Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Banking & Finance 27 (2003) 891–917
Journal of BANKING & FINANCE
www.elsevier.com/locate/econbase

Efficiency in Japanese banking: An empirical analysis

Leigh Drake *, Maximilian J.B. Hall

Department of Economics, Loughborough University Banking Centre, Loughborough University, LE11 3TU Loughborough, Leicestershire, UK

Received 10 November 2000; accepted 10 December 2001

Abstract

This paper utilises the non-parametric frontier approach, data envelopment analysis, to analyse the technical and scale efficiency in Japanese banking using a recent cross-section sample. Efficiency analysis is conducted across individual banks, bank types and bank size groups. Following Berger and Humphrey [Eur. J. Oper. Res. 98 (1997) 175], problem loans are controlled for as an exogenous influence on bank efficiency. Powerful size-efficiency relationships are established with respect to both technical and scale efficiency. Furthermore, the logic of the recent large-scale merger wave in Japan is questioned as the larger (City) banks are generally found to be operating above the minimum efficient scale and to have limited opportunity to gain from eliminating *X*-inefficiencies. The opposite result is found for the smaller banks. Finally, the results suggest that controlling for the exogenous impact of problem loans is important in Japanese banking, especially for the smaller regional banks.

JEL classification: G21; D21; G23
Keywords: Data envelopment analysis; Technical and scale efficiency; Banks

1. Introduction

The Japanese financial system is currently experiencing a phase of significant structural change and consolidation. In part, this reflects a legacy of relatively poor profitability and the problem loans associated with the bursting of the "bubble

* Corresponding author. Tel.: +44-1509-263-171; fax: +44-1509-269-232.
E-mail address: l.m.drake@lboro.ac.uk (L. Drake).

0378-4266/03/$ - see front matter
doi:10.1016/S0378-4266(02)00240-6

economy'' of the late 1980s (Hall, 1999a, 2000). It also reflects the impact of deregulation and increasing competition from abroad (Hall, 1998a,b). These pressures have produced a wave of mergers, both across hitherto fragmented segments of banking and financial markets and across the traditional Keiretsu structures. During 1999, for example, a planned three-way amalgamation was announced between the Industrial Bank of Japan (IBJ) and two large city banks, Dai-Ichi Kangyo Bank (DIK) and Fuji Bank. The planned single holding company (Mizuho) will represent the world's largest banking group (by assets) with combined assets in excess of ¥bn 141,800 (approximately $1177 bn at the exchange rate ruling on 6 June 2001), and will account for around 25% of the Japanese retail and corporate banking market. Mergers have also been announced between Sumitomo Bank and Sakura Bank, between Mitsubishi Trust and Banking and the Bank of Tokyo Mitsubishi, and between Sanwa, Asahi and Tokai Bank (although the Asahi Bank subsequently withdrew), while Sumitomo Trust and Daiwa Bank have announced an asset management alliance (see Table 1 for full details).

As for the banks' profitability, almost all measures tell the same sad story. For example, the return on assets at the major Japanese banks averaged −0.1% during the years 1993/94 to 1997/98, compared to an average of +1.2% for the 22 largest US banks. Similarly, core profitability (i.e. operating profits before provisions) over total assets for the same period was a meagre +0.7% at Japanese banks compared with a figure of +2.1% for US banks. The need for Japanese banks to increase their profitability is thus self-evident, although the intensification in competition they are facing following financial liberalisation is not making this easy. In time, it is to be hoped that stronger shareholder pressure for increased profits (until now, weak corporate governance—corporate borrowers are often major holders of bank equity—has undermined market discipline), a switch in managerial focus to risk-adjusted rates of return and return on equity (rather than balance sheet size), and astute diversification within the new deregulated environment—which widens opportunities to cross-sell products, market new products and access new customers and markets—will deliver the much needed boost in profits. But in the short run the solution to higher profitability is likely to come mainly from cost savings, which the banks hope to deliver through mergers and rationalisation.

The ability to merge is thus potentially very important to Japanese banks. Although their profitability is low, cost efficiency, especially at the major banks, is good—the cost to income ratio (overheads as a proportion of operating profits before provisions) for Japanese banks over the 1993/98 period is slightly better than the ratio achieved by US banks. If significant economies of scale do exist in the Japanese banking sector, then banks would be able to secure potential cost benefits (thereby increasing profitability) by merging and moving closer to the minimum efficient scale (MES).

This paper utilises the non-parametric frontier approach, data envelopment analysis (DEA) to investigate both the scale and technical efficiency of a recent cross-section of Japanese banks. The results provide very clear evidence of substantial economies of scale for the smaller banks but suggest that diseconomies of scale exist

Table 1
Recent mergers/alliances in the Japanese banking industry[a]

Date of announcement	Institutions involved	Nature of alliance	Stated objectives of the alliance
(1) 20 August 1999	The Industrial Bank of Japan, Dai-Ichi Kangyo Bank and Fuji Bank.	A full merger is likely, to form the biggest bank in the world by size of assets ($1,470 bn). In the short term, the "consolidation" will take the form of the establishment of a joint holding company by the end of the year 2000 (it actually took place in October), with the banks continuing to operate separately until early 2002. Logic: IBJ facing the same pressures as the Long-Term Credit Bank and Nippon Credit Bank, the other long-term credit banks recently nationalised; the city bank partners, who will provide the retail distribution network, want access to IBJ's client list and expertise in wholesale finance.	To cut costs, over a five-year period, by: (i) Slimming down the combined branch networks (of 645 branches) by 170; (ii) reducing combined annual expenses by ¥100 bn ($0.83 bn); (iii) shedding staff—up to 7000 of the existing combined workforce of 34,000 may go. Total cost savings are put at ¥220 bn ($1.83 bn). The plan is to create three entirely new companies by Spring 2002, covering retail, investment and corporate banking, with each separate entity being given individual return-on-equity targets. In this fashion it is hoped that the merged entity, to be called the Mizuho Financial Group, will become a top-five, global universal banking operation earning a 12% return on equity by fiscal 2005.
(2) 7 October 1999	Asahi Bank and Tokai Bank, both city banks.	A full merger is planned by October 2001 to create a "super-regional" bank headquartered in Nagoya. A joint holding company was formed in October 2000.	To generate operating profits of ¥450 bn ($3.74 bn) a year and to achieve a return-on-equity of 14.5% by the year 2003. This will be achieved, in part, by shedding up to 4000 of the combined labour force of 24,000, and by producing cost savings of ¥32 bn ($0.27 bn) per year.
(3) 14 October 1999	Sumitomo Bank and Sakura Bank, both city banks.	A full merger[b] is planned for 2001 (it actually occurred in April) to create the world's second largest banking group, by assets size ($950 bn). (The new group was named the Sumitomo Mitsui Banking Corporation.) No holding company structure is envisaged. In the interim, equity stakes in each other will be taken, and computer systems and some business operations combined. Most significant sign yet that "keiretsu"/"zaibatsu" relationships are crumbling.	To cut costs by: (i) cutting the combined workforce by around 30% (i.e. by 9000) by 2004; (ii) reducing the number of combined domestic branches by 151 and overseas branches by 32.

(*continued on next page*)

Table 1 (*continued*)

Date of announcement	Institutions involved	Nature of alliance	Stated objectives of the alliance
(4) 9 November 1999	Sumitomo Trust Bank and the city bank Daiwa Bank.	Under the asset management alliance announced, the two banks will set up a new trust bank joint venture—50/50 owned—to provide asset administration services (for $444 bn of assets).The new bank will be capitalised at around $50 bn. A full merger is not on the agenda.	
(5) 13 March 2000	Three city banks, Sanwa Bank, Tokai Bank, and Asahi Bank. (The last two mentioned had already agreed on a union previously—see (2) above. In June 2000, however, Asahi Bank withdrew from the tripartite alliance citing irreconcilable differences over future management strategy.)	It was announced that the three banks were considering a full merger in the year 2001. If the merger had gone ahead, it would have resulted in the creation of Japan's second largest banking group with assets of around $960 bn. The logic of the merger lay in the nature of the geographical fit—Sanwa is strong in Osaka, Asahi in Tokyo and Saitama, and Tokai in Nagoya—and the common focus on the retail and small business markets. A holding company structure was envisaged. In the event, Sanwa Bank, Tokai Bank and Toyo Trust and Banking formed a holding company, UFJ Holdings, in April 2001; and Sanwa and Tokai will merge in April 2002.	Although the lack of business overlap makes cost-cutting difficult, it was envisaged that 150 domestic branches and 30 overseas branches would be closed, but none before the year 2003. IT savings of between ¥30 bn. ($0.25 bn.) and ¥50 bn ($0.42 bn) a year were expected to result from the integration of the banks' existing systems. Up to 5000 jobs (14%) were also expected to be cut over the next five years.
(6) 19 April 2000	Mitsubishi Trust and Banking and the city bank Bank of Tokyo Mitsubishi.	A full merger was planned for 2000/1. (In the event, a holding company (named Mitsubishi Tokyo Financial Group) was established in April 2001, with the two entities—plus Nippon Trust Bank and Tokyo Trust Bank—becoming subsidiaries of it. Mitsubishi Trust and Banking will merge with Nippon Trust Bank and Tokyo Trust Bank in April 2002.)	Substantial cost savings from cutting the number of branches and staff are envisaged.

