance sheet categories (Bain, 1956, p. 301–302).

[26] Two largest commercial banks merged in 1988 (Banco de Bilbao and Banco de Vizcaya), other two in 1991 (Banco Central and Banco Hispano Americano). More recently, the latter merged to Banco Santander—which previously absorbed Banco Español de Crédito by 1993—whereas the former merged to Argentaria (a banking group including Banco Exterior), although their branches still operate separately. The CR10 increase in savings deposits in 1990 reveals also the mergers of some savings banks which brought about by two of the largest, La Caixa and Bilbao Bizkaia Kutxa.

emerge from CR3, CR5 and CR10, which show an upward tendency when comparing initial and final years, and reveal that the 10 largest firms in the industry (CR10) have a market share of more than 50% in 1997—either for assets, loans or savings deposits—; for the five largest firms it decreases partly to below 40%, and the three largest firms still rank well above 20%.

If these largest firms were the efficient firms, the overall efficiency in industry could be understated, and vice versa. Hence, according to these findings, making inference on a simple instead of a weighted basis could drive to misleading conclusions or, at least, conclusions that miss some important features in industry. In other words, if we are not indifferent to *which* firms are the most inefficient, a different type of analysis should be performed, as simple and weighted results might differ substantially.

So, in order to evaluate the distorting effects size may cause on efficiency's appraisal, and considering the nonparametric approach of this study, it could be apposite undertaking the *weighted* counterpart to the analysis performed in sections 4 and 5. Although some related stuff has already been provided—table 2 reported summary statistics on a weighted basis (weighted mean)—stronger results are achievable if we attempt to estimate both weighted density functions and transition probability matrices.

Our approach to deal with this issue follows from Goerlich (1999) and consists of estimating the weighted counterpart to equation 4, which adopts the following expression:

$$\widehat{f}(x) = \frac{1}{Sh}\sum_{s=1}^{S} \varsigma_s K(\frac{x - NES_s}{h}) \qquad (8)$$

where ς_s represents the weight of firm s (in terms of their assets relative to total assets in industry).

Figure 2 is exactly the weighted counterpart to figure 1. The main conclusion we might infer is that distributions are not, by far, the same. In this case, although multimodality still remains, probability tends to collapse more markedly around some values, and densities are much tighter. Now we appreciate a larger peak of probability mass above industry's average at all periods. The patterns, though, are not time-invariant, as tighter density functions are more clearly observed after 1985. In addition, it seems that inefficiency in industry is decreasing, as the "twin peaks" property is not very accentuated by 1997. In sum, the picture emerging differs substantially from that figure 1 shows. The basic finding is that efficiency is understated at firm-level, but on a weighted basis efficiency is far higher. Although this could suggest there are many small/medium very efficient firms, given the structure of the Spanish banking industry it seems that the largest firms are the efficient. Thus, a negative relationship could be arising between concentration and inefficiency—contrary to the alleged positive relationship—attributable to the behaviour of the largest firms, which could be conducting a wholehearted drive for efficiency to hinder entry. Or, as suggested by Wheelock and Wilson (1999), relaxation of barriers to branching, among other significant deregulations, would seem to have favoured larger banks.

We must consider, though, that still much inefficiency is appreciated, as suggested by the noticeable probability mass below the unity. This inefficiency, probably attributable to small/medium firms, should be further explained. In particular, it might be the case that the industry included small/medium specialized institutions facing little competition in their markets niches. Then, unless we compared firms only with those specializing in the same range of products and services, inefficiency would be, once more, overstated.

