School of Finance & Law
Working Paper Series

Cost efficiency, economies of scale and economies of scope in the British retail banking sector

By

John Ashton

Bournemouth University

No. 13.

1998

Bournemouth University, School of Finance & Law (Department of Accounting & Finance), Talbot Campus, Fern Barrow, Poole, Dorset. BH12 5BB.

Published 1998 by the School of Finance and Law, Bournemouth University, Talbot Campus,

Fern Barrow, Poole, Dorset, BH12 5BB.

For further details contact:

John Ashton

School of Finance & Law
Bournemouth University
Fern Barrow
Poole, BH12 5BB
United Kingdom

Tel: (00)-1202-595245

Fax: (00)-1202-595261

Email: jashton@bournemouth.ac.uk
ISBN 1-85899-074-2

A catalogue record for this publication is available from the British Library.

For further information on the series, contact Dr. K. Howell,

School of Finance & Law,
Bournemouth University.

Cost efficiency, economies of scale and economies of scope in the British retail banking sector

By

John Ashton

Abstract

In this paper the efficiency characteristics of the British retail banking sector are investigated. Distribution free cost efficiency, economies of scale and economies of scope are all measured. The study employs a one way fixed effects model with a translog specification of productive technology. Both 'production' and 'intermediation' models of bank production are employed and contrasted. A substantial distribution of cost efficiency is observed in this most important commercial sector. Overall, slight dis-economies of scale are reported for the 'intermediation' approach and substantial diseconomies of scale are recorded for the 'production' approach. A substantial dispersion of cost efficiency is observed for this sector with both model specifications.

1. Introduction

Many studies in the USA have analysed the existence of economies of scale and scope in banking. Such a study is particularly pertinent for the Britain, in the light of the substantial re-regulation of retail banking markets during the 1980s. Significant structural change is observable in this sector as banks have merged, a large number of new banks (de-mutualised building societies) have entered retail banking markets and banks are broadening their range of services. The importance of such a study is primarily the provision of previously absent empirical evidence as to whether many of the recent changes in British banking have been driven by cost factors. Such evidence is of particular importance when attempting to assess developments such as the continued external growth of banks through acquisition and merger and the future regulation of banking.

The only efficiency studies to date of UK retail banking include Allen and Rai (1996), Drake and Howcroft (1994) and Altunbas *et al* (1995). Allen and Rai (1996) who estimated a global cost function including bank observations from 15 countries including the UK. Drake and Howcroft (1995) used a DEA (data envelopment analysis) method to calculate the relative efficiency of bank branches for a British clearing bank. Optimal bank branches where deemed to be those which had total lending of between £3 - 5.25m and an average of nine employees. Altunbas *et al* (1995) estimated inefficiency scores for a pooled sample of banks and building societies in 1993 (see also Molyneux *et al*, 1996). An econometric frontier model was employed with assumptions of an exponential distribution of inefficiency imposed. The model assumed banks and building societies employed three inputs to produce earning assets, the only output from the model. Overall very low levels of inefficiency were discovered.

In these previous studies a number of distinct approaches and methods have been employed in the estimation of efficiency and economic characteristics such as economies of scale and scope. Principal differences between these studies the method of estimation, and the individual measures of efficiency used. The paper is structured in the following manner. Section two outlines the efficiency concept considered. The model specification is considered in section three. The efficiency statistics employed are reviewed in section four and the method of estimation adopted is considered in section five. The data used in the study is discussed in section six. Results are displayed and assessed in section seven. Section eight provides a summary of the paper and forwards conclusions.

2. The efficiency concept

Before deciding what methods to use in the measurement of efficiency it is important to consider the efficiency concept employed. Efficiency has been estimated using a number of efficiency concepts including production and cost. Productive efficiency is derived as the distance an individual institution has from the 'optimal' or 'best practice' institution existing on a production function. This hypothesised 'best practice' institution firm is defined with reference to all the institutions in the sample set. Assumptions of total transferable productive technology and firm objectives of output maximisation and cost minimisation are made for production or cost functions respectively.

A production function assumes that the level of output of an individual institution is dependent on the amount of inputs expended in production, random error and any other additional variables accounting for the environment or particular circumstances of individual institutions. Productive efficiency is therefore limited in its extent to considering the amount of input quantity that may be reduced to produce a specified quantity of output. This form of efficiency is commonly termed productive or technical efficiency. The derivation of production functions has been criticised both for the difficulties in providing appropriate data and the limited definition of efficiency.

Cost efficiency estimates how far the production costs of an individual institution differs from the production costs of a best practice institution or firm operating under the similar conditions and producing the same outputs. This measure is defined with reference to a cost function constructed from the observations of all institutions considered within the sample set. The cost function assumes the total production costs of individual institutions are dependent on the price of variable inputs, such as capital and labour, the quantity or value of outputs produced, random error and any other additional variables accounting for the environment or particular circumstances of individual institutions. A cost function allows the measurement of the least cost proportions of inputs in terms of input prices. This framework enables the consideration of both productive efficiency and the optimal proportion of inputs in terms of input prices or allocative efficiency.