Source: Press Reports (various); Euromoney (1999), 'Japan's New Leviathans', pp. 48–56, November; Japanese Bankers Association (2001).

[a] This table updates the information contained in Hall (1999b).

[b] Effectively, a takeover of Sakura Bank by Sumitomo Bank.

for the larger City banks. Interestingly, technical efficiency levels are found to improve with bank size, and the Trust and Long-Term Credit Bank sectors are found to exhibit the highest overall efficiency levels.

2. Brief literature review

Given the problems experienced by the Japanese financial system (and banks in particular—Hall, 1999a) in recent years, and the recent pressures for consolidation, surprisingly little academic research has been undertaken into the costs and efficiency of Japanese banks. This contrasts markedly with the wealth of research undertaken into the performance of US financial institutions. Tachibanaki et al. (1991) were among the first researchers to focus on this area. They estimated a two output Translog cost function using a sample of 61 banks between 1985 and 1987, and found evidence of economies of scale for all sizes of bank in all three years of the study.

In contrast, Fukuyama (1993) used the non-parametric technique, DEA, to analyse the overall technical efficiency (OE) of Japanese commercial banks, decomposing this into its two constituent components, pure technical efficiency (PTE) and scale efficiency (SE) (see Section 3). The cross-section sample consisted of 143 banks for the financial year 1990/91. The mean level of OE for the whole sample was found to be 0.8645 which, compared to a maximum level of unity, implies that banks could, on average, have produced the same levels of outputs with around 14% less resources or inputs. Unlike other studies, Fukuyama found evidence of only mild economies of scale, with the mean level of SE being 0.9844. Hence, most of the observed inefficiency was associated with pure technical (mean PTE score, 0.8509), rather than scale inefficiency. Interestingly, however, only 7% of the sample exhibited constant returns to scale, with the vast majority of banks (81%) exhibiting increasing returns. This implies that most of the Japanese banks were operating below the MES and hence above the minimum attainable average costs.

McKillop et al. (1996) were the next to report on efficiency in Japanese banking. They used the composite cost function developed by Pulley and Braunstein (1992) to analyse costs and efficiency in giant Japanese banks. The study relates to annual data for 5 Japanese City banks over the period 1978–1991, and McKillop et al. use the intermediation approach in a three output, three input model. The authors find evidence of statistically significant economies of scale for all banks at the sample mean. Furthermore, the estimated values of the economies of scale parameter were found to range between 1.08 and 1.28 (where in this study a figure over 1 indicates economies of scale), numbers very similar to the values found by Tachibanaki et al. (1991) in the earlier study. It is interesting to note, however, that McKillop et al. found that this pattern of economies of scale holds for all years of the sample "except for the late 1980s onwards where the results suggest that constant returns pertain for all models" (p. 1665). This accords with the work of Fukuyama (1993), who finds that, based on 1990/91 data, "the majority of the City banks exhibit constant returns to scale, implying that the City banks seemingly operate close to the MES..." (p. 1107). McKillop et al. (1996) also refer to "the persistent, and somewhat surprising, finding of increasing returns to scale for *all sizes* of Japanese banks" (p. 1652) prior to the late 1980s. Until recently, the vast majority of empirical studies in other countries have found that economies of scale are exhausted at relatively low output levels. Hence, in this context, the Japanese results are indeed surprising. It is also clear, however, that both McKillop et al. (1996) and Fukuyama (1993) find some evidence that large city banks operating in the late 1980s/early 1990s exhibited constant returns to scale.

Clearly, given the current consolidation wave sweeping Japan, the precise nature of economies of scale in Japanese banking is extremely important, both from an academic and policy perspective; and a more recent paper attempts to shed fresh light on the scale economy puzzle in Japanese banking. Altunbas et al. (2000) utilise the parametric Fourier flexible stochastic cost frontier (see Berger et al., 1997) to investigate both scale economies and *X*-efficiencies in Japanese banking. They specify three outputs (total loans, total securities and off-balance sheet items) and three inputs (labour, capital and total funds). In addition to the usual cost function specification, however, Altunbas et al. also test for the impact of risk and quality factors (clearly of potential significance given the recent banking crisis in Japan) on costs, scale economies and *X*-efficiency. It should be noted, however, that Japanese banks were renowned for concealing the true scale of their bad debt problems for most of the 1990s (see Hall, 2000). Their sample consists of 136 Japanese banks and covers the years 1993 to 1996; and they also allow for the possibility of technical change over the period via the inclusion of a simple time trend.

The authors find that economies of scale in Japanese banking tend to be overstated when risk and quality factors are not incorporated, particularly for the larger banks. Specifically, they find that "diseconomies of scale become much more widespread and optimal bank size falls from around Yen 5–10 Trillion [$42–83 bn] to Yen 1–2 Trillion [$8–17 bn] when risk and quality factors are taken into account." This contrasts with recent evidence from the US, however. Hughes and Mester (1998), for example, find evidence that banks of all sizes enjoy significant economies of scale once the risk preferences of managers are accounted for. Furthermore, they suggest that the failure to adequately take account of this factor accounts for the previous findings of constant returns in US banking.

With respect to *X*-inefficiencies, Altunbas et al. find that these range between 5% and 7%, in contrast with the levels of around 20% typically found in studies of US banks (see Berger and Humphrey, 1997). Moreover, the *X*-efficiency estimates are found to be much less sensitive to the exclusion of risk and quality factors than the economies of scale estimates.

In summary, relatively little academic research has been undertaken into costs and efficiency in Japanese banking, particularly in relation to the post-1996 era. Furthermore, there is clearly a degree of ambiguity concerning the true nature of the returns to scale/economies of scale in Japanese banking. This issue, however, is of great importance given the recent trend towards large-scale mergers and rationalisation within the Japanese banking sector.

Hence, in contrast with the study of Altunbas et al. (2000), this paper utilises the non-parametric technique, DEA, in order to offer a fresh perspective on both scale and technical efficiency in Japanese banking. Furthermore, in contrast with Fukuyama's 1993 DEA study, which used data up to the early 1990s, we utilise a large cross-section sample from 1997. Comparison with Fukuyama's results will therefore facilitate analysis of the potential impact of the 1990s banking crisis on the structure and performance of the Japanese banking sector. We also focus on the contrasts between the relative efficiency levels of the various sub-sectors of Japanese banks, such as City banks, Trust banks, etc, and on the nature of the size-efficiency relationship in Japanese banking. Finally, we follow Altunbas et al. (2000) in recognising the potential importance of lending quality in studies of banking efficiency and, in particular, Japanese banking efficiency.

3. Methodology

In academic studies of costs and efficiency in banking, two main approaches have been adopted; a parametric and a non-parametric approach. Both require the specification of a cost or production function or frontier, but the former involves the specification and econometric estimation of a statistical or parametric function/frontier, while the non-parametric approach provides a piecewise linear frontier by enveloping the observed data points. Hence, this latter technique has come to be termed DEA. Unlike the parametric approach, DEA does not require the specification of a particular functional form for the cost or production function. Hence, the derived efficiency estimates are not functional form dependent. In contrast, the accuracy of the efficiency estimates in the parametric approach are conditional on the accuracy of the chosen functional forms' approximation to the cost or production function. It should be noted, however, that, unlike the parametric stochastic frontier approach (SFA), DEA does not allow for the presence of a random error term. Hence, DEA attributes any deviation from the efficient frontier as being purely associated with inefficiency, and for this reason DEA may overstate the true levels of relative inefficiency for some units (Berger and Mester, 1997; Grosskopf, 1996).