Complementarily, table 5 describes the weighted counterpart to table 3 and, once more, we notice dissimilar patterns. Both in the case of annual and 13-year transitions, patterns are far more emphasized, particularly relating the ergodic distributions. For annual transitions (table 5.a) results are partly paralleled, as probability tends once more to abandon the main diagonal. Firms with initial relative efficiency below 80.2% of average—accounting for 8% of total assets in the 1985–97 period (column 1 in table 5.a)—shift to states 2, 3, 4 and 5 with probability 0.21, 0.15, 0.01 and 0.13, respectively. Mobility is also high for the remaining states, except for state 5; in this case, probability shifts to lower states at low rates. Diagonal entries average to 54.60%, which is very low. Accordingly, the dynamic nature of inefficiency is corroborated, although both the lower-right entries in 5×5 matrices in tables 5.a and 5.b reveal that the top-efficient assets remain in their state of relative top-efficiency, specially for annual transitions. Consequently, dynamics exist, but they affect more markedly those assets below 107.7% and 109.0% of industry efficiency average for annual and 13-year transitions, respectively.

7 Conclusions

This paper addresses the omission of several issues in studies of bank efficiency: the accurate study of the distributions of cost efficiency scores and their evolution over time, along with an explicit concern on the differing firms' sizes. To investigate this, nonparametric approaches are applied. Not only to compute efficiency scores—using ADEA-type models—but also to estimate their density functions and to consider, in a *continuous* way, how the structure of the industry might bias our findings.

Results show that efficiency gains—according to our sample, technique, and inputs and outputs definition—have been minor, in line with previous research studies. This is suggested by average efficiency, both in a simple and weighted basis. Although it has fluctuated over the sample period, initial and final values are very similar. Standard deviation has fluctuated also but, in like manner, their extreme values are very close too. Density functions and transition probability matrices provide a more comprehensive view of these patterns. The former show that multi-modality is present at several periods, as suggested by the existence of two perceptible modes. This "twin peaks" property looks likely to persist, as ergodic distributions for annual transitions are very similar to the initially chosen. It seems, though, that inefficiency is quite dynamic, as important intra-distribution movements occur.

Things differ much when the peculiar structure of Spanish banking industry is considered. Introducing weights in the kernel, density functions are much tighter, with probability strongly concentrated at values above unity. Such a pattern seems to be more accentuated over time. Complementarily, weighted probability matrices reveal also strong intra-distribution movements. However, in this case ergodic distributions probability mass accumulates well above the average.

Together, these results imply that the banking industry may be more efficient than earlier studies have suggested. Overall the results support both an accurate analysis of the cross section densities of efficiency scores but, very specially, the inclusion of the structure of the industry, if it is possible in a continuous way. These patterns are congruous with those that could emerge from an industry undergoing rapid change, where few pioneering firms might adapt quickly, and others respond more prudently.

The empirical findings should add to the knowledge of Spanish banking institutions, but also to the extant literature on efficiency and size, and even to the spirit of the persistence literature. They may help policy makers evaluate the enhancing effects of mergers among financial institutions, but they are also corroborative relative to efficiency—once regulations have been dimantled—as a highly dynamic issue.

References

Aitchison, J. and Brown, J. A. C. (1954). On criteria for descriptions of income distribution. *Metroeconomica*, 6:88–107.

Altunbas, Y. and Chakravarty, S. P. (1998). Efficiency measures and the banking structure in Europe. *Economics Letters*, 60:205–208.

Altunbas, Y., Evans, L., and Molyneux, P. (2000). Bank ownership and efficiency. Mimeo.

Aly, H. Y., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, scale, and allocative inefficiencies in U.S. banking: An empirical investigation. *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.

Bain, J. S. (1956). *Barriers to New Competition*. Harvard University Press, Cambridge, Massachussets.

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., Demsetz, R., and Strahan, P. (1999). The consolidation of the financial services industry: causes, consequences and implications for the future. *Journal of Banking and Finance*, 23:135–194.

Berger, A. N., Hanweck, G. A., and Humphrey, D. B. (1987). Competitive viability in banking: Scale, scope, and product mix economies. *Journal of Monetary Economics*, 20:501–520.

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. (1992). Measurement and efficiency issues in commercial banking. In Griliches, Z., editor, *Output Measurement in the Service Sectors*, pages 245–300. The University of Chicago Press, Chicago.