In the study, the cost efficiency concept is employed. This efficiency concept is used for three reasons. Firstly, it is deemed preferable to view efficiency in its broadest sense, incorporating both productive and allocative efficiency. Secondly, their exists a substantial quantity of anecdotal information that banks both in the UK and globally have emphasised the closer management of costs. A study by Salomon Brothers (cited in Molyneux et al, 1996, pp.4-7) suggested "… cost management was now a dominant strategic theme throughout the banking world" (see Molyneux *et al*, p.4). Through adopting a cost efficiency approach it is hoped that the performance of UK retail banks may be considered in form more closely representing that favoured by the industry itself. Lastly, most previous econometric studies of UK depository institutions use cost efficiency concepts. The use of cost efficiency concepts therefore improves the degree of comparability of analysis in this thesis with other studies.

3. Model specification

Two models of depository institution production are employed. In both models, banks are assumed to minimise costs. Two models are employed as micro and macro views of banks differ. In the micro sense, a bank is producing services that are sold in a market-place. In the macro sense, banks act as a producer of money itself. These conflicting properties have led to difficulties in the modelling of bank production. In an economy, many institutions accept or borrow funds from the public, firms and money markets. These funds are then essentially transformed and re-lent to borrowers at a rate of interest.

This form of intermediation is a principal and defining aspect of the production process of banks.

In the ‘intermediation’ approach to modelling bank production, banks borrow funds, which are transformed into loan funds, which form the principal output from the system. The bank is assumed to use deposits as a raw material or intermediate product, which are then transformed into the final products. Outputs from an 'intermediation' model are loan funds and ancillary business, the latter outside the scope of the principal intermediating process. The model thus consists of costs of production (labour and capital costs) and the costs for attracting the borrowed funds that are then re-lent. Thus costs incorporate both the operational costs and the attraction costs, including interest costs.

In this study, the ‘intermediation’ approach bank production is viewed as a transformation of three input groups (capital, labour and deposits; X_1, X_2, X_3) into two output groups; loans; Y_1, and investments Y_3. A production correspondence may be written as:

$$f(Y_1, Y_3; X_1, X_2, X_3) \quad (1)$$

The dual cost function would then be:

$$C = g(Y_1, Y_3; P_1, P_2, P_3) \quad (2)$$

This approach both includes the effect of non-monetary and monetary inputs. Monetary output is included, whilst non-monetary output is omitted. A classical emphasis on the macro and dynamic changes in assets is therefore imposed.

Whilst, the ‘intermediation’ approach implies that bank production is driven by the processes used in the transformation of funds, the ‘production’ approach specifically incorporates the functions of the bank. The provision of depository services is viewed as a form of production in its own right and therefore as an output. Retail banks are viewed with the ‘production’ approach to transform physical inputs (capital and labour; X_1, X_2) into outputs (loans, deposits, and investments); Y_1, Y_2, Y_3. Cost is denoted by non-interest costs of production (OC). A production correspondence could then be written as:

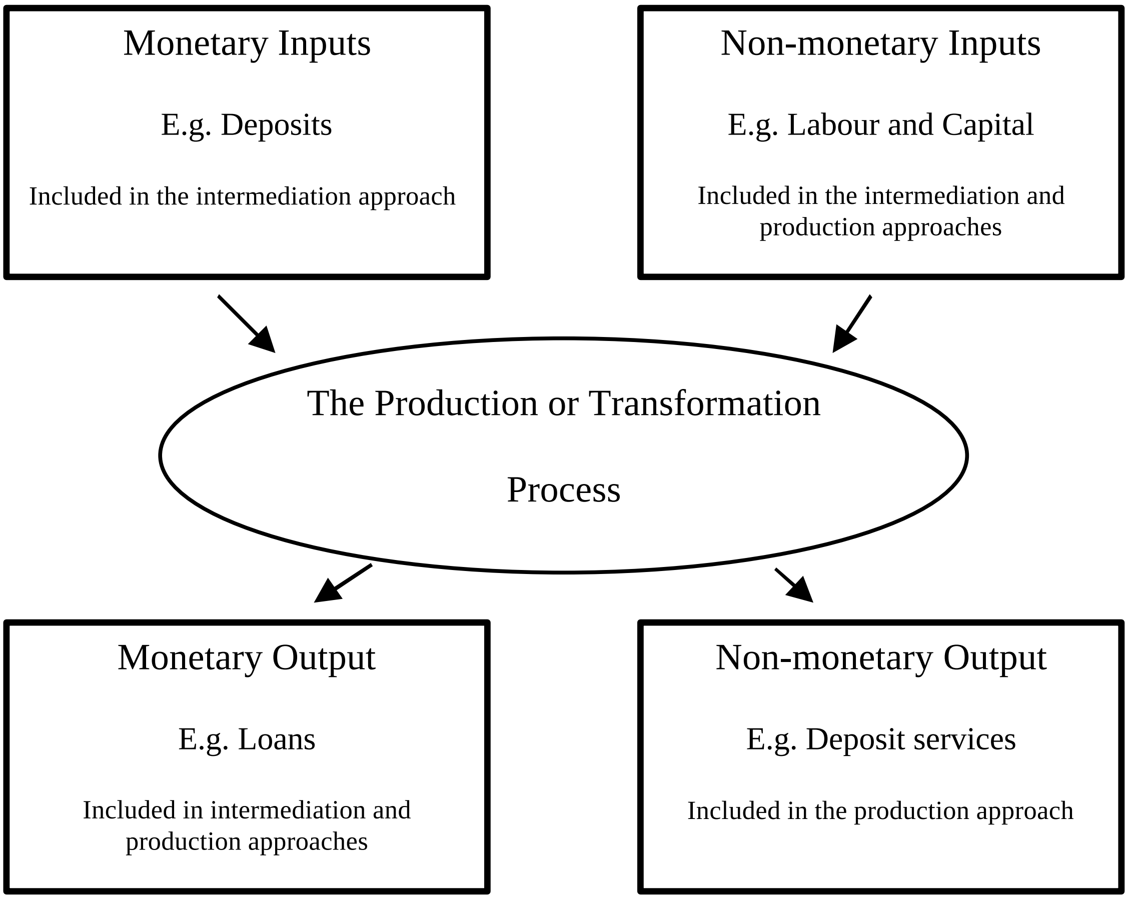
$$f(Y_1, Y_2, Y_3; X_1, X_2) \tag{3}$$

and the dual cost function would be:

