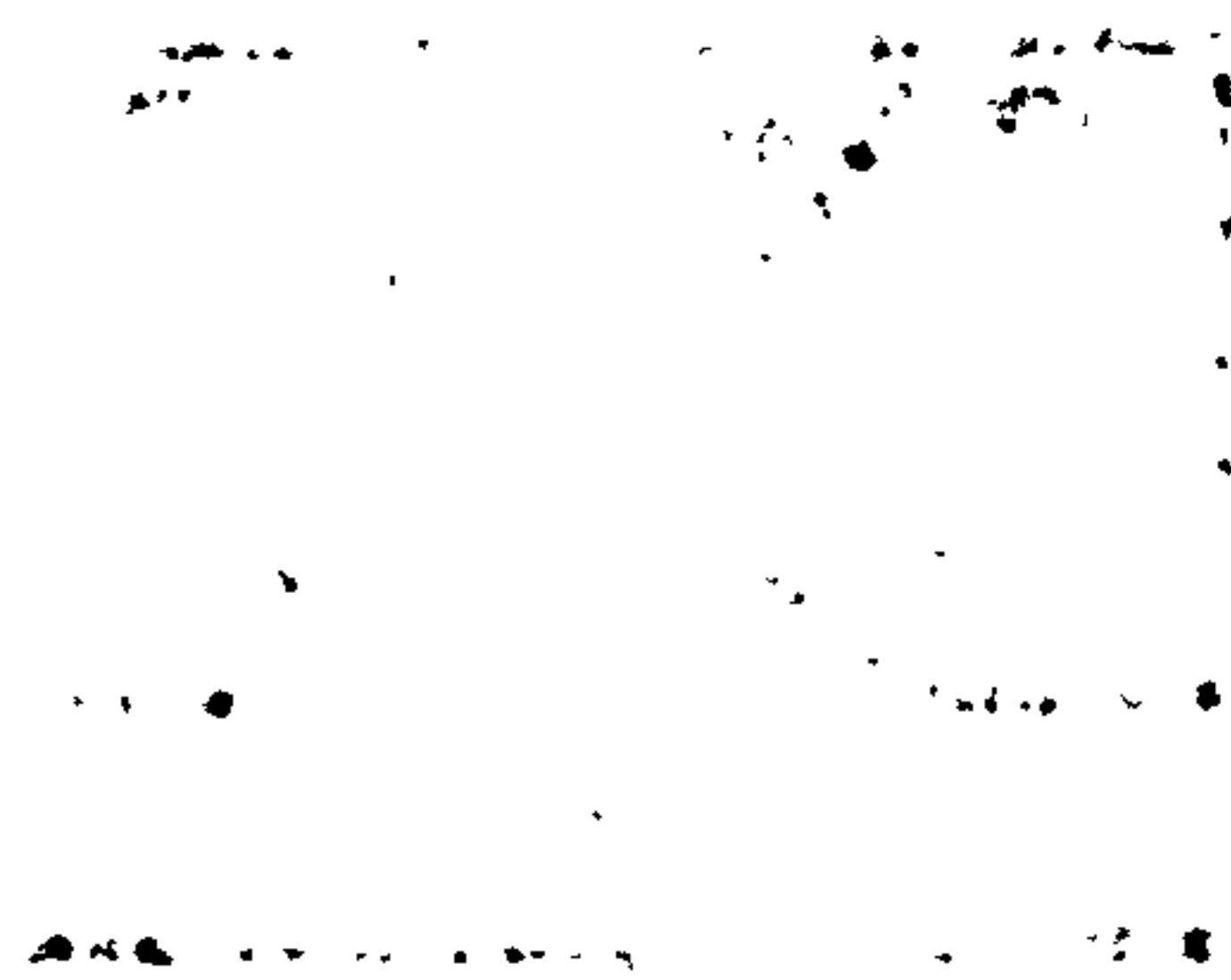


This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information from it may be published without the author's consent.

**Productivity, efficiency and competition of UK depository
institutions**

John Kevin Ashton



A thesis submitted in partial fulfilment of the requirements of Bournemouth
University for the degree of Doctor of Philosophy

June 1999

Bournemouth University

John Kevin Ashton

Productivity, efficiency and competition of UK depository institutions

In the thesis, a number of cost characteristics of the UK building society and retail banking sectors are estimated using econometric techniques. The cost characteristics considered broadly fall into the three areas of productivity, efficiency and competitiveness. The study was undertaken with the aim of considering both the magnitude and degree of change in these cost characteristics over time. This assessment is deemed to be important due to the wide-ranging changes in the regulatory, institutional and market environment of both these sectors.

For reasons of structure and clarity, the thesis is divided into two parts. The first part of the thesis provides a broad discussion of the operating environment, the model specification, variable definition, the concept of efficiency, econometric techniques, previous literature and the form of the statistics used to measure the relevant economic characteristics. The second part of the thesis contains the empirical studies in which economic characteristics are estimated. Cost efficiency is measured for both the retail banking and building society sectors using differing model forms and distinct functional forms. Fixed effects panel data techniques are employed in both studies. Both model specification and functional forms are deemed to influence the estimates produced. It is discovered that both sectors experience a degree of cost efficiency dispersion. Measures of economies of scale and product mix are also estimated with positive economies of scale and constant returns to scale found in the retail bank and building society sectors, respectively. The findings from the analysis of economies of product mix are less than clear. Nevertheless, it may be stated that the re-regulation which allowed a greater degree of product diversification in the building society sector appears to be justified in terms of cost efficiency. Very low levels of technical change and total factor productivity growth are found for both sectors. The degree of competitiveness of the building society sector was assessed using a revenue function approach. The results suggest that both retail banks and building societies operate in markets characterised by a moderate degree of monopolistic competition.

Contents

List of Tables

List of Equations

List of Figures

Acknowledgements

Chapter 1	<i>Introduction, thesis aims and framework</i>	1
Section One		
Chapter 2	<i>The operating environment of UK depository institutions</i>	11
Chapter 3	<i>Model and variable definition</i>	34
Chapter 4	<i>Previous studies of efficiency and productivity in depository institutions</i>	59
Chapter 5	<i>Measuring efficiency</i>	101
Section Two		
Chapter 6	<i>A general test of competitive conditions in UK building society mortgage market: 1990-95</i>	130
Chapter 7	<i>Economies of scale, economies of product mix and cost efficiency in the UK retail banking industry</i>	152
Chapter 8	<i>Total factor productivity growth and the characteristics of production technology in the UK retail banking sector.</i>	170
Chapter 9	<i>The cost efficiency of UK building societies</i>	186
Chapter 10	<i>Economies of scope, economies of scale, technical change and total factor productivity growth in UK building societies.</i>	208
Chapter 11	<i>Conclusions</i>	224
	<i>Bibliography</i>	249

Chapter 1	<i>Introduction, thesis aims and framework</i>	
1.1	<i>Introduction</i>	1
1.2	<i>Thesis aims</i>	2
1.3	<i>Thesis outline</i>	6
Chapter 2	<i>The operating environment of UK depository institutions</i>	
2.1	<i>Introduction</i>	11
2.2	<i>The macro-economic environment</i>	11
2.3	<i>Regulatory and supervisory changes</i>	13
2.4	<i>Changes in market structure</i>	22
2.5	<i>The growth of new products</i>	26
2.6	<i>Changes in the technology of service provision</i>	28
2.7	<i>Changing labour markets</i>	31
2.8	<i>Conclusions</i>	33
Chapter 3	<i>Model and variable definition</i>	
3.1	<i>Introduction</i>	34
3.2	<i>Production in depository institutions</i>	35
3.3	<i>The intermediation or asset approach</i>	37
3.4	<i>The production, value added or portfolio approach</i>	38
3.5	<i>National accounting measures</i>	40
3.6	<i>Delegated monitoring</i>	41
3.7	<i>Other models</i>	43
3.8	<i>Variable definition: deposit pricing</i>	43

3.9	<i>capital price</i>	46
3.10	<i>labour price</i>	49
3.11	<i>Output measures</i>	50
3.12	<i>Problems with accounting data</i>	54
3.13	<i>Conclusions</i>	55
Chapter 4	<i>Previous studies of efficiency and productivity in depository institutions</i>	
4.1	<i>Introduction</i>	59
4.2	<i>Building society studies</i>	59
4.3	<i>Non-US bank studies</i>	66
4.4	<i>Studies of savings and loan associations</i>	72
4.5	<i>Credit union studies</i>	77
4.6	<i>US banking studies</i>	82
4.7	<i>Conclusions</i>	99
Chapter 5	<i>Measuring efficiency</i>	
5.1	<i>Introduction</i>	101
5.2	<i>The efficiency concept</i>	102
5.3	<i>Estimation methods</i>	104
5.4	<i>The choice of functional form</i>	106
5.5	<i>Econometric measurement of economies of scale and scope</i>	109
5.6	<i>Econometric measurement of efficiency</i>	116
5.7	<i>What is productivity?</i>	120

5.8	<i>Productivity models</i>	122
5.8	<i>Conclusions</i>	128
Chapter 6	<i>A general test of competitive conditions in UK building society mortgage market: 1990-95</i>	
6.1	<i>Introduction</i>	130
6.2	<i>Contestable market theory and monopolistic competition</i>	133
6.3	<i>Previous studies</i>	135
6.4	<i>The Rosse-Panzar statistic</i>	139
6.5	<i>Data</i>	142
6.6	<i>Model specification</i>	144
6.7	<i>Results</i>	146
6.8	<i>Conclusions</i>	150
Chapter 7	<i>Economies of scale, scope and cost efficiency in the UK retail banking sector</i>	
7.1	<i>Introduction</i>	152
7.2	<i>Model specification</i>	154
7.3	<i>Estimation of the cost models</i>	155
7.4	<i>Measurement of cost efficiency and the data</i>	159
7.5	<i>Results</i>	161
7.6	<i>Conclusions</i>	166

Chapter 8	<i>Total factor productivity growth and characteristics of productive technology in the UK retail banking industry</i>	
8.1	<i>Introduction</i>	170
8.2	<i>Relevant literature</i>	171
8.3	<i>Model definition</i>	174
8.4	<i>Data and results</i>	179
8.5	<i>Conclusions</i>	184
Chapter 9	<i>The cost efficiency of UK building societies</i>	
9.1	<i>Introduction</i>	186
9.2	<i>Model specification employed</i>	187
9.3	<i>Long-run distribution-free cost efficiency</i>	189
9.4	<i>Data employed</i>	190
9.5	<i>The choice of functional form and results from the cost efficiency model</i>	193
9.6	<i>Preliminary results</i>	197
9.7	<i>The re-estimated efficiency models</i>	200
9.8	<i>Conclusions</i>	205
Chapter 10	<i>Economies of scope, economies of scale, technical change and total factor productivity growth in UK building societies</i>	
10.1	<i>Introduction</i>	208
10.2	<i>Model specification</i>	209
10.3	<i>Approaches to measuring productivity and characteristics of productive technology</i>	210
10.4	<i>Data and results</i>	214

10.5	<i>Conclusions</i>	221
Chapter 11	<i>Conclusions</i>	
11.1	<i>Introduction</i>	224
11.2	<i>The thesis outline</i>	224
11.3	<i>Economies of scale and product mix</i>	228
11.4	<i>A descriptive analysis of efficiency:</i>	230
	(a) <i>The use of different estimation methods and models</i>	231
	(b) <i>The growth of depository institutions over the sample period</i>	231
	(c) <i>A distinct productive technology employed by small institutions</i>	232
	(d) <i>The level of competition</i>	232
11.5	<i>A survey and assessment of the estimates of total factor productivity growth and technical change</i>	234
	(a) <i>The level of inefficiency in the sector</i>	236
	(b) <i>The adoption of new technology and structural change</i>	237
	(c) <i>The presence of low levels of competition</i>	238
	(d) <i>The service industry's 'cost disease'</i>	239
	(e) <i>Measurement problems</i>	240
11.6	<i>Problems with our conception of the 'firm'</i>	242
11.7	<i>Overall conclusions and suggestions for further research</i>	245

Tables

2.1	<i>Concentration in the retail banking and building society sectors</i>	24
6.1	<i>Interpretations of the H statistic</i>	140
6.2	<i>Parameter estimates, diagnostic statistics and results</i>	147
6.3	<i>Tests of market stability</i>	149
7.1	<i>Descriptive statistics</i>	160
7.2	<i>Parameters of the fixed effects model</i>	162
7.3	<i>Partial derivatives of the logarithm of cost with respect to the logarithm of input prices and output quantities</i>	163
7.4	<i>Economies of scale, economies of scope and cost complementarities</i>	164
7.5	<i>Firm specific estimates of cost efficiency</i>	166
8.1	<i>Parameter estimates of the time trend model</i>	180
8.2	<i>Economies of scale and partial derivatives of the logarithm of cost with respect to the logarithm of output quantities and input prices</i>	181
8.3	<i>Total factor productivity and its components</i>	182
9.1	<i>Descriptive statistics: Overall and over asset size</i>	191
9.2	<i>Descriptive statistics: Over time and over asset size</i>	192
9.3	<i>Scale factors and scaled variables for the flexible Fourier functional form</i>	196
9.4	<i>The efficiency model: parameter estimates, partial derivatives and diagnostic statistics. Overall</i>	198
9.5	<i>Distribution free cost efficiency estimates from the cost model. Overall</i>	199
9.6	<i>The efficiency model: parameter estimates, partial derivatives and diagnostic statistics. Large</i>	201

9.7	<i>The efficiency model: parameter estimates, partial derivatives and diagnostic statistics. Small</i>	202
9.8	<i>Distribution free cost efficiency estimates from the cost model. Large</i>	204
9.9	<i>Distribution free cost efficiency estimates from the cost model. Small</i>	205
10.1	<i>Parameter estimates: 1990-93</i>	215
10.2	<i>Parameter estimates: 1994-1996</i>	216
10.3	<i>Diagnostic statistics</i>	217
10.4	<i>Partial derivatives of the logarithm of cost with respect to the logarithm of input prices</i>	217
10.5	<i>Economies of scope and cost complementarities</i>	218
10.6	<i>Economies of scale</i>	219
10.7	<i>Total factor productivity growth, output growth and technical change (percentage change)</i>	219

Equations

3.1	A production correspondence for the intermediation model	37
3.2	A cost function for the intermediation model	37
3.3	A production correspondence for the production model	39
3.4	A cost function for the production model	39
3.5	A commonly applied price of capital	47
5.1	A cost function	110
5.2	Ray economies of scale (<i>OES</i>)	111
5.3	Expansion path economies of scale (<i>EPSCE</i>)	111
5.4	Augmented overall economies of scale (<i>AOES</i>)	112
5.5	Cost attributable to the j^{th} input	112
5.6	Input –specific overall economies of scale (<i>OES_j</i>)	112
5.7	Economies of scope (<i>ESC</i>)	114
5.8	Condition for economies of diversification	115
5.9	The disturbance term	117
5.10	The error term	117
5.11	Distribution free efficiency	120
5.12	A production relation incorporating time	124
5.13	The translog time-trend cost model	124
5.14	Average technical change from a time-trend model	125
5.15	Bias of technical progress	126
5.16	Total factor productivity growth	126
5.17	Technical change from a cost function shift model	127
6.1	The H statistic	139
6.2	The competitive environment test	145
6.3	The equilibrium test	145
7.1	The basic linear relationship for a panel data model	155
7.2	The one-way error component	156
7.3	The fixed effects linear model	156
7.4	The production model	157
7.5	The intermediation model	158
7.6	Restrictions imposed on the translog cost functions	158
7.7	Distribution-free cost efficiency	159
8.1	A cost function of the intermediation model	175
8.2	A production relation including a time-trend variable	176
8.3	The intermediation time-trend model	176
8.4	The cost share equations	177
8.5	The restrictions imposed on the time trend translog cost model	177
8.6	Average technical change	177
8.7	Bias of technical change	178

8.8	Total factor productivity growth	179
9.1	A cost function of the intermediation model	188
9.2	Distribution free efficiency	189
9.3	The flexible Fourier functional form	195
9.4	The likelihood ratio test	197
10.1	A cost function of the intermediation model	209
10.2	The translog cost function model	210
10.3	The cost share equations	211
10.4	The restrictions imposed on the time trend translog cost model	211
10.5	Technical change	211
10.6	Total factor productivity growth	212
10.7	Overall economies of scale (OES)	213
10.8	Economies of scope (ESC)	213

Figures

<i>1.1</i>	<i>Section One</i>	<i>7</i>
<i>1.2</i>	<i>Section Two</i>	<i>8</i>
<i>2.1</i>	<i>The combination of regulatory powers under the FSA</i>	<i>18</i>
<i>2.2</i>	<i>The number of building societies: 1970-96</i>	<i>23</i>
<i>2.3</i>	<i>Number of branches in UK retail banks</i>	<i>29</i>
<i>2.4</i>	<i>Number of branches in UK building societies</i>	<i>30</i>
<i>2.5</i>	<i>Number of cash dispensers and ATM's in UK retail banks</i>	<i>31</i>
<i>3.1</i>	<i>The relationship between the production and intermediation approaches</i>	<i>40</i>
<i>5.1</i>	<i>Components of total factor productivity growth</i>	<i>121</i>
<i>10.1</i>	<i>Change in output in the UK building society sector: 1990-96</i>	<i>220</i>
<i>10.2</i>	<i>Total factor productivity growth in the UK building society sector: 1990-96</i>	<i>221</i>

Acknowledgements

I would like to thank my thesis supervisors P. Hardwick, D. McKillop and B Howcroft for their continued patience and encouragement. The usual disclaimer applies.

Chapter 1 Introduction, thesis aims and framework

1.1 Introduction

In his seminal paper, Leibenstein (1966) suggested that “ ... at the core of economics is the concept of efficiency, (but) ... empirical evidence has been accumulating which suggests that the problem of allocative efficiency is trivial. Yet it is hard to escape the notion that efficiency in some broad sense is significant” (p. 392). This thesis is an investigation of efficiency in UK depository institutions and attempts to ascertain whether cost efficiency can be considered to be in some way ‘significant’ in this most important sector of the UK economy. Considerable change has occurred in UK depository institutions during the 1980s and 1990s. The ramifications of this change are still less than clear. An investigation of efficiency and the associated concepts of productivity and competition in the UK depository institutions sector is deemed, therefore, to be both timely and potentially very important.

The economic analysis of depository institutions has taken many forms. In macro-economic analysis, depository institutions have been considered as “ ... passive conduits through which monetary policy is effected” (Benston and Smith, 1976, p. 215). Many micro-economic approaches have emphasised specific aspects of bank production and operation. In this thesis, UK depository institutions are examined, employing techniques drawn from industrial and applied economics. Using this framework, assumptions of efficiency are tested and challenged.

This chapter introduces a number of issues that are central to the analysis of UK retail banks and building societies. In the next two sections, the aims and format of the thesis are outlined.

1.2 Thesis aims

The main aim of the thesis is to employ econometric methods and techniques to quantify the performance of UK depository institutions. When performance has been previously examined, an emphasis on cross-sectional analysis has nearly always been present. Such an approach, whilst providing a 'snapshot' of present positions, may ignore potential disequilibrium positions and the dynamic nature of efficiency. The measurement of efficiency using a dynamic framework is introduced and examined in the thesis. A multi-product specification of productive technology is adopted.

It is important to analyse depository institutions because of their central importance in the functioning of a modern economy and the potentially systemic effect of their failure. The elimination of regulatory barriers to expansion and product provision and the continuing removal of geographical restraints within the European Union (EU) are generally thought to have increased the prevailing levels of competition within depository institutions' core markets. The present form of banking institutions is also generally perceived to exhibit certain characteristics such as the "... preponderance of scale economies across a broad range of bank output in the European banking market" (European Economy, 1997, p.12).

Both these preconceptions and the effects of the significant change in the sector are assessed.

Wheelock and Wilson (1995) suggest that the measurement of bank efficiency is essential for two central reasons: the quantification of the relative performance of firms; and the assessment of the effects of regulatory change. The level (or degree) of efficiency of a depository institution provides an indication of its relative 'success' in production terms. This 'success' deserves attention as institutions in competitive markets will be better able to maintain and develop their business when they operate efficiently. In essence, the efficient depository institution will be able to avoid the high costs of failure. Support for the view that inefficient production is associated with bank failure has been provided for the US commercial banking sector in the 1980s (see Berger and Humphrey, 1992a).

The primary providers of deposit-taking services in the UK are the *retail banks* and *building societies*. These financial institutions are considered in the thesis and are viewed as 'special' in terms of their high levels of financial leverage, the significant macro-economic consequences of their collapse or failure and the correspondingly high level of regulation and supervision imposed by governments to direct their operation. These institutions, in common with most financial firms in the UK, are highly regulated. The analysis of the efficiency of production in retail banks should provide a clearer understanding of the ramifications of this regulation in terms of performance, where " ... the motivation for such regulation demands an understanding of the behaviour of these firms" (Hancock, 1991, p.2). Empirical evidence is also important when attempting to

assess developments such as the continued external growth of depository institution and the future regulation of banking.

Traditional banking business involves the making of long-term loans with short-term deposits. This activity has led to specialisation in the areas of credit evaluation, the monitoring of borrowers and the completion of transactions. The process of undertaking these activities may differ substantially between different institutions and confer differential rates of performance. Undertaking these activities may therefore affect the efficiency of production of individual institutions. Other aspects of production related to these activities may include factors such as risk and liquidity, which can be viewed as external or central to the production process. These aspects may be related to efficiency and could provide insight into the sources of efficiency

The UK retail-banking sector is distinguished by its relatively early development and traditionally high levels of concentration. UK retail banks have for a long time competed with the broader financial markets as a source for funds. This situation is linked to the highly developed financial markets in London. An array of activities ranging from core deposit and loan activities to foreign exchange services and insurance provision are performed by retail banks. Building societies' behaviour is restricted by legislation that is separate from that imposed on retail banks. The Building Societies Act 1986 (Section 5) specifies that building societies are established " ... for the purpose of raising, primarily through subscriptions of members, a stock or fund for making them advances secured on land for their residential use". Conversely, the banks do not have a function specified in

law. Building societies principally act as intermediators of savings (term, share and investment deposits) into house purchase loans. As mutual firms, they have a tradition of supplying members (the *de facto* owners of the society) with mortgage loans for house purchase.

Change in UK depository institutions has occurred concurrently with a re-regulation of the liability side of depository institutions' balance sheets. In response, depository institutions have broadened their activities and increased the number of products or services they provide. The external growth of some institutions, through mergers, has also continued. This shift, acting in tandem with the internal growth of institutions, has increased the degree of market concentration and reduced the number of institutions overall. Alteration of the form of supervision of many institutions has occurred at national, European and international levels. This transformation has been heralded under the banners of improved performance, globalisation, competition and de-regulation of financial markets.

An exploratory econometric analysis of the performance of UK depository institutions is the ultimate goal of the thesis. The cost characteristics of UK depository institutions are estimated using different model forms and econometric techniques. This approach is viewed as pertinent in light of the lack of previous work relating to UK retail banks, the potential mis-specification of previous work and the mis-interpretation of results in isolation. The substantial shift in the industrial structure and behaviour of banking sectors observed in the USA and other nations has yet to be clearly quantified for UK depository

institutions. Empirical studies of the UK retail banking industry have previously been fragmented or absent. Indeed, this study is the first of this breadth to consider such fundamental issues within this most prominent and essential area of commercial activity in the UK.

1.3 Thesis outline

The thesis is divided into eleven chapters, which may be broadly divided into two sections. Section One comprises chapters 2 to 5, which outline the principal assumptions and approaches made in the empirical studies. Section Two comprises chapters 6 to 11, which incorporate the empirical studies of efficiency, productivity and competitiveness and present the conclusions of the thesis. The sequential development of the thesis is displayed in Figures 1.1 and 1.2.

Chapter 2 introduces a number of environmental issues affecting the industry, including regulatory change and developments in technology. Chapter 3 considers the definition of the models of production employed in the thesis, in addition to the 'book-keeping' definitions of the variables employed. Model forms suggested in the literature are also reviewed in chapter 3. Chapter 4 provides a comprehensive international review of previous studies of productive characteristics of depository institutions. A critical assessment of the econometric techniques employed for the measurement of efficiency and productivity, used within previous studies is made in chapter 5. Recommendations suggesting the use of the more flexible and widely applied statistics are made.

Figure 1.1 Section One

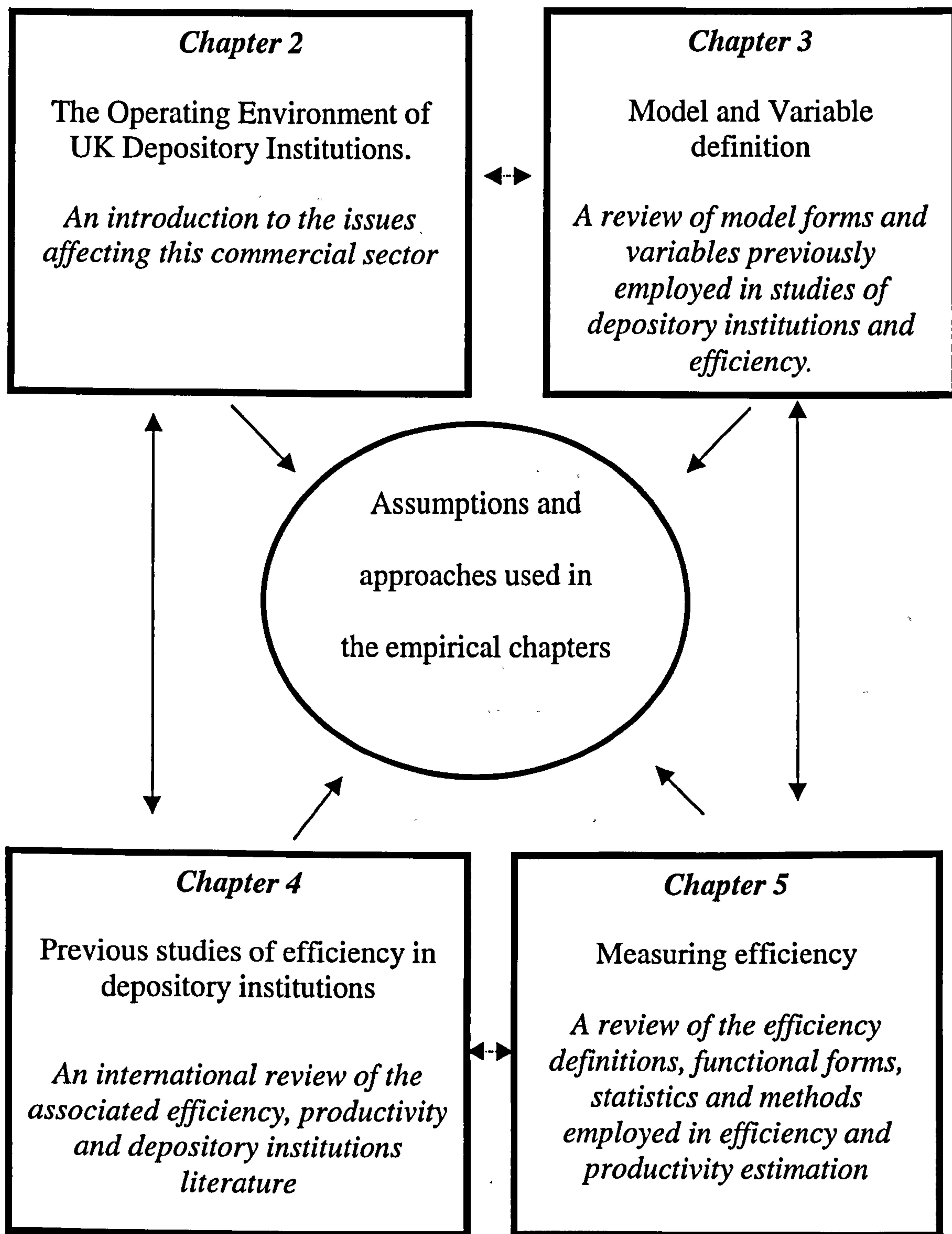
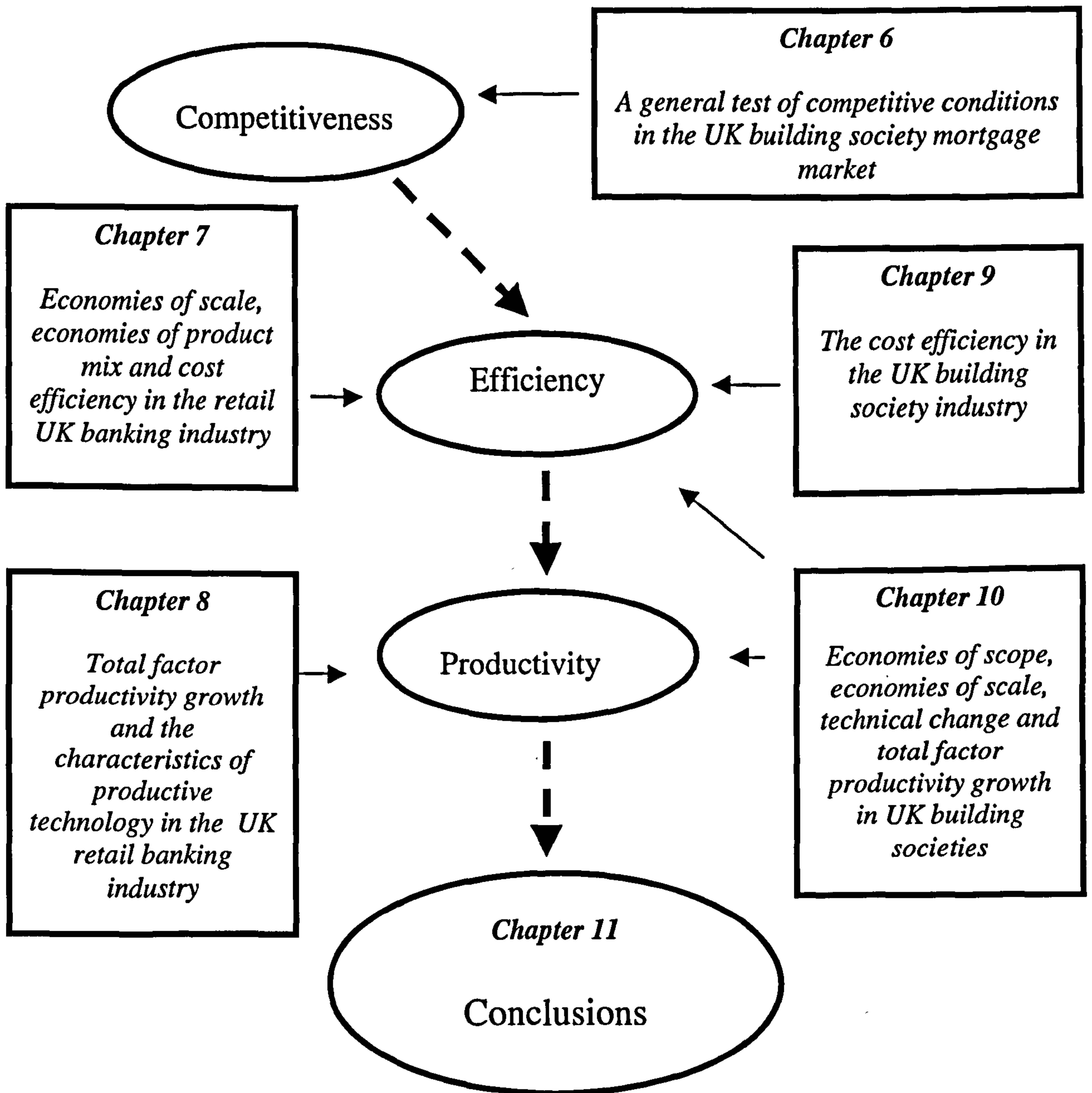


Figure 1.2 Section Two

UK Retail Banks

UK Building Societies



The chapters that constitute Section One are substantially inter-linked. Chapter 4, the review of previous studies, for example, provides a basis from which the model forms and variables, in chapter 3, may be considered. Equally, chapter 3 provides a framework through which the review of previous studies may be considered with greater clarity.