Berger, A. N., Hunter, W. C., and Timme, S. G. (1993). The efficiency of financial institutions: A review and preview of research past, present and future. *Journal of Banking and Finance*, 17:221–249.

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.

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.

Cebenoyan, A. S., Cooperman, E. S., Register, C. A., and Hudgins, S. C. (1993). The relative efficiency of stock versus mutual S&Ls: A stochastic cost frontier approach. *Journal of Financial Services Research*, 7:151–170.

Clark, J. A. (1988). Economies of scale and scope at depository financial institutions: A review of the literature. *Federal Reserve Bank of Kansas City Economic Review*, 73:16–33.

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

Elyasiani, E. and Mehdian, S. (1995). The comparative efficiency performance of small and large U.S. commercial banks in the pre- and post- deregulation eras. *Applied Economics*, 27:1069–1079.

European Commission (1988). The Economics of 1992 European Economy. 35.

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

Fukuyama, H., Guerra, R., and Weber, W. L. (1999). Efficiency and ownership: Evidence from Japanese credit cooperatives. *Journal of Economics and Business*, 51:473–487.

Gilligan, T. and Smirlock, M. (1984). An empirical study of joint production and scale economies in commercial banking. *Journal of Banking and Finance*, ???:67–77.

Goerlich, F. J. (1999). Desigualdad, diversidad y convergencia: (más) instrumentos de medida. Technical report, IVIE.

Hart, P. E. (1971). Entropy and other measures of concentration. *Journal of the Royal Statistical Society*, Ser.A,134:73–85.

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.

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

Leibenstein, H. (1966). Allocative efficiency vs "X-efficiency". *American Economic Review*, 56(3):392–415.

Leibenstein, H. (1978). On the basic proposition of X-efficiency. *American Economic Review*, 68(2):328–332.

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

Mayes, D., Harris, C., and Lansbury, M. (1994). *Inefficiency in industry*. The European Initiative. Harvester-Wheatsheaf, London.

McAllister, P. H. and McManus, D. A. (1993). Resolving the scale efficiency puzzle in banking. *Journal of Banking and Finance*, 17:389–405.

Mester, L. J. (1989). Testing for expense preference behaviour: mutual versus stock savings and loans. *The Rand Journal of Economics*, 20:483–498.

Mester, L. J. (1993). Efficiency in the savings and loan industry. *Journal of Banking and Finance*, 17(2–3):267–286.

Mitchell, K. and Onvural, N. M. (1996). Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form. *Journal of Money, Credit, and Banking*, 28:178–199.

Mountain, D. C. and Thomas, H. (1999). Factor price misspecification in bank cost function estimation. *Journal of International Financial Markets, Institutions and Money*, 9:163–182.

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 y cajas de ahorro españolas: un análisis de la frontera no paramétrico. *Revista Española de Economía*, 12(1):35–73.

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

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, 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.

Simon, H. A. (1959). Theories of decision making in economics and behavioural science. *American Economic Review*, 49(3):253–283.

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

Stennek, J. (2000). Competition increases X-efficiency: A limited liability mechanism. *European Economic Review*, 44:1727–1744.

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

Wheelock, D. C. and Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in US banking, 1984–1993. *Journal of Money, Credit, and Banking*, 31(2):212–234.

Wheelock, D. C. and Wilson, P. W. (2000). New evidence on returns to scale and product mix among US commercial banks. *Journal of Monetary Economics*, forthcoming.

Yoo, S. M. (1992). Technical efficiency over time in Korea, 1978–88: Exploratory analysis. In Caves, R. E., editor, *Industrial Efficiency in Six Nations*. MIT Press, Cambridge.