Notwithstanding this potential reservation, we prefer to use DEA for this present study as the use of SFA requires the specification of a cost function, and hence data on input prices. Unfortunately, although data on the costs of the labour input is available for Japanese banks, data on the number of employees is not. Hence, it is not possible to produce an accurate measure of the labour input price. Furthermore, in this study we elect to concentrate on measures of technical and scale efficiency. In contrast, the use of SFA produces measures of *X*-efficiency, which is composed of both technical and allocative efficiency. Clearly, the accurate measurement of the latter may be compromised by the lack of accurate input price data for labour. Furthermore, the issue of functional form dependence in respect of parametric techniques such as SFA is particularly pertinent in the context of this study, given the wide diversity across the banking institutions in Japan, both in respect of size and business mix. As Berger et al. (1997) point out, "usually a local approximation such as the translog is specified, which has been shown to be a poor approximation for banking data that are not near the mean scale and product mix (see McAllister and McManus, 1993; Mitchell and Onvural, 1996). The translog also forces the frontier average cost curve to have a symmetric U shape in logs" (p. 7). Furthermore, Mester (1997) argues that the failure to adequately take account of bank heterogeneity can result in estimates of bank cost efficiency being biased. In contrast, DEA imposes very little structure on the efficient frontier and does not require the maintained assumption that all firms face the same unknown production technology. Provided that a reasonably comprehensive set of inputs and outputs is specified, DEA simply requires the existence of an input/output correspondence in order to produce relative efficiency measurements.

Finally, it is important to note that, although SFA does allow for random error in the analysis of efficiency, the decomposition of the combined error term into the random error and inefficiency components requires an assumption concerning the appropriate distribution of the latter. As Bauer et al. (1998) argue, however, "any distributional assumptions simply imposed without basis in fact are quite arbitrary and could lead to significant error in estimating individual firm efficiencies" (p. 12).

3.1. Data envelopment analysis (DEA)

Within the DEA framework, it is possible to decompose relative efficiency performance into the categories initially suggested by Farrell (1957) and later elaborated by Banker et al. (1984) and Fare et al. (1985). The constructed relative efficiency frontiers are non-statistical or non-parametric in the sense that they are constructed through the envelopment of the decision making units (DMUs), with the "best practice" DMUs forming the non-parametric frontier. Farrell's categories are best illustrated, for the single output/two input case, in the unit isoquant diagram, Fig. 1, where the unit isoquant (yy) shows the various combinations of the two inputs (x_1, x_2) which can be used to produce 1 unit of the single output (y). The firm at E is productively (or overall) efficient in choosing the cost minimising production process given the relative input prices (represented by the slope of WW′). A DMU at Q is allocatively inefficient in choosing an inappropriate input mix, while a DMU at R is both allocatively inefficient (in the ratio OP/OQ), and technically inefficient (in the ratio OQ/OR), because it requires an excessive amount of both inputs, x, compared with a firm at Q producing the same level of output, y.

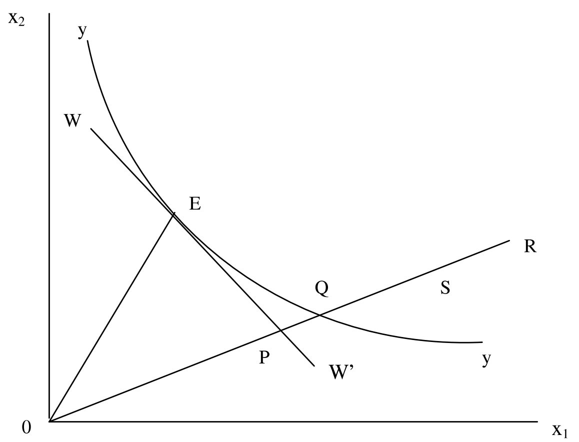

Fig. 1. Farrell efficiency.

The use of the unit isoquant implies the assumption of constant returns to scale. However, a firm using more of both inputs than the combination represented by Q may experience either increasing or decreasing returns to scale so that, in general, the technical efficiency ratio OQ/OR may be further decomposed into scale efficiency, OQ/OS, and pure technical efficiency, OS/OR, with point Q in Fig. 1 representing the case of constant returns to scale. The former arises because the firm is at an input–output combination that differs from the equivalent constant returns to scale situation. The latter, pure technical efficiency, represents the failure of the firm to extract the maximum output from its adopted input levels, and hence it may be thought of as measuring the unproductive use of resources. In summary,

$$\begin{aligned}\text{productive efficiency} &= \text{allocative efficiency} \times \text{scale efficiency} \\ &\quad \times \text{pure technical efficiency}, \\ \text{OP/OR} &= [\text{OP/OQ}] \times [\text{OQ/OS}] \times [\text{OS/OR}]. \qquad (1)\end{aligned}$$

As outlined previously, accurate data on all input prices was not available for the Japanese banks in this sample and it was, therefore, not possible to consider the issue of allocative efficiency. Nevertheless, unlike SFA, where the inefficiency measure is necessarily a composite of allocative and pure technical efficiency, in DEA it is possible and legitimate to focus on overall technical efficiency and the decomposition into scale and pure technical efficiency. Hence, concentrating on technical efficiency, Farrell suggested constructing, for each observed DMU, a pessimistic piecewise linear approximation to the isoquant, using activity analysis applied to the observed sample of DMUs in the organisation/industry in question. This produces a relative rather than an absolute measure of efficiency since the DMUs on the piecewise linear isoquant constructed from the boundary of the set of observations are defined to be the efficient DMUs.

Subsequent developments have extended this mathematical linear programming approach. If there are n DMUs in the industry, all the observed inputs and outputs are represented by the n-column matrices: X and Y. The input requirement set, or reference technology, can then be represented by the free disposal convex hull of the observations, i.e., the smallest convex set containing the observations consistent with the assumption that having less of an input cannot increase output. We do this by choosing weighting vectors, λ (one for each firm) to apply to the columns of X and Y in order to show that firm's efficiency performance in the best light.

For each DMU in turn, using x and y to represent its particular observed inputs and outputs, pure technical efficiency is calculated by solving the problem of finding the lowest multiplicative factor, θ, which must be applied to the firm's use of inputs, x, to ensure it is still a member of the input requirements set, or reference technology. That is choose

$$\begin{aligned}\{\theta, \lambda\} \text{ to} : \min\ \theta \text{ such that} : \theta x &\geqslant \lambda' X, \\ y &\leqslant \lambda' Y, \\ \lambda i &\geqslant 0, \quad \sum \lambda i = 1, \ \ i = 1, \ldots, n\end{aligned} \qquad (2)$$

To determine scale efficiency, we solve the technical efficiency problem (2) without the constraint that the input requirements set be convex, i.e. we drop the constraint $\sum \lambda i = 1$. This permits scaled up or down input combinations to be part of the DMU's production possibility set. Fig. 2 illustrates this for the case of a single input and a single output. In Fig. 2, the production possibility set under constant returns to scale is the region to the right of the ray, OC, through the leftmost input–output observation. Any scaled up or down versions of the observations are also in the production possibility set under this assumption of constant returns to scale.

Imposing the convexity constraint, $\sum \lambda i = 1$, ensures the production possibility set is the area to the right of the piecewise linear frontier VV′, which does not assume constant returns to scale, but allows for the possibility of increasing returns to scale at low output levels and decreasing returns at high output levels. The resulting overall technical and pure technical efficiency ratios, AQ/AR and AS/AR, are illustrated for one of the observations. Scale efficiency is the ratio of the two results.

In the case of programme (2), the efficiency ratios with and without the convexity constraint may be labelled θ_p and θ_o, and scale efficiency, θ_s, is then θ_o/θ_p. In the subsequent results we refer to overall technical efficiency as OE, pure technical efficiency as PTE, and scale efficiency as SE. As explained above, it follows that

$$\text{OE} = \text{PTE} \times \text{SE} \quad \text{and} \quad \text{SE} = \text{OE}/\text{PTE}. \tag{3}$$

Although the scale efficiency measure (SE) will provide information concerning the degree of inefficiency resulting from the failure to operate with constant returns to scale, i.e., at the MES, it does not provide information as to whether a DMU is operating above or below the MES. Hence, in order to establish whether scale inefficient banks exhibit increasing or decreasing returns to scale, we simply solve the technical efficiency problem (2) under the assumption of non-increasing returns to scale rather than variable returns to scale. If these two measures of PTE differ, this indicates that the bank is operating in the region of increasing returns to scale. Conversely, if the two measures coincide then the bank is operating in the region of

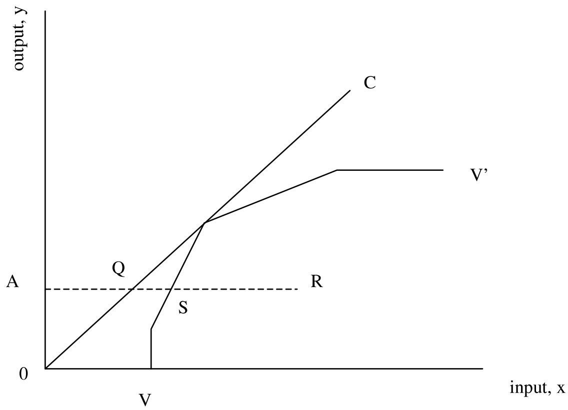

Fig. 2. Scale and technical efficiency.

decreasing returns to scale. Hence, in the context of Fig. 2, it is possible for two banks to exhibit the same degree of scale inefficiency (deviation from the CRS frontier), but for one bank to exhibit increasing returns while the other exhibits decreasing returns.