$$OC = g(Y_1, Y_2, Y_3; P_1, P_2) \tag{4}$$

This approach uses non-monetary inputs and ignores the effect of monetary inputs, whilst monetary and non-monetary outputs are included. The relationship between the ‘intermediation’ and ‘production’ is displayed in Figure 1.

Figure 1 The relationship between the production and intermediation approaches



Outputs are quantified by their value. Y_1 denotes the aggregate of loans issued by the bank in that year, Y_2 denotes the total deposits held by the bank Y_3 denotes investments held by the bank. The price of labour (P_1) is proxied by the total wage bill divided by the number of full-time equivalent employees. The price of capital (P_2) is proxied by the aggregation of property and equipment rentals and depreciation divided by the book value of physical capital, multiplied by 1000 to provide measure of capital cost for every £1000 of physical capital. The price of deposits (P_3) is total interest payable divided by the book value of deposits multiplied by 1000, to provide a measure of the interest cost of every £1000 of deposits. The level of liquid assets and the level of provisions for bad and doubtful loans are incorporated as variables to amend for the omission of asset quality and liquidity factors in the model. The level of liquid assets (LA) and provision for bad and doubtful loans (PROV) are quantified by their value (for further discussion of these issues see Berger and Mester, 1997).

4. Measures of cost efficiency and data

In this study, economies of scale, economies of product and cost efficiency are estimated. Economies of scale proxied by ray scale elasticity, economies of product mix are measured using economies of scope and cost efficiency is quantified using 'distribution free' efficiency techniques.

Ray scale elasticity, outlined by Baumol *et al* (1982), measures elasticity of cost relative to scale. This measure assumes proportional increases or decreases of cost in relation to scale and the constant composition of outputs in relation to costs. These restrictions limit the measurement of elasticity to a single constant ray emerging from the origin. Such a restriction simplifies the problem at hand where fixed proportions are assumed to exist within the measures, effectively producing an aggregate or composite output. The measure considers changes in scale in isolation of changes in product mix.

Ray scale economies are the first derivative of cost with respect to output evaluated for a representative institution (usually assumed to be a mean value for a set of institutions).

Elasticity’s greater than unity imply dis-economies of scale and values less than unity imply economies of scale; unity denotes constant returns to scale.

Economies of scope, outlined by Baumol *et al* (1982) quantify the cost savings from producing quantities of the two outputs jointly within a single institution relative to specialised production by two institutions. This measure thus represents the economies of simultaneous production relative to specialised production. The magnitude of economies of scope of output (or *y*) relative to the product set *T* (possible production combinations) may be measured by

$$Scope\ economies = [\{y^T + y^{U-T}\} - y^T] / C(y) \qquad (5)$$

where *y* is an arbitrarily chosen value of output *j* (usually the sample mean to denote the multi-product 'representative firm'), $y = y^{U-T} - y^T/y^T$ represents complete specialisation within one output and y^{U-T} represents specialisation within remaining outputs. Values > 0 imply economies of scope and values < 0 imply dis-economies of scope.

The measurement of economies of scale and economies of product mix involves the consideration of the characteristics of a function or frontier. Efficiency measurement differs from the measurement of economies of scale and product mix in that efficiency of an individual institution is viewed to be a distance from the ‘best practice’ function or frontier.

This distance from a cost function or frontier is estimated or captured by a disturbance term. This disturbance term may be viewed as having two components, inefficiency and random error. Random error is assumed to be distributed as a symmetrical two-sided term. Efficiency is represented by a one-sided disturbance component where efficiency represents all the effects of the data that are within the control of the model.

The modelling of this disturbance has inspired a range of debate on what is an appropriate distribution to model the one-sided disturbance term. Principal criticisms that have been considered focus on the actual dispersion of costs within the 'raw' data and the distributional assumptions imposed on the error term. A potential solution to ease this problem is the use of long term distribution free efficiency measures.

Distribution-free cost efficiency measures have two main advantages over alternative measures. These measures overcome the distributional problems previously outlined and quantify long-term efficiency. The measurement of efficiency over a period of time is important in that it reduces the possibility of reporting an isolated or unrepresentative set of estimates from an individual year or couple of years. The use of panel data therefore enables cost efficiency to be considered over time providing a long-run average efficiency measure for the relevant time period. Berger and Mester (1997), considering the example of US banks, suggest a panel data approach is preferable to cross-sectional studies. They suggest this is the case as cost efficiency should, “ …estimate how well a banks tends to do relative to its competitors over a range of conditions over time, rather than a firm’s relative efficiency at any one point in time” (pp. 919-920).