In Section Two the related concepts of efficiency, productivity and competitiveness are assessed for the UK building society and retail banking sectors. All the studies in Section Two draw heavily on the assumptions and approaches recommended in Section One.

The chapters in Section Two include the empirical studies of efficiency, productivity and competitiveness. Within these chapters a number of cost characteristics (discussion of which is developed in chapter 5) are estimated for both the UK retail banking and building society sector. Economies of scale, economies of scope, cost complementarities, technical change, total factor productivity growth and cost efficiency are estimated for both the retail banking and building society sectors. Due to the differing characteristics of the data samples employed and cost characteristics considered, a number of different estimation procedures are used in the studies. Chapter 6 is an analysis of the competitiveness of the mortgage market for building societies. Chapter 7 considers economies of scale, economies of scope and cost efficiency in retail banks between 1984 and 1997. Chapter 8 quantifies technical change and total factor productivity growth in retail banks across the same time period. Estimates of cost efficiency and the influence of differing regulatory regimes in terms of cost efficiency for building societies are estimated and discussed in chapter 9 for the sample period 1990-96. Chapter 10 presents

estimates of economies of scale, economies of product mix, total factor productivity growth and technical change for building societies for the same sample period as used in chapter 9. A summary of the thesis and conclusions are drawn in chapter 11.

SECTION ONE

Chapter 2 The operating environment of UK depository institutions

2.1 Introduction

This chapter outlines the environment within which UK depository institutions operate. The sector to be investigated primarily consists of retail banks and building societies. Many features of the operating environment may have substantial implications for the performance of both these industries. In this chapter, the broader macro-economic environment, central regulatory and supervisory changes, the changing market structure, the growth of new products, the development of the technology of service provision and labour market changes are considered.

2.2 The macro-economic environment

The 'headline' performance of retail banks and building societies in the last few years has been impressive. After extricating themselves from the bad debts incurred during the 1989-93 recession, the retail banks and building societies have continued to move along the path of restructuring. During the sample period of 1984 to 1997, the broader performance of the UK economy may have had a significant effect on both the performance and the behaviour of the retail banking and the building society sectors. Between 1984 and 1988, the UK experienced a strong consumer boom, which

contributed to a significant rise in the price of housing property. This period was characterised by more intense competition between the building societies and the retail banks in the mortgage and savings markets, and by a number of other financial institutions entering the markets for the first time. The sharp decline in property prices, after the peak of the property boom in 1988, did not lead to a large reduction in lending. Instead, competitors, particularly within the mortgage market, loosened their restrictions on both security and income in the provision of mortgage loans, and the retail banks became involved to a greater degree in lending to the commercial sector.

During the first years of the 1990s, the effects of the bursting of the property bubble created significant problems for depository institutions. An unprecedented number of home-owners were removed from their properties. The spectre of 'negative equity', the problem of losses on the realisation of assets when repossessed homes were sold, became a reality at this time. These circumstances, whilst forcing many issues of regulatory and structural developments in financial institutions to the fore, provided a less than sympathetic environment for change. This was a period of substantial losses across the sector. The building societies and banks suffered considerable losses on some activities, such as the investment in estate agency, which resulted in losses of approximately £0.25bn (Barnes and Ward, 1997). The lending decisions of the retail banks have also been increasingly viewed as miscalculated, with lending made for major commercial projects, such as Canary Wharf, particularly criticised by some commentators. Significantly, some building societies suffered such large losses that they were taken over

(such as the Town and Country in 1992) or moved towards merger to replenish their reserves.

As the economic position gradually stabilised towards the end of 1993, the profits of retail banks and building societies improved. These profits, from the increasing interest spread on the institutions' core business, were used to offset many of the previous losses. Barnes and Ward (1997) suggested that " ... building societies were able to do this because their competitors, the retail banks and specialised lenders who had also incurred heavy losses, either pulled out of the market (mostly the specialised lenders) or increased their own margins (the retail banks), thus sacrificing part of their market share to the building societies" (p.21).

2.3 Regulatory and supervisory changes

A change in both the level of structural regulation and the re-regulation of supervision, including the level of investor protection and the conduct of business, has occurred at the national, European and international levels. Banks and buildings societies are in many ways regulated on similar lines. These institutions are regulated within a legislative framework enforced by a central regulator. During the sample period, the banking regulator, enfranchised in the 1987 Banking Act, was the Bank of England¹. The building society regulator, enfranchised by the 1986 Building Societies Act, was the Building Societies Commission. Both regulators are duty bound to enforce the compliance of institutions to existing legislation and their own guidelines, including capital adequacy

requirements and a degree of depositor protection. It has been suggested by Miles (1994) that whilst the structure of regulation appears similar, the temper of such regulation does differ between the sectors. For example, the Bank of England had greater informational requirements for banks. Other regulatory differences in the sample period included the limitations on the proportion of funds raised from non-depositor sources by building societies, the restrictions on building societies relating to the proportion of funds that may be held that are not first mortgages, and the maximum level of unsecured lending. These restrictions have been altered over the sample period.

Domestic re-regulation of depository institutions has enabled the provision of a wider range of services from both banks and building societies. The Building Society Act of 1986 also allowed the demutualisation or conversion of building societies. This and other legislation effectively removed the restrictions on providing a wider range of savings products, such as insurance. The 1986 Building Society Act was introduced to improve the competitive position of building societies in relation to retail banks and other centralised mortgage providers. The market share of building societies in the mortgage market had been consistently eroding throughout the 1980s, as new mortgage providers exploited their relatively privileged regulatory position. The Building Society Act (1986) initiated a number of changes to building societies with the intention of amending this position.

The Building Society Commission was established as a regulator to ensure investor protection, financial stability and to advise the government. This regulatory overhaul was

established with the dual objective of achieving the broader requirements of the European legislation. The scope of products that building societies could supply was increased with limited amounts of advances secured on land and unsecured advances allowed to customers. The provision of banking, insurance and investment products within this restrictive framework was enfranchised in the 1988 amendment to the Building Societies Act. Provision of corporate lending still remains limited, as does expansion in the ownership of insurers and European operations. The Financial Services Act of 1986 loosened many restrictions related to the provision of investments. The regulation of such provision was gathered under the auspices of the Securities and Investment Board.

Further revision of the 1986 Building Societies Act is still an area of speculation. Recent secondary amendments of the 1986 legislation have allowed societies to undertake life insurance business and have increased the size of unsecured loans allowed and the wholesale funding limit. The continued development of legislation concerning building societies seems inevitable, particularly in the areas of continued extensions to wholesale funding limits, improved accountability of society management to their members and amendments to the merger process. A Treasury draft building society bill review document for industry consultations was issued on 18 March 1996. The suggestion of 'ring fencing' building societies from potentially hostile mergers was made. The review document includes a variety of alterations to existing practice. The purpose of building societies was specified as the making of "... loans that are secured on residential property and are funded substantially by members". This new legislation, it is hoped, "... will remove barriers to what they (building societies) can do and strengthen bonds with

members” (Treasury Draft Building Society Review Document, March, 1996). A number of measures have been introduced to improve the accountability of the management to the membership. These have included a broadening of the membership, the expansion of potential activities and a clearer definition of mutuality.

The primary aims of the legislation may be viewed as the increased accountability of societies. This has been sought through measures to improve the information received by members concerning elections to the board and new business ventures. The increased transparency of the rights of members, of the electoral process and the annual report has additionally been posited as an aim of the legislation. The funding of 50 per cent of activities through shares and a maximum 25 per cent of assets consisting of non-residential mortgages was proposed to assist in improving the viability of the mutual status and purpose of building societies. This change will allow an expansion into other areas of business both within the UK and Europe. The Building Societies Commission has also worked with the Treasury to produce a voluntary code of practice. Additionally, the rules relating to speculation by the public for payments from building society mergers have been tightened.

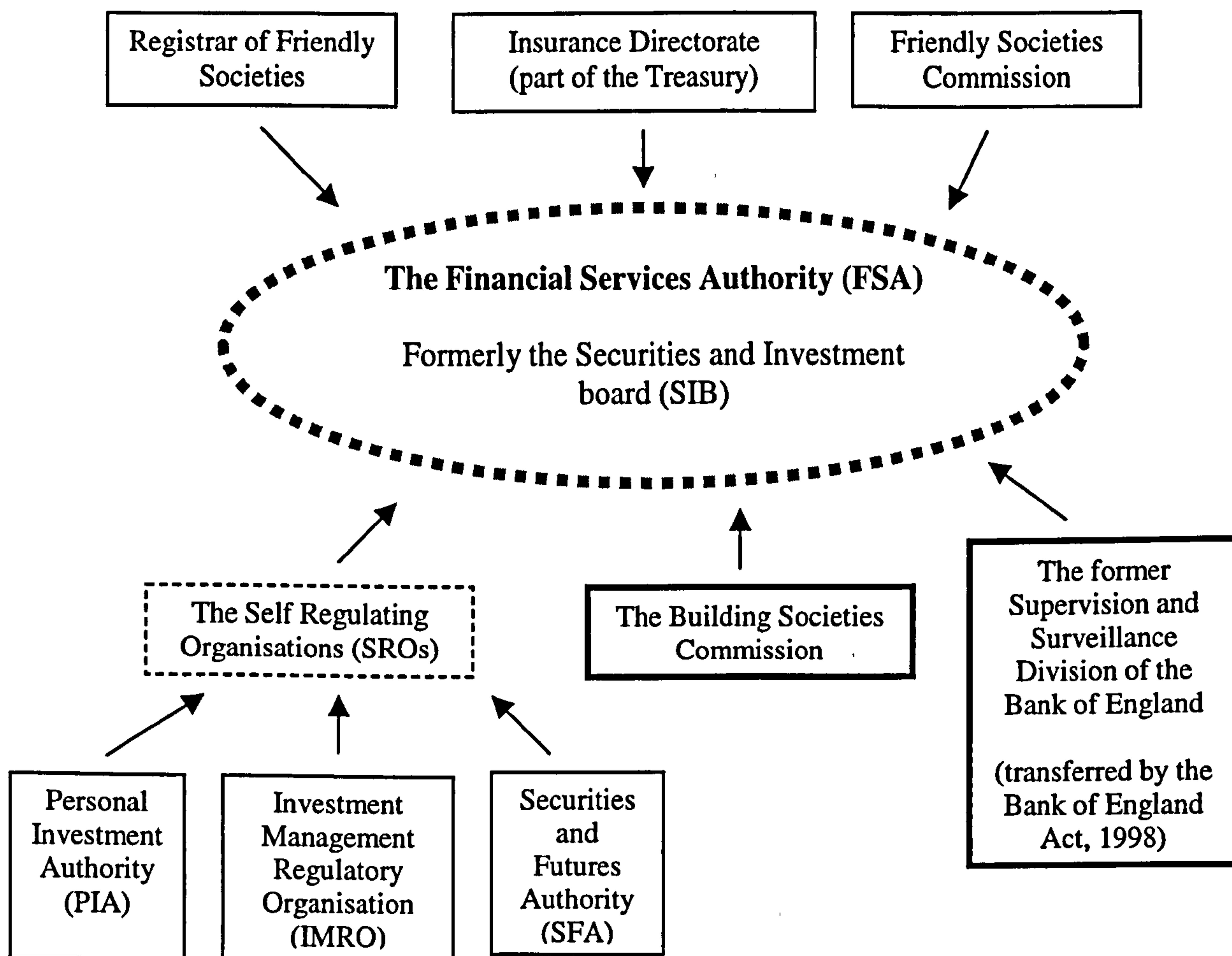
In March 1997 the Building Society Act 1997, was introduced. This act had the explicit aim of amending the 1986 Building Society Act, particularly in relation to the amalgamation of depositor protection schemes. Under this Act, the Building Societies Investor Protection Board and Deposit Protection Board and the Building Societies Investor Protection Fund and the Deposit Protection Fund were combined. Also in March

1997, the Building Societies (Distribution) Act, 1997 was introduced. This Act had the specific aim “ ... to amend the law in respect of distribution of assets on the take-over or conversion of a building society” (p.1, Building Societies Act (Distributions) 1997). This Act introduced a number of changes including amendment of the rights of Trustee Savings Accounts². The removal of the distinction between retail shareholders (members) and depositors (non-members) and voting rights for mortgagees were previously suggested as ways of expanding the membership of building societies³.

Recently, regulation for both the retail banks and building societies has been shifted to the recently established Financial Services Authority (FSA). This new ‘super regulator’, established by the government in October 1997, aims to bring regulation of financial service provision, including insurance and banking, under a single authority. The aims of the new institution have emphasised the importance of consumer protection and safe and sound conduct within financial markets. Of primary importance in this on-going reform is the merger of a number of financial services regulators. In justification of this move, it was stated that there “ ... has been a blurring of the distinctions between different kinds of financial services business: banks, building societies, investment firms, insurance companies and others. This has added further to the complexity of financial regulation. The Government believes the current system is costly, inefficient and confusing to both regulated firms and their customers. It is not delivering the standard of supervision and investor protection that the public has a right to expect” (p. 8, *The Financial Services and Markets Bill: A Consultation Document*, 1998). To amend for this ‘complexity’, the FSA

will take over responsibility from nine financial services regulators. This process is displayed in Figure 2.1.

Figure 2.1 The combination of regulatory powers under the FSA



The Bank of England Act (1998) came into force on the 1st June, 1998. This act transferred the prudential supervision of banks from the Bank of England to the FSA. The draft Financial Services and Markets Bill lays the basis of other functions of the FSA by providing it with “ ... the comprehensive and coherent set of powers needed to respond to

an industry that transcends geographical and sectoral boundaries.” (Brown, 1998, from the Financial Services and Market Bill, foreword). The bill is also responsible for the establishment of a single ombudsman, compensation scheme and appeals tribunal.

The development of the internal market within the European Union has also potentially significant effects on the performance of UK depository institutions. The European Union aims to promote free movement of goods and services among 290 million consumers in 15 countries. The expansion of the European Union also seems probable with talks in 1997 concerning the incorporation of the Czech Republic, Poland and Hungary.

Lohéac (1991) implies that the three principal aims of the European movement are minimum harmonisation, mutual recognition and home country control. The principal of minimum harmonisation indicates that legislation between member states is to be eventually brought into a common format. Progress towards this goal has been greatly complicated due to the distinct traditions and legislative frameworks of the member states. Mutual recognition puts forward the goal of overall acceptance of regulators. Such a shift has been criticised for introducing competition between national regulators at a European level. Home country control concerns the proposed single license system, which would be acceptable for trade in all member states.

A raft of European legislation has been developed in the attempt to achieve the first goal of harmonising supervision and regulation. Molyneux *et al* (1996) outlines the principal European contributions of this re-regulation. It is suggested that the First Banking

Directive (77/80/EEC) 1977, 1985 EU White paper, the Second Banking Co-ordination Directive (89/646/EEC) 1988, the Own Funds Directive (89/299/EEC) 1989, the Solvency Ratio Directive (89/647/EEC) 1989, the Money Laundering Directive (91/308/EEC) 1991, the Large Exposures Directive (92/121/EEC) and the Deposit Guarantee Scheme (1993) are of primary importance. Other legislative changes relating to the development of a single market for securities and investment have been developed over the sample period. In 1993, a range of legislation was phased in, including the Second Banking Co-ordination Directive, granting EU incorporated banks the right to establish branches and offer services within other member states. Other European level regulatory developments include the Solvency Ratio Directive, which established common reserve requirements. The definition of capital for regulatory purposes is imposed by the Own Funds Directive and the Large Exposure Directive. These changes have, in turn, led to the introduction of limits on banks' potential exposure from single sources. The Depositor Guarantee Scheme attempts to set minimum standards for deposit protection across the EU.

A number of changes in the European regulation and supervision of banking and finance were either introduced or proposed in 1996. Of these changes, the international declaration on co-operation and supervision and the complementary Memorandum of Understanding and Agreement, the introduction of the EU Capital Adequacy Directive and the Investment Services Directive (1st January, 1996) dominate. The Capital Adequacy and Investment Services Directives enable amendments to be made to the existing solvency and safety regulation of investments and banking in Europe. New rules

have also been introduced with the aim of limiting risk and bank exposure for a range of activities, such as foreign currency investment, domestic investment (with limited exceptions) and options trading. The early introduction of these rules in Britain continues with the Bank of England's policy of 'super-equivalence'. However, the directive has not created a level playing field across Europe due to differing interpretations by central banks. The Capital Adequacy Directive, for example, was introduced in Germany in January 1997, a year later than the UK. These directives will require amendment with the introduction of the new Basle proposals in 1998.

A number of influences have affected the international regulation of financial services, in general, and depository services, in particular. The development within the world of a number of trading areas, such as the EU, has continued apace. The North American Free Trade Agreement between Canada, Mexico and the USA, the Andean Pact, between a number of South American countries and ASEAN, within the eastern pacific region, emphasise the promotion of free trade between their member states. These free trade blocs could be viewed as part of a global shift towards greater regional control of trade.

International legislation has emphasised the importance of increasing collaboration between national supervisors and the improvement of information transferral and sharing. The establishment of mechanisms for the sharing of information between market authorities has been proposed as a central goal of international co-operation. Co-operation is encouraged when triggered by events such as large exposures of banks and firms. Such developments all assist in reducing potential credit squeezes that have

followed re-regulation of financial sectors. This has been previously observed in Japan⁴, Italy⁵ and the Nordic countries⁶.

The different regulatory and supervisory structures of banks and building societies support the suggestion that banks and building societies should be considered separately. Further, it is important that reference to the time both macro-economic and regulatory events occurred during the sample period be made when considering changes in productive technology and efficiency over time.

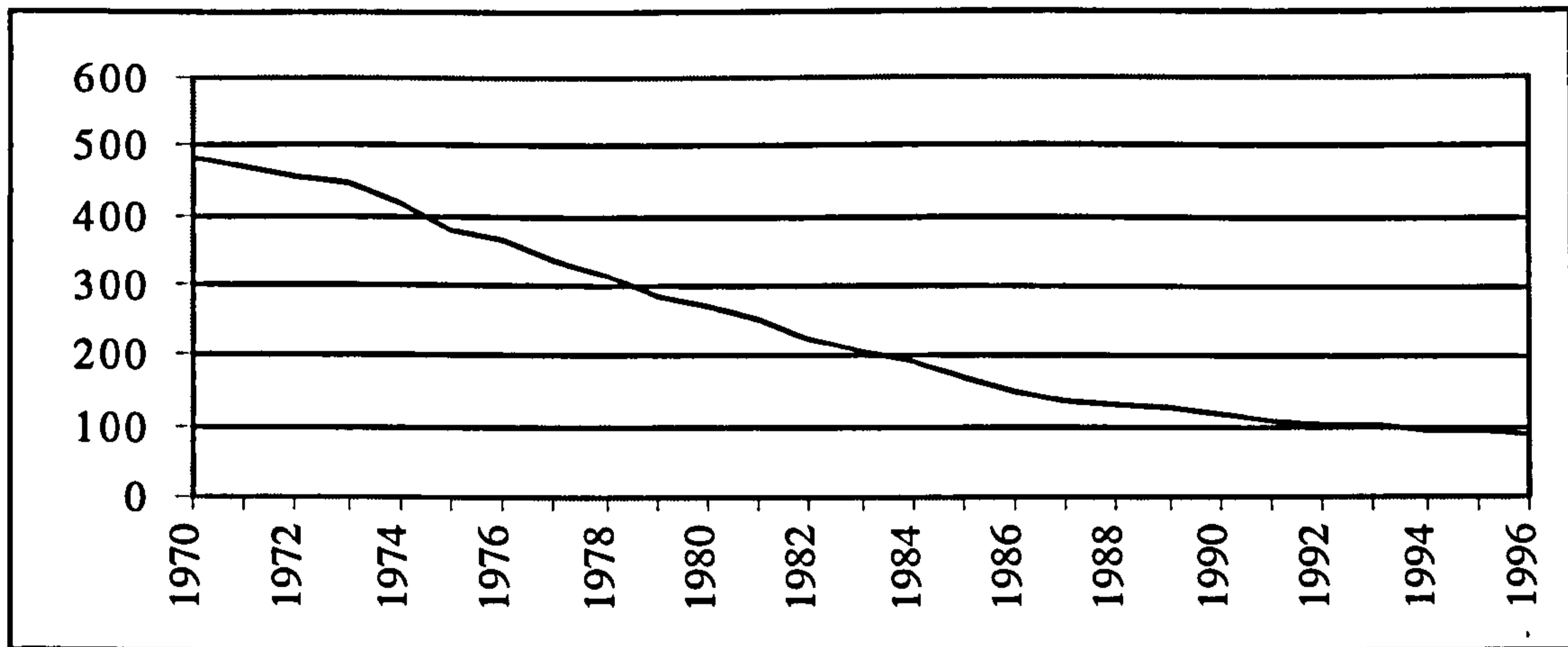
2.4 *Changes in market structure*

During the sample period, the processes of increasing concentration and de-mutualisation have continued. Concentration in retail banks and building societies has increased over the sample period. The processes of merger and take-over have led to contraction of depository institution numbers, particularly in the building society sector. This development is illustrated in Figure 2.2.

Concentration in the industry and the market power of individual building societies can also affect the motivation of individual societies to minimise costs. Concentration can be quantified with a number of differing measures; the Herfindal-Hirschman index (or *HH* index), for example, is an index of market share calculated as the sum of the squares of the market shares of individual depository institutions. This measure of concentration is

bounded by [0, 1] with one indicating pure monopoly and zero indicating an industry with an infinite number of firms.

Figure 2.2 The number of building societies: 1970-96



Source: Annual Reports of the Chief Registrar of Friendly Societies and Annual Reports of the Building Societies Commission.

Another measure of concentration is the entropy measure. This is computed as the sum of individual shares multiplied by the natural logarithm of the reciprocal of market share. This measure varies from zero for an industry with just one firm to $\ln(n)$ (here approximately 6.53) for a very low levels of concentration. The change in concentration for total advances and deposits, measured by the *HH* index and entropy measures is displayed in Table 2.1. The levels of concentration reported are low for both deposits and advances in banks and building societies. The degree of concentration is greater for retail banks than for building societies. This result is not unexpected due to the relatively small number of retail banks⁷. Over time, a slight increase in the level of concentration is recorded.

Table 2.1 Concentration in the retail banking and building society sectors

Year	Retail banks				Building societies			
	HH	HH	Entropy	Entropy	HH	HH	Entropy	Entropy
	Advances	Deposits	Advances	Deposits	Advances	Deposits	Advances	Deposits
1990	0.18	0.17	1.94	1.97	0.11	0.11	2.85	2.87
1991	0.15	0.15	2.09	2.09	0.10	0.11	2.82	2.81
1992	0.15	0.15	2.10	2.09	0.11	0.10	2.78	2.79
1993	0.15	0.14	2.08	2.11	0.11	0.10	2.77	2.78
1994	0.14	0.14	2.11	2.13	0.11	0.10	2.76	2.76
1995	0.14	0.13	2.13	2.15	0.15	0.15	2.53	2.54
1996	0.14	0.14	2.10	2.14	0.13	0.16	2.32	2.50
1997	0.16	0.16	2.02	2.00	-	-	-	-

The literature on the causes of differing levels of concentration in a market is both extensive and diverse (see chapter 6 and Gilbert, 1984, for further discussion of some of these issues). The central features of both the advances and deposit markets for retail banks and building societies are the increasing level of concentration and the low initial level of concentration. The increasing level of concentration observed in both markets may have resulted from a number of possible factors, although the changing operating environment may be of particular importance. Depository institutions have undergone substantial changes both in the form and organisation of work and technology used in the workplace. These shifts during the sample period may have resulted in a movement towards more capital-intensive production. Such a technological shift may have resulted in both increasing economies of scale over time and acted as a motivation for the on-going merger and take-over of depository institutions. The presence of shifting levels of

economies of scale will be considered more closely in the empirical chapters in Section Two.

Possible explanations for the initial low level of concentration may be closely linked with the development of the operating environment and past conduct of incumbent institutions. Both sectors have until recently been quite closely supervised in terms of the scope of operation. These restrictions may have represented a substantial barrier to entry allowing sub-optimal behaviour by incumbents. This sub-optimal behaviour may have materialised as scale inefficiencies, where small incumbents may have been able to survive in an industry characterised by economies of scale or production inefficiencies or both. Clarification of and reference to both these eventualities will be made in the empirical chapters in Section Two. Coles (1992) and Hardwick (1996) have also provided discussion of concentration in the building society sector.

In the USA, the Department of Justice merger guidelines (1992) explicitly enabled mergers of financial institutions to proceed unchallenged when there were possible or potential efficiency benefits, that may not be gained from other means (Berger and Humphrey, 1992b). Across the globe, such 'efficiency' criteria have been used as strong justification for mergers and take-overs involving depository financial institutions. It is clearly important to review the evidence at both an institutional and at a societal level. Commenting on the existence of economies of scale in large banks, Berger and Humphrey claim that "... contrary to the claims made in the trade press and elsewhere, the extant academic literature on scale efficiencies in banking suggest that there are little

or no cost savings to be made simply by increasing bank size through mega-mergers” (p.546).

Demutualisation, either through conversion to bank status or by take-over by a bank, is a shift from mutual ownership to proprietary ownership. Whilst the retail banking and building society sectors are similar in many respects, their ownership form has been proposed as a possible reason for their differential performance. This view has led to differing approaches to both the regulation and operation of these institutions.

2.5 *The growth of new products*

The loosening of restrictions on the products that depository institutions may provide has increasingly blurred the distinction between banks and building societies. The regulatory authorities swiftly identified the beginnings of this shift. The governor of the Bank of England declared in 1984, “ ... you will of course be well aware of the move by banks into the housing market, previously very much the domain of the building societies. The banks have also started – but with limited success – to seek to attract funds much more vigorously from individual savers. The building societies have responded by forays into what might earlier have been considered the preserve of banks, in some cases with the co-operation of the banks themselves”. These moves occurred simultaneously with longer-term changes. Customers of depository institutions, including households, corporations and other financial institutions, may have increased their sensitivity to the differentials in yields from investments over time. Increasing awareness of rates of return on investments

after the fluctuations of interest rates in the UK since the 1970s may have driven this process.

The greater awareness of the costs of other sources of finance has increased demand for products, such as commercial paper, mutual funds and securitised loans that are provided by depository institutions and by other institutions. The effect of this dis-intermediation, as personal investors have increasingly made use of pension funds and life insurance, has been a movement of savings away from traditional sources such as deposit accounts or building society share accounts. For example, pensions and life insurance assets as a percentage of total household assets increased from 39.9 per cent in 1980 to 49.9 per cent in 1985 and 53.7 per cent in 1990. More aggressive competition for deposits and an expansion of the range of products available is thought to have resulted.

A response to dis-intermediation and economic re-regulation has been the expansion by banks into a broad range of non-core product areas. Banks with extensive experience and organisational knowledge in the sphere of assessing an individual's risk have prospered in a range of new markets, especially mortgages. Measurement of success in other product areas, such as life assurance, remains problematic, though the considerable gain of market share⁸ provides some indication of success. Assessment of performance in this area is further complicated by the substantial alteration in the levels of market share for certain products held by banks, building societies and other financial institutions. For example, diversification into some areas such as estate agencies turned out to be temporary in many cases. Change of this magnitude may be driven by a number of cost

considerations. In the provision of new and many more services, differing productive technologies may be emerging. Additionally, the provision of more services may provide economies of product-mix. It is important in a study of depository institutions, therefore, to consider both the potential economies of product-mix and changes in productive technology.

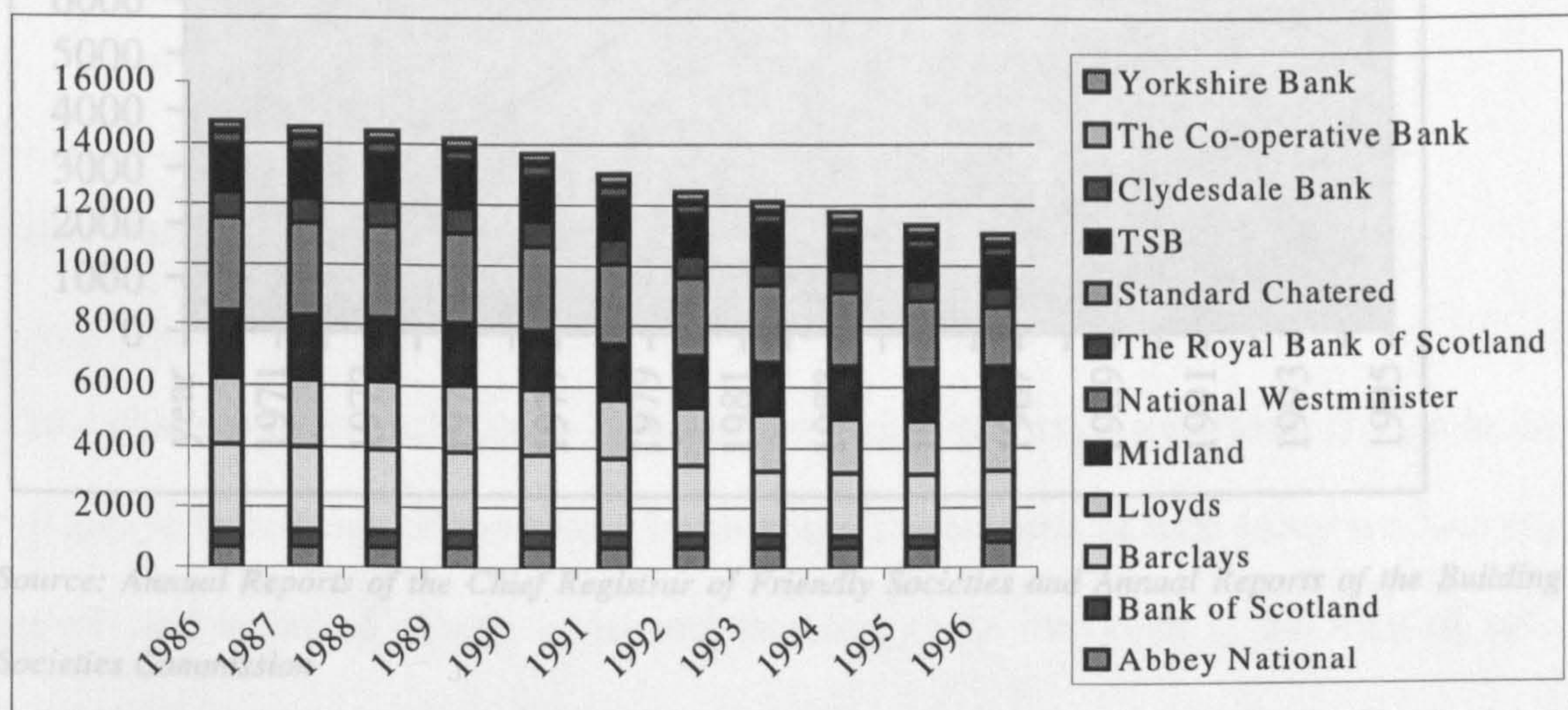
2.6 Changes in the technology of service provision

The introduction of innovations, particularly in the area of distribution, has been widespread. Long-term aims of this change include the establishment of unmanned branches, the improvement of customer information and increased quantities of customer services handled either through automated machinery or centralised call centres. The success of these innovations will very much depend on whether they are technology driven, like the ill-fated Prestel system of the 1980s, or market driven, like the successful First Direct telephone banking system.