Table 1: Definition of the relevant variables (1997)

Variable	Definition	Mean	Std Dev
Outputs			
y_1	Loans[‡]	375536	698227
y_2	Other earning assets[‡]	382912	927361
y_3	Savings deposits[‡]	376079	689750
Inputs			
x_1	Total labour expenses[‡]	10780	20423
x_2	Funding[‡]	721068	1525598
x_3	Physical capital[‡]	19931	39705
Inputs' prices			
ω_1	Price of labour $= \frac{x_1}{\text{number of employees}}$	5.122	1.141
ω_2	Price of funds $= \frac{\text{financial costs}}{x_2}$	0.060	0.183
ω_3	Price of physical capital $= \frac{\text{amortizations+other non-interest expenses}}{x_3}$	0.499	0.420

[‡]In millions of 1990 pesetas.

Table 2: Efficiency evolution, banking firms (1985–97)

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Commercial banks	Simple mean	78.88	75.08	79.11	80.81	81.37	80.98	80.07	81.43	80.67	78.00	81.90	78.10	77.12
	Weighted mean	90.20	87.77	88.51	89.69	90.58	89.14	90.56	91.08	83.07	89.67	88.42	89.40	89.04
	Standard deviation	16.61	14.95	15.84	15.08	13.92	16.18	16.21	15.86	15.72	17.02	14.39	17.22	17.61
Savings banks	Simple mean	85.57	82.59	84.08	82.17	81.63	83.85	82.29	85.06	84.28	83.53	88.02	88.17	87.76
	Weighted mean	86.24	87.03	88.17	88.75	89.03	90.37	85.77	90.37	89.05	89.65	88.95	93.23	93.24
	Standard deviation	9.63	9.55	9.70	10.29	10.39	10.02	11.22	9.80	9.62	9.57	8.08	8.11	8.57
Total	Simple mean	82.09	78.69	81.50	81.46	81.49	82.36	81.14	83.18	82.40	80.66	84.84	82.94	82.24
	Weighted mean	88.95	87.50	88.39	89.34	89.99	89.65	88.88	90.81	85.26	89.66	88.99	90.92	90.73
	Standard deviation	14.11	13.19	13.48	13.01	12.35	13.65	14.08	13.42	13.27	14.21	12.17	14.52	14.98

Frontiers have been estimated for each bank and each year, considering jointly commercial banks and savings banks.

Table 3: Transition probability matrices, normalized efficiency scores

	Upper endpoint:				
(Number)	0.802	0.885	0.964	1.077	1.173
(248)	0.69	0.21	0.07	0.01	0.03
(251)	0.21	0.47	0.22	0.06	0.04
(249)	0.05	0.28	0.41	0.21	0.05
(251)	0.03	0.04	0.25	0.54	0.15
(249)	0.01	0.04	0.06	0.16	0.74
Ergodic distribution	0.20	0.21	0.20	0.19	0.20

a) Annual transitions

	Upper endpoint:				
(Number)	0.781	0.895	0.985	1.090	1.124
(21)	0.19	0.48	0.10	0.14	0.10
(21)	0.14	0.48	0.05	0.10	0.24
(20)	0.10	0.25	0.35	0.20	0.10
(21)	0.19	0.05	0.48	0.14	0.14
(21)	0.19	0.10	0.14	0.14	0.43
Ergodic distribution	0.16	0.29	0.19	0.14	0.22

b) 13-year transitions

Table 4: Concentration ratios, banking firms (%)

		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Assets	CR1	6.81	6.38	6.39	10.11	9.48	8.98	10.17	10.35	11.73	9.80	8.47	10.11
	CR3	19.48	17.74	17.31	22.13	21.43	23.08	26.76	27.82	28.74	26.17	21.29	27.03
	CR5	30.17	28.01	27.69	32.96	32.36	34.24	39.04	40.52	40.92	38.18	32.77	38.74
	CR10	48.23	47.81	46.91	49.23	48.87	49.37	52.46	54.61	53.53	53.90	53.39	53.48
Loans	CR1	9.08	7.69	6.72	9.83	9.06	8.22	11.00	10.82	10.42	9.54	8.47	8.66
	CR3	21.59	20.18	19.51	22.01	21.72	20.83	24.99	24.98	24.80	24.12	21.29	23.82
	CR5	32.91	31.47	30.20	32.93	32.44	31.94	36.08	36.75	35.30	34.63	32.77	35.01
	CR10	48.26	47.64	47.04	48.41	48.67	47.57	50.90	52.41	50.70	51.78	49.70	51.02
Savings deposits	CR1	7.70	7.37	6.33	8.80	8.22	7.79	9.98	10.59	10.48	8.96	9.00	8.94
	CR3	21.03	20.09	18.48	20.99	19.08	20.83	26.36	26.96	27.85	26.46	25.88	25.80
	CR5	31.07	30.31	29.09	31.41	29.11	31.19	38.27	39.05	39.72	37.90	38.10	37.83
	CR10	46.99	45.98	44.32	45.13	44.52	48.25	52.00	53.18	51.37	51.89	51.77	51.44