Efficiency is derived directly from the individual effects produced by the fixed effects model, where the individual effects, u_i, would include the "… unobservable entrepreneurial or managerial skills of the firm's executives" (Baltagi, 1995, p.9). This is a development of the approach initially forwarded by Berger (1993) who employed the average residuals from cross-section regressions for a ten-year period to provide estimates of relative and distribution free efficiency. The approach assumes efficiency is constant over time and bias in efficiency may be removed through averaging over time. The individual effects (u_i) may be employed as an indicator of non-negative cost efficiency. Thus distribution free efficiency may written:

$$Efficiency_i = exp\,(min[Ln\ u_i]-Ln\ u_i) = \mathrm{Min}[u_i]/\ u_i \qquad (6)$$

for i building societies, where $min(u_i)$ *is* assumed to be the most efficient depository institution in the sample. Efficiency is bounded by [0, 1] with 1 indicating 100 per cent efficiency. For further discussion of this measure see Allen and Rai (1996, 1997) and DeYoung (1997).

5. *Estimation of the cost models*

To accommodate the estimation of both long term economic characteristics and efficiency a one-way 'fixed effects' panel data model is estimated. Panel data models attempt to amend for differences in time and firm specific change, which are not considered with pooled data analysis. The fixed effects model assumes that the period-varying effects are constants for each firm. This assumption implies that the sample is drawn from a population with finite boundaries with the fixed effects model operating through conditional inference. To estimate a cost function over a data panel (including bank observations both over time and across a cross section of institutions) a one component fixed effects model is used. 'Effects' models aggregate both period invariant and individual invariant variables with individual time varying variables. The basic linear relationship, may be defined,

$$Y_{it} = u_i + b'X'_{it} + v_{it} \qquad (7)$$

Where Y_{it} are the time and firm estimates dependent on; b' the parameters of the K explanatory variables within the model, X'_{it}, the it^{th} observation of the K explanatory variables and v_{it}, the disturbance term; for all $i = 1,....,$ N; $t = 1, ...,$ T. u_i represents individual specific effects of the banks and is used to the capture non-random disturbance between the banks. v_{it} is employed to capture random error within the model. X'_{it} and v_{it} are assumed to be independent for i and t observations of the K variables within the model. The procedure for estimation is set out in detail within Greene (1993, 1995), Baltagi (1995) and Intriligator *et al* (1996).

A translog functional form is employed. Guilkey *et al* (1983) found the translog functional form to broadly display better global behaviour than other Diewert flexible forms such as the generalised Box Cox and the Generalised Leontief functional forms. A non-decreasing relationship between inputs and outputs, concavity of the cost function and homogeneity of degree one of input prices is assumed. The expansion of the cost function into a Diewert flexible, second-order translog cost model allows the potential benefits of multi-product production to be estimated. The relaxation of the output homogeneity restrictions enables measurement of economies of scale. The cost function model is assumed to be separable by restriction.

The production and intermediation translog models may be written:

$$LnOC = \sum_j a_j LnY_j + \sum_r b_r Ln P_r + 1/2 \sum_j \sum_s c_{js} LnY_j LnY_s + 1/2 \sum_r \sum_q w_{rq} Ln P_r Ln P_q + \sum_j \sum_r d_{jr} LnY_j Ln P_i + gLnLA + hLnPROV + u_i + v \quad (8)$$

for $j, s = 1, \ldots, 3$. r, $q = 1, \ldots, 2$ for the production model

$$LnC = \sum_j a_j LnY_j + \sum_r b_r Ln P_r + 1/2 \sum_j \sum_s c_{js} LnY_j LnY_s + 1/2 \sum_r \sum_q w_{rq} Ln P_r Ln P_q + \sum_j \sum_r d_{jr} LnY_j Ln P_i + gLnLA + hLnPROV + u_i + v \quad (9)$$

and for $j, s = 1, \ldots, 3$. r, $q = 1, \ldots, 3$ for the intermediation model.

Following established cost and production theory restrictions are imposed to ensure symmetry ($c_{js} = c_{sj}$ and $w_{rq} = w_{qr}$). Linear homogeneity in input prices of degree one, where linear homogeneity suggests if all input prices are doubled then costs exactly double, requires:

$$\sum_{r} b_r = 1 , \quad \sum_{rq} w_{rq} = 0 , \quad 1 < r <, n, \quad \sum_{jr} d_{jr} = 0 , \quad 1 < j < m \qquad (10)$$

Where *Ln* denotes logarithm, *LA* indicates liquid assets, *PROV* signifies provisions against bad and doubtful debts, *OC* indicates operational costs, *C* represents total costs, *Y* represents outputs, *P* denotes input prices and *a, b, d, c, g, h* and *w* are parameters to be estimated. The error term represented here as $u_i + v$ includes both random and firm specific effects, incorporating both the intercept and the error term.

6. Data

The sample has been constructed with data from the Annual Reports and Accounts of 12 UK retail banks from 1985 to 1997. All of the 12 banks are recorded, over the period, in the Annual Abstract of Banking Statistics produced by the British Bankers Association. The data are deflated using the Retail Price Index to 1985 prices. The banks included in the sample are the Royal Bank of Scotland, Standard Chartered, TSB, Barclays, Clydesdale, The Co-operative Bank, Lloyds, Midland, Natwest, Bank of Scotland, Abbey National and Yorkshire. Descriptive statistics are presented in Table 1.