Branch networks have been substantially rationalised over the sample period. The provision of a full range of services in a broad catchment area has been posited as an aim of such change. Such a strategy would aim to create a hierarchy of branches defined through their level of service provision. This differs from the previous strategy of full provision of services at most branches. As part of this change, many functions previously performed at branches have been transferred to larger centralised and specialised centres of operation. Cost cutting may also be viewed as a motive for such change.

Rationalisation has resulted in the re-location of staff as well as increasing responsibility and autonomy of those within the lower-tier branches. The rationalisation of branch services from a network geared towards serving customer and company requirements for transactions and administration towards a service and more 'sales' orientated approach has continued. This has led to the closure of branches for all the banks, which may be observed in Figure 2.3.

Figure 2.3 Number of branches in UK retail banks

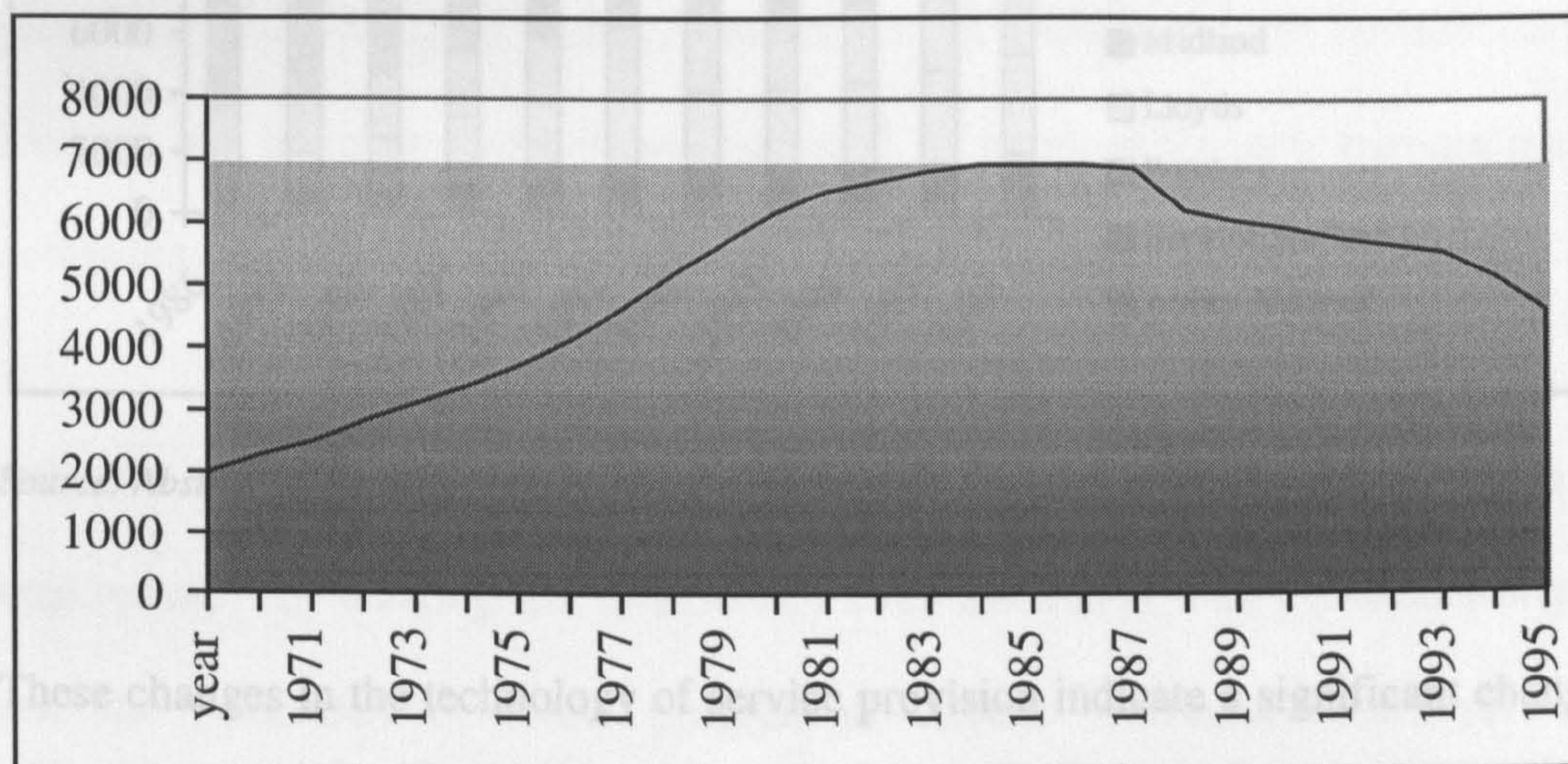


Source: Abstract of Banking Statistics The British Bankers Association: various editions.

A similar process has occurred for most building societies with a substantial rise in the number of building society branches until the mid-1980s when an equally large decline in branch numbers was recorded. This is displayed in Figure 2.4. Rationalisation of branch numbers has occurred in tandem with programmes of refurbishment of branches. Institutions have attempted to improve the sales environment to accommodate changing

customer service and corporate requirements. This change is mirrored by the equally significant rise of alternative outlets from which banking services are provided. The rise in the number of automatic cash dispensers and automated teller machines (ATM's) is illustrated in Figure 2.5.

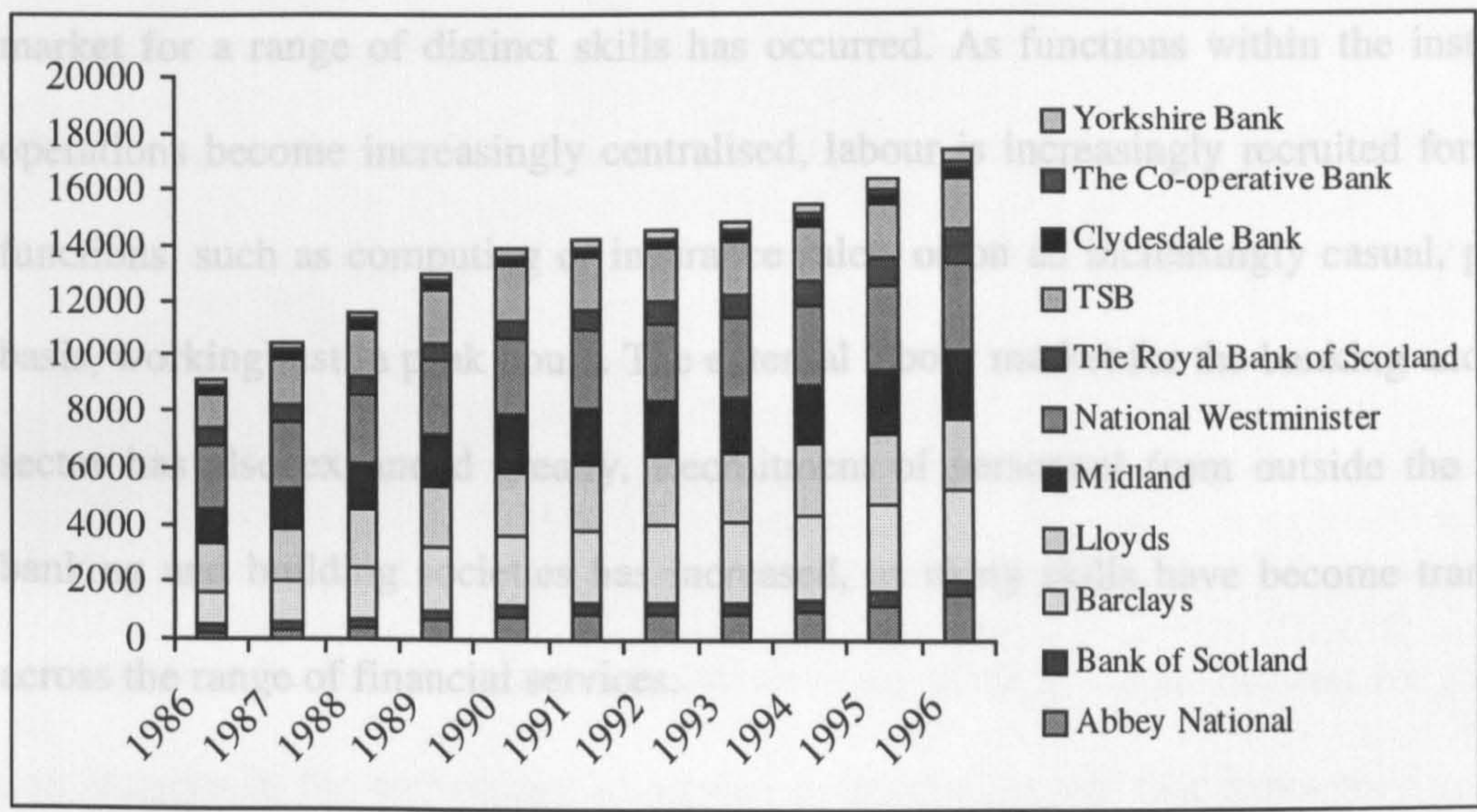
Figure 2.4 Number of branches in UK building societies



Source: Annual Reports of the Chief Registrar of Friendly Societies and Annual Reports of the Building Societies Commission

The growth and greater availability of information technology has substantially reduced the costs of certain information-intensive activities. Traditional banking activities of credit rating and transactions have been particularly affected with both changes in the organisation and practice of work. Whilst the most information-intensive activities may not be within the capabilities of many competitors, the growth of computing capacity and statistical tools for activities such as credit rating has increased the potential for entrance into core product provision.

Figure 2.5 Number of cash dispensers and ATM's in UK retail banks and 'stand-up'



Source: Abstract of Banking Statistics The British Bankers Association: various editions.

These changes in the technology of service provision indicate a significant change in the productive technology of depository institutions. The analysis of total factor productivity growth and technical change is deemed therefore to be important in the light of such developments.

2.7 Changing labour markets

In the early 1990s, banking was described as “... the steel industry of tomorrow” due to its dis-equilibrium position of over-manning within an expanded branch network. An overall reduction of staff and the fragmentation of internal labour markets have occurred. The implications are widespread with a notable resurgence in re-training, particularly for the provision of more specialised services. Banking and building societies were until

recently characterised by tightly defined internal labour markets and 'stand-up' recruitment policies. In the sample period, movement towards a sector-wide labour market for a range of distinct skills has occurred. As functions within the institutions' operations become increasingly centralised, labour is increasingly recruited for specific functions, such as computing or insurance sales, or on an increasingly casual, part-time basis, working just in peak hours. The external labour market for the banking and finance sector has also expanded greatly. Recruitment of personnel from outside the areas of banking and building societies has increased, as many skills have become transferable across the range of financial services.

O'Reilly (1992) bleakly elaborates the implications of such a transformation. For employees, " ... banking has traditionally been associated with permanent employment and the possibility of career progression through well-established internal labour markets". The previous " ... stable comfortable atmosphere in the banks has been shaken up by increasing competition and deregulation, (as) ... the introduction of information technology has had serious consequences for the restructuring of employment practices and the services banks, and other financial services firms, can offer. The heavy burden of mundane paper-work has been replaced by the use of information technology and computer terminals" (p.45).

2.8 *Conclusions*

In conclusion, changes in the market environment are likely to have had a significant effect on the performance of depository institutions over the sample period. The effects of increases in concentration, changes in market structure and differing ownership forms suggest that the separate treatment of retail banks and building societies is correct. Economies of scale and cost efficiency will be measured to assist in the assessment of whether increasing concentration across the sector has been driven by cost considerations or other criteria. Further growth of new services, shifts in labour markets for this sector and changes in the technology of service provision suggest that depository institutions have altered their productive technology. Testing for economies of product-mix, technical change and total factor productivity growth within the sector is therefore important. The macro-economic environment and regulatory and supervisory changes are significant exogenous effects over the sample period. Thus, it is also important to assess the competitiveness of depository institutions in their core markets, in light of the re-regulation of depository institutions.

¹ The responsibility for banking supervision was moved to the Financial Services Authority in 1998, following the Bank of England Act 1998, see p.17.

² Trustee Saving Account holders were previously disenfranchised when building societies converted to bank status and paid compensation to their membership.

³ This amendment was initially suggested in the Treasury Draft Building Society Review Document, March, 1996.

⁴ Japanese banks announced \$86bn of provisions for bad and doubtful debts on March 21st, 1996.

⁵ The Italian government issued a \$2bn rescue package for Banco di Napoli also in March, 1996.

⁶ IMF estimates suggest that rescuing ailing banks has cost between 2.8 per cent and 4 per cent of the Nordic countries' GNP.

⁷ A sample of 11 retail banks was considered for all years. A sample of 99 societies was used in 1990. The number of societies declines over the sample period due to wastage. The number of building societies operating in 1996 was 76.

⁸ 14 per cent of the life assurance market was held by bancassurers in 1994.

Chapter 3 *Model and variable definition*

3.1 *Introduction*

This chapter reviews the approaches that have been previously employed in the modelling of production in depository institutions. Recommendations as to what model forms and variables are deemed appropriate for use in the empirical studies are made. This chapter is also closely linked to chapters 4 and 5 where the model forms referred to here will be developed. In the following sections, a broad range of studies are considered. These studies have examined many different forms of depository institutions, including banks, building societies, credit unions and savings and loan associations (S & L's). These studies are discussed in a review of the general approaches to modelling production, efficiency and productivity in depository institutions.

The chapter is structured as follows. Section 3.2 provides an introduction to the principal issues involved in the modelling of production in depository institutions. Sections 3.3 to 3.8 consider model forms that have been employed in previous studies of depository institutions. Section 3.3 introduces and outlines the *intermediation* model of production and section 3.4 examines the principal features of the *production* model. Sections 3.5 and 3.6 consider the *delegated monitoring* and *national accounting* approaches to modelling depository institutions, respectively. Other approaches that have been employed to model depository institutions are assessed in section 3.7. A summary of the differing variable definitions used in these models is presented in sections 3.8 to 3.11. This summary includes consideration of labour, deposits, loans and capital variables. Possible

difficulties encountered when using accounting data are discussed in section 3.12. Section 3.13 provides a summary of the chapter and proposes recommendations for model and variable definitions in the empirical chapters of the thesis.

3.2 *Production in depository institutions*

Micro and macro views of depository institutions differ. In the micro sense, a depository institution produces services that are sold in a marketplace. In the macro sense, depository institutions act as a producer of money itself. These conflicting properties have led to difficulties in the modelling of a depository institution's 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 process, termed financial intermediation, is very widespread and performed by a broad range of financial institutions, firms and specialist government agencies. This form of intermediation is a principal and defining aspect of the production process of depository institutions.

Production in depository institutions creates services as opposed to recognisable goods. Depository financial institutions use both physical and non-monetary inputs (i.e. physical capital and labour) and monetary inputs (i.e. funds from a number of sources) to produce services. Similarly, output consists of non-monetary services, such as the manipulation and administration of customers' accounts, and monetary services, such as loans (Hancock, 1985).

Due to the distinct nature of inputs and outputs and the absence of clear variable definitions, a wide range of approaches to modelling this production process have been suggested. A number of issues have arisen in the modelling of production. These include the lack of directly applicable data and the distinction between the depository institutions' functions at a micro and macro level. These issues present difficulties in the quantification and pricing of variables where " ... a substantial part of their (*depository institutions*) costs are incurred in providing services as partial payment for an input" (Sealey and Lindley, 1977, p.1254).

The objective function of depository institutions in most studies has been the provision of a range of non-monetary and monetary services at minimum cost, maximum profit or with the most productive use of inputs. In some studies, the objective function of certain institutions has been considered differently. The objective function of credit unions, for example, has been viewed as the maximisation of the level of service provided to depositors and loan holders (Fried, *et al*, 1993). The objective function of a depository institution has also been viewed as providing service and convenience to customers, where production is " ... the transformation of financial assets suited to the needs of savers into a different set of financial assets suited to the (*needs of*) borrowers" (Nelson, 1985, p.177). The objective function may also be constrained by the necessity to hold risk at an acceptable level. This is often achieved through the devices of loan portfolio diversification and minimum capital requirements.

3.3 *The intermediation or asset approach*

In the *intermediation* approach to modelling production, depository institutions are assumed to borrow funds (*deposits*), which are transformed into loan funds (*advances*). The depository institution uses deposits as a raw material or intermediate product, in combination with non-monetary inputs (e.g. capital and labour). This input is then transformed into the final product. Outputs from an *intermediation* model could include loan funds and ancillary business. The model thus consists of factors of production or non-monetary inputs, such as labour and capital, and borrowed funds that are then re-lent. Thus, costs include both production costs and the attraction costs, which include interest costs and the costs of non-price competition.

In a simple *intermediation* model, production might be viewed as a transformation of three input groups (capital, labour and deposits; X_1, X_2, X_3) into one output group (loans; Y_1). A production correspondence could be written as:

$$f(Y_1; X_1, X_2, X_3) \tag{3.1}$$

The dual cost function would then be:

$$C = g(Y_1; P_1, P_2, P_3) \tag{3.2}$$

Where P_1 is the price of capital inputs, P_2 is the price of labour inputs, and P_3 is the price of deposit inputs. This approach includes the effects of both non-monetary and monetary inputs. Monetary output is included (loans), whilst non-monetary output is omitted. A classical emphasis on the macro and dynamic changes in assets is therefore imposed. The definition of monetary inputs and outputs ignores the effect of retaining liquidity. Including the level of liquid assets in some form may be a way of overcoming such a problem.

3.4 *The production, value added or portfolio approach*

The intermediation approach implies that the processes used in the transformation of funds drive production in depository financial institutions. The production approach specifically incorporates the functions of the depository financial institutions. These functions are viewed from an operational, 'technical' or 'economic' perspective (Sealey and Lindley, 1977). Researchers who have placed a greater importance on what depository financial institutions directly produce have adopted this approach. Benston, (1972) suggested that "... financial institutions produce services rather than readily identifiable physical products" (p. 320). Sealey and Lindley, stated that the "... financial firm's output in a technical sense is thus a set of financial services to the firm's depositors (creditors) and borrowers." (p.1252). The provision of depository services is viewed as a form of production in its own right and is therefore defined as an output.

A simple production model might be viewed as the transformation of two input groups (capital and labour; X_1 , X_2) into two output groups (loans and deposits; Y_1 , Y_2).

Production or non-interest costs would represent total cost in such a model. A production correspondence could then be written as:

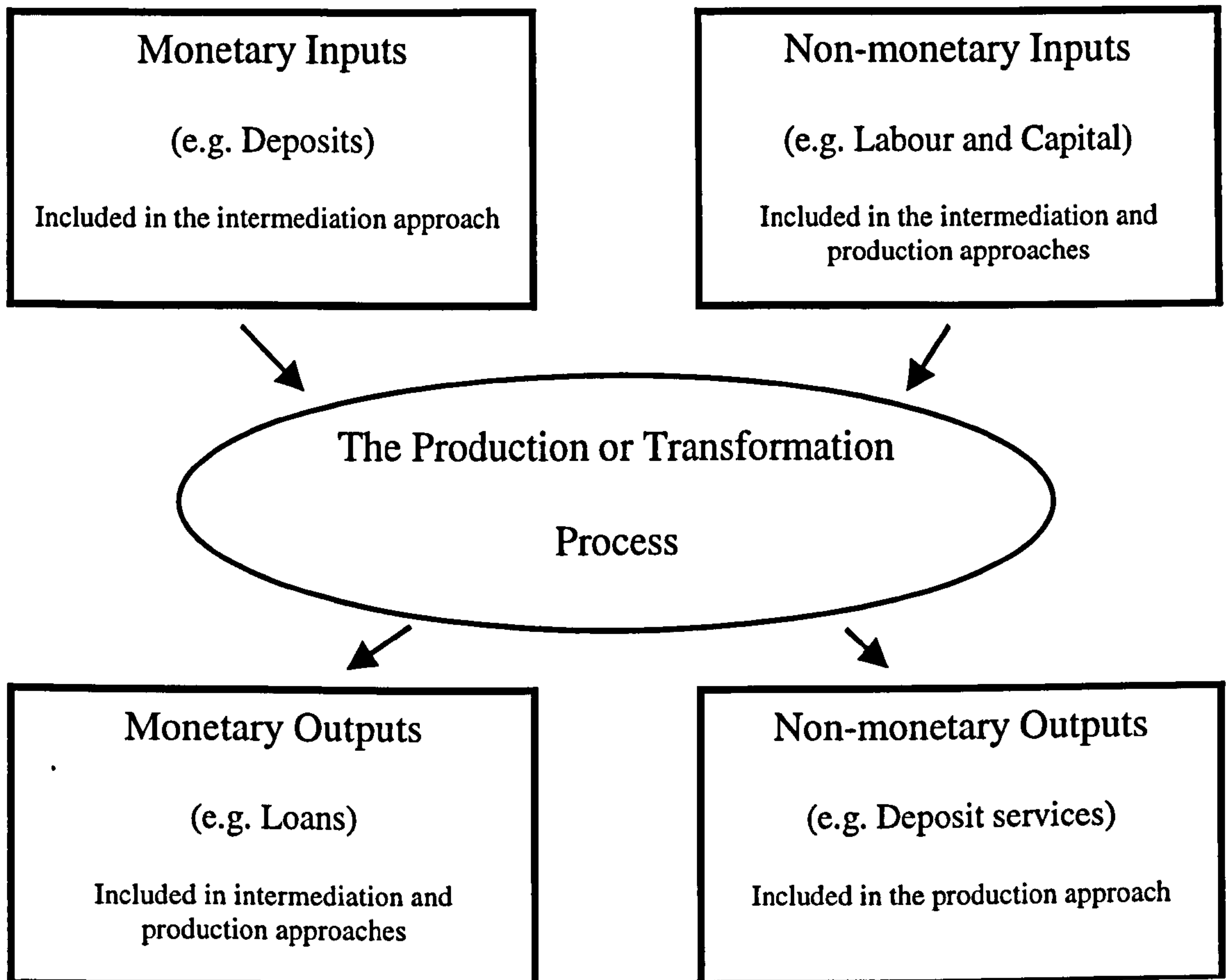
$$f(Y_1, Y_2; X_1, X_2) \quad (3.3)$$

and the dual cost function would be written as:

$$C = g(Y_1, Y_2; P_1, P_2) \quad (3.4)$$

This approach uses non-monetary inputs and ignores the effect of monetary inputs. Monetary and non-monetary outputs are included. The relationship between the production and intermediation approaches to modelling depository institution production is outlined in Figure 3.1.

Figure 3.1 The relationship between the production and intermediation approaches



3.5 National accounting measures

Colwell and Davis (1992) propose the use of *national accounting* measures to model the production of depository institutions. These accounts “... seek to measure the value added by different sectors of the economy, reflecting in turn the profits and income from employment arising in each sector” (p.112). Significant problems with this approach include the movement of the focus of analysis away from the firm to the sector and not fully incorporating the costs of attracting funds. Both these omissions could result in

significant model mis-specification. Other criticisms could include the difficulties of variable definition and the use of aggregate data. The use of aggregate industry level data commonly produced by national accounts may also be open to criticism. By comparing industries to gain relative efficiencies only very general structural measures of efficiency are approximated. Such measures are limited for intra-industry and firm level analysis.

3.6 *Delegated monitoring*

The *delegated monitoring* framework for incorporating risk in the measurement of production characteristics is attributed to Diamond (1984). A depository institution making loans has to monitor debtors both at inception and for the duration of the contract to reduce the potential for loan default. As intermediation is victim to imperfect information about potential customers, the costs of monitoring a debtor are real. The production of depository services allows for the collection of information on both individual debtors and a large number of debtors in the broader economy. The gross costs of assessing risk and acquiring information may be reduced through the use of deposit provision to assess potential customers before the provision of loans (i.e. by holding records of a customers' deposit account, their future behaviour may be assessed more clearly). Thus a "... financial intermediary must choose an incentive contract such that it has incentives to monitor the information, make proper use of it and make sufficient payments to depositors to attract deposits" (Diamond, 1984, p.395). The process of intermediation, therefore, necessitates the viability and continued provision of deposit services by the financial institution. Important aspects to be included in the modelling of

this conception of the production process of depository institutions include the analysis of incentives to depositors, expected returns from debtors beyond monitoring costs and the presence of 'dead-weight penalties' in production.

Attempts to incorporate aspects of *delegated monitoring* in studies of production in depository institutions have been fragmented and piecemeal. Aspects of depository institution behaviour include the level of diversification of loan providers. Diversification reduces the risk of individual contracts and for the institution as a whole. This capture of information allows a cheaper provision of loans and a reduction in the amount of resources dedicated to monitoring individual contracts. This conception of a depository institution suggests a model in which loans and deposits are joint products or at least strong cost complements. Loans and deposits are thus viewed as outputs from the production process, with similarities to the production approach. McAllister and McManus (1993) include aspects of such an approach in their study. They suggest that the reduction of risk is an integral part of the bank production process. Proxies for risk include the size of the institution, the financial portfolio and the level of diversification displayed within the balance sheet of individual banks. Deposit pricing improvements have been assisted by the work of Barnett (1978) and Fixler and Zieschang (1992) who have focused upon the development of opportunity, rental or 'user' costs of deposits. These measures contain an element, which incorporates the costs of adapting to risk. Monitoring of the firm and the development of relationships between groups involved in depository institution production are generally long-term. Such time effects may therefore have real ramifications for contemporary costs.

3.7 *Other models*

Modigliani and Miller (1958) assume perfect capital markets, thereby disputing the *raison d'être* of depository institutions and emphasising their lack of importance within the wider economy. This approach assumes that the “... average costs of capital to any firm is completely independent of its capital stream and is equal to the capitalisation rate of a pure equity capital stream of its class” (p. 268). The capital asset pricing model assumes that maximising their portfolios' rates of return and minimising variance determine the financial firms' optimal outputs. Saunders and Ward (1976) used such a model to analysis British banks between 1965 and 1975. Normality in the distribution of rates of return was assumed. Inputs were viewed as exogenously determined in a perfect capital market, reducing the expected performance to a calculation of the utility function.

3.8 *Variable Definition: deposit pricing*

In the following section, the definition of variables used in models of production are considered, with emphasis placed on the definition of variables for use in cost functions. The rationale for this decision is discussed in greater detail in chapter 5. Variable definitions to be used in Section Two will be made on the criteria of what appropriate model formats require, the function of the variable within the production process, approaches that have been adopted in previous studies and what is practical considering issues of data availability and format.

Different model forms have considered deposits both as inputs and as outputs. To represent deposits as an output, a number of differing variables have been used. These definitions have often emphasised the role played by deposits in the production process. Deposits are often viewed as a service the depository institution performs to attract funds and potential customers and are frequently offered as a free service or at a nominal price. Deposits are also essential for depository institution production, enabling both fund creation and the assessment of customers for potential lending. The benefits of deposit provision for the bank may therefore be seen as a source of funds and the improved credit rating of individuals to assist in the bank's lending business.

These perspectives of the role of deposits enable two definitions to be suggested. First, the *value* of deposits would provide an output measure that would emphasise the importance of deposits as a source of funds. Secondly, by considering the *number* of deposit accounts the function of deposits as a screening device for customers would be stressed. Both these definitions have been used as output proxies in previous studies.

The benefits customers gain from deposits may differ from those gained by depository financial institutions. A customer uses deposits for purposes of safely holding money. Deposits are used as convenient platforms from which transactions are made. Since the 1950s, a concerted shift has occurred towards making payments of salaries and wages through bank current accounts. This has reinforced the broader social function of this form of banking service. The provision of records to customers is a principal form of asset and liability accounting for individuals with huge benefits for customers. A limited

amount of research has attempted to incorporate these broader social benefits. Deposit services have been measured as the quantity of depositors serviced. This approach considers deposit output as a distinct 'activity' relating to what the bank does in terms of providing for customers. Other aspects that have been considered include the individual product characteristics or *hedonic attributes* to provide a measure of customer-related output from deposits.

As previously stated, deposits have also been used as an input in production and cost models. To model the price of deposits, a number of definitions have been employed. Murray and White (1983) suggested the interest paid on demand deposits divided by the average dollar quantity of demand deposits as an 'effective' interest rate or price of deposits. Clark (1984) suggested that interest on deposits measured in terms of total interest paid on deposits divided by the asset sum of loan funds produced would be an appropriate measure. The use of the deposit interest rate as a measure of deposit price has been used by a number of researchers.

A range of other amendments may be proposed that incorporate features of deposits such as product diversity and the relationship between deposits and other outputs. Mester (1991) uses the interest rate paid on deposits net of service charges for a range of specific savings products, NOW accounts, passbook accounts and accounts earning in excess of the normal rate for deposits to represent the costs of deposits. This approach attempts to accommodate the differing cost structures and distinct turnovers of different types of savings products, which may be affected by distinct benefits, regulations and conditions.

Yue (1992) breaks the costs of bank deposits into costs that are considered 'transactional' and 'non-transactional'. Transactional deposit costs include interest expenses (for federal funds, the purchase and sale of securities, interest on demand notes and other borrowed funds). Non-transactional deposit expenses relate to the provision of depository services. Mester (1996) in a study of US banks includes the costs of loan capital (subordinate liabilities) and federal funds as additional components of deposit cost. The approach attempts to accommodate for funds provided from a number of sources.

Other measures of deposit cost have focused upon the opportunity costs of money. The US Bureau of Economic Analysis (BEA) defines the opportunity cost of money as the effective interest charged on loans. This crude statistic was developed by the United Nations Statistical Office (UNSO) to consider the opportunity cost of money. This is viewed as the average of the rate charged for loans and paid for deposits (Fixler and Zieschang, 1992).

3.9 *Capital prices*

The definition of the cost and price of capital has been problematic principally due to a lack of directly applicable data. The differing formats of data sources for different types of institutions have produced a broad range of distinct measures. This has led to inconsistencies and difficulties in the comparison of studies of both distinct types of depository institutions and depository institutions operating in different countries.