Table 5: Transition probability matrices, weighted efficiency scores

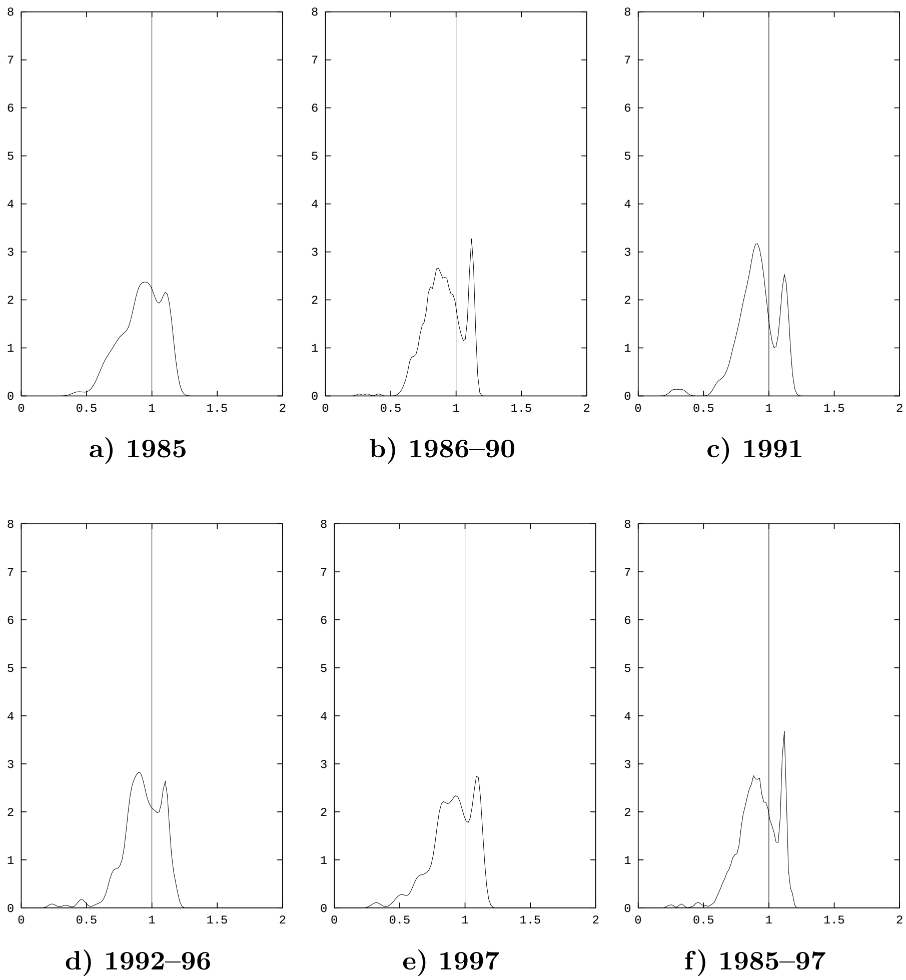
	Upper endpoint:				
(Probability)	0.802	0.885	0.964	1.077	1.173
(0.08)	0.50	0.21	0.15	0.01	0.13
(0.13)	0.18	0.50	0.18	0.05	0.10
(0.15)	0.02	0.23	0.39	0.15	0.21
(0.17)	0.04	0.02	0.16	0.55	0.23
(0.47)	0.00	0.03	0.07	0.11	0.79
Ergodic distribution	0.07	0.13	0.15	0.18	0.46