Table 1 Descriptive Statistics

Year		Operating Cost (£m)	Loans (£m)	Deposits (£m)	Labour price (£)	Capital price (£)	Deposit price (£)
1985-1989	Mean	1403.7	32417.3	36988.0	33451.9	271.9	79.7
	Std. Dev.	***1023.7***	***24036.4***	***25430.0***	***50232.1***	***253.9***	***13.6***
1990-1993	Mean	1074.0	25976.5	29670.1	30915.5	470.3	85.8
	Std. Dev.	***1113.1***	***24537.3***	***27567.6***	***58695.3***	***507.9***	***22.0***
1994-1997	Mean	1071.8	30271.6	31971.4	39813.8	499.8	50.3
	Std. Dev.	***1077.5***	***26375.1***	***27313.5***	***83362.5***	***545.5***	***12.1***
Overall	Mean	1145.9	28728.9	31965.0	34381.1	430.0	71.6
	Std. Dev.	***1056.8***	***24535.0***	***26399.0***	***64403.4***	***476.4***	***28.1***

Over the sample period, many substantial changes have occurred in the UK retail-banking sector. The levels of operating cost have consistently fallen throughout the sample period. The levels of deposits and loans (by value) initially fell from a high in the 1985-1989 period to a low point in 1990-1993. This low period was marked by economic depression and increased competition within the retail-banking sector, both between retail banks and new entrants such as the building societies. Recovery from this position is observed in the 1994-1997 period. Input prices also vary substantially over the sample period. The price of labour, initially falling in the 1990-1993 period rose to a high in the 1994-1997 period. The price of capital has risen consistently through the sample period, perhaps providing an indication of the levels of investment undertaken in this sector. Deposit prices have consistently fallen through the period 1985-1997. This decline in deposit price may be regarded as a reflection of the long-term decline of interest rates over the sample period and the consequent reduction in interest costs. Further discussion of the performance of the UK retail-banking sector between 1970 and 1990 is provided by Colwell (1991).

7. Results

The parameter estimates and diagnostic statistics are presented in Table 2. The estimates of economies of scale are contained in Table 3. The estimates of economies of scope are displayed in Table 4. Estimates of cost efficiency are presented in Table 5. Results for economies of scale and scope are presented overall, over distinct time periods and across a range of total asset sizes. Six asset size groups are defined: Group 1 includes banks with less than £5,000m in total assets, group 2 includes banks with between £5,000 and £25,000m in total assets and group 3 includes banks with between £25,000m and £50,000m in total assets. Banks with between £50,000m and £75,000m in total assets are contained in group 4, group 5 includes banks with between £75,000m and £100,000m in total assets and group six includes banks with excess of £100,000m in total assets.

Acceptable levels of diagnostic statistics are recorded for both the 'production' and 'intermediation' model forms. The relatively low level of significant T statistics for the ‘intermediation’ model suggests that this model may appreciate higher levels of approximation error than the ‘production’ model. Positive estimates of the partial derivative of cost with respect to prices are recorded for all asset sizes and time periods for the 'intermediation' model and for all time periods and most asset sizes with the 'production' model. This indicates that the ‘production’ model may suffer a higher level of specification error than the 'intermediation' model. These results are presented in Appendix 1.

Table 2 Parameters of the fixed effects models

	Intermediation		Production			Intermediation		Production	
	Coeff.	Standard Error	Coeff.	Standard Error		Coeff.	Standard Error	Coeff.	Standard Error
β_1	1.025	(0.5080*	3.458	(0.405)*	ω_{33}	0.208	(0.083)*	-	-
β_2	0.144	(0.361)	-2.458	(0.405)*	ω_{12}	0.000	(0.036)	-0.262	(0.031)*
β_3	-0.169	(0.443)	-	-	ω_{13}	-0.135	(0.062)*	-	-
α_1	-0.287	(0.593)	5.892	(1.684)*	ω_{23}	-0.078	(0.029)*	-	-
α_2	-0.056	(0.240)	-5.862	(1.697)*	δ_{11}	-0.053	(0.056)	-0.032	(0.101)
α_3	-	-	0.700	(0.207)*	δ_{12}	-0.014	(0.037)	-0.501	(0.173)*
χ_{11}	0.077	(0.053)	-0.591	(0.138)*	δ_{13}	0.033	(0.030)	-	-
χ_{22}	-	-	0.664	(0.143)*	δ_{21}	0.023	(0.064)	0.214	(0.044)*
χ_{33}	-0.013	(0.009)	0.035	(0.009)*	δ_{22}	0.000	(0.025)	0.076	(0.021)*
χ_{12}	-	-	0.034	(0.062)	δ_{23}	0.012	(0.028)	-	-
χ_{13}	0.010	(0.018)	0.340	(0.095)*	δ_{31}	-	-	-0.293	(0.104)*
χ_{23}	-	-	-0.482	(0.100)*	δ_{32}	-	-	0.538	(0.166)*
ω_{11}	-0.031	(0.064)	-0.096	(0.033)*	γ	-0.104	(0.025)*	-0.117	(0.028)*
ω_{22}	0.037	(0.041)	0.358	(0.040)*	η	0.023	(0.014)	0.023	(0.015)