In defining the price of capital, many studies have used a crude division of the total capital expenditure of the institution by a measure of scale for the institution. This measure has been the basis for most capital price measures used in the literature, with many variants of this measure proposed. A potential problem with this definition is that the price of capital is related to the scale of the institution. This can result in bias, as larger firms tend to produce relatively lower capital prices.

The most 'common' measure of the price of capital used in previous studies may be written as:

$$\textit{Price of Capital} = \textit{cost of capital} / \textit{total fixed assets} \quad (3.5)$$

Studies of UK building societies have employed a number of proxies to represent the price of capital. Hardwick (1989, 1990) used annual expenditure on office accommodation and equipment, plus a measure for depreciation, divided by mean assets. Drake (1992) considered the charge for office accommodation and other expenditure, including depreciation, as the measure of capital cost, divided by the net book value of fixed assets.

For the savings and loan sector, a similar range of measures has been used. Mester (1987, 1991) suggested the sum of the costs of rent, depreciation, utilities and expenditure on equipment and furniture (the sum of capital expenditure) divided by the average dollar of deposits held within the study year. Le Compte and Smith (1990) followed a similar

procedure by dividing capital expenditure by the total fixed assets held by the firm. Hermalin and Wallace (1994) used the number of branches or the average value of equity capital for the denominator in place of total fixed assets. Such a measure is limited to use in proprietary or branching institutions.

US banking studies have also provided a broad array of measures. Grabowski *et al* (1993) suggested capital expenses divided by the book value of the banks' total assets as an appropriate measure of capital price. Clark (1984) defines the capital price as capital expenses divided by the banks' net book value of total assets minus a level of depreciation of physical capital. The rented cost of a square foot of office space for a specified geographic area of operation was used by Gilligan *et al* (1984). Lawrence (1989) suggested that the average retail cost per hour for use of a central processing unit (CPU) multiplied by the average number of weekly CPU hours used by the bank, multiplied by fifty-two, provided an original annual figure for the capital price.

Studies of non-US banking differ in the measure of scale used as a denominator of physical capital expenses. Kim and Weiss (1989) use the area of office space used by the institution as a measure of institution scale. Dietsche (1993) considered the quantity of borrowed funds, and Mulder and Sassenou (1993) proposed the size of 'physical capital' as an appropriate measure of scale. Further developments were suggested by Glass and McKillop (1991) who used the book value of net bank premises, furniture and fixtures as a scale measure in their definition of a capital price. A distinctive feature of the time series study by Glass and McKillop (1992) was the use of a capital expense measure

derived for the analysis of the Canadian telecommunications sector over time by Denny *et al* (1981). This measure views capital expenses as the aggregation of occupancy and equipment costs, deflated by a weighted average of prices for buildings and equipment. The definition of this weighted average of prices attempts to accommodate for the opportunity costs of capital in relation to other forms of investment.

3.10 Labour price

The definition of the price of labour is an area of consensus in most of the studies reviewed. The most commonly used measure is an aggregate of wages, salaries and benefits paid to employees of the institution divided by the number of full-time equivalent employees. This measure has been used by most of the studies reviewed, although a number of minor amendments to this definition have been proposed. This measure again is defined with reference to what variable definition is possible in relation to the data available.

Building society studies have predominantly used wages, salaries and other benefits divided by the number of full time equivalent employees (see Hardwick, 1989, 1990, Drake, 1992, Drake and Weyman-Jones, 1996). Non-US banking studies (e.g. Dietsche, 1993 and Muldur and Sassenou, 1993) have also used this proxy. Kim and Weiss (1989) included the cost of computer services in addition to labour costs in their measure. US banking studies have used similar measures.

Some savings and loan association (S & L's) studies have used proxies, which are distinct from those previously suggested. Mester (1987, 1991) employed a weighted average of wage rates within counties of operation of the S & L's as a measure of labour price, though Mester (1993) does not use this approach, employing instead the quarterly average labour expense divided by the number of full-time equivalent employees. All these measures exclude any distinctions between types of workers.

3.11 Output measures

Output definition has been a primary problem in the analysis of depository institutions. Previous attempts to define output have often been steered as much by the limitations of the tools of analysis and data availability as by theoretical considerations. Initially, single output production measures were employed in empirical studies. As the sophistication of the tools of analysis developed, the use of multiple output measures has become common. Examples of single product models include output proxied by an overall measure such as total assets or earning assets. These measures are reliant on assumptions of homogeneity and the rejection of real differences in costs between distinct outputs in multi-product production. Benston (1972), employing an alternative approach, recognised the multi-product aspects of bank production by estimating separate production functions for the principal products produced by the bank.

A third approach has been to create output indices enabling a number of outputs to be incorporated within the measure. Longbrake and Haslem (1975) employed a weighted

index of output, constructed from the number of accounts, the account size and the number of offices of the bank. These three factors were aggregated to represent aspects of firm or 'plant' size, the type of customer and individual firm structure. Clark (1984) describes such a procedure where the "... weights employed are usually determined by the regressing of bank revenue from interest earning assets against the composition of earning assets, ... thus the resulting index represents total revenue from earning assets adjusted for any inter-bank price differences which may arise from the existence of imperfect markets for bank output" (p. 54). This approach moves closer to modelling multi-product behaviour. Clark creates three such derived measures. The first measure is produced by regressing revenue on the gross sum of earning assets held by individual banks. The second approach measures output by regressing revenue on lending and non-lending income. This measure is therefore comparable with the first measure plus the difference between total operating income and operating income from lending. The third measure of output proposed by Clarke is the sum of total earning assets.

The definition of output in a multi-product cost study is related to our previous discussion of model definition. The actual variables used in previous studies do vary significantly in terms of the number of outputs considered and the measure used to represent the quantity of this output. In practice, the number of outputs is a function of model specification and the access to data. Whilst some databases provide breakdowns of quantities of products produced or services provided (such as the Functional Cost Analysis programme in the USA), most data is presented in an aggregated form from company accounts. Availability is affected by the nationality of the study and the sector considered. For example, Hunter

and Timme (1995) used a FCA data set in a study of US banks: the outputs included wholesale loans, average dollar balances of commercial loans, industrial loans, security loans, the average dollar balance of credit cards and personal loans, real estate loans to agriculture and other real estate loans, and off-balance sheet activities proxied by the income produced from loan sales, letters of credit, securitisation, swaps and clearing services received on transaction and non-transaction accounts. Hardwick (1990), using UK building society accounts data, employed the number of outstanding mortgage loans and the number of share deposit accounts. This difference in the level of detail is a major difference of the non-US bank studies, conducted in countries where data is less freely available. This difference is compounded by the greater degree of contractual heterogeneity observed in loan products, where variation in output may also occur due to the format of data.

As previously stated, the definition of output, in common with the definition of capital and labour prices, has been an area where development has been severely restrained by limitations in the data availability and construction. In contrast to the definition of input prices, a significant development in the definition of output has occurred.

The *hedonic attributes* approach to modelling depository institution production emphasises possible differences that may exist between the individual productive technologies of firms. As the activities that depository institutions undertake may incur differing costs, it is assumed that the quality of the outputs produced may vary or be heterogeneous.

The measurement of different qualities and attributes for depository and loan products has been incorporated in a range of studies. Fried *et al* (1993) uses a variety of proxies to represent *hedonic attributes*, Flannery (1983) employs the public consideration of services and Nelson (1985) applies the location of branches relative to customers. Shaffer (1993) considers *hedonic attributes* as qualitatively determined differences in bank operations which include differing funding strategies, deposit mix, average deposit size, off-balance sheet activities, asset quality, target clientele, asset mix and average loan size. Thus the potential hedonic attributes that could be considered are both numerous and diverse in form. Overall, in previous studies, most attributes considered are related to the type of service provided, the provision of liquidity by a product, observable by the provision of withdrawal services, and the terms and conditions of the product. Such attributes should have a noticeable effect on the costs of operation for depository financial institutions.

Fried *et al* (1993) in a study of US credit unions provides a good example of this modelling approach. The hedonic attributes used in this study include the number and level of services provided. Proxies are used to measure components of quantity, price and variety of service relating to both loan and deposit services. Variety aspects are quantified through consideration of a range of factors. Other attributes used in similar studies have included the number of loan or deposit products offered by individual institutions, the number of products or services offered or combinations of the above. Nelson (1985) includes the costs of transportation for the customer to the nearest branch, the time taken

to cash and clear cheques and the amount of information required for customer decision-making. Such aspects of production may affect the true cost function of the institution. Branch level cost functions are defined to amend for these potential difficulties. This approach can be seen to be increasingly applicable for output definition for cost functions in depository financial institutions.

3.12 Problems with accounting data

The information taken from company accounts, as used in this thesis and most non-US studies, poses certain difficulties. The changing regulation of accounting practice and difficulties in the translation of 'true and fair' may reduce the accuracy, objectivity and consistency of data. Other problems could include the shifting data format and regulations affecting information disclosure. More generally, the form of data produced by accountants places its primary emphasis on bookable, tangible, concrete historical values. This emphasis may conflict with the economic interpretations of models employing this data. Annual accounts may place a greater degree of emphasis on the legal, as opposed to economic, position of the firm, often for 'cosmetic' reasons. Such information is produced according to regulatory requirements and as a method of quantifying explicit, verifiable, market transactions. This formulation of value is divergent from the broader conception of value accorded and desired by economists who require information on both explicit and implicit market transactions. The emphasis on historical values as opposed to opportunity cost and market values of activities and resources could possibly produce problems.

3.13 Conclusions

In this chapter, basic model forms and variables applied in previous studies have been reviewed. The discussion of the different approaches to modelling bank production is central to providing a broader understanding of efficiency and productivity. The production and intermediation approaches appear to represent most clearly, an economic interpretation of production. Both these approaches provide a relevant economic and distinct representation of production. As no substantial advantage may be seen in employing one of the model approaches in isolation, both will be employed in the empirical study of economies of scale and product mix in UK retail banking (chapter 7). By determining the level of fit and compliance with established cost and production theory, a choice between these models will be made for use in other empirical studies, acknowledging the limited breadth of these criteria for choosing between models. Drawing from some of the other approaches outlined, risk and liquidity are included in some of the empirical chapters to investigate their potential impact on the depository institution production process.

A measure of the price of capital should quantify how much it costs a depository institution to employ an amount of capital in the production process. A number of issues surround this definition. First, why should the price of capital vary between institutions and secondly, what are the factors that determine this variation, if it exists. A third issue is what would be an appropriate model for representing differences in capital prices in the UK.

The UK is quite small geographically and fairly homogeneous in terms of prices. Whilst it could be assumed that a constant value for the price of capital could exist for such a market, a number of factors could influence capital price. Small but noticeable variation in capital prices exists in the UK. For example, land and building prices for a building society located in the South East, a prosperous and highly populated area, would be expected to be higher than those in mid-Wales, a sparsely populated area with an emphasis on primary production. The cost of funding capital projects may also vary between institutions where larger institutions may be able to gain better terms for financing capital investments. A degree of capital price variation is therefore expected.

The derivation of an appropriate model for representing the price of capital raises a number of issues. What factors should be included in such a model? Will the estimated price of capital represent only this price and not, for example, display the variation in the efficiency of capital use across the industry? Additionally, an optimal model would consider the price of capital as a current economic cost or opportunity cost, as opposed to a historical accounting cost. To conclude, the use of appropriate and developed models to estimate the price of capital, for models of production in depository institutions, has been limited. The study by Glass and McKillop (1992), which considered a user cost measure of capital, is one of the few studies to consider the capital price on the lines outlined. In common with the more advanced methods of deposit pricing, such as the user cost methods outlined by Barnett (1978), this approach emphasises the opportunity cost of capital.

The price of labour has been represented in most other studies of depository institutions as the total labour cost divided by the number of employees. This measure may misrepresent the 'true' level of labour price amongst depository institutions and through considering the aggregate cost of labour removes any consideration of the distribution of wages amongst employees. This approach thus removes any analysis of who is being paid what to do a certain task. Such a generalisation may again lead to the possible mis-specification in the final model as a blunt definition of labour price may pick up what could be deemed inefficiency if a 'better' or alternative proxy of labour price was employed.

A problem with all price proxies that use a measure of assets or institutional size is that price may be affected by the size of the institutions considered. As the size rises, the denominator grows and the proxied price may fall, resulting in possible mis-specification. Lower prices for larger firms are a particular problem in the analysis of holding companies. While the entire holding company may be required for analysis to reduce local mis-specification of efficiency, the additional activities undertaken by a larger firm may skew the price measures.

Another general problem that presents difficulties in the definition of variables has been the rapid expansion in the number of products offered by depository institutions. As re-regulation has continued throughout the 1990s, depository financial institutions in many countries have offered an expanded range of services. This phenomenon of increasingly

differentiated products produced by multi-product depository financial institutions has been deemed to be of importance to the US Congress. The Board of Governors of the Federal Reserve system is required to report annually on the availability of retail banking services and fees provided and charged by S & L's and commercial banks (US Congress, 1989, *Financial Institutions Reform, Recovery and Enforcement Act*, section 1002). Analysis of these reports (Hannan, 1994) has suggested that not only do the services provided and fees charged vary significantly between S & L's and commercial banks. Significant variation also occurs between individual S & L's and commercial banks in the range of services provided.

Chapter 4 Previous studies of efficiency and productivity in depository institutions

4.1 Introduction

This chapter reviews a number of empirical studies of cost efficiency, technical change and productivity in depository institutions. The chapter is closely linked with chapters 3 and 5. In chapter 3 the model forms referred to in this chapter are considered in more depth. In chapter 5, many of the approaches and assumptions made by the studies reviewed here are outlined in detail. In this chapter, previous studies on the characteristics of efficiency and measures of productivity and technical change are reviewed. Data sets, methods of estimation and model specifications are surveyed and the empirical evidence is assessed in light of the approaches and methods used.

The chapter is structured in the following manner. Section 4.2 reviews studies of the UK building society sector. Sections 4.3, 4.4 and 4.5 outline the principal studies of non-US banks, savings and loan associations (S & L's) and credit unions respectively. Studies of US banking, which provide the bulk of the literature on depository institutions, are reviewed in section 4.6. A summary of the chapter and conclusions are presented in section 4.7.

4.2 Building society studies

The building society studies included in this survey are all from the UK and Australia and, as with most studies of different countries, variations in the regulatory systems make

direct comparisons of the results difficult. The main function of building societies in both countries is to smooth the passage of funds from household savers to borrowers who wish to acquire long-term mortgage loans for house purchase secured on property. In the UK, since the Building Societies Act (1986), some of the mutual building societies have been engaging in new activities, such as money transmission services, unsecured lending, and estate agency and insurance business, and have been allowed to raise a proportion of their funds from wholesale sources. As discussed in chapter 2, some UK societies have also taken up the option provided for in the 1986 Act to convert to PLC status and so become banks. Indeed, the industry has been characterised by two interrelated structural developments: a decline in the number of societies (through mergers and conversions) and an increase in the degree of market concentration. These changes have acted as a spur to research into the efficiency of UK building societies. In Australia, research into the performance of permanent building societies has been undertaken under quite different industrial conditions. During the 1970s permanent building societies operated within a fairly relaxed state-based regulatory system. These regulatory conditions placed permanent building societies in an advantageous position relative to Australian banks, which were subject to quite strict central bank control. Re-regulation of banking in Australia in the early 1980s helped to eliminate this advantage. This process of equalising the regulatory conditions of Australian banks and building societies continued with stricter building society regulations introduced in 1990.

Of the UK studies, a number of approaches have been used to estimate inefficiencies. These include simple linear regression (Gough, 1979); the estimation of a translog cost

function and its variants (Hardwick, 1989, 1990, Drake, 1992 and McKillop and Glass, 1994); non-parametric DEA techniques (Field, 1990, Drake and Weyman-Jones, 1992 and Piesse and Townsend, 1995) and a comparison of *DEA* and stochastic frontier techniques (Drake and Weyman-Jones, 1996). Data were obtained for all of these studies principally from annual reports and accounts, with some additional statistics provided by the Building Societies Association and the Building Societies Commission. In general, variations in the datasets and the models used have led to a mixed set of results.

An early UK study was undertaken by Gough (1979) who used data for 1972 and 1977 and found no significant relationship between assets and average management expenses to test for evidence of economies of scale. However, Hardwick (1989, 1990) did find evidence of economies of scale. Using an intermediation model with a single output variable, Hardwick (1989) used a sample of 97 building societies in 1985 to estimate a translog cost function jointly with derived input cost share equations and found significant 'augmented' economies of scale for societies with assets of less than £280 million and significant augmented diseconomies of scale for societies with assets in excess of £1,500 million. In a later paper, but using the same dataset, Hardwick (1990) undertook the first multiproduct study of UK building societies, using the production approach and estimating a translog cost function with two outputs (measured by the number of outstanding mortgage accounts and the number of outstanding share and deposit accounts) and two inputs (labour and capital). This time, significant overall economies of scale were found for societies with assets up to £5,500 million with constant costs for larger societies. Hardwick (1990) also found evidence of product-

specific economies of scale for both outputs and diseconomies of scope for societies with assets up to £1,500 million. In both studies, Hardwick found that economies attributable to the employment of labour were greater than those attributable to the employment of any other input.

Drake (1992) also estimated a multiproduct translog cost function for 76 building societies with 1988 data using an intermediation approach. He found only small economies of scale for one asset group (societies with assets of between £120 million and £500 million) and constant returns to scale elsewhere. Further, he found no evidence of economies of scope. Using the same data, Drake (1995) developed an expense preference model and this time found no evidence of either economies of scale or scope in the industry. Clearly, the question of whether economies of scale and scope exist in the UK building society industry is an issue still to be resolved.

McKillop and Glass (1994) estimated a hybrid translog cost function with an intermediation approach. The data, used in the study was taken from annual reports for 89 building societies in 1991. The average cost function estimated suggested constant economies of scale and diseconomies of scope were present. McKillop and Glass also developed their cost model to test for the impact of the geographic focus of branch networks on economies of scale and scope. A variable was employed to represent the differing geographical focus in operations of individual building societies. The building society sample was subdivided into three groups, which had either a national, regional or local branch network. It was reported that the geographic focus of the branch network has

a significant influence on both the economies of scale and economies of scope of UK building societies.

The three DEA studies of UK building societies (Field, 1990, Drake and Weyman-Jones, 1992, and Piesse and Townsend, 1995) all used separate datasets and produced very different results. Field used a sample of 205 building societies in 1981 in a production model and estimated that only 14 per cent of the societies in the sample were productively efficient: it was concluded that the disparities in inefficiency levels between societies were attributable to differing levels of managerial incompetence, rather than scale. Drake and Weyman-Jones used the same sample as Drake (1992, 1995) in an intermediation model and found that about 37 per cent of the societies were efficient, with about 61 per cent exhibiting pure technical efficiency and 41 per cent exhibiting scale efficiency. Piesse and Townsend constructed five separate models with different objective functions for a sample of 57 building societies in 1992. In their first model, which followed the production approach (unusually using profit as a single output measure), only six societies were on the efficient frontier and for those not on the frontier, the major factor was scale inefficiency: 77 per cent of the sample were found to be operating with decreasing returns to scale and 14 per cent with increasing returns. The other models followed the intermediation approach, but differed in respect of the included output and input variables. In these models, a much higher proportion of the societies were found to be both technically and scale efficient.

Drake and Weyman-Jones (1996) used both, non-parametric *DEA* techniques and a translog stochastic frontier approach to estimate scale, technical and allocative inefficiencies. They used 1988 data for 46 UK building societies and employed an intermediation model. Using the non-parametric *DEA* technique, they found considerable variability in inefficiencies across the building societies in the sample, with allocative inefficiency dominating technical and scale inefficiencies. The cost frontier results, on the other hand, suggested very little allocative or technical inefficiency and constant returns to scale at the mean. However, the efficiency ranking provided by the two approaches were found to be remarkably similar.

Of the three Australian studies of building societies, Crapp (1982) estimated a series of separate Cobb-Douglas (log-linear) cost functions for each year from 1976 to 1980. The more recent study by Esho and Sharpe (1995) used pooled time-series and cross-section data (for 20 building societies from 1974 to 1990) to estimate a dynamic multiproduct translog cost function with an associated input share equation. Worthington (1998) used a stochastic frontier model of building society production.

Crapp found evidence of economies of scale in New South Wales permanent building societies, but observed marginal decreases in the magnitude of the scale economies over the period of the study. Esho and Sharpe's pooled model had the advantage that it allowed technical change to be incorporated into the analysis. They estimated a dynamic cost function, which allowed the adjustment costs incurred in moving from one equilibrium position to another, to be taken into account. Their results indicated that the

adjustment of costs to their long-run levels was in fact quite fast, with over 80 per cent of the total adjustment occurring in the first period. Finally, they found evidence of small economies of scale for small building societies and diseconomies of scale for large societies.

Worthington (1998) estimated a translog stochastic frontier for 22 Australian building societies between 1992 and 1995. Quarterly data from the Australian Financial Institutions Commission was employed with an intermediation model specification. A truncated normal distribution of firm-specific inefficiency is assumed. Six explanatory variables were included in the model, including total assets, total capital, the number of branches and agencies, a time trend dummy and total commercial loans held. Average inefficiency was shown to decline throughout the sample period from 0.239 in 1993 to 0.120 in 1995. Worthington suggested that both the number of branches and agencies and increased asset size had a substantial effect on the level of prevailing inefficiency

Some preliminary conclusions may be drawn from this review of building society studies. It is suggested that the model form employed in efficiency studies appears to have a substantial effect on the results estimated. The sample period also appears to affect the estimates. Studies employing cross-sections of data from different years yield different sets of results, as would be expected in a developing commercial sector. Finally, the size and market share of building societies may have an effect in the determination of efficiency characteristics.

4.3 *Non-US bank studies*

Cost efficiency studies of non-US banks include studies of Australia (Edgar *et al*, 1971); Israel (Kim and Weiss, 1989); Japan (Tachibanaki *et al*, 1991; Fukuyama, 1993; and McKillop *et al*, 1994); Ireland (Glass and McKillop, 1991 and Lucey, 1993); France (Dietsche, 1993); the Nordic countries (Berg *et al*, 1993); Germany (Lang and Welzel, 1996); Italy (Resti, 1997, Favero and Papi, 1995); Canada (Nathan and Neave, 1992); Greece (Pavloppoulos and Kouzelis, 1990, Karafolas and Mantakis, 1996); Finland, (Kolari and Zardkoohi, 1992, Zardkoohi and Kolari, 1994); Allen and Rai (1996) estimated a global cost function from banks operating in fifteen countries; and a comparison of European countries was provided by Altunbas and Molyneux (1996). The majority of the researchers have estimated multiproduct translog cost functions. Exceptions are Edgar *et al* (1971) who estimated a Cobb-Douglas cost function, Berg *et al* (1993), Fukuyama (1993) who used a non-parametric *DEA* technique, Lucey (1993) who estimated a translog profit function, and McKillop *et al* (1994) who estimated a composite cost function.

For Australia, Edgar *et al* (1971) estimated a single-product Cobb-Douglas cost function using a sample of 8 banks over the period 1947-68 and found evidence of economies of scale for seven of the banks. In a later study for Israel, Kim and Weiss (1989) estimated a multiproduct translog cost function for 17 banks between 1979 and 1982 and found evidence of large and statistically significant economies of scale for both small and large

banks: these economies were considered to be the main causes of total factor productivity growth observed during the period.

For Japan, the three studies, that were reviewed, gave similar results. Using a sample of 61 banks between 1985 and 1987, Tachibanaki *et al* (1991) estimated a two-output translog cost function and found evidence of economies of scale for all banks in all three years, but evidence of cost complementarities only in 1986 and 1987. Fukuyama (1993) employed an intermediation model, in a *DEA* study of 145 banks, found average technical inefficiency equal to 0.86, and evidence of slight economies of scale but no evidence of economies of scope. McKillop *et al* (1994) used a composite cost function model, originally suggested by Pulley and Braunstein (1992), which combined a log-quadratic input price structure with a quadratic output structure. Following an 'intermediation' approach in a three-output, three-input model, they found evidence of economies of scale, but no evidence of economies or diseconomies of scope.

For Ireland, Glass and McKillop (1991) used a multiproduct translog model with two outputs and two inputs for the period 1972-90. They found no evidence of economies of scale, and while diseconomies of scope were reported in the earlier part of the time period, significant economies of scope were found in the late 1980s. Lucey (1993) used 17 Irish banks over the period 1988-91 to estimate a profit function to measure technical and allocative inefficiency. Technical inefficiency was reported to be 0.83 or 83 per cent, dominating allocative inefficiency which was estimated to be 0.18 or 18 per cent. Lucey also reported evidence of economies of scope.

For France, Dietsche (1993) estimated a four-output, three-input translog model using data for 345 French banks in 1987. Evidence of economies of scale for both small and large banks, and cost complementarities between deposits and loans, long-term investments and inter-bank activities, and loans and inter-bank activities were reported.

For the Nordic countries of Sweden, Finland and Norway, Berg *et al* (1993) used samples of 126 Swedish, 503 Finnish and 130 Norwegian banks in a non-parametric *DEA* study. The authors compared the relative efficiency of the three countries' banks using two models, one with constant returns to scale and one with variable returns to scale. In both models, they found that most banks in Finland were relatively inefficient, but they reported similar efficiency results for Swedish and Norwegian banks. Of the banks on the efficiency frontier, the greatest proportion were resident in Sweden.

The study of German co-operative banks by Lang and Welzel (1996) used both fixed and random effects one-way component panel data models (panel data models are considered in more detail in chapters 5, 7 and 9). A translog functional form specification of productive technology is applied with a sample of 757 Bavarian co-operative banks from 1989 to 1992. Statistically significant economies of scale and product mix are reported for both the fixed and random effects models. Product mix economies were reported with both economies of scope and expansion-path sub-additivity measures. Negative technical change and significant cost efficiencies are observed over the sample period.

Resti (1997) used a panel sample of 270 Italian banks to estimate cost efficiency with both econometric and non-parametric methods. A production model was employed. Resti suggested that inefficiency in Italian banks may exist at levels of approximately 68 per cent for a non-parametric constant returns model and 75 per cent with a non-parametric variable returns model. The econometric model also indicated that a high level of inefficiency existed. These levels of efficiency reduced slightly over time (from 69.4 per cent in 1988 to 69.8 per cent in 1992), across geographic area (the highest and lowest efficiency scores are recorded for the north-west and centre of Italy respectively) and across asset size. A high correlation existed between the results from the econometric and non-parametric models. Favoro and Papi (1995) used a non-parametric DEA approach with a number of models, including production and intermediation specifications, to analyse a sample of 174 Italian banks from 1991. Large banks were seen to be more efficient than smaller banks, whilst banks from southern Italy were less efficient than those from outside this area.

Nathan and Neave (1992) used a translog cost function to analyse economies of scale and cost complementarities with both production and an intermediation models. Data was used from the Federal Office of the Superintendent of Financial Institutions on approximately 60 Canadian banks was used for the sample period 1983 - 87. Slight economies of scale were found with the production model and constant economies were estimated using the intermediation model of bank production. It was suggested that these findings implied that “ ... Canada’s concentrated banking system exploits and exhausts available sources of scale economies and cost complementarities. Hence mergers and

acquisitions that lead to concentrated financial systems may not imply any substantial increases in costs” (p. 272).

Pavloppoulos and Kouzelis (1990) employed a translog cost function approach to estimate long-run marginal costs of deposit and loan production for a single Greek bank. Branch level data was used (data for 362 branches) from 1983. Following the techniques forwarded by Mullineaux (1978), U-shaped marginal cost curves were estimated for both loan and deposit production. Karafolas and Mantakis (1996), using a time-trend translog model, considered a sample of 11 Greek banks between 1980 and 1989. Significant economies of scale were reported with a production approach. The levels of technical change over the sample period were positive, yet statistically insignificant.

Kolari and Zardkoohi (1992) used an intermediation model with a reduced¹ translog cost function model to assess the efficiency characteristics of Finnish co-operative and savings banks. Sample data for 1983-84 was employed for 369 co-operative and 255 savings banks. These results suggested that cost curves tend to be L-shaped at the plant or branch level and U-shaped at the firm level. This finding was considered to be a result of “... the widespread use of branch facilities (as a method to) pay an implicit service return” (p. 450) when the financial sector was tightly regulated. Diseconomies of scope in the joint production of bill and advances were also suggested. Zardkoohi and Kolari (1994) estimated economies of scale and scope for 615 bank branches from 34 Finnish savings banks for 1988. An intermediation model with a translog functional form was used. It was reported that relatively large bank branches tended to be more efficient than

smaller bank branches and cost efficiency was improved by membership of a larger as opposed to smaller branch network. It was further suggested that economies of scope were not significant at the branch level.

Using a sample of 194 banks from 15 countries Allen and Rai (1996) estimated a global cost function with a translog cost function specification of an intermediation approach. Significant dis-economies of scale were found for the larger banks, whilst smaller banks displayed economies of scale. Average input efficiency was observed to exist at 27.5 per cent and average cost efficiency was recorded at 15 per cent of total costs. The global cost function was estimated with a fixed-effects panel data model.

Finally, in a comparative study of European banking, Altunbas and Molyneux (1996) estimated translog cost functions in France, Germany, Italy and Spain. They employed the intermediation approach and, for each country, included two output variables (total loans and total securities) and two input prices. Unfortunately, the same input price variables were not available for all four countries so that the comparisons made may not be valid. They found some evidence of economies of scale in all four countries (though these were not statistically significant for Spain and Germany) and mixed evidence of economies of scope. Clearly, more research is called for in comparing cost inefficiencies among countries.

4.4 *Studies of Savings and Loan Associations*

A large number of cost efficiency studies of saving and loan associations (S & L's or thrifts) have been undertaken in the USA since the late 1970s. S & L's are financial institutions that traditionally have provided finance for residential loans in addition to offering investment and deposit accounts. They have taken on both mutual and proprietary (stock) ownership forms with a large number of conversions from the mutual to the stock form since 1980. The early studies of S & L's were motivated mainly by the large number of mergers in the industry during the 1970s and early 1980s, while later research has been largely a response to the serious crisis in the industry in the later 1980s precipitated by financial re-regulation, a crisis described by many commentators as the 'S & L debacle' and by Hermalin and Wallace (1994) as a 'mass extinction'. It was a period during which over 600 thrifts failed, over \$300 billion in real estate investments had to be liquidated and over \$110 billion was paid to bail out the S & L insurers, the Federal Home Loan Bank Board. The causes and serious effects of the US S & L deregulation experience are discussed fully in White (1991).

Cost studies in this area have all used similar data sources: the quarterly reports from the Federal Home Loan Bank Board and semi-annual reports from the S & L's, with additional information from the Office of Thrift. A variety of approaches for estimating efficiency have been adopted. These include a single-product linear regression model (Brigham, 1964); a Cobb-Douglas cost function (Benston, 1972); a single-product translog cost function (McNulty, 1982 and Dowling and Philippatos, 1990); a multi-

product translog cost function (Mester, 1987, 1991 and Le Compte and Smith, 1990); a stochastic cost frontier approach (Mester, 1993); and a multi-product, non-parametric approach (Hermalin and Wallace, 1994).