In the interests of brevity, information regarding the nature of returns to scale is not provided for each individual banks. However, Section 5 discusses the salient features of the size efficiency relationship and the nature of returns to scale across bank types.

4. Data

The sample used in this study consists of data for 149 Japanese banks for the financial year ending March 1997. Furthermore, the data covers the full spectrum of "ordinary" bank types operating in Japan, embracing City banks, Regional banks, and Second Association Regional banks, *plus* certain "specialised" banks i.e. the Long-Term Credit Banks (LTCBs) and Trust banks. The City bank group comprises the nation wide branching institutions. By tradition the suppliers of short-term funds to large corporations, they have recently begun to focus on developing the longer-term end of their business. Additionally, they have been forced to cultivate clients from amongst the smaller corporate fry and from the personal sector because of the downturn in their traditional customers' demand for bank credit. Apart from borrowing from the Bank of Japan, they fund themselves in the deposit and short-term financial markets. They are widely engaged in securities-type operations (the more so since the implementation of the "Big Bang" package of reform, see Hall, 1998b), both at home and abroad, and most have a significant international dimension to their operations. The largest, by assets size, are some of the largest banks in the world.

The second grouping of ordinary banks comprises the Regional banks, which are divided into Regional banks proper and Second Association Regional banks (the latter tending to be the smaller operators of the two groupings). They are smaller in scale than the City banks, and usually confine their operations to the principal cities of the prefectures in which their head offices are located. Accordingly, their local ties are strong, with the bulk of their lending going to small and medium-sized companies in the locality. The bulk of their deposits are time deposits of an initial term of one year or more, with over 50% being accounted for by deposits from individuals. Apart from making business loans, they also invest heavily in the local stock market, and they are important lenders in the local money market.

The first group of specialised banks, the LTCBs, were established under the Long-Term Credit Bank law of 1952 to engage in long-term finance in order to achieve a separation between short- and long-term finance and to reduce the long-term funding burden imposed on the ordinary banks by the demands of industry. They are traditionally distinguishable from ordinary banks in their funding operations, their lending operations and in the size of their branch networks. On the funding side, they alone are permitted (until recently) to issue debentures, but they must confine their deposit taking to their borrowing clientele, public corporations, corporations which entrust bond subscriptions to them, and other clients; on the lending side, the average term to maturity of their loans is considerably longer; and the size of their branch networks is considerably smaller. Two of the three original LTCBs, the Long-Term Credit Bank and the Nippon Credit Bank, have recently been sold to private consortiums following temporary nationalisation in 1998. The other group of specialised long-term banking institutions, the Trust banks, are ordinary banks which have been allowed to concurrently engage in trust operations. They obtain most of their funds from trusts (particularly the loan trusts peculiar to Japan) and, in respect of their trust accounts, satisfy much of the large Japanese corporations' capital investment finance needs. They also offer savings and deposit accounts, and tap the money markets. Apart from their trust and banking operations, they are active in funds management, and also provide real-estate broking and stock transfer services. These days, they are significant investors in overseas markets, where they are also engaged in securities underwriting and distribution as well as in lending operations.

With reference to these categories of Japanese bank, the sample consists of 10 City banks, three LTCBs, 7 trust banks, 64 Regional banks and 65 member banks of the Second Association of Regional Banks (Second Association Banks).

We follow the intermediation approach in modelling the bank production function (see Sealey and Lindley, 1977). Hence, the initial DEA model estimated consists of three outputs (Y) and three inputs (X), as follows:

$Y1$ Total loans and bills discounted,
$Y2$ Liquid assets and other investments in securities,
$Y3$ Other income;

$X1$ General and administrative expenses,
$X2$ Fixed assets (premises and equipment),
$X3$ Retail and wholesale deposits.

$Y3$ is included in the analysis to reflect the fact that banks around the world have been diversifying, at the margin, away from traditional financial intermediation (margin) business and into "off-balance sheet" and fee income-generating business. Hence, it would be inappropriate to focus exclusively on earning assets as this would fail to capture all the business operations of modern banks. Furthermore, Stiroh (2000) finds, in the context of US bank holding companies, that the efficiency estimates are particularly sensitive to the failure to incorporate non-traditional activities in the output specification. The inclusion of other income is therefore intended to proxy the non-traditional business activities of Japanese banks, and $Y3$ consists of net fee and commission income and other net non-interest operating income. (It should be noted that a universal banking style of operation, embracing wide-ranging securities and insurance services, only became possible after 1999 following implementation of the 'Big Bang' reforms (see Hall, 1998b, 1999b).)

With respect to inputs, the specification reflects the standard intermediation approach, in the sense that capital and labour inputs are used to intermediate deposits into loans and other earning assets. In this case, the capital input is proxied by the level of fixed assets, while the labour input is proxied by general and administrative expenses, which are typically dominated by personnel expenses. The use of this proxy is necessitated due to the unavailability of data on employee numbers across the sample. Clearly, it is possible that the use of personnel expenses rather than employee numbers could result in some bias against those banks which hire high quality, and therefore relatively high cost staff. This potential bias should be mitigated, however, given that banks with higher quality staff should expect to see some benefit in output terms. Hence, providing that the high quality staff are sufficiently productive, such banks will not be disadvantaged from a relative efficiency perspective. [1]

As mentioned previously, it may be important to account for risk and lending quality (problem loans) in the assessment of banking efficiency, particularly in the context of Japanese banking. Whether these factors should be controlled for in efficiency analysis is a controversial issue, however. As Berger and Humphrey (1997) argue:

> "Whether or not it is appropriate to control for problem loans depends on which is the dominating explanation for the observed negative relationship between measured efficiency and problem loans... If 'bad luck' dominates, then problem loans are mostly exogenous and should be controlled for in efficiency models. If 'bad management' dominates, then problem loans are essentially endogenous to financial institution efficiency and should not be controlled for in the analysis of efficiency" (p. 194).

An important and well recognised problem, however, is that it is typically very difficult to determine the extent to which problem loans are exogenous or endogenous, and hence whether or not they should be controlled for in the efficiency analysis. Following Berger and Humphrey (1997), however, a potential solution is to recognise that, if problem loans should be controlled for, then their impact on efficiency is essentially outside the control of bank management. Hence, in the context of DEA, we choose to measure the impact of problem loans as an additional uncontrollable input within the DEA model and use the provisions for loan losses as an indicator of the extent of problem loans. Hence, provisions for loan losses are acting as a proxy for the resources and expenses involved in dealing with problem loans.

In terms of the DEA literature, we are following the approach of Charnes et al. (1990) in incorporating loan loss provisions as an input in the DEA analysis. The

[1] We are grateful to an anonymous referee for raising this issue of staff quality. This is also an issue in cost function/SFA studies where inputs prices are typically proxied by personnel costs divided by employee numbers.

important innovation, however, is to recognise that this input should not be modelled as a choice input, but as an uncontrollable input reflecting the exogenous impact of problem loans. In respect of the linear programming problem itself, therefore, the uncontrollable input is taken into account in the DEA analysis but, as the input cannot be varied at the discretion of the bank, all inputs, other than the uncontrollable input, are subject to the multiplicative factor, θ, in respect of Eq. (2) (see Banker and Morey, 1986). It must be recognised, however, that, in practice, banks do have a certain degree of discretion with regard to provisioning for problem loans. As mentioned previously, Japanese banks were renowned for concealing the true scale of their bad debt problems for most of the 1990s.