*T ratios in brackets * = significant at 10%*

Diagnostic Statistics for the Intermediation model

F statistic for model [30, 102] = 462.00 prob. = 0.0000

adj. R^2 = 0.9906

F statistic for the restrictions = [3, 99] 8.544 prob = 0.0000

Diagnostic Statistics for the Production model

F statistic for model [29, 103] = 315.80 prob. 0.0000

adj. R^2 = 0.9858

F statistic for the restrictions [4, 99]= 17.917 prob = 0.0000

Table 3 Economies of scale

	Intermediation		production	
	Coefficient	Standard Error	Coefficient	Standard Error
Overall	1.195	(0.538)	1.632	(0.644)
1985-1989	1.232	(0.549)	1.686	(0.647)
1990-1993	1.184	(0.527)	1.653	(0.643)
1994-1997	1.175	(0.539)	1.567	(0.644)
Total asset size in £m				
Group 1	1.069	(0.280)	1.126	(0.124)
Group 2	1.329	(0.357)	1.048	(0.105)
Group 3	1.426	(0.386)	1.345	(0.142)
Group 4	1.471	(0.398)	0.930	(0.103)*
Group 5	1.549	(0.416)	0.798	(0.092)*
Group 6	1.549	(0.421)	0.912	(0.098)

*Standard errors in brackets * = significant at 10%*

Overall, slight diseconomies of economies of scale are recorded with the ‘intermediation’ model and substantial diseconomies of scale are estimated with the ‘production’ model. Overall, a value of 1.195 is recorded for the ‘intermediation’ model and a value of 1.632 is recorded for the ‘production’ model. Over asset size distinct estimates are recorded for the two model forms. Constant returns to scale are recorded with the intermediation model for banks between £0 and £5000m in total assets (a value of 1.069 is recorded), with diseconomies of scale rising with increases in total asset size there after. The ‘production’ model produces quite peculiar results over asset size. Banks with between £0 and £25,000m in total assets broadly display constant returns to scale. Diseconomies of scale are reported for banks in the £25,000m - £50,000 total asset group, a finding not disassociated from expectation. Banks with greater than £50,000m in total assets appear to appreciate economies of scale. This rather peculiar result may be produced by a number of reasons. Whilst never discounting potential specification or approximation error, such a finding could indicate an ‘S’ shaped cost function. More plausibly, the results could also be a product of the presence of distinct productive technologies existing in larger and smaller banks.

Economies of scope estimates are recorded for both model forms, overall, over time and across a range of asset sizes. The ‘intermediation’ model with two outputs considers the cost advantage of producing loans and investment jointly as opposed to separately. The ‘production’ model with three outputs (loans, investments and deposits) allows three measures, loans specific, investment specific and deposit specific scope economies to be estimated. For example, loan specific scope economies quantify the cost advantage of producing loans, deposits and investments jointly as opposed to producing loans separately from deposits and investments. Similarly, deposit specific scope economies would indicate the cost advantage of producing deposits jointly with loans and investment relative to producing deposits separately from loans and investments.

Overall, overtime and across asset sizes substantial and statistically significant diseconomies of scope are recorded with the ‘intermediation’ model. A value of –0.5321 for economies of scope was recorded, overall. This indicates banks that produce loans and investments separately could gain cost economies. The magnitude of the diseconomies of scope appears to slightly decline as total asset size increases. A value of

–0.7076 for the banks in the £0 - £5,000m total asset group is recorded, falling to -0.4310 for banks within the greater than £100,000m total asset group.

Table 4 Economies of scope

Intermediation	Estimate	T ratio	Asset group	Estimate	T ratio
Overall	-05321	(0.1001)*	Group 2	-0.5482	(0.1006)*
1985-1989	-0.5376	(0.0958)*	Group 3	-0.4909	(0.1167)*
1990-1993	-0.5454	(0.0960)*	Group 4	-0.4478	(0.1236)*
1994-1997	-0.5163	(0.1098)*	Group 5	-0.4201	(0.1217)*
Group 1	-0.7076	(0.0748)*	Group 6	-0.4310	(0.1185)*

Production	Estimate	T ratio	Estimate	T ratio	Estimate	T ratio
	Loan specific		Investment specific		Deposit specific	
Overall	1.2497	(0.6390)*	-0.0042	(0.0020)*	0.0044	(0.0591)
1985-1989	0.6369	(0.3895)	-0.0042	(0.0020)*	-0.0430	(0.0579)
1990-1993	0.6513	(0.4078)	-0.0044	(0.0021)*	-0.0451	(0.061)
1994-1997	0.5970	(0.0374)	0.0042	(0.0020)*	-0.0430	(0.058)
Asset group						
Group 1	0.9800	(0.6794)	-0.0067	(0.0039)	-0.0695	(0.0927)
Group 2	0.6321	(0.4020)	-0.0044	(0.0021)	-0.0455	(0.0611)
Group 3	0.6745	(0.4039)	-00042	(0.0019)*	-0.0415	(0.0559)
Group 4	0.6193	(0.3871)	-0.0039	(0.0019)*	-0.0403	(0.0543)
Group 5	0.3909	(0.2695)	-0.00326	(0.0014)*	-1.0658	(0.0706)*
Group 6	0.5285	(0.3121)*	-0.0033	(0.0014)*	-1.2001	(0.0917)*

*Standard errors in brackets * = significant at 10%*

A variety of results are produced with the ‘production’ model. Substantial loan specific economies of scope are recorded both over time and asset size. Overall the level of loan specific economies of scope is 1.2497. The size of loan specific economies of scope is seen to fall both over time and with increasing total asset size. A value of 0.6369 is reported for the 1985-1989 period rising to a value of 0.5970 in the 1994-1997 period. A value of 0.98 is reported for banks in the £0 - £5,000m asset group declining to 0.5285 for banks in the greater than £100,00m asset group. Reported investment specific economies of scope estimates are insubstantial. Overall investment specific economies of scope are –0.0042. Estimates of deposit specific economies of scope are slight both overall and over time. Overall levels of deposit specific economies of scope are 0.0044. Substantial deposit specific diseconomies of scope appear as the total asset size of banks rises above £75,000m increasing from –0.0695 for banks in the £0 - £5,000m total asset group to –1.2001 for banks in the over £100,000m total asset group.