One of the first studies of economies of scale in S & L's was provided by Brigham (1964). Brigham used linear regression techniques to estimate a cost model for a number of individual data cross-sections for the sample period, 1956-62. The dependent variable was average cost (measured as total operating costs divided by average assets) and the main independent variables were the asset size of the institutions and their annual rate of growth of assets. Brigham concluded the S & L industry did not appear to be subject to very significant economies of scale. Benston (1972), on the other hand, did find evidence of scale economies in the industry. He used data for a sample of 3,159 S & L's in California from 1962 to 1966 to estimate a Cobb-Douglas cost function, using three separate output measures (the average number of loans serviced each year, the number of loans made, and the average number of service accounts serviced). He calculated the average elasticity of total expenses with respect to the three output measures over the five years to be 0.923, implying that, on average, a 10 per cent increase in asset size would raise costs by 9.23 per cent.

McNulty (1982) and Dowling and Philippatos (1990) both estimated single-product translog cost functions using total assets as a measure of output. McNulty used a sample of 117 S & L's in Florida in 1980 and found significant economies of scale for associations with less than \$500 million in assets, confirming his findings reported in an

earlier study (see McNulty, 1982). Dowling and Philippatos used cross-section data for each of the years of the period 1973-83 and also found evidence of cost reductions through asset expansion for associations up to a certain size: this 'minimum efficient scale' was estimated to have declined from about \$2,500 million in 1979 to about \$2,000 million in 1983.

Mester (1987) was the first researcher (to the author's knowledge) to estimate a multiproduct translog cost function for S & L's. Data for 149 associations operating in California in 1982 were used to estimate an 'intermediation' model, which had three outputs (mortgage loans, other loans, and cash, securities and real estate investments) and four inputs (labour, capital, demand deposits and term deposits). A growth variable and the number of branch offices were also included as explanatory variables. The research found only small cost reductions were possible from increasing the scale of operation and no evidence was provided concerning economies or diseconomies of scope. However, in a later study using the same dataset, but running separate regressions for stock and mutual companies in a two-output, three-input translog model, Mester (1991) found evidence of significant economies of scale. She also found that while the stock S & L's tended to operate with an efficient output mix, mutuals operated with significant diseconomies of scope. Le Compte and Smith (1990) estimated both a multiproduct translog cost function and a multiproduct miniflex Laurent translog cost function to investigate scale economies and cost complementarities. Using data for 431 S & L's in 1978 and 1983, they found evidence of economies of scale for small associations and diseconomies of scale for large associations. They also found evidence of significant cost

complementarities for each pair of outputs (mortgage loans, consumer loans and investments) in 1978, but no significant cost complementarities in 1983.

More recent studies have investigated general inefficiency in S & L's. Mester (1993) employed the stochastic cost frontier approach, and estimated separate translog cost frontiers for 807 mutual and 208 stock S & L's in 1991. The model had three outputs (mortgage loans, other loans and investments) and three inputs (labour, capital and deposits). The results suggested nearly constant returns to scale in the industry, but significant economies of scope between the three outputs for both types of organisational form. The measures of inefficiency indicated that stock associations were on average less efficient than mutuals.

Hermalin and Wallace (1994) provided a significant development to the literature through considering the both the measurement of inefficiency and the sources of inefficiency and insolvency. A WAPM (weak axiom of profit maximisation), non-parametric methodology from Varian (1985) is used for efficiency measurement. All data is taken from semi-annual reports for 1982 and quarterly financial reports for 1986 to 1988. In this study, the ownership form of the S & L (either mutual or proprietary), the amount of foreclosed property held by the S & L, the area of operation, the institutions current activities and total assets are considered to be potentially determinants of efficiency. It was found that the ownership of the S & L's (whether the association was mutual or proprietary) and the portfolio of assets, held by the S & L were influential in

the determination of efficiency. Hermalin and Wallace estimated an average saving and loans association is operating nearly thirty per cent below the efficient level.

Gropper (1995) considered economies of scale and economies of scope for S & L's in 1988. A generalised quadratic flexible functional form is applied with an intermediation model specification. Gropper found small, yet statistically significant economies of scale and scope for the S & L's in 1988.

Finally, in a recent study of efficiency characteristics in S & L's, Stiroh (1997) adopted an intermediation approach, estimated a translog cost function, including a time trend variable, to provide estimates of economies of scale, economies of scope and technical change. Data was taken from financial reports for 899 S & L's in the eastern US states between 1990 and 1995. Significant technical change was reported in the early 1990s, until a sudden halt in 1993. This was deemed to be the result of substantial association defaults. Constant economies of scale were reported for S & L's over the sample period. Economies of scope appeared to exist in the joint production of mortgage and consumer loans, although their magnitude reduced over the sample period.

To conclude, the analysis of the efficiency characteristics of the savings and loans association industry has employed a number of approaches that have been influential in considering the broader literature considering depository institutions. The study by Hermalin and Wallace (1994) on the determinants of inefficiency and insolvency extended both the breadth and focus of the literature. A number of later studies (e.g. the

analysis of US commercial banks by Berger and Mester, 1997) have expanded on this general approach. The use of distinct 'Diewert' functional forms by Le Compte and Smith (1990) showed that the choice of functional form may be another aspect of the modelling process that can lead to distinct efficiency estimates. Preliminary conclusions are generally supported by the review of S & L studies.

4.5 *Credit union studies*

Cost inefficiency studies of credit unions are complicated by the peculiarity of this type of depository institution. Credit unions are often established by employers or friendly societies, with the aim of providing a cheap source of funds and depository services for specific groups, such as employees or residents within a particular geographical area. Basically, they are non-profit-maximising co-operatives whose members are both lenders (*i.e.* shareholders) and borrowers. The co-operative nature of credit unions leads to a possible conflict between new and existing members as the influx of new members can affect the return to existing members in various ways. For example, extra saving by new members may cause the dividend rate to fall; similarly, extra demand for loans by new members may cause the loan rate to rise. This potential conflict between new and existing members would be lessened by the existence of economies of scale and worsened by the existence of diseconomies of scale. This was one of the principal motives for investigating scale economies in credit unions.

The credit union studies reviewed concern the US, Canadian and UK credit union sectors. Five approaches of measuring efficiency have been used in the reviewed studies including a multiple regression model applied by Taylor (1972), a Cobb-Douglas cost functions used by Koot (1979) and Wolken and Navratil (1980), a translog cost function employed by Murray and White (1983) and Kim (1986), a non-parametric approach adopted by Fried *et al* (1993) and a paired difference analysis used by McKillop *et al* (1995).

The earliest studies (by Taylor, 1972, Flannery, 1974, Koot, 1979 and Wolken and Navratil, 1980) mainly concerned credit unions in the USA. Taylor (1972) concentrated on large credit unions (those with assets over \$1 million) and developed a model in which average cost (defined as total cost divided by the total value of assets) depended on the squared logarithm of the total value of assets, together with a number of other explanatory variables intended mainly to capture the influence on costs of subsidies and the average size of shares and loans. The set of explanatory variables, however, did not include factor prices. Taylor's results indicated the presence of significant economies of scale in the credit union industry. But when Flannery (1974) estimated a number of cost functions for credit unions (not just large ones), he found evidence of decreasing returns to scale. However, Flannery's cost functions did not make any allowance for the subsidies provided to credit unions (mostly in the form of voluntary labour services from members and donated office space and equipment from sponsors), and Flannery himself believed that once these were taken into account, it would be likely that increasing

returns to scale would be found. This would follow, he argued, because in general smaller unions benefit more from subsidies than larger unions.

Koot (1979) estimated a log-linear cost function in an attempt to test this hypothesis. A sample of 380 credit unions from all size groups and four alternative physical output measures (the number of loans granted during the year, the number of outstanding loan accounts, the number of outstanding share accounts and the number of outstanding loan *and* share accounts) was used. In spite of including subsidies in the analysis (by including subsidies provided by sponsors in the definition of total cost and adding the annual number of hours supplied by voluntary workers as an explanatory variable), Koot obtained essentially the same result as Flannery, *i.e.* strong evidence of decreasing returns to scale.

Wolken and Navratil (1980) criticised the methods employed by Koot and Flannery and showed that, when an input price was included as an independent variable and when all subsidies were included in the measurement of total cost, the empirical evidence did support the hypothesis of increasing returns to scale in the credit union industry. The main lessons to be learned from these early credit union studies include the desirability of using as representative a sample as possible and the importance of including all the relevant influences on cost (including input prices) as explanatory variables in the cost function.

A more recent study of US credit unions was undertaken by Fried *et al* (1993) who developed a model of credit unions in which the institutions were assumed to aim for maximum benefit for their members. These benefits were defined to include the saving and loan services which credit unions supply and, in each classification, three types of services were identified: the number of accounts, interest rates and 'convenience' features offered to lenders and borrowers. Thus, output consisted of six services altogether. In addition, two categories of resources (labour and other resources) were included in the model. Non-parametric, non-stochastic *DEA* techniques were used to construct the 'free disposal hull' of the data (*i.e.* the production possibilities set of the credit unions in the sample). This was then used to estimate the degree of productive inefficiency; it was found that, on average, inefficient credit unions were supplying about 20 per cent less in terms of services than the 'best practice' credit unions. Finally, a regression analysis was used to investigate whether the variations in inefficiency were related to twenty environment characteristics of credit unions. It was found that overall efficiency was generally higher for credit unions having common bonds, sponsorship, a state charter, no branch offices, a high investment to loan ratio and few real estate loans.

Studies of non-US credit unions include Murray and White (1983), Kim (1986) and McKillop *et al*, (1995) who analysed Canadian and UK credit unions. Murray and White (1983) estimated a translog cost equation with a multiproduct specification. The multiproduct specification enabled them to test for economies of scale, economies of scope and factor substitution. Output was represented by three variables: mortgage lending, non-mortgage lending and investments in excess of the minimum liquidity

requirement. The inputs were assumed to be capital, labour, demand deposits and term deposits and the prices of these four inputs were included as explanatory variables. The results reported by Murray and White were, first, that the Cobb-Douglas and CES functional forms were rejected in favour of the more general translog specification; secondly, that there was no evidence to support homotheticity, unitary elasticity of substitution or constant returns to scale in credit unions; thirdly, that there was evidence to support the existence of economies of scale for most of the credit unions in the sample; and finally, that there was strong evidence for economies of scope between mortgage and other lending activities.

Kim (1986) replicated Murray and White's study and additionally computed estimates of product-specific economies of scale. Kim's study confirmed Murray and White's findings of overall scale economies and found product-specific economies of scale associated with mortgage loans and investment activities, but product-specific diseconomies of scale associated with non-mortgage lending.

McKillop *et al* (1995) due to limited data availability used paired difference analysis to estimate whether economies of scale were prevalent among UK credit unions. Strong evidence of economies of scale among UK credit unions was reported. A sample of nearly 500 credit unions for 1992 was used, with data was taken from annual reports and accounts.

A consideration of the differing regulatory structures and histories of the US, Canadian and UK credit union sectors may assist in the explanation of national differences in results recorded. Other differences between studies are perhaps a function of the breadth of approaches, techniques, time periods and model specifications used within the individual studies reviewed.

4.6 *US banking studies*

Studies of economies of scale and scope and technical and allocative efficiencies in US commercial banking have been by far the most numerous and have provided most of the innovations in the area. In general, the studies tend to support the view that there are economies of scale for small banks and diseconomies of scale for large banks, but are not in such agreement concerning the existence of economies or diseconomies of scope. With regard to technical and allocative efficiencies, most studies agree that such inefficiencies exist and differ significantly between banks, but there is no agreement on the best technique for estimating these inefficiencies or the average level of banking inefficiency.

The principal data sets employed in the reviewed studies are the Functional Cost Analysis programme (*FCA*) and the Reports on Condition and Income Tapes, or call reports. The *FCA* data is produced by the Federal Reserve Bank of New York and provides voluntarily forwarded results on predominantly small and medium sized banks. The Reports on Condition and Income Tapes are produced for all banks and held by the Comptroller of the Currency.

These data sets through their emphasis have been associated with the two main model specifications: the production model and the intermediation model respectively. The production specification studies (e.g. the studies by Gilligan *et al*, 1984, Berger *et al*, 1987, Lawrence, 1989, Pulley and Braunstein, 1992, Ferrier *et al*, 1993, McAllister and McManus, 1993, and Jagtiani *et al*, 1995) have principally used the functional cost analysis programme data sets to emphasise the cost differences on a range of bank activities. Alternatively, the call reports emphasis the static and dynamic asset positions of banks and have been employed more with an intermediation specification (e.g. the studies by Berger *et al*, 1987, Noulas *et al*, 1990, Aly *et al*, 1990, Berger and Humphrey, 1991, Gropper, 1991, Shaffer and Edmond, 1991, Yue, 1992, English *et al*, 1993, Grabowski *et al*, 1993, Pi and Timme, 1993, Elyansiani *et al*, 1994, Karaparkis *et al*, 1994, Hunter and Timme, 1995, Elyasiani and Mehdian, 1995, Jagtiani *et al*, 1995, Mester, 1996, Millar and Noulas, 1996, Jagtiani and Khanhavit, 1996 and Mahajan *et al*, 1996).

Functional forms adopted in econometric US commercial bank studies have also varied greatly. Functional forms employed have included the Cobb-Douglas cost function (Benston, 1972, Longbrake and Haslem, 1975), the translog cost function (Flannery, 1983, Gilligan *et al*, 1984, Berger *et al*, 1987, Noulas *et al*, 1990, Gropper, 1991, Shaffer, 1993, Evanoff and Israilevich, 1994, Karaparkis *et al*, 1994, Hunter and Timme, 1995, Jagtiani *et al*, 1995 and Mester, 1996, Mahajan, *et al*, 1996, Jagtiani and Khanthavit, 1996, DeYoung, 1998), Box-Cox generalised functional forms (Clark, 1984, Kilbride *et al*, 1986, and Lawrence, 1989), and flexible Fourier functional forms (McAlister and

McManus, 1993, Mitchell and Onvural, 1996, Berger and DeYoung, 1997, Berger and Mester, 1997, DeYoung and Hasan, 1998). Non-parametric methods have been also adopted in a number of studies (Berger and Humphrey, 1991, Yue, 1992, English *et al*, 1993, Ferrier *et al*, 1993, Grabowski *et al*, 1993, Elyansiani *et al*, 1994, Elynsiani and Mehdian, 1995, and Millar *et al*, 1996). To summarise, difficulties in the comparison of studies are exaggerated by the use of differing model specifications, functional forms, differing sample periods, the scope of data both in quantity and area of analysis and a multitude of variants in the estimated functions.

Among the earliest studies were the classic papers of Benston (1965) and Bell and Murphy (1968) who used data from the *FCA* programme to estimate separate cost functions for the various operations of a bank, using Cobb-Douglas functional forms. The banking operations identified in the two studies were demand deposits, time deposits, mortgage loans, business loans and instalment loans. Both studies found evidence of small economies of scale for the main banking operations and higher costs for branch banking compared with unit banking. Longbrake and Haslem (1975) also estimated a Cobb-Douglas cost function, but for demand deposits only, using a three-component measure of output (consisting of the number of demand deposit accounts per branch office, the average size of account and the number of offices). Their main conclusions were that the number of offices had little effect on average costs, but that an increase in the number of demand deposits per office would lead to a decline in average costs for all banks.

Linear and log-linear cost functions have been criticised for being too restrictive, and it was for this reason that attention was directed in the 1970s to the development and application of flexible functional forms. By far the most popular of these in studies of US banking has been the translog cost function. One of the earliest applications of the translog function in a study of economies of scale in US banking was undertaken by Benston *et al* (1982). They used data from the FCA programme and applied a 'production' model to four separate years (1975-8) with cross-section data for a sample, which varied from 747 to 852 banks. Whilst five main bank services were identified (demand deposits, time and savings deposits, real estate loans, instalment loans, and commercial and industrial loans) only a single output variable was included (obtained by aggregating the outputs of the five main banking services using a Divisia index). A translog cost function was then estimated. No attempt was made to estimate the cost function simultaneously with the cost share equations. In defence of this decision, the authors referred to Monte Carlo experiments by Guilkey and Knox-Lovell (1980) which showed that while there are gains in efficiency through joint estimation, these gains are relatively small. In addition to calculating a measure of scale economies, Benston *et al* calculated an 'augmented' scale economy measure to make an allowance for the fact that branch banks are able to expand by opening new offices rather than by adding new customers and accounts to existing offices. Their results indicated that economies of scale were experienced by small branch state banks and diseconomies by large unit state banks. In both cases, either U-shaped or upward-sloping average cost curves were derived. The minimum cost size was achieved by bank offices with between \$10 million and \$25 million of deposits.

The study by Benston *et al* (1982) was followed in quick succession by Flannery (1983) and Gilligan *et al* (1984). Flannery used an approach similar to that of Benston *et al*, but included correspondent service costs (*i.e.* the costs incurred by banks in purchasing services from other banks) in the definition of total cost. Flannery obtained results similar to those of Benston *et al*, but concluded that studies which omit correspondent service costs tend to overestimate the extent of branch bank scale economies by a small but statistically significant amount. To the authors knowledge, Gilligan *et al* (1984) was the first published study to estimate a multiproduct translog cost function for US banks. In a two-output 'production' model with the number of deposit accounts and the number of loan accounts as the output measures, they found evidence of overall economies of scale for small banks and overall diseconomies of scale for larger banks. They also found significant positive economies of scope, but product-specific diseconomies of scale for both outputs.

Many other studies using the translog cost function followed. These included a single-product study of large banks by Hunter and Timme (1986), who found evidence of scale economies across a broad range of bank sizes, and a study by Noulas *et al* (1990) who used data for 308 large branch banks (with assets over \$1 billion) taken from the 1986 Call Report. Noulas *et al* estimated a four-output, four-input 'intermediation' translog cost function jointly with the derived share equations and found evidence of small economies of scale for banks with assets between \$1 billion and \$3 billion, and small diseconomies of scale for banks with assets between \$3 billion and \$6 billion. Thus, they

found no evidence based on cost analysis for the emergence of 'megabanks' in the USA. This conclusion was supported by Shaffer (1993) who used a fitted translog cost function to simulate mergers between pairs of large banks and found cost savings in only half of the cases considered. More recent studies which also used the translog functional form have considered the effects of different organisational forms on cost efficiency (Newman and Shrieves, 1993); the effects on economies of scale and scope of introducing quasi-fixed inputs (such as core deposits and physical capital) into the cost function (Hunter and Timme, 1995); and the effects of introducing off-balance-sheet activities into the cost function (Jagtiani *et al*, 1995).

Some studies attempted to use generalised functional forms to test for the underlying technology. For example, Clark (1984) used Call Report data for the period 1972-77 for 1,205 banks to estimate a Box-Cox function and test the assumption of an underlying Cobb-Douglas technology. Clark was unable to reject the assumption of a Cobb-Douglas technology and found that economies of scale estimates were insensitive both to the choice of log-linear or generalised functional form, and to the choice of bank output measure. He reported evidence of small economies of scale for the industry. Kilbride *et al* (1986) also used Call Report data, but for a sample of 1,858 banks between 1979 and 1983. Using Clark's Box-Cox generalised model they found that they were able to reject the assumption of a Cobb-Douglas technology for samples of unit banks, independent banks and holding company affiliates. They reported evidence of economies of scale, which decreased as the level of output increased for all three samples. Lawrence (1989) estimated a more general version of the Box-Cox function in a multiproduct model,

which had the Cobb-Douglas and translog functions as special cases. He was also able to reject the Cobb-Douglas function, mainly for failing to capture cost complementarities in the multiproduct model, and he concluded that translog equations do provide an “ ... adequate fit of the banking cost data” (p. 377).

Following quickly on from these studies of economies of scale and scope in US banks were a number of econometric studies of technical and allocative inefficiencies. These studies also tended to use translog functional forms in estimating stochastic cost frontiers. Possibly the most important of these studies were Berger and Humphrey (1991) and Kaparakis *et al* (1994). Berger and Humphrey (1991) developed a ‘thick frontier’ approach. Using the authors’ words, “ ... the thick frontier is estimated using data from banks in the lowest average cost quartile and then is compared to an estimated cost function for banks in the highest average cost quartile” (p. 118). Differences in fitted costs which cannot be explained by the exogenous variables in the model are assumed to represent the ‘inefficiency residual’. This residual consists of both technical and allocative inefficiencies and in separate ‘intermediation’ studies of 7,653 branch banks and 6,298 unit banks, they found that these inefficiencies were approximately 25 per cent of average costs and dominated the scale and product mix effects. Kaparakis *et al* (1994) estimated an ‘intermediation’ model with four outputs, four variable inputs and one quasi-fixed input, using data for 5,548 banks taken from the 1986 Call Report. A translog cost frontier was specified with a composite error term in which the one-sided inefficiency disturbance was assumed to have a truncated normal distribution. They found that banks tended to become less efficient with increasing size: for example, for

banks with over \$10 billion in assets, average inefficiency was 17 per cent, which was 70 per cent higher than the average of 9.8 per cent for all banks in the sample.

In spite of its wide use, the translog functional form does have serious drawbacks. These are principally concerned with its failure to be well behaved globally. These potential problems, which are discussed by McAllister and McManus (1993) and within chapter 5, have undoubtedly encouraged bank efficiency studies based on non-parametric DEA techniques. One of the first non-parametric programming studies of US banks was undertaken by Aly *et al* (1990) who used a five-output, three-input 'intermediation' approach with data from the Call Reports for 322 banks. They found " ... relatively low levels of overall efficiency" (p. 218), with most inefficiency being technical, rather than allocative or scale. This result, that technical inefficiency dominates scale inefficiency, helped to confirm the findings of Berger and Humphrey (1991) and was also reported by Ferrier *et al* (1993) who estimated " ... a large degree of technical inefficiency, a lesser degree of scale inefficiency, and diseconomies of diversification" (p. 248).

English *et al* (1993) employed a sample of 473 small banks in 1982 to estimate a translog output distance function using linear programming techniques and found evidence of " ... considerable technical inefficiency" (p. 363). Comparisons of the relative inefficiencies of small and large banks using DEA were undertaken by Elyasiani and Mehdiian (1995) and Miller and Noulas (1996). Miller and Noulas found that larger banks had higher levels of technical efficiency in the period 1984-90. Elyasiani and Mehdiian also found that larger banks were more efficient using 1986 data, but reported no significant

efficiency differences between small and large banks using 1979 data. Thus, it seems that larger banks may have become more efficient during the 1980s.

More recently, Mahajan *et al* (1996) quantified both radial and non-radial measures of economies of scale. The study attempted to ascertain whether there were differences in terms of cost inefficiency and economies of scale and non-radial expansion-path sub-additivity between multinational and domestic banks. A pooled data set of 1987 to 1990 of US multinational and domestic banks was considered. Significant diseconomies of scale were found for domestic and multinational banks for all asset classes at plant level. At the firm level, domestic banks exhibited economies of scale for all but the smallest banks. Multinational banks exhibited diseconomies of scale, increasing with their asset size at the firm level. Domestic and multinational banks were therefore viewed as having significantly different cost functions. Mahajan *et al*, (1996) found that both multinational and domestic banks had negative expansion-path sub-additivity. These diseconomies seemed to depreciate with increases in the size of the institution. Mahajan *et al* (1996) further found that allocative inefficiencies seemed to dominate both sample sets. The efficiency of multinational banks appears to be greater than domestic banks.

Jagtiani and Khanthavit (1996) used a sample of 120 large US banks with significant off-balance sheet activities over the period 1984-1990. A translog cost function specification of an intermediation approach was used to measure radial and non-radial economies of scale and the change in radial and non-radial economies of scale. The change in economies of scale was viewed as a measure of technical change over the sample period.

Overall, significant structural change was identified. Diseconomies of scale seemed to exist for a range of banks in the latter part of the 1980s. The reaction to such a change was not to reduce output as would be expected. Jagtiani and Khanthavit (1996) find significant cost complementarities between off-balance sheet activities and balance sheet activities. Cost complementarities between individual balance sheet activities were not evident. This result contradicted previous work by Gilligan *et al* (1984). Jagtiani and Khanthavit (1996) measured both expansion-path sub-additivity and the change in expansion-path sub-additivity with respect to time. Indications of economies of product mix were viewed to move within the sample period (the 1980s) from economies to diseconomies. Overall, a significant shift in the cost structure or productive technology was observed.

Using the translog functional form, Evanoff and Israilevich (1991) considered the difference between measures of scale efficiency and measures of scale elasticity used to proxy economies of scale. Data from 1987 for 164 banks held by holding companies and in the top 500 banks for the last 20 years was used. The holding company affiliation and the number of branches were assumed to be exogenously determined. Economies of scale (from elasticity measures) were found for smaller banks. Dis-economies of scale are observed after banks reached approximately \$3.3 billion in output. Evanoff and Israilevich suggested that a major under-estimation of potential scale inefficiency is possible whenever the elasticity measure differs from one. Further discussion of the differences between local and global measures of 'economies of scale' are contained in chapter 11.

Hunter and Timme (1995) incorporated aspects of both the equilibrium distribution of inputs (in which inputs were assumed to be fully utilised and variable) and a disequilibrium distribution of inputs (in which core deposits were assumed to be quasi-fixed inputs to the system and thus may not be fully used due to adjustment lags) in their model of banking production. The model therefore attempted to account for the dramatic changes occurring in this industry which might have resulted in over-supply or under-utilisation of particular resources. This approach was developed from a technique proposed by Flannery (1983). Call report data from 1983 to 1990 for 317 banks was employed and the banks had at least \$1 billion in assets. A multi-product translog cost function with a production specification was estimated. Economies of scale were found for both models, with the variable model indicating economies of scale for banks with up to \$10 billion in assets. Banks with over \$10 billion in assets were seen to exhibit dis-economies of scale. Employing the quasi-fixed model, constant returns to scale were found for banks with assets between \$2 billion and \$25 billion. Banks with assets above and below this level experienced slight economies and dis-economies of scale, respectively.

DeYoung (1998) used a translog functional form with an intermediation approach to estimate a 'thick' cost frontier. Data for 39 banks were extracted from the Report of Condition and Income for 1993. The efficiency results from this study were used to identify the possible relationship between cost efficiency and the exam results of managers working in these banks, i.e. to test the hypothesis that good managers run their

banks efficiently. It was found that banks deemed to be poorly managed, in terms of their managers' exam results, were 29 per cent less efficient than those banks associated with better managers, supporting the hypothesis posed.

Previous studies of depository institutions employing flexible Fourier functional forms have been limited in number. The earliest study (to the author's knowledge) to employ this form for analysis of depository institutions was McAllister and McManus (1993) using the example of US commercial banks. Other studies, which have used this functional form, have also analysed US commercial banks, including Mitchell and Onvural (1996), Berger and DeYoung (1997), DeYoung and Hasan (1998) and Berger and Mester (1997). Mitchell and Onvural and McAllister and McManus, both compared the performance of the flexible Fourier functional form with that of the translog functional form.

McAllister and McManus (1993) suggested that the inconsistent results presented on bank efficiency were caused by “... a combination of econometric mis-specification and faulty model logic”, where the “... choice of technique is the main cause of the puzzling result” (p.390). Particular criticism is made of the global application of locally defined translog models. McAllister and McManus used call report data on 4,550 banks for the sample period 1984 to 1990. A production model specification was adopted. After amendment for the cost of capital to incorporate some element of risk (this is discussed in greater detail in chapter 3), scale elasticities were estimated. The model was estimated employing two distinct functional forms, the widely applied translog functional form and

the relatively under-used flexible Fourier functional form. Using a combination of the flexible Fourier function form and kernel estimation, the results showed that banks with assets greater than \$10m had average scale efficiency of 65 per cent. An average scale efficiency of 98 per cent was reported for banks with \$500 million in assets. These results were markedly different from previous studies.

Mitchell and Onvural (1996) used both production and intermediation models to estimate cost efficiency characteristics for US banks between 1986 and 1990. Data was taken from the Call and Income reports for 306 large banks. Both radial and non-radial measures of economies of scale and product mix were estimated. Constant returns to scale were found with slight diseconomies of scope recorded for the joint production of real estate, consumer and commercial and industrial loans. The authors suggested that "... the industry cost function for large banks does not have a Translog form for either the production or intermediation approaches to bank production ... (and) ... estimated Translog and FF (flexible Fourier) cost equations lead to different conclusions about scale and scope economies" (p.197).

Berger and Mester (1997) used data for 6,000 commercial banks over the sample period, 1990–95. An intermediation model was employed to measure cost, profit and 'alternative'² profit functions. These functions were all estimated using a flexible Fourier functional form and a fixed-effects panel data model. The level of off-balance sheet activities, physical capital and financial equity capital were incorporated into the functions as fixed quantities. The levels of non-performing loans and the weighted

average of the proportion of loans issued by banks were included as environmental variables. The addition of these variables was made to accommodate the substantial operating heterogeneity in the data set employed. Additionally, output quantities and costs were deflated by fixed capital equity, to control for heteroskedasticity and scale bias in the measurement of firm-specific efficiency. The efficiency estimates were then compared with a range of potential determinants of efficiency to determine which may be significant. Overall, substantial levels of inefficiency were reported with all functions, although far lower levels of cost efficiency (0.868) appear to exist. Reported profit efficiency was 0.549 and 'alternative' profit efficiency is 0.463, suggesting 54.9 per cent and 46.3 per cent profit efficiency respectively. Substantial scale economies were reported, and were suggested to be a result of regulatory change, technological improvement and lowering of the interest rate levels. The analysis of the potential correlates or determinants of efficiency was less successful, leaving most inefficiency unexplained. This study is considered in more detail in Chapter 5.

Berger and DeYoung (1997) employed a flexible Fourier functional form to analyse the relationship between asset quality, denoted by problem loans, and cost efficiency. A data set of commercial banks operating between 1985 and 1994, including in total 69,742 observations, was used. Ten separate cost frontiers were estimated employing a production approach. The relationship between the estimated cost efficiency and asset quality was assessed using Granger causality tests. Average cost efficiency fell over the sample period from 0.947 to 0.878 indicating 94.7 and 87.8 per cent cost efficiency respectively. The analysis broadly indicated that rises in the level of problem loans lead

to reductions in cost efficiency. This decline in cost efficiency, in turn, leads to a decline in asset quality. This finding therefore suggested that cost efficiency may be viewed as an important indicator of problem banks and the level of provisions for bad and doubtful loans may be viewed as important in the estimation of cost efficiency.