Finally, in order to facilitate the subsequent analysis of the size-efficiency relationship in Japanese banking, the sample is further sub-divided into six size classes. For ease of exposition, this sub-division is undertaken according to total lending ($Y1$). The size groups are as follows, where all the data is expressed in Y billion:

	$Y1$ Range	No. of banks
Group 1	0–1000	56
Group 2	1001–2000	42
Group 3	2001–3000	18
Group 4	3001–6000	12
Group 5	6001–10,000	6
Group 6	Above 10,000	11

It is clear that the chosen size groups are not equal in terms of the number of banks. This reflects, in the main, the predominance of relatively small banks in Japan. To produce groups of roughly equal size would entail identifying an additional three or four groups of small banks, or combining the larger bank groups, which would considerably reduce the discrimination in the analysis, particularly in respect of differences between the very large banks (predominantly City banks) and the smaller banks. Hence, we decided to segment the bank size groups on the basis of what appear to be "natural breaks" in the bank size spectrum. In recognition of the fact that any such sub-division is to a degree arbitrary, however, we also analyze the size efficiency relationship using rank correlation analysis based on the whole sample.

5. Results

In order to provide a basis for comparison, and due to the uncertainty regarding the appropriateness of controlling for problem loans, we first present the results from the basic DEA model without the incorporation of provisions for loan losses. We then contrast these results with those obtained when we control for the exogenous impact of problem loans.

5.1. Relative efficiencies across Japanese banks and bank types

In the interests of brevity, summary statistics for overall technical efficiency (OE), pure technical efficiency (PTE) and scale efficiency (SE) are provided in Table 2 for the full sample. Table 2 also provides these statistics by bank type, i.e., City bank, Trust bank, etc. The full set of results is available from the authors on request.

It is clear from Table 2 that Japanese banks exhibit considerable overall inefficiency, with a sample mean for OE of 72.36 and a minimum figure of 53.37. This suggests that the majority of banks could make significant reductions in input usage (given output levels) and thereby achieve significant cost savings. It is interesting to note, however, that the bulk of this inefficiency is attributable to pure technical inefficiency rather than scale inefficiency. Whereas the mean DEA score for the latter

Table 2
DEA efficiency results

	OE	PTE	SE	$Y1$
Full sample				
Mean	72.36	78.11	92.78	3844.19
SD	11.62	11.19	7.51	7926.47
Min	53.37	60.40	53.37	105.42
Max	100.00	100.00	100.00	43751.86
City banks				
Mean	87.09	95.56	91.27	28341.98
SD	4.61	5.68	4.40	12543.64
Min	80.01	85.80	82.18	6971.25
Max	93.65	100.00	97.27	43751.86
Long-Term Credit Banks				
Mean	100.00	100.00	100.00	17551.58
SD	0.00	0.00	0.00	7898.34
Min	100.00	100.00	100.00	9080.48
Max	100.00	100.00	100.00	24713.57
Trust banks				
Mean	100.00	100.00	100.00	5345.28
SD	0.00	0.00	0.00	3275.95
Min	100.00	100.00	100.00	853.71
Max	100.00	100.00	100.00	9301.23
Regional banks				
Mean	68.47	71.65	95.55	2146.00
SD	7.84	7.17	4.72	1552.65
Min	56.74	60.40	73.13	226.90
Max	100.00	100.00	100.00	8432.23
Second association regional banks				
Mean	69.54	78.42	89.00	813.17
SD	8.35	8.22	8.75	543.73
Min	53.37	65.09	53.37	105.42
Max	95.30	100.00	99.50	2415.95

is 92.78, the mean score for PTE is only 78.11. This is very much in line with recent US evidence (see, for example, Berger and Humphrey, 1997) which typically finds that *X*-inefficiency (failure to minimise costs for a given output vector) is a much more serious problem than scale inefficiency (failure to operate at the MES). The results are in marked contrast, however, to previous results on Japanese banking efficiency. Fukuyama's (1993) DEA study, for example, finds evidence of only mild economies of scale, with a mean SE score of 98.44. This contrasts with our finding of 92.78. We also find much higher levels of pure technical inefficiency, with a mean PTE score of 78.11 in contrast to Fukuyama's estimate of 85.09.

Of potentially more interest than the mean efficiency levels, however, is the evidence obtained from the breakdown of efficiency across bank categories. It is clear from Table 2, for example, that the Trust and Long-Term Credit Banks form by far the most efficient sectors in Japanese banking (notwithstanding the subsequent nationalisation of two LTCBs in 1998 (see Hall, 1999c)), with all of these banks exhibiting both scale and pure technical efficiency (and thereby overall efficiency). This result implies that all the Trust and Long-Term Credit Banks are on the efficient frontier. Although more research is clearly warranted in respect of these institutions, part of the explanation may be found in the fact that these two sets of institutions are relatively specialist and narrowly focused in their business operations, have limited branch networks and can support assets from sources other than deposits. If this is part of the explanation for the high efficiency levels, however, then it is significant to note that the trend in banking around the world is away from specialisation and focus, and towards financial conglomeration and universal banking.

In contrast, the City banks all exhibit evidence of decreasing returns to scale (DRS), with a mean SE score of 91.27. Not surprisingly, the degree of scale inefficiency does appear to increase with bank size (as proxied by $Y1$). The smallest City bank, for example, has an SE score of 97.27, while the largest exhibits a score of 82.18. This evidence is in line with the findings of Altunbas et al. (2000) who find significant diseconomies of scale for the largest Japanese banks, although both sets of results are at odds with the typical earlier findings of pervasive scale economies in Japanese banking. As outlined previously, even studies which used data into the 1990s (such as Fukuyama (1993) and McKillop et al. (1996)) found that the largest City banks tended to operate with constant returns to scale, but found no evidence of decreasing returns to scale for these large banks.

Clearly, our finding of decreasing returns for all the City banks does call into question the commercial logic of the large-scale Japanese bank mergers (typically involving the City banks) announced in recent years. These mergers would be likely to exacerbate scale inefficiencies and hence their success, at least in efficiency terms, would have to be founded either on potential cost cutting benefits and improvements in *X*-efficiency, or on potential economies of scope via the increased scope for diversification within the larger financial institutions. Although we can only capture the technical rather than allocative aspect of *X*-efficiency in this study, the evidence in Table 2 does suggest that the City banks exhibit relatively low levels of *X*-inefficiency, particularly in contrast with the regional banks. The mean PTE score for the City banks is 95.56 in contrast to the mean levels for the Regional and Second Association Regional Banks of 71.65 and 78.42 respectively.

It should be noted, however, that, as emphasised by Berger et al. (1999), cost efficiency is not the only motive behind financial consolidation. In respect of what is viewed as the primary motive for consolidation, value maximisation, Berger et al. stress that this can be achieved either by increasing efficiency or by increasing market power in the setting of prices. Furthermore, Berger et al. stress that there are many non-value maximising motives behind financial consolidation, such as empire building objectives designed to enhance managerial status and executive compensation. They argue that this motive is particularly likely in banking where corporate control may be relatively weak and, as alluded to previously, this is particularly true in the case of Japanese banking. Finally, it must be recognised that the recent consolidation wave witnessed in the Japanese financial system is strongly associated with the severe banking crisis which developed in the early 1990s.

With respect to overall efficiency, the Trust banks and Long-Term Credit Banks both exhibit mean efficiency levels of 100% and are clearly the most efficient sectors in Japanese banking, as outlined previously. The City banks exhibit a mean OE score of 87.09, which is composed of relatively high scale inefficiency (SE = 91.27) but relatively high pure technical efficiency (PTE = 95.56). In contrast, the Regional banks exhibit an overall efficiency score of only 68.47 and this is composed of relatively high levels of technical inefficiency (PTE = 71.65), but more modest levels of scale inefficiencies (SE = 95.55). Not surprisingly, given that these regional banks are much smaller than the City banks and the Trust and Long-Term Credit Banks, they are overwhelmingly characterised by increasing returns to scale (IRS). Furthermore, a clear size efficiency relationship is evident even within this sector, as the largest and smallest of the Regional banks exhibit SE levels of 99.82 and 73.13 respectively.

Finally, the Second Association Regional banks exhibit a mean OE level of 69.54, which is slightly higher than that of the larger Regional banks. In contrast to the latter, however, the Second Association Regional banks exhibit a higher mean level of technical efficiency (PTE = 78.42), but higher levels of scale inefficiency (SE = 89.0). Again, this seems to be indicative of a powerful size efficiency relationship in Japanese banking with the SE levels for the largest and smallest Second Association banks varying between 97.05 and only 53.37 respectively.

Although we have so far been focusing on the full sample results and on differences in efficiency levels across bank types, it is clear from Table 2 that there does appear to be a powerful size efficiency relationship evident in Japanese banking. We consider this in more detail in the next section.