Table 5 Cost efficiency

	Production Efficiency	Intermediation Efficiency	Production Average Costs*	Intermediation Average Costs*
Barclays	0.741	0.715	0.079	0.024
Co-operative	1.000	1.000	0.090	0.038
Clydesdale	0.939	0.988	0.108	0.039
Lloyds	0.756	0.745	0.134	0.067
Midland	0.760	0.736	0.088	0.026
Natwest	0.710	0.732	0.089	0.022
Royal Bank of Scotland	0.804	0.806	0.079	0.019
Yorkshire	0.847	0.959	0.086	0.030
Bank of Scotland	0.917	0.817	0.087	0.021
TSB	0.913	0.757	0.057	0.023
Abbey National	0.876	0.787	0.048	0.006

* = Average costs defined as the average total cost divided total assets

Firm specific distribution free cost efficiency estimates are made for both ‘production’ and ‘intermediation’ specifications of bank production. An average efficiency of 0.845 is recorded for the ‘production’ model and an average of 0.822 is provided for the ‘intermediation’ model. This indicates that present levels of output could be provided at a cost requiring 15.504 and 17.769 per cent less inputs for the ‘production’ and ‘intermediation’ models respectively. A Pearsons correlation coefficient of 0.739 and a Spearman Rank Correlation coefficient of 0.818 are estimated for the two sets of efficiency results, indicating a substantial degree of association in terms of efficiency between the two model forms. Levels of dispersion in efficiency are similar between the model forms with a standard deviation of 0.907 and 0.103 estimated for the ‘production’ and ‘intermediation’ models respectively. Such variation in efficiency does not to correlate with what would be expected with reference to the dispersion of 'raw' costs, here represented by average costs. The Pearsons correlation coefficients measuring the association between average costs and efficiency are 0.167 and -0.057 for the ‘intermediation’ and ‘production’ models respectively. The smaller retail banks appear to relatively more efficient than their larger counterparts. This provides further support for the suggestion that substantial diseconomies of scale exist for British retail banks with greater £5,000m in total assets.

8. Conclusions

In this paper cost efficiency, economies of scale and economies of scope in the British retail-banking sector are estimated. This analysis differs from previous studies by considering the retail-banking sector in isolation from other institutions and over time. Two models of bank production are estimated, the ‘production’ and the ‘intermediation’ approaches. A panel data fixed effects model is employed to consider cost efficiency over a number of years removing any potential bias from dis-equilibrium effects that may have influenced estimation made by previous cross sectional studies. A translog specification of productive technology is employed. Cost efficiency is estimated employing ‘distribution free’ methods removing potential bias that may occur when distributional assumptions on the distribution of efficiency are imposed.

Overall diseconomies of scale for banks with excess of £5,000m in total assets are indicated for both models. This finding is distinct from the evidence presented for Italian, Spanish, German and French banks by Altunbas and Molyneux (1996). This may be a ramification of far fewer retail banks operating in the UK market and the sample period considered. The implication of this finding is that little empirical support is provided for the continuing process of external growth in this sector. This finding implies that the continuing growth of retail banks has been driven by factors other than cost and efficiency considerations. Other potential factors are numerous, although reference to the merger movement amongst US commercial banks may provide some useful insights. Benston et al (1986) for example suggested political reasons could be influential, as banks desired to a reach a size where regulators viewed them as ‘too large to fail’, in terms of the possible economic and political ramifications of such a failure.

A variety of economies of scope estimates have been made. These broadly indicate that diseconomies of scope exist in the joint production of loans and investments, although the potential cost advantage to be gained from separate production may be trivial. The presence of statistically significant loan specific economies of scope with the production model provides empirical support for the joint provision of loans with investments and deposits.

A high level firm specific cost efficiency dispersion exists indicating substantial cost efficiency improvements may be made within this industry, particularly by the larger retail banks. A substantial distribution of efficiency is recorded. This result is distinct to that reported by Altunbas *et al* (1995). This may be attributed to the different econometric techniques employed.

It may be stated that the estimation of economies of scale may be affected by the model form and techniques applied. Whilst this conclusion indicates the fallibility of econometric cost analysis, it may also enable a broader view of the problem to be considered. The use of dissimilar definitions of cost is a primary difference between the two models. The ‘production’ approach does not recognise the cost of borrowed funds, whilst with the ‘intermediation’ approach the interest costs may come to dominate other sources of cost, potentially distorting results. The presence of substantially lower diseconomies of scale with the intermediation approach indicates that the inclusion of retail and non-retail funds as inputs significantly alters the production problem considered. The larger banks may gain access to funds at a cost advantage due to their size and are able to substitute such funds for labour and capital within the production of loans. Additionally the 'intermediation' approach appears to display a higher degree of approximation error and the 'production' model a higher degree of specification error. Such a finding suggests that the production approach whilst ‘fitting’ the data comfortably, does not concur fully to what would be expected in accordance to theory. This in turn provides support for the modelling of bank production employing the intermediation model appears to be preferable in this example.