a) Annual transitions

	Upper endpoint:				
(Probability)	0.781	0.895	0.985	1.090	1.124
(0.13)	0.05	0.30	0.04	0.05	0.56
(0.19)	0.07	0.44	0.01	0.09	0.38
(0.09)	0.06	0.31	0.38	0.18	0.07
(0.13)	0.24	0.01	0.31	0.27	0.18
(0.47)	0.02	0.16	0.03	0.16	0.62
Ergodic distribution	0.07	0.23	0.11	0.16	0.43

b) 13-year transitions

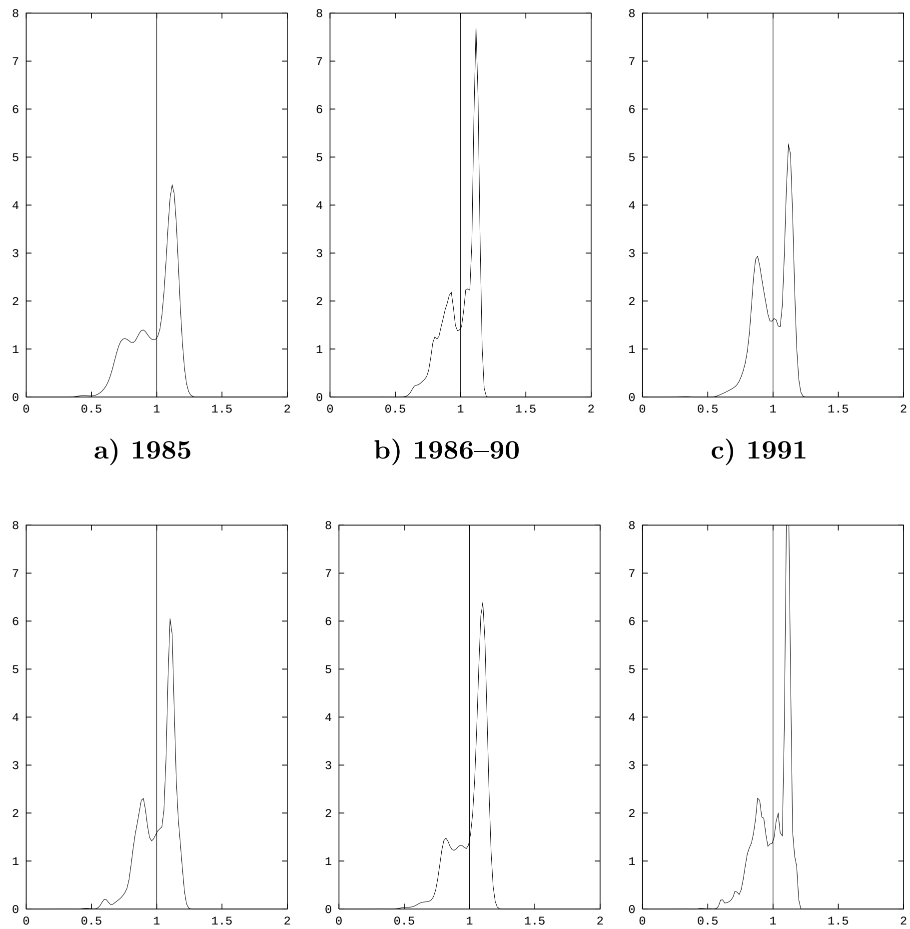
Figure 1: Normalized efficiency densities, banking firms



a) 1985 b) 1986–90 c) 1991

d) 1992–96 e) 1997 f) 1985–97

Figure 2: Normalized efficiency densities, banking firms (weighted)



a) 1985 b) 1986–90 c) 1991

d) 1992–96 e) 1997 f) 1985–97