5.2. The size efficiency relationship

As outlined previously, the sample of Japanese banks was split into six size groups (ranked by $Y1$) in order to facilitate the analysis of the size efficiency relationship. The mean levels of OE, PTE and SE for each group, together with other summary statistics, are presented in Table 3, with Group 1 representing the smallest banks and Group 6 the largest. Focusing first on scale efficiency, it is clear that the largest degrees of scale inefficiencies are evident for the smallest banks (Group 1 SE = 86.68) and the largest banks (Group 6 SE = 92.31). Furthermore, all but one of the Group 1 banks exhibited increasing returns (the other exhibiting constant returns), while all but two of the largest banks exhibited decreasing returns to scale. Additionally, it is interesting to note that the two Group 6 banks exhibiting constant returns to scale (CRS) were not City banks but Long-Term Credit Banks.

It is evident that Table 3 provides clear support for a relatively flat "saucer shaped" average cost curve in Japanese banking, as scale efficiency levels gradually improve from Group 1 up to Group 5 (SE = 99.52) before deteriorating again for the largest Group 6 banks. Drake (2001) also finds evidence of this type of unit cost curve in a study of UK banking efficiency. Furthermore, if we contrast the SE levels for Group 1 and Group 6 banks (86.68 and 92.31, respectively) it would appear that

Table 3
Size efficiency relationship

	OE	PTE	SE	*Y*1
Group 6 banks				
Mean	88.99	96.46	92.31	29092.98
SD	6.79	5.66	5.28	10277.32
Min	80.01	85.80	82.18	10671.24
Max	100.00	100.00	100.00	43751.86
Group 5 banks				
Mean	94.96	95.41	99.52	8291.37
SD	9.03	8.87	1.10	1093.63
Min	77.71	77.85	97.27	6889.94
Max	100.00	100.00	100.00	9301.23
Group 4 banks				
Mean	79.25	79.68	99.42	4628.45
SD	10.61	10.39	0.40	697.08
Min	68.31	69.14	98.67	3605.56
Max	100.00	100.00	100.00	5978.86
Group 3 banks				
Mean	68.67	70.71	97.08	2425.35
SD	6.35	6.09	0.98	326.16
Min	60.44	62.86	95.24	2009.71
Max	80.18	81.33	99.24	2981.79
Group 2 banks				
Mean	69.70	72.28	96.26	1447.11
SD	9.85	8.95	2.34	284.17
Min	56.97	60.40	90.77	1009.74
Max	100.00	100.00	100.00	1982.75
Group 1 banks				
Mean	68.43	79.16	86.68	536.74
SD	9.18	8.76	8.42	231.81
Min	53.37	62.10	53.37	105.42
Max	100.00	100.00	100.00	988.28

the unit cost curve is asymmetric in the sense that the potential economies of scale available to the smallest banks are greater than the potential diseconomies facing the largest banks. This would suggest considerable scope for mergers and rationalisation amongst the smaller, mainly regional, banks in Japan.

More formal analysis of the size–efficiency relationship, particularly in respect of the size/scale efficiency relationship, is problematic due to the evident non-linearity in the relationship. Hence, rather than using regression analysis, we prefer to use non-parametric rank correlation analysis in order to investigate these relationships further. It is clear from Table 4, for example, that, based on the full sample, there is a very strong positive rank correlation (0.70) between size, as measured by $Y1$, and SE, but no apparent relationship between PTE and size (rank correlation 0.06). Consequently, a more modest positive association exists between size and OE (rank correlation 0.46). With respect to the strong positive rank correlation between size and scale efficiency, however, it may well be that this relationship is dominated by the very many Regional and Second Association Regional Banks which overwhelmingly exhibit IRS. Furthermore, as we have seen, the scale inefficiencies exhibited by these banks appear to be more severe than the scale inefficiencies associated with the City banks displaying DRS.

In order to gain a more accurate perspective on the size efficiency relationship, therefore, in Table 4 we also divide the sample into those banks exhibiting IRS and those exhibiting DRS, excluding those banks with CRS. The latter, with only one exception, are the Long-Term Credit Banks and the Trust Banks which exhibit overall efficiency. It is clear from this analysis that, for those banks exhibiting IRS, there is a very strong positive rank correlation between size and SE (rank correlation 0.877, significant at the 5% level), which confirms the evidence provided in Table 3 in respect of the two sets of Regional Banks. In contrast, however, there is an equally strong and statistically significant negative rank correlation between size and SE for those banks displaying DRS (rank correlation, −0.826,). These contrasting size/scale efficiency relationships are illustrated very clearly in Figs. 3 and 4, where size is measured in terms of $Y1$ (Y bn). Furthermore, this confirms the evidence from Table 3 that the scale inefficiencies are much more severe for the smaller banks than for the larger banks.

Table 4
Rank correlations with $Y1$

OE	PTE	SE
Full sample		
0.46*	0.06	0.70*
Banks exhibiting (DRS)		
0.09	0.79*	−0.83
Banks exhibiting (IRS)		
0.23	−0.40*	0.88*

*Significant at the 5% level.

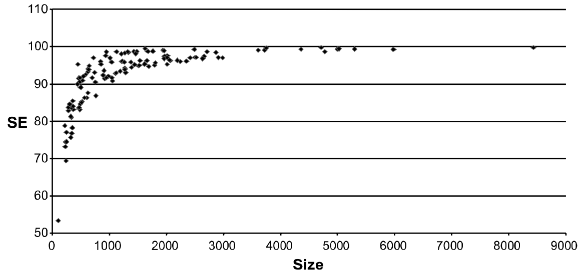

Fig. 3. Scale efficiency (IRS banks).

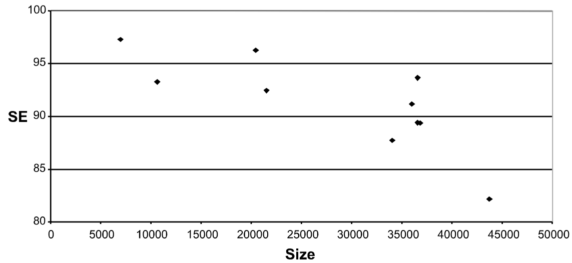

Fig. 4. Scale efficiency (DRS banks).

With respect to the MES in Japanese banking, it is clear from Fig. 5 (based on the full sample) that the MES is attained in the range ¥ 3000–4000 bn ($25–33 bn), in terms of *Y*1. Furthermore, with the exception of two large Long-Term Credit Banks, the CRS region generally extends up to around ¥ 10,000 bn ($8.3 bn). This evidence is echoed in Table 3 where the Group 4 and 5 banks exhibit mean SE scores of 99.42 and 99.52 respectively, and have mean asset levels of ¥ 4628 bn ($39 bn) and ¥ 8291 bn ($69 bn). Interestingly, with only one exception, all of the Group 5 banks exhibited constant returns to scale (CRS) (or very close to CRS in the case of one bank). Significantly, the one exception was a City bank, which was actually smaller than the mean size for the group. This suggests that size may not be the only important influ-

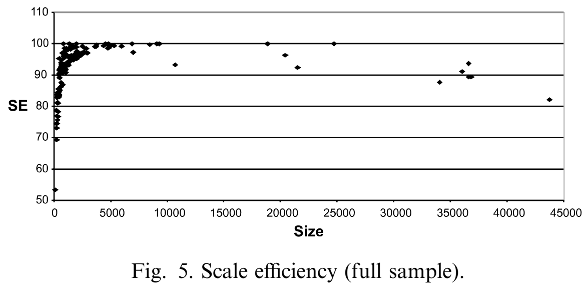

Fig. 5. Scale efficiency (full sample).

ence on scale efficiency/economies of scale. This hypothesis is given further credence by the fact that all of the Long-Term Credit Banks were found to be scale efficient, despite the fact that their mean size was *Y* 17552 bn ($146 bn). This is clearly well above the MES indicated by the Group 4 and 5 banks in Table 3. Indeed, as emphasised previously, two of the Long-Term Credit Banks feature in the Group 6 banks, and are the only large banks to exhibit CRS rather than decreasing returns.

In summary, therefore, although bank size clearly has an important influence on scale efficiency, it would appear that scale efficiency is also strongly influenced by bank type and business structure. It may be the case, for example, that the scale efficiencies sustained by the relatively large Long-Term Credit Banks is associated with their more specialised and focused business structure. A similar result was established by Drake (2001) in the case of UK banks. Specifically, it was found that the more specialised banks such as the Abbey National (which had previously converted from a mutual Building Society to a stock bank) exhibited CRS at asset sizes well above the level where the established clearing banks began to encounter DRS. Clearly, this issue merits further investigation, although this is beyond the scope of this paper.