In conclusion, this analysis of British retail banks over the sample period 1985-1997 indicates a variety of results distinct from the findings made in previous studies of this sector. Diseconomies of scale appear to be present in British retail banks. A high level of efficiency dispersion is observed. Smaller retail banks appear to appreciate the highest levels of relative cost efficiency further supporting the estimates of diseconomies of scale for most British retail banks.

References

Allen, L. and A. Rai, (1996) "Operational efficiency in banking: An international comparison" *Journal of Banking and Finance* vol.20 655-672.

Allen, L. and A. Rai, (1997) "Operational efficiency in banking: An international comparison. Reply to the comment" *Journal of Banking and Finance*, vol.21, 1451-1455.

Altunbas, Y., D. Maude and P. Molyneux (1995) Efficiency and mergers in the UK (retail) banking market” Institute of European Finance Working Paper, 1-40.

Altunbas, Y. and P. Molyneux, (1996) "Economies of scale and scope in European banking" *Applied Financial Economics* vol.6 367-375.

Ashton, J. K. and P. Hardwick, (1997) “Cost efficiency in UK building societies: A panel data study, 1990-1995” presented at the *4th Annual Conference of the Multinational Finance Society*, June 25-29, 1997, Thessaloniki, Greece.

Baltagi, B. H. (1995) *Econometric Analysis of Panel Data* Chichester: John Wiley & Sons.

Benston, G. J., R. A. Eisenbeis, P. M. Horvitz, E. J. Kane, and G. G Kaufman, (1986) *Perspectives on Safe and Sound Banking. Past, Present, and Future* American Bankers Association Washington: MIT Press.

Berger, A. N. (1993) “Distribution free estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions” *The Journal of Productivity Analysis* vol.4 261-292.

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

Baumol, W., J.C. Panzar, and R.D. Willig, (1982) *Contestable Markets and Industry Structure* New York: Harcourt Brace Jonanovich Inc.

Colwell, R. (1991) “The performance of major British banks, 1970-90” *Bank of England Quarterly,* November, 508-515.

DeYoung, R. (1997) “Comment on ‘Operational efficiency in banking: An international perspective’” *Journal of Banking and Finance* vol.21 1325-1329.

Drake, L. and B. Howcroft, (1994) "Relative efficiency in the branch network of a UK bank: An empirical approach" *International Journal of Management Science*, vol.22, no.1, 83-90.

Greene, W. H. (1993) *Econometric Analysis* 2nd edition Prentice Hall Englewood Cliffs New Jersey.

Greene, W. H. (1995) *LIMDEP Version 7.0 User’s Manual* Econometric software Inc New York.

Guilkey, D. K., C. A. Knox Lovell, and R. Sickles, (1983) "A comparison of the performance of three flexible functional forms" *International Economic Review* vol.24 591-616.

Intrigator, M. D., R. G. Bodkin, and C. Hsiao, (1996) *Econometric Models, Techniques, and Applications* Second Edition Prentice Hall NewJersey.

Molyneux, P., Y. Altunbas, and E. Gardener, (1996) *Efficiency in European Banking* Chichester: John Wiley and Sons.

Appendix 1 Partial derivatives of cost with respect to input prices

	intermediation Coeff.	intermediation S.E.	Production Coeff.	Production S.E.	intermediation Coeff.	intermediation S.E.	intermediation Coeff.	intermediation S.E.	production Coeff.	production S.E.
	Partial derivative of cost with respect to labour price				Partial derivative of cost with respect to deposit price		Partial derivative of cost with respect to capital price			
Overall	0.099	(0.052)*	0.105	(0.043)*	1.321	(0.320)*	0.120	(0.243)	2.351	(0.245)*
1985-1989	0.033	(0.052)	0.027	(0.049)	1.489	(0.382)*	0.082	(0.236)	2.300	(0.242)*
1990-1993	0.088	(0.048)*	0.120	(0.045)	1.381	(0.345)*	0.113	(0.244)	2.382	(0.248)*
1994-1997	0.162	(0.065)*	0.156	(0.041)*	1.121	(0.244)*	0.158	(0.248)	2.358	(0.246)*
0.-5,000	0.144	(0.095)	0.229	(0.073)*	1.337	(0.353)*	0.175	(0.295)	2.654	(0.275)*
5,000-25,000	0.053	(0.088)	0.016	(0.059)	1.260	(0.351)*	0.080	(0.244)	2.223	(0.235)*
25,000-50,000	0.063	(0.054)*	-0.004	(0.043)	1.274	(0.352)*	0.077	(0.275)	2.456	(0.261)*
50,000-75,000	0.135	(0.060)*	0.204	(0.051)	1.251	(0.325)*	0.080	(0.271)	2.543	(0.263)*
75,000-100,000	0.086	(0.050)*	0.252	(0.044)*	1.385	(0.369)*	0.020	(0.227)	2.116	(0.220)*
>100,000	0.060	(0.084)	-0.081	(0.059)	1.269	(0.324)*	0.033	(0.253)	2.219	(0.238)*

*Standard errors in brackets * = significant at 10%*