DeYoung and Hasan (1998) used a flexible Fourier 'alternative' profit function to assess the profit efficiency of 5,435 small, urban commercial banks for 1988, 1990, 1992 and 1994. Banks were included that had less than \$500m in total assets, had headquarters in a metropolitan area, had been in existence for more than a year and provided both loans and deposit services. An intermediation model of bank production was applied with non-performing loans included in the model to control for asset quality. Overall, low levels of profit efficiency were estimated. These measures of efficiency improved over the sample period, rising from 0.477 in 1988 to 0.545 in 1994, suggesting 47.7 and 54.5 per cent profit efficiency respectively. These firm-specific profit efficiency measures were then used in an analysis of the determinants of efficiency. Overall it was suggested that small, urban banks "... do not exploit multiple branch locations or purchased fund financing as well as established banks" (p.585) and profit efficiency of small urban banks levels of between 9 and 14 years of operation.

A number of U.S. studies investigated technical change in the banking sector. Previous studies may be broadly delineated into studies performed using econometric estimation techniques and studies employing non-parametric index number methods. Previous econometric studies have fairly consistently estimated very low levels of technical

change and similarly low or negative levels of total factor productivity growth. Hunter and Timme (1986) reported both significant levels of positive technical change in US banking and emphasised the importance of economies of scale in determining the levels of technical change. A sample of 91 large US banks, between 1972-82 were considered in this study. Criticism of this study has focused upon the use of total assets and total deposits combined as the single measure of output. Humphrey (1993) analysed the technical change in US commercial banks between 1977 and 1988 reporting an array of findings. Between 1977-80, positive technical change was reported for the US commercial banking sector. This progress was reversed as negative technical change was reported for the 1980-82 period. Between 1983 and 1988 minimal technical change was reported. This study also considered the technical change of banks with different asset sizes. Smaller banks, defined as those between \$200 and \$300m in assets, experienced substantial negative technical change over the sample period relative to larger banks.

Elyasiani and Median (1990) employed a non-parametric analysis of large (banks with assets in excess of \$300m) US commercial banks. A significant level of positive technical change was observed within the sample period of 1980-85 for the 191 large commercial banks in their sample. Wheelock and Wilson (1996) discovered that some US commercial banks had experienced both technological improvement and productivity gains between 1984 and 1993. In this study, smaller banks, with assets less than \$300 million, experienced lower levels of technical change. Daniel and Tirtiroglu (1998) quantified total factor productivity growth for US commercial banks between 1935 and 1991. This study initially calculated total factor productivity growth employing an index

number approach, with a non-parametric Tornquist index, and then attempted to determine what components of total factor productivity growth may be attributed to trend and cyclical components through employing a Kalman filter technique. A declining trend of total factor productivity growth was observed over the sample period, averaging 2.27 per cent.

In this review of efficiency studies of US banks, a number of issues and innovations seem to appear. Overall, technical inefficiency seems to dominate allocative inefficiency, particularly in the US commercial banking sector. Larger US banks also appear to be, on average, more efficient than smaller US banks. Estimates of economies of scope appear to be both contradictory with a variety of results reported. Other more general results include the rejection of Cobb-Douglas functional forms in modelling bank production and evidence that questions the use of translog functional form for modelling bank productive technology. Also, new methods such as expansion path economies, thick frontier estimation techniques and the inclusion of off-balance sheet outputs are suggested in this substantial literature. None of the previous conclusions relating to functional form, model specification, estimation technique and the sample period appear to be rejected by this review of US bank efficiency studies.

4.7 Conclusions

Overall, a broad range of studies are reviewed. A number of significant conclusions may be drawn. The use of differing model specifications and variable definitions and methods

of estimation have all had a considerable effect on the efficiency results produced. Depository institutions operating at different times, in different countries and under different regulatory regimes are also likely to display quite distinct efficiency characteristics.

More research is called for to investigate the factors that may influence or bias estimates of efficiency. Whilst most studies have attempted to measure the extent of inefficiencies and to rank institutions in order of inefficiency, there has been little work on what could be the possible sources of inefficiencies or forms of bias that may influence the estimation of efficiency. Thus, more research into the main causes of economies of scale and scope and technical and allocative inefficiencies in depository institutions would be useful.

Although this survey has covered a large number of non-US depository institutions, it is still true that work in the USA dominates the field and there remains scope for the application of the available techniques to some institutions for the first time. A particularly useful research project would be one, which attempted more international comparisons of inefficiency, such as those undertaken by Allen and Rai, (1996), and Molyneux *et al*, (1996).

Finally, the analysis of the differing functional forms used to model depository institution production has also been illustrative. The use of constrained functional forms, such as the Cobb-Douglas, has been rejected. Many of the later studies that employ the flexible

Fourier functional form have strongly criticised the use the ‘Diewert’ functional forms, such as the translog. Discussion of all these factors will be continued in both chapter 5 and in the empirical studies of Section Two.

It is clear that, although an enormous literature has developed concerned with the investigation of production and cost inefficiencies in depository institutions, there remains much work to be done in defining the theoretical underpinnings of the research³, refining old techniques, devising new techniques and comparing methods and results. In short, many avenues remain to be explored.

¹ The model is reduced through dropping both the squared and interaction terms within the translog functional form. This reduction was made to reduce approximation error and multicollinearity.

² The profit function considers that profits are dependent on levels of output prices and input prices. The ‘alternative’ profit function differs in that it considers that profits are dependent on output quantity and input prices.

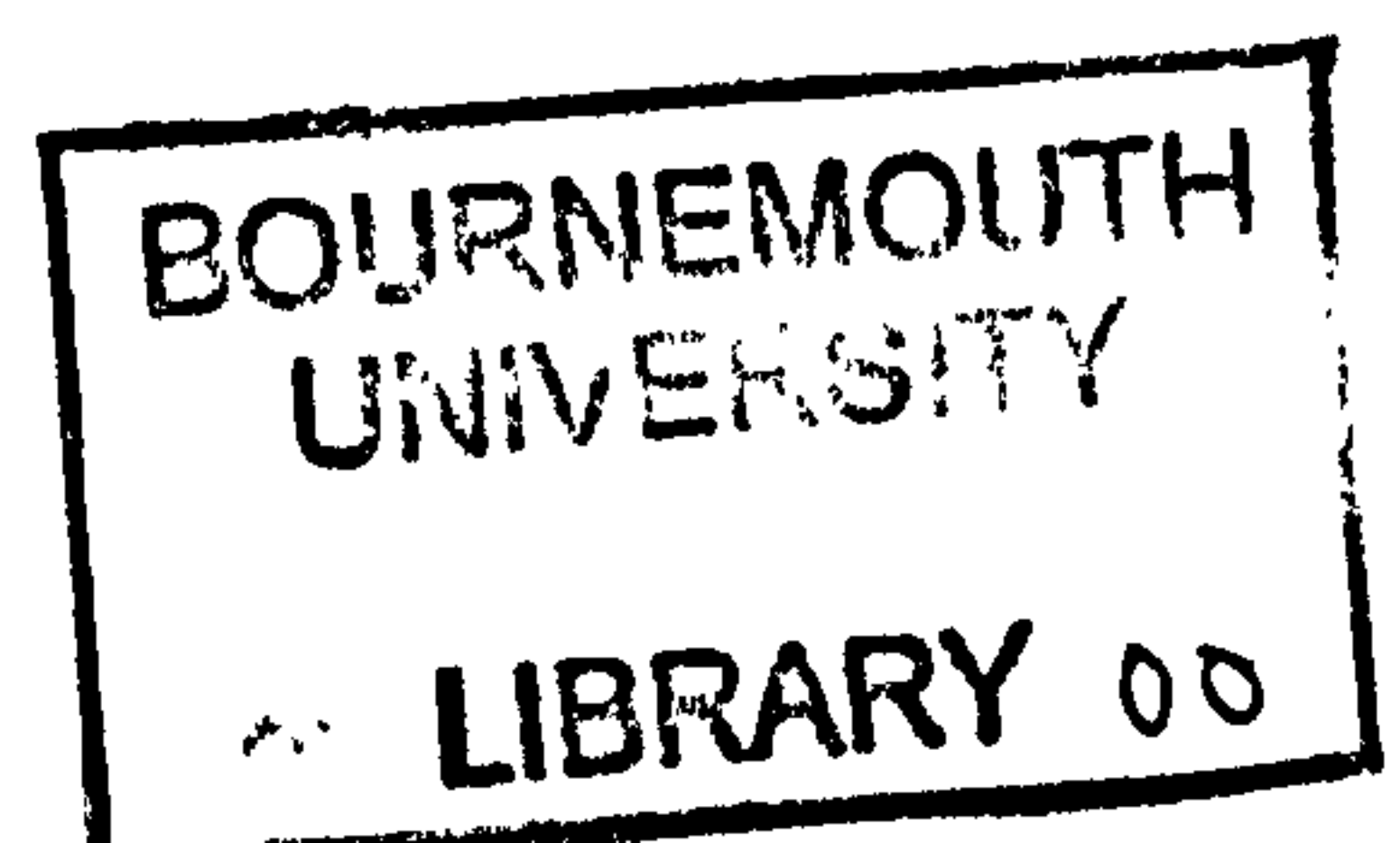
³ For example the monograph by Hancock, (1991).

Chapter 5 Measuring efficiency

5.1 Introduction

This chapter outlines a range of measures that may be used to quantify economic efficiency and productivity. In chapters 3 and 4, it was pointed out that a range of different approaches had been used in many studies to estimate efficiency and productivity. Principal differences between these studies include the efficiency concept used, the method of estimation, the functional form employed and the individual measures of efficiency to be estimated. In this chapter, these differing approaches will be assessed. Recommendations will be made to determine what estimation methods, efficiency concepts and efficiency and productivity statistics will be employed in the empirical studies of efficiency (chapters 7 and 9) and productivity (chapters 8 and 10).

The chapter is organised as follows. The choice of efficiency concept is introduced and discussed in section 5.2. Section 5.3 outlines the advantages and disadvantages of the two principal estimation methods employed in the literature. The choice of functional form used to represent productive technology is discussed in section 5.4. Econometric methods of measuring of economies of scale and economies of product mix are reviewed in section 5.5. A critical review of the differing econometric methods of cost efficiency estimation is provided in section 5.6. The definition of productivity is discussed in section 5.7 and section 5.8 provides an assessment of two productivity models for use in the empirical chapters. A summary of the chapter and conclusions are provided in section 5.9.



5.2 *The efficiency concept*

Before deciding what methods to use in the measurement of efficiency, it is important to consider the definition of efficiency employed. Previous studies (see chapter 4) have used a number of different definitions of efficiency, including productive efficiency, cost efficiency and profit efficiency. Assumptions of totally transferable productive technology and firm objectives of output maximisation, cost minimisation or profit maximisation for production, cost and profit functions, respectively, are made. These assumptions are discussed further in chapter 11.

Productive efficiency or technical efficiency is defined as the distance, in terms of output produced, between an individual institution and the 'optimal' or 'best practice' institution. This hypothesised 'best practice' institution is defined with reference to all the institutions in the sample set. Such an 'optimal' institution would exist on the production function or frontier. The distance a sample institution is from this 'optimal' institution or the production function is viewed as productive inefficiency. A production function assumes that the level of output of an individual institution is dependent on the amount of inputs employed in production, plus random error and any other variables that account for the environment or the particular circumstances of individual institutions. Productive efficiency is therefore limited in its extent to considering the input quantity that may be reduced to produce a specified quantity of output.

Cost efficiency studies estimate how far the production costs of an individual institution differ from the production costs of a best practice institution operating

under similar conditions and producing the same outputs. This measure is defined with reference to a cost function constructed from the observations of all institutions in the sample set. This cost function assumes that the total production costs of individual institutions are dependent on the prices of variable inputs, such as capital and labour, the quantity or value of outputs produced, random error and any other variables that account for the environment or the particular circumstances of individual institutions. A cost function allows the measurement of the least-cost proportions of inputs in terms of input prices. Cost efficiency is dual to productive efficiency allowing measurement of both productive efficiency and the optimal proportion of inputs in terms of input prices or allocative efficiency, also known as price efficiency.

Studies of profit efficiency attempt to quantify the degree to which an institution is yielding maximum possible profit. Profit efficiency measures are derived from a profit function or frontier, which assumes that profits are dependent on the level of variable output prices, variable input prices, random error and other variables that account for the environment or particular circumstances of individual institutions. Researchers adopting this efficiency concept attempt to measure the degree to which output prices may be varied. This variation is expected to influence revenue, assuming that output prices are determined by factors outside the boundaries of the model. A profit efficiency measure may therefore be defined as the ratio of actual profits achieved to the estimated maximum profits attainable for a 'best practice' institution.

In the thesis, the *cost efficiency* concept is employed for four central reasons. Firstly, it is deemed preferable to view efficiency in its broadest sense, incorporating both productive and allocative efficiency. Secondly, profit efficiency measures assume that institutions maximise profits. Whilst this assumption may appear to be reasonable for many firms, the objective function of a depository institution may not be as clear-cut compared to other industries. Thirdly, there exists a substantial quantity of anecdotal evidence that banks both in the UK and globally have emphasised the need to manage costs more closely. For example, a study by Salomon Brothers (cited in Molyneux *et al*, 1996, pp.4-7) suggested that “ ... cost management was now a dominant strategic theme throughout the banking world” (Molyneux *et al*, 1996, p.4). Through adopting a cost efficiency approach in the thesis, it is hoped that the performance of UK depository institutions may be examined in a form more closely representing that favoured within the industry itself. Lastly, most previous econometric studies of UK depository institutions (see chapter 4) use cost efficiency concepts. The use of cost efficiency concepts therefore improves the degree of comparability of analysis in this thesis with other studies.

5.3 *Estimation methods*

The estimation methods used in the studies considered in the literature review (chapter 4) broadly fall into econometric and non-parametric approaches. The difficulty of choosing an appropriate estimation method is compounded by the fact that different methods appear to produce different estimates.

A limited number of studies have used simulation analysis to assess which technique provides the 'better' estimator. For example, Banker *et al* (1988) estimated production frontiers from a data sample of known characteristics using a translog production function and non-parametric *DEA* (Data Envelopment Analysis) model. It was suggested that whilst the *DEA* techniques provided more reliable estimates of both technical and scale efficiency than the translog, this reliability was only achieved under certain conditions. It was reported that the non-parametric *DEA* model performed poorly with a small sample size and where the true production function did not concur with established production theory.

Econometric estimation methods are favoured for use in this thesis for three reasons. Firstly, econometric methods allow for the accommodation of random error in the estimates. Conversely, non-parametric methods do not account for random error, potentially allowing random error to be measured as technical inefficiency¹. Berger and Humphrey (1997) report that on average non-parametric studies of efficiency in US banking indicate lower levels of efficiency. This could be consistent with the misspecification of random error as technical inefficiency. Secondly, Berger and Mester (1997) suggest that efficiency measures estimated using econometric approaches correspond more closely to the economic understanding of efficiency. This greater degree of agreement between economic concepts and econometric approaches occurs due to the inclusion of more information, such as input prices, and the incorporation of a range of related efficiency aspects, such as allocative efficiency. These features are not always fully accounted for in non-parametric approaches. Berger and Mester in developing this point indicated that "... non-parametric techniques typically focus on *technological* optimization rather than *economic* optimization and do not

correspond to the cost and profit efficiency concepts” (p. 905). Lastly, this thesis employs a relatively small sample of institutions with an unknown technology. These features combined with the high level of dispersion over asset size may pose problems. According to Banker *et al* (1988), small samples with unknown technologies are poorly accommodated by *DEA* techniques. For example a small sample with a high level of dispersion in terms of asset size may lead to a high proportion of observations being defined as efficient and effectively being self-identifiers, reducing the validity of the overall analysis.

5.4 *The choice of functional form*

In this section, the choice of functional form to represent productive technology in the empirical studies is discussed. In previous studies, a considerable number of functional forms have been used, but by far the most frequently used has been the transcendental logarithmic or translog functional form (Christensen *et al*, 1971, 1973). In this brief discussion, the translog and the related flexible Fourier functional form are considered. A broader review of functional forms used in such studies is provided by Molyneux *et al* (1996).

In selecting a functional form, researchers face a choice between greater flexibility and better global behaviour across a spectrum of observations. Simple functional forms, such as the Cobb-Douglas form, satisfy many conditions or properties of a cost function over a broad range of observations. Their simplicity enables properties within the function to be consistently, if bluntly and robustly, applied. Difficulties with the use of simple forms centre on the limited scope of productive technologies

that may be represented. Modelling a more sophisticated productive technology, as envisaged in UK depository institutions, requires a greater degree of flexibility; this in turn enables a wider range of productive characteristics to be represented. This flexibility may be obtained by employing a more complex functional form, such as the translog functional form.

It was previously stated that the translog has been the most frequently employed functional form. This Diewert flexible functional form provides a cost function that can accommodate a second-order differential approximation to an arbitrary twice continuously differentiable cost function. However, such a cost function will exhibit linear homogeneity in input prices only within certain parameters or an 'admissible domain' (Diewert, 1971). Thus the translog, whilst thought by many researchers to be appropriate for the analysis of depository institutions, may only be quantified or estimated consistently within a certain range of observations or within a 'specified domain', leaving the possibility of specification error in estimation.

There have been a limited number of studies that have considered the workings of the translog functional form in applied work. Wales (1977) and Caves and Christensen (1980) have undertaken investigations into the viability of approximation, as the number and variability of observations is increased. Wales (1977) found the behaviour of the translog to deteriorate substantially as the true substitution elasticity departs from unity. Caves and Christensen (1980) found the translog to be better behaved over a broader range of observations than the generalised Leontief function.

Guilkey (1980) and Guilkey *et al* (1983) undertook Monte Carlo tests to assess the performance of different functional forms. These tests examined the ability of the translog functional form and other functional forms to represent properties or characteristics of productive technology using pre-defined test data. Guilkey (1980, 1983) examined the productive characteristics, such as economies of scale and factor substitution, of the translog and other Diewert flexible forms. Tests were performed with data representing varying degrees of technology. Guilkey (1980) found that the translog functional form performed well over a broad range of technologies. Guilkey *et al* (1983) found that the translog behaved better globally than other Diewert flexible forms such as the generalised Box Cox and the generalised Leontief, where “... the translog form provides a dependable approximation to reality provided that reality is not too complex” (p.614).

Ivaldi *et al* (1996) in a comparison of the translog and the flexible Fourier functional forms with a panel data model employing data from French farms, indicated that although both functional forms provided equivalent descriptions of the productive technology, the flexible Fourier functional form was able to represent a wider range of cost structures. Whilst it may be concluded that use of the translog form may be appropriate when substantial dispersion of data is not present, the potential for specification error when considering a diverse data set may present difficulties.

The flexible Fourier functional form (see Gallant, 1981, 1984) is a second-order polynomial in the explanatory variables together with a combination of sine and cosine functions in the explanatory variables. This form is Sobolev flexible (see Gallant, 1981) and therefore can estimate elasticities consistently and has negative

prediction bias, thus removing the potential for specification bias in the representation of productive technology. The flexible Fourier form has the translog form nested within it and should provide a similar interpretation of productive technology whilst reducing the potential for specification error. The expansion of the functional form to fit the scale of the data set is performed through the inclusion of additional trigonometric terms. Eastwood and Gallant, (1991) suggested that the total number of parameters should equal the number of observations raised to the two-thirds power. This novel approach of fitting the size of the functional form to the sample size differs from the method employed when using Diewert flexible forms, such as the translog, where the functional form is assumed to provide *a priori* a representation of the true cost function.

Both the widely used translog and flexible Fourier functional are used in the empirical chapters. The possibility of mis-specification with the translog functional form is also discussed in the pertinent empirical chapters and in the conclusions (chapter 11).

5.5 *Econometric measurement of economies of scale and scope*

In the studies reviewed in chapter 4, a number of efficiency measures were employed. The most frequently employed measures include economies of scale, economies of product mix and measures of cost efficiency. As we saw in chapter 4, most studies concerned with investigating the existence of economies or diseconomies of scale and scope in depository institutions have employed statistical cost analysis. In these studies, it is assumed that depository institutions operate like manufacturing firms in the sense that they attempt to minimise the cost of using various inputs to produce an

output or set of outputs which is then sold to consumers. Thus, the total cost function facing these firms relates minimum cost to the firms' outputs, input prices and other independent variables. In general terms, such a cost function may be written as:

$$C = C(y, p, z) \quad (5.1)$$

where y is a vector of outputs, p is a vector of input prices, z is a vector of other explanatory variables and C is the minimum cost of producing y for given values of p and z . The cost function is normally assumed to be linearly homogeneous, monotonically increasing and concave in input prices. These conditions then ensure that it is dual to the transformation function $T(x, y)$ where x is a vector of inputs.

Early economies of scale studies estimated single-product functions only, either by assuming that a depository institution's outputs could be aggregated into a single product or by estimating separate functions for the different products. The theoretical groundwork developed by Panzar and Willig (1977), Brown *et al* (1979) and Baumol *et al* (1982) soon led to the estimation of multiproduct functions and the calculation of multiproduct production/cost characteristics, such as overall economies of scale, product-specific economies of scale and economies of scope. These measures are considered in turn.

Ray economies of scale are cost savings resulting from proportional increases in the quantities of all outputs produced. The extent of overall economies of scale (*OES*) at any particular output combination is usually measured by the elasticity of total cost with respect to composite output. This measures the proportionate change in total cost

as all outputs are changed by the same proportion. Thus, for a firm producing m outputs, we can write, using natural logarithms (Ln):

$$OES = \sum_i (\partial Ln C / \partial Ln y_i) \quad (5.2)$$

$i = 1, 2, \dots, m$. According to this radial measure, there are overall economies of scale (and therefore decreasing *ray* average cost) if $OES < 1$, overall diseconomies if scale (increasing *ray* average cost) if $OES > 1$ and constant *ray* average cost if $OES = 1$.

Berger *et al* (1987) devised a measure, known as *expansion path economies of scale*, calculated as the elasticity of incremental costs with respect to incremental outputs, which takes into account changes in the composition of output as firms grow. Expansion path scale economies incorporate both scale costs and costs of distinct production technologies adopted at different scales. Emphasis is placed on the difference in operating costs between two institutions. The difference in output between the banks is divided by the difference in costs between the two outputs.

$$EPSCE^{AB} = \{ \sum [(y_i^B - y_i^A) / y_i^B] [\delta Ln C(p, y^B) / \delta Ln y_i] \} / \{ [C(p, y^B) - C(p, y^A)] / C(p, y^B) \} \quad (5.3)$$

where y_i^B and y_i^A are the quantities of the i^{th} output at banks A and B respectively and $C(p, y^A)$, $C(p, y^B)$ are the operating costs at banks A and B respectively. Measures are interpreted in a similar manner to the radial measures previously outlined.

Measuring overall or expansion path economies of scale is appropriate only as long as the other explanatory variables included in the cost function remain unchanged as outputs vary. A measure of ‘augmented’ overall economies of scale has occasionally been computed to take into account the indirect effects of induced changes in other explanatory variables. For example, Benston *et al* (1982) used the following measure of augmented economies of scale (*AOES*) to allow for the induced change in the number of branch offices (*B*) as banking outputs changed:

$$AOES = \sum_i (\partial \ln C / \partial \ln y_i) + (\partial \ln C / \partial \ln B) \left(\sum_i (\partial \ln B / \partial \ln y_i) \right) \quad (5.4)$$

In general, there has been very little work on the *sources* of economies or diseconomies of scale. Some researchers (see, for example, Hardwick, 1990) have attempted to identify the main sources by measuring the cost saving attributable to the j^{th} input as the firm expands. This can be done by writing the logarithm of the cost of employing the j^{th} input ($\ln C_j$) as:

$$\ln C_j = \ln S_j + \ln C \quad (5.5)$$

where S_j is the j^{th} input’s cost share. Input-specific overall economies of scale (OES_j) may then be written as:

$$OES_j = \sum_i (\partial \ln S_j / \partial \ln y_i) + OES \quad (5.6)$$

While this formula can indicate whether cost savings are attributable to the employment of any particular input, it can give little guidance as to the real reasons for the existence of economies or diseconomies of scale. This is clearly an area where more research is called for.

Product-specific economies of scale measure the effect on the i^{th} product's incremental cost of a change in the quantity of product i , with the quantities of the other products remaining unchanged. Assuming that a reasonably wide-ranging data set is available so that accurate estimates of incremental cost can be obtained, the extent of scale economies attributable to product i can be measured by the elasticity of the i^{th} product's incremental cost with respect to the output of the i^{th} product. Unfortunately, data sets relating to depository institutions rarely allow accurate estimates of incremental costs (which require an estimate of total cost when the quantity of product i is zero) to be calculated. Some researchers, therefore, have investigated local product-specific economies of scale by examining the behaviour of the i^{th} product's marginal cost. In particular, the value of $\partial^2 C / \partial y_i^2$, the gradient of the i^{th} product's marginal cost, has been used to judge whether local product-specific economies or diseconomies of scale exist. A negative gradient suggests the existence of product-specific economies of scale, while a positive gradient suggests the existence of product-specific diseconomies of scale.

The ray economies of scale (*OES*) measure will be employed in the empirical studies. This statistic is the most frequently used statistic for estimating economies of scale.

Through employing this statistic, increased comparability with other studies may be possible.

Economies of scope may be defined as cost savings achievable from the joint production of two or more goods or services within a single enterprise, compared with their separate production by specialised firms. In other words, they are economies, which arise from the range of products produced, rather than the scale, of a firm's operations. For a firm producing two outputs, y_1 and y_2 , the magnitude of scope economies (*ESC*) may be measured by:

$$ESC = [C(y_1, 0) + C(0, y_2) - C(y_1, y_2)] / C(y_1, y_2) \quad (5.7)$$

ESC measures the relative increase in cost that would result from a separation of the firm into two specialist firms. There are economies of scope if $ESC > 0$ and diseconomies of scope if $ESC < 0$. Unfortunately, this formula requires estimates of total cost when the quantity of each of the products in turn is set equal to zero and, as we saw above, such estimates are not easy to obtain. An alternative way of investigating the existence of local economies of scope is to test for *cost complementarities* by checking whether the second derivative, $\partial^2 C / \partial y_1 \partial y_2$, is significantly less than zero. If it is, so that the marginal cost of one good decreases as the output of the other good increases, then there are cost complementarities and therefore (local) economies of scope. Another approach was suggested by Berger *et al* (1987); they devised a formula for *expansion path subadditivity* which measures the extent to which lower costs result from the joint production of an output bundle at a depository institution (whose output composition is representative for its size)

compared with specialised production by two smaller institutions (whose combined outputs are equal to those of the bigger institution).

A more general measure, known as *economies of diversification*, was developed by Ferrier *et al* (1993) to allow for the situation in which at least one output is produced by all firms, while the production of other outputs is disjoint. Thus, for two firms, *A* and *B*, and three outputs, *1*, *2* and *3*, economies of diversification would be said to exist if:

$$C(y_1^A, 0, y_3^A) + C(0, y_2^B, y_3^B) > C(y_1^A, y_2^B, y_3^A + y_3^B) \quad (5.8)$$

Although there have now been a large number of studies of economies of scope and subadditivity in depository institutions, none have addressed the issue of the *sources* of these cost savings and so this represents another important area where more research is required.

The economies of scope statistic will be employed as a measure of economies of product mix in the empirical studies. This statistic is the most frequently used statistic for estimating economies of scope in the broader literature. Through employing this statistic, therefore, increased comparability with other studies may be possible. Cost complementarities will also be estimated as a further 'test' of local economies of scope.

5.6 *Econometric measurement of efficiency*

In the previous section, approaches to measuring economies of scale and economies of product mix were reviewed. The measurement of these characteristics 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 the efficiency of an individual institution is viewed to be a distance from the 'best practice' function or frontier. In this section, the development of both econometric frontier techniques and panel data methods to measure both overall and then firm-specific efficiency are reviewed. Recommendations for the empirical chapters are made. Following previous recommendations on the definition of efficiency to be used, the discussion is stated in terms of cost efficiency.

The evolution of econometric frontiers to measure technical, allocative and overall efficiency was initiated by the development of deterministic frontiers. These frontiers considered all distance from the frontier (both random error and inefficiency) to be inefficiency. Such measures have been criticised due to the absence of any statistical properties and the high degree of sensitivity to outliers and mavericks in the data. Attempts to amend for this have included the development of probabilistic frontiers. Timmer (1971) developed a probabilistic frontier by excluding a proportion of the data set where " ... the frontier is estimated in probabilistic fashion by constraining X per cent of the observations to fall outside the frontier surface" (p.778).

Aigner *et al* (1977) and Meeusen and van de Broek (1977) have both been attributed with the development of the stochastic frontier for the measurement of technical and

allocative efficiency. Stochastic frontiers are distinguished from probabilistic and deterministic methods by their statistical estimation, as opposed to calculation of the frontier where “ ... it seems preferable to incorporate the possibility of measurement error and of other unobservable shocks, in a less arbitrary fashion.” (Aigner *et al*, p.23). A cost function frontier is considered where the disturbance term is split into two components. This may be represented as:

$$C = C(y, p; \beta) + \varepsilon \quad (5.9)$$

where y denotes outputs, p represents input prices, β are the parameters to be estimated and ε is the error term. The error term may be subdivided into

$$\varepsilon = u + v \quad (5.10)$$

The components of the error term include v , a symmetrical two-sided term, which includes the effects of data outside the control of the model. The other component, u , is a one-sided disturbance included to represent all the effects of the data that are in the control of the model and may be used to derive measures of inefficiency.

The two-sided residual term v is assumed to be symmetrical and independently and identically distributed. This may be estimated with familiar statistical methods. For example, the normal distribution, may be imposed upon the error component, such that $v \sim N(0, \sigma_v^2)$. The one-sided disturbance term u is assumed to be distributed independently of v and must satisfy $E(u) = 0$ for a cost frontier. The non-negative disturbance u reflects the positioning of all institutions on or above the cost function

frontier. The modelling of this disturbance has inspired a debate on what is the most appropriate distribution to apply.

Several one-sided distributions have been used to model the inefficiency component. The most popular is the half-normal distribution suggested by Aigner *et al* (1977). Others include the exponential distribution (also suggested by Aigner *et al*, 1977), the truncated normal distribution suggested by Stevenson (1980) and the two-parameter Gamma distribution proposed by Greene (1990). Evidence suggests that the choice of distributional assumption has only a small effect on the inefficiency estimates and hardly any effect on the efficiency rankings of firms. For example, Kaparakis *et al* (1994) obtained very similar inefficiency results using the half-normal and truncated normal distributions. Greene (1990) reported similar rankings using the half-normal, truncated normal, exponential and Gamma distributions, but suggested that applications of the single-parameter distributions may yield higher overall estimates of inefficiency than the two-parameter Gamma distribution.

These results suggest that the choice of the one-sided distribution may have a minor, albeit significant, effect on the inefficiency results. The impact of differing distributions compounded with the use of potentially inappropriate flexible functions (such as the translog functional form) suggests that results from econometric stochastic frontier models need to be interpreted with care as the use of any one-sided probability distribution (for estimating firm-specific inefficiencies) has been criticised by some commentators. For example, Berger (1993) suggested that the distribution of inefficiency in US banks was similar to a symmetrical normal distribution where “ ... if, in fact, inefficiencies do have something close to a symmetric distribution this

raises questions about any single period, composed error method that tries to separate inefficiency from random error, since random error is also generally believed to follow a symmetric distribution” (p.290).