Turning now to pure technical efficiency, it is clear from Table 3 that the larger Japanese banks tend to be considerably more efficient than their smaller counterparts, with Group 5 and Group 6 banks exhibiting mean PTE scores of 95.41 and 96.46 respectively. In contrast, all the other size groups exhibit mean levels of PTE of less than 80. This is consistent with the US evidence which suggests that larger banks tend to be more *X*-efficient than smaller banks (see Berger and Humphrey, 1997).

It is clear from Table 3, however, that a straightforward relationship does not exist between size and PTE in Japanese banking, as PTE appears to initially deteriorate with size. Whereas Group 1 banks exhibit a mean PTE score of 79.16, the corresponding levels for Groups 2 and 3 are 72.28 and 70.71 respectively. Thereafter, PTE appears to improve with size as Groups 4, 5 and 6 exhibit PTE scores of 79.68, 95.41 and 96.46 respectively. This non-linear relationship is confirmed more formally by rank correlation analysis. For all the banks in size groups 1, 2 and 3, the rank correlation between size and PTE is −0.46 (significant at the 5% level). In contrast, the corresponding rank correlation for all the banks in size groups 4 to 6 is 0.66 (significant at the 5% level). A similar result can be seen in Table 4 where, for those banks exhibiting IRS (the smaller banks), there is a negative rank correlation between size and PTE (−0.40, significant at the 5% level). In contrast, for the larger banks exhibiting DRS, there is a very strong positive rank correlation between size and PTE (0.79, significant at the 5% level). This clear non-linear relationship probably explains the lack of a significant rank correlation between size and PTE for the whole sample (0.06) evident in Table 4.

This result has a very important policy implication as it suggests that, while the smallest Japanese banks may reap very significant cost savings from expansion and mergers via economies of scale, these gains may be offset, at least partially, by increasing levels of *X*-inefficiency. In turn, this suggests that any planned mergers between relatively small Japanese banks should be firmly grounded on cost saving and rationalisation strategies if they are to be effective in cost terms. This would clearly be enhanced if the least efficient of the smaller Japanese banks were to be acquired by their more efficient counterparts.

Finally, in terms of overall efficiency, the Group 1 and Group 3 banks clearly exhibit the lowest levels of OE (68.43 and 68.67 respectively). For the Group 1 banks this is attributable to a mixture of scale and technical inefficiencies, whereas for the Group 3 banks it is attributable primarily to pure technical inefficiency. At the other end of the spectrum, the Group 5 banks are clearly the most efficient Japanese banks in terms of OE. Not only do these banks operate close to MES but they also exhibit relatively high levels of PTE. These factors combine to produce an OE score of 94.96, well ahead of the figure of 88.99 exhibited by the larger Group 6 banks.

5.3. Controlling for problem loans

Having established the basic DEA results, we now analyse the potential impact of risk and problem loans on Japanese banking efficiency. As outlined previously, these results are obtained by modifying the initial DEA model to incorporate an additional (but non-discretionary) input in the form of provisions for loan losses. As previously, we do not present these results for all banks but in the form of summary statistics for the full sample, the bank categories and the bank size groups. This allows us to focus clearly on the impact of loan loss provisions on the efficiency analysis.

Table 5 presents the summary statistics for the full sample and the individual bank sub-sectors. It is immediately apparent that controlling for problem loans raises the mean pure technical efficiency level of all banks considerably from 78.11 to 89.38, and also results in a more modest increase in the mean level of scale efficiency from 92.78 to 96.59. The combination of these two factors is such that the mean overall efficiency level increases from 72.36 to 86.32.

These results suggest that potential economies of scale (i.e., the extent of the increasing returns to scale which appear to dominate Japanese banking) may well be overestimated when risk factors are excluded, and this is in line with the findings of Altunbas et al. (2000). In contrast, however, we find that the pure technical efficiency estimates are much more sensitive than the scale efficiency estimates to the ex-

Table 5
Mean efficiency levels (controlling for problem loans)

	OE	PTE	SE
Sample mean	86.32	89.38	96.59
Sample min	57.15	69.20	76.73
Sample max	100.00	100.00	100.00
City bank's mean	92.08	96.15	95.74
LTCB's mean	100.00	100.00	100.00
Trust bank's mean	100.00	100.00	100.00
Regional bank's mean	85.87	88.43	97.20
Second association bank's mean	83.66	87.57	95.56

clusion of risk factors, whereas Altunbas et al. find that economies of scale estimates are much more sensitive than X-efficiency estimates.

Turning now to the impact on bank types, Table 5 indicates that, in contrast to the sample mean results, controlling for problem loans has relatively little impact on the technical efficiency levels of the City banks (PTE increases from 95.56 to 96.15). It does appear, however, that the exclusion of risk factors exaggerates the extent of the scale diseconomies experienced by the large City banks as SE increases from 91.27 to 95.74. Furthermore, whereas all the City banks were previously found to experience decreasing returns, when problem loans were incorporated two City banks were found to be operating with constant returns to scale.

It is interesting to note that the Trust and Long-Term Credit Banks are clearly the most efficient banking sectors, as all the member banks exhibit both technical and scale efficiency, irrespective of whether or not we control for problem loans.

Turning now to the two sets of regional banks, the overall efficiency of the larger Regional banks increases markedly from 68.47 to 85.87, while that of the second association banks increases by slightly less, from 69.54 to 83.66. In terms of the relative decompositions of the changes in overall efficiency, however, the Second Association banks exhibit substantial increases in both technical efficiency (78.42–87.57) and scale efficiency (89.00–95.56), whereas the Regional banks exhibit a modest increase in scale efficiency (95.55–97.20), but a very substantial increase in PTE (71.65–88.43).

These results confirm the evidence from the sample means, and indicate that technical efficiency levels can be highly sensitive to the exclusion of risk and problem loan factors, particularly for the smaller regional banks. Furthermore, this makes sense in terms of our approach to controlling for exogenous risk factors, as these (relatively) smaller regional banks would tend to be much more exposed to regional specific factors than the larger and more diversified (both geographically and possibly also in business terms) City, Trust and Long-Term Credit Banks. As Berger and Humphrey (1997) emphasise:

> "If problem loans are generally caused by 'bad luck' events exogenous to the bank, such as regional specific downturns, then measured cost efficiency may be artificially low because of the expenses of dealing with these loans (e.g., extra monitoring, negotiating workout arrangements, etc.)" (p. 194).

Finally, the evidence from the regional banks, and particularly the smaller Second Association banks, supports our earlier assertion (based on the sample mean evidence) that the exclusion of risk factors does appear to exaggerate the extent of the increasing returns evident for the majority of Japanese banks, and hence to exaggerate the potential economies of scale and gains from mergers for these smaller banks.

In order to provide a more definitive analysis of the impact of problem loans on the size–efficiency relationship in Japanese banking, however, Table 6 presents the revised mean efficiency levels for the six bank size groupings. In line with the evidence gleaned from the different bank types, controlling for problem loans appears to flatten the "saucer shaped average cost curve" alluded to previously. More

Table 6
Mean efficiency levels (controlling for problem loans)

Size group	OE	PTE	SE
Group 6	93.52	96.99	96.37
Group 5	95.96	96.40	99.54
Group 4	90.95	94.37	96.47
Group 3	84.45	86.65	97.54
Group 2	85.16	87.33	97.55
Group 1	84.39	88.53	95.31

specifically, the extent of the increasing and decreasing returns for the smallest and largest banks respectively, appears to be reduced with the incorporation of loan loss provisions. The mean SE level for Group 1 banks, for example, increases from 86.68 to 95.31, while the corresponding figure for the Group 6 banks increases from 92.31 to 96.37. This also implies that Japanese banks in general are operating much closer to MES than the initial DEA results would suggest.

With respect to the actual level of the MES in Japanese banking, Altunbas et al. (2000) found that:

> "... diseconomies of scale become much more widespread and optimal bank size falls from around Yen 5–10 Trillion to Yen 1–2 Trillion when risk and quality factors are taken into account".

In contrast, our DEA results indicate that the MES in Japanese banking appears to increase following the incorporation of risk factors. It is evident from Table 6, for example, that the MES now appears to be clearly located in the size range Yen 6–10 Trillion ($50–83 bn) in terms of $Y1$, as represented by the Group 5 banks which exhibit a mean SE level of 99.54. It is also evident that the extent of the decreasing returns appears to be reduced, rather than increased, once problem loans are controlled for, as the mean SE level falls to only 96.37 for the Group 6 banks, as compared to 92.31 in the initial results. Furthermore, once risk is controlled for, over a third of the Group 6 banks actually exhibit constant returns to scale, rather than decreasing returns.

These results suggest that the potential diseconomies resulting from the recent large-scale merger wave in Japan may be less serious than the initial DEA results would suggest. If we take the case of the three-way merger between IBJ, DIK and Fuji bank, for example, our initial results suggested SE levels of 100, 93.65 and 87.72 respectively. In contrast, once we control for problem loans the corresponding figures are 100, 96.15 and 89.21.

Finally, if we consider the impact of problem loans on the size-pure technical efficiency relationship, Table 5 indicates that the impact is fairly minimal for the largest banks but very substantial for the smaller banks. For the Group 5 and 6 banks, for example, we find that the mean PTE levels increase from 95.41 to 96.40 and from 96.46 to 96.99 respectively. For the smaller banks, however, the increase varies between around 9 and 16 percentage points. In the case of the Group 3 banks (the worst performing group according to the initial results), for example, the mean PTE score increases from 70.71 to 86.65. These results appear to confirm that the smaller, and predominantly regional, banks appear to be more exposed to the exogenous impact of problem loans than their larger competitors.

Finally, notwithstanding the favourable impact of controlling for problem loans on the efficiency of the smaller Japanese banks, it remains the case that the larger banks (Groups 5 and 6) exhibit the lowest levels of technical inefficiency. It follows from this, however, that these banks also have the least to gain, in terms of potential X-efficiency gains, from mergers.

6. Conclusions

Notwithstanding the potentially large short run cost savings achievable through rationalisation of the branch network and computing systems and staff cuts, the results presented in this paper cast strong doubts on the commercial logic of the recent wave of planned mergers amongst Japan's largest banks, at least in respect of economies of scale. Although controlling for problem loans does modify the results somewhat, our evidence suggests that these banks are typically operating well above MES and hence any such mergers would tend to exacerbate diseconomies of scale within these larger combined banking groups. Furthermore, the DEA results also suggest that these large banks (and particularly the LTCBs) have the least potential X-efficiency gains as they tend to exhibit the lowest levels of pure technical inefficiencies of all Japanese banks.

There is, however, very clear evidence of potentially significant economies of scale for the smallest Japanese banks, which is suggestive of considerable cost benefits from mergers and rationalisation across these smaller banks, at least for those from the more prosperous regions and possessing high quality human capital. Once again, however, the extent of these potential economies of scale is reduced once we control for the impact of problem loans. Furthermore, our results suggest that the levels of pure technical efficiency tend to deteriorate with size up to the middle ranking banks. Hence, mergers between relatively small banks will need to be carefully planned and managed to ensure that any potential cost savings from economies of scale are not eroded by increased levels of X-efficiency.

Although our results do suggest powerful size efficiency relationships, both with respect to PTE and SE, it is clear that business structure is a potentially more important influence on scale efficiency than size. The Trust and Long-Term Credit Banks, for example, were found to be unambiguously the most efficient banking groups, both in terms of scale and technical efficiency. Furthermore, all of these banks exhibited CRS despite the fact that some of the LTCBs were larger than some of the City banks, which tended to exhibit decreasing returns to scale (particularly when problem loans were not controlled for). Similarly, all of the Trust banks exhibited CRS despite the fact that some of them operated at size levels well below the indicative MES established by the Group 4 and 5 banks in Table 3.

Although controlling for risk is a controversial issue in respect of efficiency analysis, we adopt an innovative approach in the context of DEA, by attempting to control only for the exogenous impact of problem loans on efficiency, as suggested by Berger and Humphrey (1997). Our results suggest that controlling for this exogenous impact is potentially very important as it produced marked changes in both the scale and technical efficiency results. Specifically, the extent of both decreasing and increasing returns were reduced (and hence mean SE levels increased), and technical efficiency levels were also found to increase, particularly for the smaller regional banks.

In spite of the incorporation of loan loss provisions, however, the central results of the paper remained: the Long-Term Credit Banks and the Trust banks were found to be the most efficient banking sectors; the majority of banks displayed increasing returns, but the larger City banks typically displayed decreasing returns; and the MES was found to lie in the size range of the Group 5 banks.

Finally, it is interesting to note that much of the rationalisation and merger activity within Japanese banking in recent years has taken place across banking groups and across financial services boundaries (see Hall, 1999b). The clear differences which emerged between efficiency levels across banking groups, however, suggest that it will be very difficult to predict the outcomes of such cross sector mergers, particularly those involving some of the Trust banks and LTCBs combining with City or Regional banks. If the efficiency of the former is attributable to good management practices, for example, then it is possible that this may improve the performance of the enlarged banking group. If the efficiencies have more to do with focus and specialisation, however, then our results suggest that the trend towards enlargement and financial conglomeration in Japan (as well as in many other countries around the world) may well lead to reduced levels of scale efficiency and, possibly, also X-efficiency.

Hence, accounting for the sources of the differences in efficiency levels across different banking groups in Japan is an extremely important issue for future research.

References

Altunbas, Y., Liu, M.-H., Molyneux, P., Seth, R., 2000. Efficiency and risk in Japanese banking. Journal of Banking and Finance 24, 1605–1628.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30, 1078–1092.

Banker, R.D., Morey, R.C., 1986. The use of categorical variables in data envelopment analysis. Management Science 32, 1613–1627.

Bauer, P.W., Berger, A.N., Ferrier, G.D., Humphrey, D.B., 1998. Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. Journal of Economics and Business 50, 85–114.

Berger, A.N., 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., Leusner, J.H., Mingo, J.J., 1997. The efficiency of bank branches. Journal of Monetary Economics 40, 141–162.

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

Berger, A.N., Demsetz, R.S., Strahan, P.E., 1999. The consolidation of the financial services industry: Causes, consequences, and implications for the future. Journal of Banking and Finance 23, 135–194.

Charnes, A., Cooper, W.W., Huang, Z.M., Sun, D.B., 1990. Polyhedral cone—ratio DEA models with an illustrative application to large commercial banks. Journal of Econometrics 46, 73–91.

Drake, L., 2001. Efficiency and productivity change in UK banking. Applied Financial Economics 11, 557–571.

Euromoney, 1999. Japan's New Leviathans, November, pp. 48–56.

Fare, R., Grosskopf, S., Lovell, K.C.A., 1985. The Measurement of the Efficiency of Production. Kluwer Nijhoft, Boston.

Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (part III), 253–281.

Fukuyama, H., 1993. Technical and scale efficiency of Japanese commercial banks: A non-parametric approach. Applied Economics 25, 1101–1112.

Grosskopf, S., 1996. Statistical-inference and nonparametric efficiency—A selective survey. Journal of Productivity Analysis 7, 161–176.

Hall, M.J.B., 1998a. Financial Reform in Japan: Causes and Consequences. Edward Elgar., December.

Hall, M.J.B., 1998b. Financial reform in Japan: 'Big Bang'. Journal of International Banking Law 13, 58–70.

Hall, M.J.B., 1999a. Current banking problems in Japan: How serious are they and how might they be resolved? Research in Financial Services: Private and Public Policy 11, 3–33.

Hall, M.J.B., 1999b. Japan's Big Bang: the likely winners and losers. Journal of International Banking Law 14, 204–216.

Hall, M.J.B., 1999c. Deposit insurance reform in Japan: better late than never? Journal of Financial Services Research 15, 211–242.

Hall, M.J.B., 2000. What is the scale of the Japanese banks bad debts? Is the situation manageable? Journal of Financial Services Research 17, 69–91.

Hughes, J.P., Mester, L.J., 1998. Bank capitalization and cost: Evidence of scale economies in risk management and signaling. Review of Economics and Statistics, 314–325.

Japanese Bankers Association, 2001. The Banking System in Japan. Tokyo, March.

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

McKillop, D.G., Glass, J.C., Morikawa, Y., 1996. The composite cost function and efficiency in giant Japanese banks. Journal of Banking and Finance 20, 1651–1671.

Mester, L.J., 1997. Measuring efficiency at US banks: Accounting for heterogeniety is important. European Journal of Operational Research 98, 230–242.

Mitchell, K., 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.

Pulley, L.B., Braunstein, Y.M., 1992. A composite cost function for multiproduct firms with an application to economies of scope in banking. Review of Economics and Statistics 74, 221–230.

Sealey, C.W., Lindley, J.T., 1977. Inputs, outputs and the theory of production and costs at depository financial institutions. Journal of Finance 32, 1251–1265.

Stiroh, K.J., 2000. How did bank holding companies prosper in the 1990s? Journal of Banking and Finance 24, 1703–1746.

Tachibanaki, T., Mitsui, K., Kitagawa, H., 1991. Economies of scope and shareholding of banks in Japan. Journal of the Japanese and International Economies 5, 261–281.