Another criticism levelled at the estimation of firm-specific inefficiencies from econometric frontier models has been that the actual dispersion of costs within the ‘raw’ data and the distributional assumptions imposed on the error term often do not concur. For example, the one-sided normal distribution imposes a structure in which inefficiency is assumed to be clustered around the frontier. This structure suggests that only a limited deviation in costs is possible from the hypothesised most efficient level. This *a priori* assumption has been criticised as the dispersion of costs for financial institutions may often be quite large. This problem has been termed the ‘Greene’ problem within the literature. Potential suggestions to ease this problem have included the use of other distributions to represent the error term or the use of a thick frontier as proposed by Berger and Humphrey (1992b). Surveys of many of the distributions adopted for the error components are considered by Stevenson (1980) and Greene (1993).

To overcome the strong *a priori* distribution assumptions made by stochastic frontier models, alternative measures of efficiency have been proposed. Distribution-free cost efficiency is a relative measure of firm-specific efficiency. The efficiency of a sample of institutions is derived through reference to the efficiency of the most cost efficient depository institution in the sample. The theoretical advantage of this approach is that it removes many of the strong distributional assumptions used in stochastic frontier analysis. Efficiency is derived directly from the individual effects produced by the

fixed-effects model, where the individual effects, v , 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 that efficiency is constant over time and bias in efficiency may be removed through averaging over time. The individual effects (v) may be employed as an indicator of non-negative cost efficiency. Thus distribution-free efficiency may be written:

$$\text{Efficiency} = \exp(\min[\text{Ln}v] - \text{Ln}v) = \min[v] / v \quad (5.11)$$

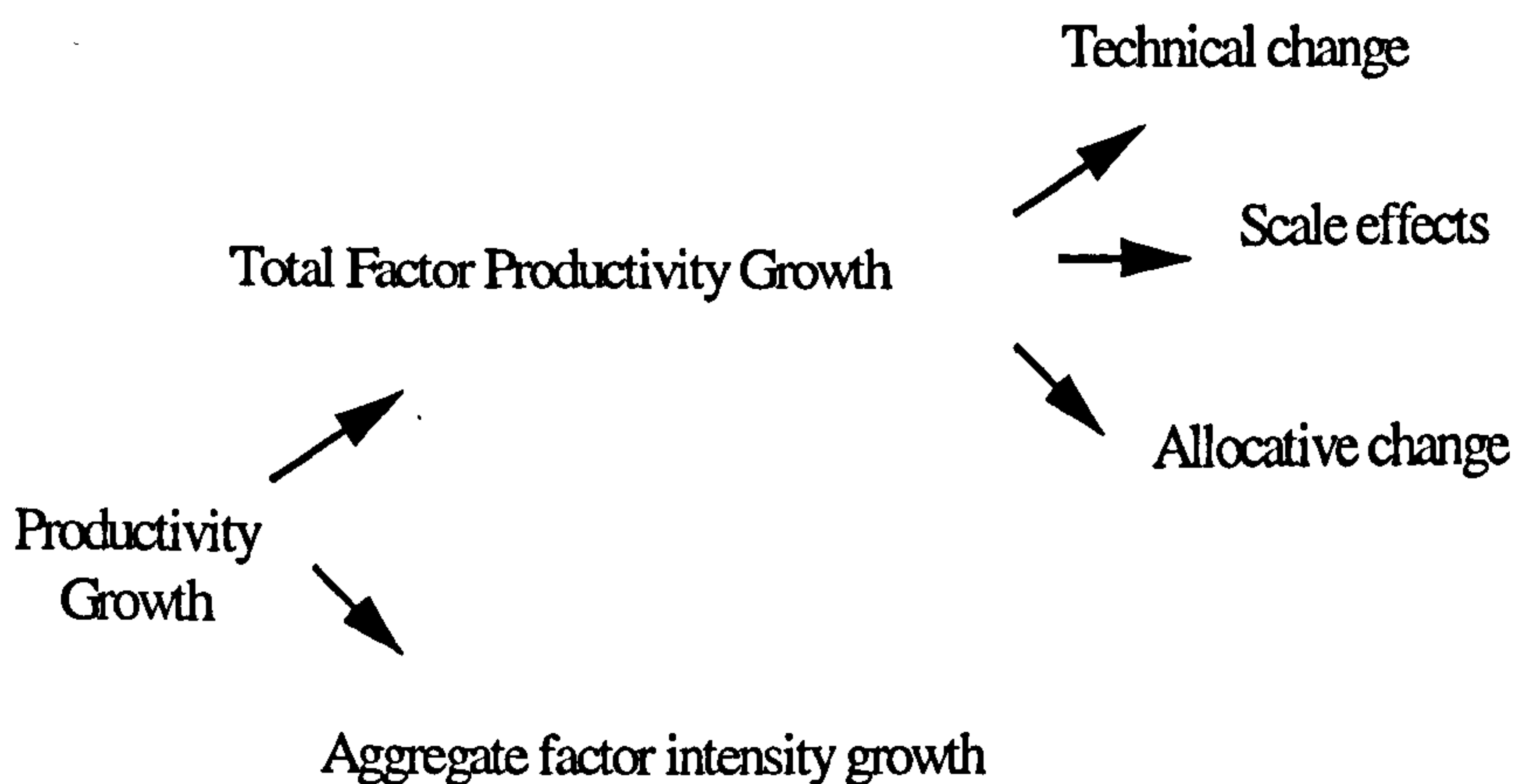
For the i^{th} institution, where $\min(v)$ is assumed to be the individual effect of the most efficient institution in the sample. For further discussion of this measure, see Allen and Rai (1996, 1997) and DeYoung (1997). Overall, due to the strong *a priori* assumptions imposed with econometric frontier techniques, distribution-free cost efficiency measures are selected for use in the empirical studies.

5.7 What is productivity?

The definition of productivity is fraught with difficulties. The term has been considered synonymous with labour productivity in manufacturing; similarly, productivity is generally associated with yields in agricultural economics. Productivity measures the dynamic relationship between the use of inputs for the creation of outputs. The quantification of the amount of resources or inputs required

for the production of a specified bundle of outputs provides a numéraire for the measurement of productivity. Productivity, in this context, is a relative concept in that the productivity of a firm only exists in relation to other time periods. The measurement of productivity growth over time is deemed to be associated with the *change* in productive technology for individual firms. Productivity may also be viewed as a generalised index containing a range of components. The relationship of productivity to these components may be observed in Figure 5.1.

Figure 5.1 Components of total factor productivity growth



Technical change has been defined as the change in the productive technology of the firm over time or over a homogeneous group. Solow (1957) viewed technical change to be any shift of the production function or productive technology. Technical change may therefore be viewed as the common rate of input reduction whilst outputs are held constant. Thus the effect of technical progress is to shift the productivity curve to the right, where “... every change in technology displaces the production frontier (of a given capital stock); every gain in technical knowledge displaces it outwards”, or “

... conversely for a dual cost function a shift to the left would be expected” (Hicks, 1939, p.271).

In this thesis, the total factor productivity growth of institutions is estimated. Total factor productivity growth is total in that all factor inputs are included in its formulation. The productivity measures used in the thesis will model productive technology both in a multi-input and multi-output form. Partial measures are not considered due to the high possibility of their mis-interpretation. Total factor productivity growth is viewed as the change in productive capability or productivity resulting from a change in the factors of production. Aggregate factor intensity growth refers to the change in the capability of the productive system caused by factors outside the firm or the productivity growth ultimately determined by factors exogenous to the firm. Scale effects are viewed as the influence of scale factors on the efficiency of production. These factors may in turn influence productivity, a measure of efficiency over time. Allocative change is inefficiency that results from a less than optimal allocation of inputs to produce a specified amount of outputs.

5.8 *Productivity models*

This section outlines two productivity models, which are compatible with the econometric cost function approach. The main aim is to select appropriate models for the empirical studies of total factor productivity growth in chapters 8 and 10.

A number of methods for measuring productivity growth have emerged within the literature. These methods fall into the econometric and index number approaches.

Diewert (1992) further categorised approaches to measuring technical change into four groups, which include the econometric estimation of cost and production functions, Divisia indices, exact number indices and non-parametric methods using linear programming.

Most research work in productivity measurement has been carried out either by estimating cost and production functions or by calculating index numbers of either Divisia or exact varieties. The development of these index number measures has emphasised the need for fewer assumptions in the measurement of productivity and technical change. The development of broader index-based measures through a range of measurement and estimation techniques has also been attempted to loosen the taut theoretical impositions inherent in the index number approaches.

The production and cost function approaches to measuring efficiency may be used to estimate the technological possibilities in an industry. The technological possibilities in period T may then be compared with those in period $T+1$ to provide a relative measure of productive capability. Causes of technical change are diverse. Unit costs in an industry can change over time as technical innovations shift the minimum cost frontier. The levels of cost dispersion may be an indicator of the variability of market conditions and alter average efficiency. Regulatory change and other exogenous phenomena may be reflected in both shifts in the frontier and changes in the dispersion in costs. For a further discussion of the possible causes and sources of technical change, see chapter 11.

The most frequently used method of econometric productivity analysis for depository institutions is the time-trend model (e.g. Stiroh, 1997, Lang and Welzel, 1996, Esho and Sharpe, 1995, Humphrey, 1993, Baltagi and Griffin, 1988). The time-trend in this model is quantified through the inclusion of a variable to represent the relevant time period. The level of technical change for a multi-product framework may be represented by the partial derivative of the logarithm of total cost with respect to time.

A production relation could be represented as:

$$y = f(L, K, t), \quad \text{i.e.} \quad y(t) = f(L(t), K(t), t), \quad (5.12)$$

where L represents labour, K represents capital and t represents time. Time is incorporated in a time trend model through the use of a time specific dummy (i.e. t for 0, 1, 2, ..., T years). This approach has been developed to incorporate both quadratic time terms and to incorporate the relationships between both time and input prices. For example, employing a translog functional form and an intermediation model, with three inputs and two outputs, the time-trend cost model (for simplicity the t subscripts are omitted) may be written:

$$\begin{aligned} \ln C = & \alpha_0 + \sum_i \beta_i \ln Y_i + \sum_j \alpha_j \ln P_j + \eta T + 1/2 \varphi(t)^2 + \\ & 1/2 \left(\sum_i \sum_l \gamma_{il} \ln Y_i \ln Y_l \right) + 1/2 \left(\sum_j \sum_k \omega_{jk} \ln P_j \ln P_k \right) + \\ & \sum_i (\kappa_i t \ln Y_i) + \sum_j (\lambda_j t \ln P_j) + \sum_i \sum_j \delta_{ij} \ln Y_i \ln P_j + \varepsilon \end{aligned} \quad (5.13)$$

for $i, l = 1, \dots, 2$ and $j, k = 1, \dots, 3$.

where C represents total costs, Y represents outputs, P represents input prices, T represent time, ε represents the error term and $\alpha, \beta, \gamma, \delta, \lambda, \kappa, \varphi, \eta$ and ω are parameters to be estimated. From this model, both estimates of technical change and its components may be derived.

Average technical change is viewed as the (negative) first derivative of the logarithm of total cost with respect to time. The average technical change from the translog cost model, may be written as:

$$-[\eta + \varphi T + \sum_i \kappa_i \ln Y_i + \sum_j \lambda_j \ln P_j] \quad (5.14)$$

Average technical change can be decomposed into radial (or neutral) technical change, non-neutral technical change and scale augmenting technical change. Radial change ($-[\eta + \varphi T]$) is equivalent to Hick's neutral technical change where the marginal substitution between factors is unchanged. Radial technical change quantifies the shifting of the cost function geometrically towards the origin. Non-neutral or dis-embodied technical change ($-\sum_j \lambda_j \ln P_j$) represents the change in the efficiency or quality of the factor inputs in the production process. This would represent a shift towards the cost function or homogenisation within the sample. Scale augmenting technical change ($-\sum_i \kappa_i \ln Y_i$) quantifies the technical change linked to changes in the scale of institutions in the sample.

An associated measure which may also be derived from a time-trend model is the 'bias of technical progress' (see Stiroh, 1997). This measure quantifies how technical change influences the cost shares of inputs in the production process. Using Shephard's lemma (see Shephard, 1970) the 'bias of technical progress' may be written as:

$$\text{Bias of technical progress} = \partial S_j / \partial T = \partial^2 \text{Ln} C / \partial \text{Ln} P_j \partial T \quad (5.15)$$

where S_j is the cost share of input j for $j = 1, \dots, 3$. The value of the bias of technical progress indicates that the technical progress employs more or less of the input in the production process when the value is greater or less than zero respectively.

From the time-trend model, the percentage total factor productivity growth may be viewed as the negative technical change plus one minus the elasticity of cost with respect to outputs multiplied by the change in outputs. This may be written as:

$$\text{Total Factor Productivity growth} = -\text{Technical change} + (1 - \epsilon_{scale}) \dot{Q} \quad (5.16)$$

This measure is an approximation to a total factor productivity growth index (for further discussion of this measure see Nelson, 1990). Assumptions implicit in the measure include perfect competition in the input and output markets and constant returns to scale. Additional problems include potential mis-specification if input prices or outputs are correlated with time. When this occurs, the distinctions between neutral and non-neutral technical change may be ambiguous. For example, when

outputs are correlated with time, scale-augmenting effects and neutral technical change may be blended.

Dissatisfaction has also emerged with time-trend models due to the slowly changing form of productivity implicitly assumed by this measure. This may not accurately represent change year on year due to an inherent averaging process. This approach may therefore mis-specify the degree of variation in technical change over time presenting a smoothed increase or decrease through interaction with the time dummy employed to capture the effect of change over time (Humphrey, 1993). Despite the wealth of information that can be obtained from such a model, it is suggested that this model be used only where data constraints rule out the use of alternative models.

Another econometric cost approach to estimating productivity growth and its components is the cost function shift model (Humphrey, 1993). This model examines how the differences between cross-sectional cost functions or frontiers can illustrate changes in productive technology over time. The model relies on a number of cost functions estimated for specified time periods, which are then used to construct hypothesised measures of average cost. Average cost, in this context, may be viewed as total (fitted) cost divided by some measure of institution scale. Once estimated, average cost may be used in the estimation of technical change. Technical change may be viewed as the proportional decrease in predicted average cost. This may be written as:

$$\text{Technical change} = -[AC_{t+1} - AC_t] / AC_t \quad (5.17)$$

Once a value for technical change has been calculated, total factor productivity growth may be estimated in a similar manner to the time-trend model (equation 5.15). Whilst this measure overcomes many of the problems of smoothing faced by the time trend model, the cost function shift approach also faces a number of potential problems. To estimate such a model, a sector needs to have a substantial number of observations or incumbents for the estimation of a number of cost functions over time. Conversely, the time-trend model may be used for industries with a far lower number of incumbents as the cross-sectional and time-series observations used in the estimation of this model are pooled. Additional disadvantages are that the cost function shift model provides far less information about the components of productivity and may provide more volatile results.

5.9 *Conclusions*

In this chapter, a number of issues have been addressed. First, the different methods of estimation, the choice of functional form, efficiency concepts and measures of efficiency identified are discussed and reviewed. The cost efficiency concept, whilst being less flexible than the profit efficiency concept, is adopted for use in the empirical chapters, due to both the breadth of efficiency considered and to improve comparability with previous studies. Econometric estimation techniques rather than non-parametric (*DEA*) methods of estimation are considered to be more appropriate for use in the empirical chapters. Theoretical, statistical and practical considerations are all forwarded in support of this choice.

To amend for potential mis-specification in the translog functional form, both this form and the flexible Fourier functional forms will be employed in the thesis. The flexible Fourier functional form is limited in application due to the requirement of a substantial data set (approximately 300 observations are considered to be a minimum). This functional form will be employed in chapter 8 where the building society data set will accommodate the practical requirements of this form. In other empirical studies, (chapters 7, 8 and 10) the translog functional form will be employed, although both acknowledgement of, and testing for, possible mis-specification will be undertaken and reported.

A range of statistics used to estimate economies of scale, economies of product mix and cost efficiency were used in the studies reviewed in chapter 4. It is proposed to estimate both measures of cost efficiency, overall economies of scale (*OES*), economies of scope and cost complementarities will be estimated in the empirical studies to test for the presence of economies of scale and economies of product mix. Distribution-free techniques will be used to estimate cost efficiency due to the less restrictive assumptions imposed by these techniques.

Finally, in this chapter, two models of productivity measurement are suggested, the time-trend model and the cost function shift model. Due to the differing data requirements of these models, the productivity study of retail banking (chapter 8) employs a time-trend model due to the small number of incumbent institutions in the market. However, the cost function shift model is used for the productivity study of building societies (chapter 10), where a sufficient number of institutions exist.

¹ A notable exception was provided by Varian (1985).

SECTION TWO

Chapter 6 A general test of competitive conditions in UK building society mortgage market: 1990-95

6.1 Introduction

In chapter 2, a number of changes to the operating environment of UK depository institutions were considered. These changes included the re-regulation of the retail banks and building societies, changes of market structure in both sectors, the development of new technology for delivering banking services and the changing form of branch networks held by depository institutions. All these factors may have had a substantial influence on the level of competition prevalent in UK depository institutions.

Previous regulation limiting the scope of services that depository institutions may provide and the degree of price competition that could be undertaken was repealed during the 1980s and 1990s. These restrictions may have led depository institutions to adopt a number of distinct competitive strategies, including an expansion of their branch networks in an attempt to provide customers with an implicit service return (see for example the discussion of Kolari and Zardkoohi, 1992 and Zardkoohi and Kolari, 1994, in chapter 4). The expansion of branch networks under more restrictive regulation is an example of depository institutions following an objective function that may conflict with cost minimisation and efficiency maximisation. When these regulatory restrictions were repealed, banks and building societies reduced the

number of branches. It is commonly thought that that change led to a greater degree of price competition in the building society sector.

The introduction of new technology to deliver banking services also affected the competitive strategies of UK depository institutions. ATMs, for example, have greatly increased in number and usage, providing both increased service quality and convenience for customers. The use of ATMs particularly in the UK has displayed a willingness on the part of depository institutions to compete in terms of this new technology. Initially, ATMs were established by banks and building societies in isolated networks. During the sample period, a number of competitive alliances between providers were established (e.g. between a number of building societies in the LINK network, or between Natwest and Midland banks), increasing the scale of competing ATM networks. Whilst most other European countries initially experienced a degree of ATM network competition during the 1980s, most have now moved towards a greater degree of co-operation, with France, Belgium, Denmark, Norway and Italy having an integrated or nearly integrated network of ATMs. This level of co-operation has not occurred in the UK where only limited alliances and bilateral agreements exist between depository institutions (see Vesala, 1995, for further discussion of this issue)

The new entrants to the mortgage market in the 1980s, particularly the specialised mortgage lenders and retail banks, may have contributed to the reduced profitability and the large losses that dominated the building society sector in the early 1990s. These entrants may be characterised both by their proprietary form and to a lesser extent the brevity of their presence in the mortgage market.

One important reason for considering the competitive conduct of institutions in a thesis primarily concerning efficiency and productivity is that a lack of competitive pressure may reduce the pressure on managers to perform optimally. The implications of such a 'quiet life' have been considered by Berger and Hannan (1998) who suggest that "... there may also be higher cost per unit output in concentrated markets because of slack management". (p.454). Thus 'slack' managers may follow private objectives or 'satisfice' through following alternative objectives to cost minimisation and efficiency maximisation, such as fostering greater market power at the expense of cost minimisation. Previous analysis of competition in the UK retail bank sector suggests that only a limited degree of contestability exists in this market, implying a lack of long-run competitive equilibrium (Molyneux *et al*, 1996).

In this chapter, the competitive conditions in the UK building society mortgage market will be assessed. This is important and timely for three reasons. First, a study of the competitive conditions in the building society industry is an omission within the literature. Secondly, this analysis should provide greater clarity in determining the competitive conduct of building societies. Lastly, the study is important in accommodating the recommendations made in chapter 4, that both potential sources and biases of efficiency be investigated further. It is posited that the competitive conditions of depository institution markets may be a primary cause of sub-optimal behaviour.

The chapter is structured as follows. Contestable market theory and monopolistic competition are outlined in section 6.2. Section 6.3 provides a review of studies that

have previously considered competition in depository institutions and section 6.4 explains the Rosse-Panzar statistic. The data and variables employed in the study are described in section 6.5. The model specification and results are provided in sections 6.6 and 6.7, respectively. A summary of the chapter is provided and conclusions are drawn in section 6.8.

6.2 *Contestable market theory and monopolistic competition*

Contestability theory (Baumol *et al*, 1982) can be viewed as a special case of classical competitive market theory. The theory was proposed to generalise differences in market structure and as a powerful tool for improving the regulation of industry. It is suggested that “... perhaps the most noteworthy implication of contestability theory is that a wide difference in appearance between a particular market and the form of perfect competition need not deprive the invisible hand of its power to protect the public interest” (Baumol *et al*, 1982, p. 447). The influential nature of this theory in economics has been described as “... a ‘rebellion’ which does without benefit of the conjectural variation, reaction functions, and other paraphernalia of standard oligopoly analysis” (Baumol, 1982, p. 1).

Contestability emphasises the assumption that an ‘imperfect’ industrial structure may allow a long-run competitive equilibrium to form. This is hypothesised to occur through the potential entry of competing firms to the market during disequilibrium. Anticipated competition, both real or imaginary, is viewed to engender competitive behaviour among the incumbents in a market. Central aspects of contestability may be defined as the static form of the model, hypothesised free entry and exit of institutions to the marketplace with no consideration of a possible time lag for retaliation or sunk

costs, and the assumption that potential entrants to the market are price takers, freely accepting the present incumbents' previous entry prices. Thus " ... the critical feature of a contestable market is its vulnerability to hit and run entry" (Martin, 1993, p. 300).

Applying the theory of contestability to the theory of monopolistic competition, developed by Chamberlin (1933), may initially be viewed as contradictory. Chamberlin suggested that product differentiation is the distinguishing characteristic of this market conduct where differentiation is achieved through such factors as " ... quality, design, color, or style", where " ... in so far as these and other intangible factors vary from seller to seller, the 'product' in each case is different" (p. 56). Baumol *et al* (1982) amended for this by suggesting that an " ... entrant can closely or exactly duplicate the product design of the firm depicted" (p. 332), or if each variant is sold by at least two suppliers, perfect contestability will lead to marginal cost pricing. Martin (1993) considers this revision in greater depth.

Firms operating in a monopolistically competitive market selling differentiated products can be viewed to be qualitatively indistinguishable from classical profit maximising monopolists at the firm level. To elaborate this point, Chamberlin (1933) emphasised that as " ... long as the substitutes are to any degree imperfect, he (*the firm*) still has a monopoly of his own product and control over its price within the limits imposed upon any monopolist - those of demand" (p. 67). The mortgage provider could therefore be assumed to provide a financial service with distinct characteristics. The provision of a mortgage with these exact characteristics is possible only through the one mortgage provider. Thus at the firm level a monopolistically competitive firm and a monopolist are indistinguishable. Following Baumol (1982), whilst characteristics may be viewed as distinct on a firm level, at the

industry level substitute characteristics, deemed equivalent by the consumer, may be incorporated within the service by other incumbent providers. A clear distinction between monopolistic competition and profit maximising monopoly may then be observed at the group or industry level. Analysis of monopolistic competition can be seen to be concerned with both an individual equilibrium and group equilibrium (Panzar and Rosse, 1987). To test for this situation the comparative static approach of the Rosse-Panzar statistic is employed.

6.3 *Previous studies*

A number of previous approaches have been used to quantify competition in banking. These may be broadly divided into the Structure Conduct Performance (SCP) paradigm (from Bain, 1951) and New Economic Industrial Organisation (NEIO) approaches (see Bresnahan, 1989).

The SCP paradigm considers how the observable characteristics of a market or industry may affect the conduct and performance of participants within the market. Development of this approach was partly driven by difficulties experienced in the empirical measurement of concepts used by economic theory (such as marginal cost and elasticity) and partly by the desire to build a theory of sub-optimal behaviour.

Previous SCP studies have used a range of proxies to represent market structure and market performance. Market structure has been represented by the relationship between buyer and seller costs, the degree of product differentiation, the degree of concentration within a market place, the degree of market share, and the entry

conditions for potential new firms. Profitability, the relation of rates of return to assets, the scale of the costs of selling and efficiencies have been used to represent market performance. Other factors, including risk, leverage, buyer and seller concentration and foreign competition and macro-economic factors, have all been considered in various SCP studies.

Principal findings of SCP studies include the rejection of the concept of long-term equilibrium, a link between the performance of the firm and the concentration of the market and the connection between performance and the market power of individual firms. Problems with the SCP paradigm have been numerous, although the central problem of whether high profits indicate good or bad performance has been a serious block on further research.

A number of empirical studies have tested hypotheses relating to the SCP framework for depository financial institutions. Lloyd-Williams and Molyneux (1994) considered the effect of governmental and regulatory pressures upon banks to merge and the consequent effects upon market structure and conduct in a study of Spanish banks. Assuming that a higher degree of market concentration will engender collusion among firms in the industry, it follows that within an industry of increasing concentration Spanish banks may receive substantially larger profits, irrespective of other changes within the industry, such as efficiency. This hypothesis was tested using a sample of 92 banks for the 3-year period 1986-1988. The results suggested that the concentrated market had reduced the cost of collusion and led to higher profits for all incumbents.

Goldberg and Rai (1996) considered the relationships between market structure and performance for a number of European banks between 1988-91. The traditional SCP hypothesis and the structure-efficiency hypothesis were tested using a range of performance indicators and a translog cost model specification with an intermediation approach. Little support was found for either hypothesis.

Molyneux and Forbes (1995) consider the SCP paradigm for banks in 18 countries. The sample taken between 1986-89 includes banks operating in a broad range of regulatory environments. The study provides empirical support for the traditional SCP paradigm concluding that the degree of concentration has an effect on the level of competition within the industry. Other studies have also been conducted by Bourke (1989), Berger and Hannan (1989) and Molyneux *et al* (1996). A review of many of the SCP studies (pre-1984 studies, primarily from the US) is provided by Gilbert (1984)

The New Economic Industrial Organisation paradigm (NEIO) (see Bresnahan, 1989) considers market power in single markets or industries. Two central ideas dominate the NEIO paradigm. First, it is assumed that price-cost margins are not directly observable (i.e. marginal cost is not directly observed), but may be inferred from firm behaviour. Secondly, the view that individual industries have important idiosyncrasies is assumed. This point implies that the institutional features of an industry may influence the conduct of individual firms. Firm and industry conduct are viewed as unknown and have to be estimated.

Previous studies of competition in depository institutions and financial markets, using techniques other than the Rosse-Panzar statistic, have been limited. Notable examples include Spiller and Favaro (1984), Suominen (1994), and Vesala (1995). Spiller and Favaro (1984) considered the form of competition that was present in the Uruguayan banking sector. The analysis suggested that banking firms do not have the same conjectures, or respond differently to actions of other banking firms within the same market. Size was viewed to be a determinant of differing conjectures. Suominen (1994) considered a two product test for competition in the Finnish banking industry between 1986 and 1990. It was concluded “... that some monopoly power was present in the pricing of banking services during the late 1980s” (p. 107). Vesala (1995), in a research monograph, estimated a number of NEIO models, including the Rosse-Panzar statistic, also for the Finnish banking industry during the late 1980s and early 1990s. A broad range of issues were considered, including the nature and level of oligopolistic behaviour, the measurement of price competition over time, and an analysis of the interdependence between deposit and loans market.

Studies that apply the Rosse-Panzar statistic in relation to depository institution markets include Nathan and Neave (1989), Molyneux *et al* (1994) and Vesala (1995). Nathan and Neave (1989) applied the Rosse-Panzar statistic to a sample of Canadian financial institutions. Cross-sectional samples of 14 schedule A and 58 schedule B banks and 39 trust companies were considered for 1982 and 1984. Monopolistic behaviour was indicated for Canadian financial institutions. Molyneux *et al* (1994) used the Rosse-Panzar statistic to assess competitive conditions in a number of European banking markets. This broad ranging study incorporated a number of variables to control for risk, cost and size characteristics of the institutions considered.

A sample of German, French, Italian, Spanish and UK banks were considered for the period 1986-89. The study concluded that monopolistic competition existed in the UK banking market (a result of 0.628 was estimated for the Rosse-Panzar statistic). Similar results were obtained for the other European markets. Vesala (1995) used a similar approach to assess the levels of competition in Finnish banks between 1985 and 1992. Cross-sections were estimated for every year. A substantial increase in the level of competition of Finnish banking was observable over the sample period, with the H statistic estimates rising from 0.182 in 1985 to 0.620 in 1992. This increase in competition coincided with the substantial re-regulation of the Finnish banking sector in 1986.

6.4 *The Rosse-Panzar statistic*

The Rosse-Panzar statistic may be used to test for long-run competitive equilibrium (or perfect competition), monopoly (or perfect cartel conditions) and monopolistic competition (or long-run Chamberlinian equilibrium). The testing procedure is undertaken in two stages. Validity of the overall or competitive equilibrium test requires that the sample be in long-run equilibrium. Thus, presence of long-run equilibrium has to be tested first. The competitive environment statistic (H), which may then be estimated, can be viewed as the sum of firm level elasticities of total revenue with respect to input prices. This can be written as:

$$H = \sum w_i \partial R / \partial w_i / R \quad (6.1)$$

where w = input price, R = the revenue function, for i firms. The differing interpretations of the H statistic are displayed in Table 6.1.

How the statistic enables testing for distinct forms of market conduct and behaviour may be explained intuitively. The H statistic quantifies the responsiveness of total revenue to a proportional increase in all input prices. Cost is assumed to be linearly homogeneous in input prices, so that a one per cent increase in input prices will inflate costs by one per cent for all output levels. The symmetry assumption is imposed *a priori* and presupposes that the quantity of output produced will not vary with differing forms of market conduct.

Table 6.1 *Interpretations of the H statistic*

<i>Estimated H</i>	<i>Industry equilibrium / competitive equilibrium</i>
$H \leq 0$	Monopoly equilibrium. H is a decreasing function of the perceived demand elasticity
$0 < H < 1$	Monopolistic competition (Chamberlinian equilibrium) H is an increasing function of the perceived demand elasticity
$H = 1$	Perfect competition

Source: adapted from Molyneux *et al* (1994) and Vesala (1995).

A profit maximising monopolist will produce where marginal cost equals marginal revenue. If factor prices increase, there will be a proportionate rise in marginal cost (assuming linear homogeneity). This in turn will lead to a reduction in the quantity of output and a rise in price. As the monopolist will only operate on the elastic portion of

the demand curve, the elasticity of total revenue with respect to input prices will be non-positive (i.e. $H \leq 0$).

Under long-run competitive equilibrium (or perfect competition), an increase in factor prices will lead to a proportionate rise in average and marginal costs. This will, in the short-term, reduce revenues and lead to the exit of incumbents. This exit will increase the demand facing the remaining incumbents. Following established theory, in the long-run, an unchanged equilibrium level of output is expected. This is due to the infinitely elastic demand curve faced by firms in a perfectly competitive market. The rise in costs will equally lead to a proportional increase in total revenue for the remaining incumbents so that the elasticity of total revenue with respect to input prices will equal unity (i.e. $H = 1$).

Under monopolistic competition or large group Chamberlinian equilibrium, the firm will produce at a long-run level of output and price that is determined by the tangency between average cost and the demand function¹. The difference between the average cost at which the firm operates and the minimum average cost (where a firm would produce in perfect competition) provides an indication of the level of market power exercised by the firm. A rise in factor prices will increase average cost, which in turn will lead to a fall in both output and revenue. Through imposing the *a priori* assumption that the elasticity of demand of a firm under symmetric monopolistic competition will increase with the number of substitutes for a product, the degree of 'competitiveness' of the market may be quantified. Thus the value of the H statistic between zero and unity should indicate the degree of control incumbent firms possess over their differentiated product markets and therefore the contestability of the market. A lower value of H will indicate a higher level of control over differentiated

product markets or a lower level of contestability. Conversely, a higher value for H will suggest a lower level of control over differentiated product markets and a higher level of contestability.

A number of potential difficulties with the statistic have been suggested. Shaffer (1982) emphasised the importance of considering firms operating in the same market. He also stressed that the presence of many small firms may disguise the presence of disequilibrium. This would cause the estimate of H to fall and make a negative value more likely, regardless of the conduct prevalent within the market. Following recommendations made by Shaffer, the stability of market shares over the sample period should be examined to test if such bias is occurring. When the market shares display stability, a bias by the small firms effect may be rejected. When instability is recognised, a small firms effect may be present leading to lower than expected H statistic results. Additionally, the limited definition of the production process denoted by the truncated functional form, may be viewed as a blunt approximation of the true productive technology (Perrakis, 1991).

6.5 Data

A balanced data set of 77 building societies between 1990 to 1995 is employed. The data, drawn from the Annual Reports and Accounts of the building societies, are pooled for two periods of 1990-1992 and 1993-1995 and deflated to 1993 prices by the Retail Price Index. An incomplete data set is used (the total number of building societies in 1990 was 99) in order to obtain a balanced and contiguous set.

The factor prices used in the empirical model are defined following the recommendations provided in chapter 3. The price of labour is defined as the total yearly labour costs of a building society divided by the number of full-time equivalent employees. The price of capital is defined as the total annual expenditure on capital goods and infrastructure in addition to the level of depreciation divided by the level of fixed assets held by the building society. Criticisms of this definition and possible bias that may consequently occur are discussed in chapter 3. The price of deposits is defined as total interest paid divided by the level of deposits held by the individual building societies.

A number of environmental variables are included in the revenue function to control for firm specific and external factors that may be associated with revenue. By controlling for factors that may systematically vary with the dependent variable estimation bias may be reduced. Total assets (*Ass*) are used to control for different building society sizes and potential economies of scale. Other control variables include the loans to assets ratio (*LAR*) and the level of provisions for bad and doubtful debts (used as a proxy for the level of risky behaviour and the differing levels of asset quality). This variable (denoted by *PD*) is included to control for the potentially higher or lower profits that may be associated with risky behaviour.

Average levels of building society profit are variable over the sample period. An average profit of approximately £16m was recorded between 1990 and 1991 across the 77 societies. A considerable rise in average profits is recorded for 1994 and 1995, when levels of £22m are achieved. Interest payable and receivable, both overall and specifically on retail deposits and non-retail deposits have displayed a gradual decline. This trend mirrors the underlying interest rate prevailing within the economy

as a whole. The differential between interest received and interest payable rises over the period from £46m in 1991 to £64m in 1995 indicating an overall rise of nearly 40 per cent over the entire period. This change is perhaps an indication of the instability of interest rates, a reduction in the level of competition or a move towards greater internal reserve generation.

6.6 Model specification

To investigate competitive conditions, a revenue function is specified, assuming an intermediation model of bank production (see Sealey and Lindley, 1977). Further discussion of the intermediation model is contained in chapter 3 and chapter 7. It is assumed that mortgage loans are produced using labour, capital and deposits. Total revenue from mortgages (TR) is the interest receivable on mortgages. The model form employed follows the approach adopted by Molyneux *et al* (1994). This allows comparison of our results with the estimations for the UK retail-banking sector.

The equilibrium test employed is similar to that used in previous studies (for example Molyneux *et al*, 1994, and Nathan and Neave, 1989). The equilibrium test is based on the assumption that in equilibrium, long-run competitive capital markets will equalise risk-adjusted rates of return across financial institutions. It would therefore be expected that, in equilibrium, the rates of return should not be correlated with input prices. This is tested by imposing return on assets (ROA) as the dependent variable in the regression equation. If a significant value is noted in the sum of parameter coefficients for factor prices, a level of correlation or association between the rate of return and factor prices is reported. This would be an indication of dis-equilibrium conditions. If the sum of parameter coefficients for factor prices is not significantly

different from zero, association or correlation between the rate of return and factor prices is rejected. This result would suggest that equilibrium conditions are prevalent in this market.

The empirical model may therefore be represented as:

Competitive environment test

$$\text{LnTR} = \alpha + \beta \text{LnP}_L + \chi \text{LnP}_K + \delta \text{LnP}_F + \varphi \text{LnAss} + \phi \text{LnPD} + \eta \text{LnLAR} \quad (6.2)$$

Equilibrium test

$$\text{LnROA} = \gamma + \kappa \text{LnP}_L + \lambda \text{LnP}_K + \mu \text{LnP}_F + \nu \text{LnAss} + \pi \text{LnPD} + \tau \text{LnLAR} \quad (6.3)$$

Where:

Ln = Natural logarithm

TR = Total mortgage interest revenue (i.e. total revenue)

ROA = Return on assets (the ratio of profits after tax to total assets)

P_L = Labour expense per full time employee

P_K = Capital expenses per pound of fixed assets

P_F = Ratio of interest payable on retail funds to total retail funds (i.e. the unit price of retail funds)

Ass = Total assets

LAR = Mortgage to assets ratio

PD = Ratio of provision for bad and doubtful debts to total assets

and $\alpha, \beta, \chi, \delta, \varphi, \phi, \eta, \gamma, \kappa, \lambda, \mu, \nu, \pi$ and τ are parameters to be estimated.

6.7 Results

Parameter estimates, diagnostic statistics and H statistics are displayed in Table 6.2.

The majority of t statistics for the competitive environment test are statistically significant, suggesting a low level of estimation error. The model fit diagnostic statistics (the adj. R^2 and F statistics) indicate a high level of model fit for the 1990-92 period with this statistic. A lower level of model fit (although still statistically significant) is recorded for the 1993-95 period. The log likelihood statistic² is substantial for both time periods, indicating that a linear form of the equation is rejected. The Durbin-Watson statistic is a test for autocorrelation and may be regarded as a test of mis-specification in this model. The presence of substantial autocorrelation for either time period is rejected. The majority of t statistics for the equilibrium test are also statistically significant, suggesting a low level of estimation error. The diagnostic statistics for model fit, with the equilibrium test, differ between the two time periods. The 1990-92 period exhibits a low, but, statistically significant level of model fit, whilst in the 1993-95 period, model fit may be rejected. Linearity and substantial autocorrelation are both rejected for the equilibrium statistic.

The coefficient for the price of labour is positive for both periods yet significant for the 1993-1995 period. Estimates of the capital price coefficient are insignificant in both equations and both time periods. The deposit price coefficient is significant and positive for both periods indicating the relative importance of this input in the revenue function.

Table 6.2 Parameter estimates, diagnostic statistics and results

Competitive	Environment test		1993-1995	
	1990-1992			
α	-0.704	(0.681)	-1.083	(0.128)*
β	0.004	(0.005)	0.008	(0.005)*
χ	-0.008	(0.008)	0.002	(0.008)
δ	0.561	(0.022)*	0.494	(0.035)*
ϕ	-0.009	(0.002)*	0.010	(0.002)*
φ	0.505	(0.040)*	0.493	(0.037)*
η	0.004	(0.001)*	0.016	(0.002)*
Equilibrium test				
γ	-2.195	(0.484)*	-5.720	(0.713)*
κ	-0.089	(0.038)*	-0.044	(0.029)
λ	-0.038	(0.056)	-0.023	(0.045)
μ	0.477	(0.158)*	-0.167	(0.019)*
ν	-0.077	(0.016)*	0.004	(0.011)
π	1.597	(0.285)*	0.280	(0.204)
τ	-0.008	(0.008)	0.003	(0.009)
	Competitive environment test		Equilibrium test	
1990-1992	Adj. R ² = 0.840		Adj. R ² = 0.185	
1993-1995	Adj. R ² = 0.565		Adj. R ² = 0.016	
1990-1992	F statistic = 208.01*		F statistic = 9.910*	
1993-1995	F statistic = 52.200*		F statistic = 1.640	
1990-1992	Durbin Watson = 1.954		Durbin Watson = 2.121	
1993-1995	Durbin Watson = 2.010		Durbin Watson = 2.192	
1990-1992	Log Likelihood = 311.652		Log Likelihood = -153.150	
1993-1995	Log Likelihood = 330.071		Log Likelihood = -76.824	
	Competitive environment test		Equilibrium test	
1990-1992	0.557 (0.023)*		0.3505 (0.164)*	
1993-1995	0.504 (0.036)*		-0.2338 (0.201)	

Note: standard errors are in brackets. * = Significantly different from zero at 10% significance.

The control for total asset size (*Ass*) appears indecisive, shifting sign between the time periods for both equations, perhaps suggesting a changing relationship between the scale of building societies and the other variables. The *PD* coefficient is significantly positive for both time periods and in both equations. This result indicates that the level of provisions for risky loans and revenue and the levels of provisions for risky

loans and return on assets may both be positively correlated. The *LAR* coefficient is positive and significant for both periods with the competitive environment test.

The results from the equilibrium test are mixed. The sum of parameter coefficients for factor prices is significantly different to zero in the 1990-92 period indicating that dis-equilibrium may be present. The sum of parameter coefficients of factor prices are not significantly different from zero in the 1993-1995 period suggesting that equilibrium is present in 1993-95 period. The results indicate only weak inference may be drawn for the dis-equilibrium period.

The competitive environment *H* test allows rejection of long-run competitive equilibrium or monopoly equilibrium for both time periods. The value of the *H* statistic falls from 0.565 to 0.504 between 1990-92 and 1993-95. This indicates that a increasing level of market power is observed over the sample period. To summarise, an increasing level of monopolistic competition may be observed for the building society sector between 1990 and 1995.

Following the recommendations made by Shaffer (1982), the presence of market stability is also tested. The degree of market stability of incumbents' market share is quantified with three methods. First, a sum of absolute changes in market shares for the building society mortgage market for the top 5, 10 and 15 building societies is taken, following the method used by Hardwick (1996). Secondly, the correlation of market shares and ranks in the mortgage market in different years is made. Thirdly, a test of independence of market shares in the mortgage market between different years

is performed using the Wilcoxon-Mann-Whitney test³. Results of the tests are presented in Table 6.3.

Table 6.3 Tests of market stability

		Sum of absolute changes in market shares				
		Top 5	Top 10	Top 15		
1990-1991		-0.171	0.000	0.102		
1991-1992		-0.136	0.013	0.005		
1992-1993		-0.041	0.035	0.014		
1993-1994		-0.077	0.013	-0.002		
1994-1995		0.323	0.119	0.069		
Correlation of market shares and market share ranks						
Rank	1990	1991	1992	1993	1994	1995
Market share						
1990	-	0.999	0.998	0.997	0.996	0.994
1991	0.999	-	0.999	0.998	0.997	0.995
1992	0.998	0.999	-	0.999	0.998	0.995
1993	0.998	0.999	0.999	-	0.999	0.997
1994	0.998	0.998	0.998	0.998	-	0.997
1995	0.984	0.978	0.998	0.980	0.984	-
The Wilcoxon-Mann-Whitney test						
Z	1990	1991	1992	1993	1994	1995
1990	-	-0.042	-0.030	-0.077	-0.103	-0.364
1991	-0.042	-	-0.056	-0.042	-0.061	-0.310
1992	-0.030	-0.056	-	-0.101	-0.115	-0.387
1993	-0.077	-0.042	-0.101	-	-0.028	-0.295
1994	-0.103	-0.061	-0.115	-0.030	-	-0.279
1995	-0.364	-0.310	-0.387	-0.295	-0.279	-

The sum of absolute changes in market shares indicates a low level of absolute change in market shares amongst the largest incumbent building societies during the sample period. The correlation of ranks and market shares are presented in the middle of the table. These statistics indicate high levels of correlation between different years in the building society mortgage market and only a slight decline over time. The Wilcoxon-Mann-Whitney test is used to test whether the market shares for a specific year have been drawn from the same population as another year. The null hypothesis

cannot be rejected for any pair of years with 0.01 significance. These results support the hypothesis that market shares were stable during the sample period.

6.8 Conclusions

This study suggests that monopolistically competitive conduct was present in the UK building society mortgage market during the early 1990s. The level of market power of incumbent building societies also exhibits a slight rise over the sample period. This is reflected in the rising level of interest spread experienced by building societies throughout the sample period. Such a result occurs at a time of rising profitability of the sector as a whole.

Bias in the Rosse-Panzar statistic resulting from instability in market share of building societies is rejected. The UK building society mortgage market appears to be operating under similar competitive conditions to the UK retail bank loan market, as reported by Molyneux *et al*, (1996). A slightly lower level of market power in the building societies mortgage market was indicated. It may be concluded that the building society mortgage market does not appear to be particularly competitive. The estimated values for the H statistic indicate a level of market power approximately half-way between monopoly and perfect competition.

The implications of these findings for the efficiency and productivity of the sector are substantial. As building societies operate with a degree of market power and not at minimum average cost, opportunities for 'slack' management to follow alternative objectives to cost minimisation and efficiency maximisation may exist. This would result in a reduction in the level of efficiency and productivity below what would

otherwise be expected. Possible avenues for future research into the measurement of levels of competition between depository institutions will be considered in chapter 11.

¹ See Varian, 1984, for further discussion of these specific market forms.

² As the model is estimated with an Ordinary Least Squares estimator, the log-likelihood test only tests for log-linearity and not for model fit, as is possible when a Maximum Likelihood estimator is employed (see Greene, 1993 and chapter 9 for further discussion).

³ The Wilcoxon-Mann-Whitney test is a non-parametric statistic, which is used to test if two groups of ordinal or continuous data have been drawn from the same population. The null hypothesis of the test is that the two samples have the same distribution. The alternative hypothesis is that one sample (X) is stochastically larger than the other sample (Y), i.e. it is a directional hypothesis. The two samples are deemed to be part of the same distribution if the 'bulk' of the sample X, is similar to the 'bulk' of sample Y. The test is conducted by considering the average ranks of the two samples (X or Y). If the average ranks are similar then the null hypothesis is not rejected. Conversely, if the average ranks do differ the null hypothesis is rejected and the alternative would not be rejected. Significance levels are determined in relation to Wilcoxon-Mann-Whitney distribution (from tables). See Siegal and Castellan (1988) for further discussion of this statistic.

Chapter 7 *Economies of scale, economies of product mix and cost efficiency in the UK retail banking sector*

7.1 *Introduction*

This chapter provides estimates of distribution-free cost efficiency, economies of scale, economies of scope and cost complementarities in the UK retail banking industry. Two distinct approaches to modelling bank production, the production and intermediation approaches, are employed. A fixed-effects panel data model with a translog specification is used to provide estimates over the sample period 1984-1997. To the author's knowledge, this is the first study to estimate the cost characteristics of the UK retail-banking sector over time.

The properties of the production and intermediation models were discussed in chapter 3. Theoretically, both models appear to represent the activities of financial institutions adequately but differences may exist in terms of statistical estimation. Following recommendations forwarded in chapter 3, both models are estimated. Differences in terms of model fit, approximation error and specification error will be assessed in order to suggest which modelling approach is superior in terms of the statistical criteria (though it must be emphasised that neither model can be rejected in terms of its economic interpretation).

Many studies in the USA and Europe (but not the UK) have examined the existence of economies of scale and scope in banking (see chapter 4). The importance of a cost study of UK retail banks lies primarily in the provision of empirical evidence as to

whether many of the recent changes in British banking have been driven by cost factors. The only previous efficiency studies of UK retail banking are Allen and Rai (1996), Drake and Howcroft (1994) and Altunbas *et al* (1995)¹. Allen and Rai (1996) estimated a global cost function with bank observations from 15 countries, including the UK. Drake and Howcroft (1995) used a non-parametric DEA (data envelopment analysis) method to calculate the relative efficiency of bank branches for a UK clearing bank. Optimal bank branches were found to be those branches which had total lending of £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 (also see Molyneux *et al*, 1996). An econometric frontier model was employed with assumptions of an exponential distribution of inefficiency imposed. The model assumed that banks and building societies employed three inputs to produce earning assets, the only output included in the model. Overall, very low levels of inefficiency were found.

The chapter is structured as follows. Section 7.2 outlines the model specification, while the methods used in the estimation of the cost models are outlined in section 7.3. The data set is described in section 7.4 and results from the cost models are presented in section 7.5. A summary of the chapter and conclusions are provided in section 7.6.

7.2 Model specification

As mentioned in section 7.1 and following the suggestion in chapter 3, both the production and intermediation models of depository institution production are estimated in this chapter.

In the *intermediation* approach to modelling bank production, banks are assumed to borrow funds and transform them into loan funds, which form the principal output from the productive system. Banks are assumed to use deposits as raw materials or intermediate products within the production process. These inputs are transformed into the final product, loans and ancillary business (the latter output being outside the scope of the principal intermediating process). Total cost consists of operational cost (i.e. labour and capital cost) and the interest cost of attracting the borrowed funds that are then re-lent. This definition of monetary inputs and outputs ignores the effects of differing asset quality and risks associated with lending.

In the *production* approach, depository institutions may be viewed as transforming two input groups (capital and labour; X_1, X_2) into three output groups (loans, investments and deposits); Y_1, Y_2, Y_3 . Cost is defined as operating cost (OC).

In both models, outputs are quantified by their values at the end of the financial year. Y_1 denotes the aggregate of loans issued by the bank at the end of that year, Y_2 represents the total deposits held by the bank, and Y_3 denotes investments held by the bank. The price of labour (P_1) is proxied by the total annual wage bill divided by the number of full-time equivalent employees. The price of capital (P_2) is proxied by the

aggregation of annual property and equipment rentals and depreciation divided by the book value of physical capital, multiplied by 1000, to provide a 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.² The level of provisions for bad and doubtful loans ($PROV$) is included as a variable to control for the effects of risk and differing levels of asset quality in the model. $PROV$ is measured by the value of provisions at the end of the financial year. Previous studies that have included the level of provisions for bad and doubtful debts include Berger and Mester (1997) who suggest that this variable may act as a control for different asset qualities. Although previous work into the effect of provisions of bad and doubtful loans on asset quality has generally been inconclusive (see, for example, Berger and DeYoung, 1997), a negative partial derivative of $\log C$ with respect to $\log PROV$ would be expected as a higher provision would represent poorer management leading to lower cost efficiency.

7.3 *Estimation of the cost models*

The model is estimated using a one-way fixed-effects model and the results are used to estimate measures of economies of scale and economies of scope. Panel data models attempt to amend for differences in time and firm-specific changes, which are not considered in pooled data analysis.

The basic linear relationship for a panel data model, may be defined as

$$Y_{it} = X_{it}\beta + u_{it} \quad (7.1)$$

where $i = 1, 2, \dots, n$ denotes, for example individual firms, and incorporates the cross-sectional dimension of the model; $t = 1, 2, \dots, T$ denotes time and the time series dimension of the model. Y_{it} is the it^{th} observation of the dependant variable and X_{it} is the it^{th} observation of the explanatory variables. β represents the coefficients of the explanatory variables. For a one-way error component model, the error term u_{it} may be written:

$$u_{it} = \mu_i + v_{it} \quad (7.2)$$

where μ_i represents the time invariant individual specific effects and v_{it} denotes the remaining error. In the fixed-effects model, it is assumed that μ_i are constant or fixed parameters to be estimated, whilst v_{it} are independent and identically distributed stochastic terms. These terms may be collected to represent the fixed-effects linear relationship as:

$$Y_{it} = \mu_i + X_{it}\beta + v_{it} \quad (7.3)$$

This model may be estimated using a least squares dummy variable (LSDV) model. The procedure for estimation is set out in detail in Greene (1993, 1995), Baltagi (1995) and Intriligator *et al* (1996). 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. The fixed-effects model thus

operates through conditional inference. 'Effects' models aggregate both period invariant and individual invariant variables with individual time varying variables.

Due to the limited number of observations in the retail banking data set, a translog functional form (rather than a flexible Fourier functional form) is used to represent productive technology. Further explanation of this decision is given in chapter 5. A non-decreasing relationship between inputs and outputs, concavity of the cost function and homogeneity of degree one in input prices is assumed. The expansion of the cost functions into Diewert flexible, second-order translog cost models allows the potential benefits of multi-product production to be estimated. The relaxation of the output homogeneity restrictions allows 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 as follows, where for simplicity the *i* and *t* subscripts are omitted.

The production model

$$\begin{aligned} \ln OC = & \sum_j \alpha_j \ln Y_j + \sum_r \beta_r \ln P_r + \frac{1}{2} \sum_j \sum_s \chi_{js} \ln Y_j \ln Y_s + \frac{1}{2} \sum_r \sum_q \omega_{rq} \ln P_r \ln P_q + \\ & \sum_j \sum_r \delta_{jr} \ln Y_j \ln P_r + \eta \ln PROV + \frac{1}{2} \lambda (\ln PROV)^2 + v + v \end{aligned} \quad (7.4)$$

for $j, s = 1, 2, 3$ and $r, q = 1, 2$.

The intermediation model

$$\begin{aligned} \text{Ln}C = & \sum_j \alpha_j \text{Ln}Y_j + \sum_r \beta_r \text{Ln}P_r + \frac{1}{2} \sum_j \sum_s \chi_{js} \text{Ln}Y_j \text{Ln}Y_s + \frac{1}{2} \sum_r \sum_q \omega_{rq} \text{Ln}P_r \text{Ln}P_q + \\ & \sum_j \sum_r \delta_{jr} \text{Ln}Y_j \text{Ln}P_r + \eta \text{Ln}PROV + \frac{1}{2} \lambda (\text{Ln}PROV)^2 + v + v \end{aligned} \quad (7.5)$$

and for $j, s = 1, 2$ and $r, q = 1, 2, 3$.

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

$$\sum_r \beta_r = 1, \quad \sum_r \omega_{rq} = 0, \quad \sum_j \delta_{jr} = 0, \quad (7.6)$$

In the equations, *PROV* represents provisions against bad and doubtful debts, *OC* indicates operating cost, *C* represents total cost, *Y* represents outputs, *P* denotes input prices and $\alpha, \beta, \delta, \chi, \lambda, \eta$ and ω are parameters to be estimated (the estimates will of course differ for the two models and will be reported separately). The error term represented here as $v + v$ includes both random error and firm specific effects, incorporating both the intercept and the error term.

7.4 Measures of cost efficiency and the data

Distribution-free cost efficiency is a measure of long-run, firm-specific efficiency. The specific efficiency of individual retail banks is derived through reference to the efficiency of the most cost efficient retail bank within the sample. The advantage of this approach is that it removes many of the strong distributional assumptions of efficiency imposed in alternative techniques such as econometric frontiers (see Drake and Weyman-Jones, 1996, and chapter 5, for further discussion of this issue). Efficiency is derived directly from the individual effects produced by the fixed-effects model, where the individual effects, v , would include the “... unobservable entrepreneurial or managerial skills of the firm's executives” (Baltagi, 1995, p.9). The approach assumes that efficiency is constant over time and random variation in efficiency may be removed through averaging over time. Distribution-free efficiency may written:

$$\text{Efficiency} = \exp(\min[v] - v) = \min[v]/v \quad (7.7)$$

for the i^{th} retail bank, where $\min(v)$ is assumed to be the fixed-effect of most efficient bank in the sample. The measure is bounded by $[0, 1]$ where 1 represents 100 per cent relative cost efficiency. For further discussion of this measure, see DeYoung (1997).

The sample was constructed with data from the Annual Reports and Accounts of 11 UK retail banks from 1984 to 1997. All of the 11 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, TSB, Barclays, Clydesdale, the Co-operative Bank, Lloyds, Midland, Natwest, Bank of Scotland, Abbey National and Yorkshire. Descriptive statistics are presented in Table 7.1.

Table 7.1 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
	<i>Standard Deviation</i>	<i>1023.7</i>	<i>24036.4</i>	<i>25430.0</i>	<i>50232.1</i>	<i>253.9</i>	<i>13.6</i>
1990-1993	Mean	1074.0	25976.5	29670.1	30915.5	470.3	85.8
	<i>Standard Deviation</i>	<i>1113.1</i>	<i>24537.3</i>	<i>27567.6</i>	<i>58695.3</i>	<i>507.9</i>	<i>22.0</i>
1994-1997	Mean	1071.8	30271.6	31971.4	39813.8	499.8	50.3
	<i>Standard Deviation</i>	<i>1077.5</i>	<i>26375.1</i>	<i>27313.5</i>	<i>83362.5</i>	<i>545.5</i>	<i>12.1</i>
Overall	Mean	1145.9	28728.9	31965.0	34381.1	430.0	71.6
	<i>Standard Deviation</i>	<i>1056.8</i>	<i>24535.0</i>	<i>26399.0</i>	<i>64403.4</i>	<i>476.4</i>	<i>28.1</i>

Over the sample period, many substantial changes occurred in the UK retail-banking sector. The levels of operating cost consistently fell. The levels of deposits and loans (by value) initially fell from a high in the 1985-89 period to a low point in 1990-93. This low period was marked by economic recession and the possible increased competition in the retail-banking sector, both between retail banks and new entrants, such as the building societies. Recovery from this position is observed in the 1994-97 period. Input prices also varied substantially over the sample period. The price of labour, initially falling in the 1990-93 period, rose to a high in the 1994-97 period. The price of capital has risen consistently throughout the sample period, perhaps providing an indication of the levels of investment undertaken in this sector. Deposit

prices fell throughout the period 1985-97. This decline in deposit price may be regarded as a reflection of a long-term decline of interest rates and the consequent reduction in interest costs. Further discussion of the performance of the UK retail-banking sector is provided in chapter 2 and by Colwell (1991).

7.5 *Results*

The parameter estimates and diagnostic statistics are presented in Table 7.2. Partial derivatives of the logarithm of cost with respect to the logarithm of input prices and output quantities are displayed in Table 7.3. The estimates of economies of scale, economies of scope and cost complementarities are shown in Table 7.4, and estimates of cost efficiency are presented in Table 7.5. Results for the partial derivatives of the logarithms of cost with respect to the logarithm of input prices and output quantities and economies of scale and scope are presented for the entire sample period and for the periods, 1984-89, 1990-93 and 1994-97. Statistics are not considered over different asset size groups due to the small sample size.

The majority of t statistics are statistically significant for the intermediation model. The relatively low number of significant t statistics for the production model, however, suggests that this model may experience higher levels of approximation error than the intermediation model. The diagnostic tests concerned with the degree of model fit (the adjusted R^2 and F statistic) both indicate high levels of model fit for both the intermediation and production models. The F test for the restrictions (equation 7.6) is significant with the intermediation model and is not significant with the production model. This finding suggests that the data employed may not 'fit' the

production model and specification error is possible. This finding is important, as the assessment of model fit is one of the objectives stated in section 7.1.

Table 7.2 Parameters of the fixed effects models

	Intermediation		Production			Intermediation		Production	
	Coeff.	Standard Error	Coeff.	Standard Error		Coeff.	Standard Error	Coeff.	Standard Error
α_1	-5.916	(0.541)*	-1.634	(2.191)	ω_{33}	0.817	(0.076)*	-	-
α_2	-8.624	(0.832)*	-0.250	(0.274)	ω_{12}	-0.809	(0.070)*	-0.073	(0.054)
α_3	-	-	1.893	(2.060)	ω_{13}	-0.806	(0.114)*	-	-
β_1	-11.334	(0.356)*	0.793	(0.585)	ω_{23}	0.083	(0.028)*	-	-
β_2	-2.037	(0.345)*	0.207	(0.585)	δ_{11}	-0.322	(0.066)*	-0.283	(0.394)
β_3	15.540	(1.010)*	-	-	δ_{12}	0.170	(0.031)*	0.063	(0.112)
χ_{11}	0.981	(0.013)*	0.968	(0.121)*	δ_{13}	0.081	(0.036)*	-	-
χ_{22}	0.030	(0.006)*	0.007	(0.007)	δ_{21}	-0.192	(0.064)*	-0.091	(0.050)*
χ_{33}	-	-	-0.074	(0.086)	δ_{22}	0.073	(0.021)*	-0.003	(0.017)
χ_{12}	-	-	0.955	(0.128)*	δ_{23}	0.190	(0.027)*	-	-
χ_{13}	-0.011	(0.014)	-0.911	(0.047)*	δ_{31}	-	-	0.409	(0.376)
χ_{23}	-	-	0.054	(0.091)	δ_{32}	-	-	-0.095	(0.106)
ω_{11}	0.718	(0.086)*	0.082	(0.034)*	η	-0.219	(0.041)*	0.038	(0.041)
ω_{22}	-0.002	(0.036)	-0.010	(0.032)	λ	0.057	(0.009)*	-0.002	(0.009)

Standard errors in brackets * = significant at 10%

Diagnostic Statistics for the Intermediation model

F statistic for model = 158.92, prob. = 0.000, adj. R² = 0.972

F statistic for the restrictions = 115.79 prob. = 0.000, Log Likelihood = 22.147

Diagnostic Statistics for the Production model

F statistic for model = 696.07, prob. = 0.0000, adj. R² = 0.994

F statistic for the restrictions = 6.598, prob. = 0.23, Log Likelihood = 120.521

Table 7.3 *Partial derivatives of the logarithms of cost with respect to the logarithms of input prices and output quantities*

	Intermediation		Production		Intermediation		Intermediation		Production			
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.		
	Partial derivative of the logarithm of cost with respect to the logarithm of labour price				Partial derivative of the logarithm of cost with respect to the logarithm of deposit price				Partial derivative of the logarithm of cost with respect to the logarithm of capital price			
Overall	0.376	(0.039)*	0.706	(0.041)*	3.501	(0.027)*	-0.079	(0.214)	0.057	(0.193)		
1984-1989	0.169	(0.053)*	0.779	(0.034)*	4.254	(0.359)*	-0.123	(0.209)	0.094	(0.028)*		
1990-1993	0.246	(0.052)*	0.695	(0.037)*	3.961	(0.324)*	-0.023	(0.217)	0.123	(0.029)*		
1994-1997	0.704	(0.059)*	0.647	(0.065)*	2.567	(0.197)*	-0.096	(0.197)	0.062	(0.194)		
	Partial derivative of the logarithm of cost with respect to the logarithm of loan quantity				Partial derivative of the logarithm of cost with respect to the logarithm of deposit quantity				Partial derivative of the logarithm of cost with respect to the logarithm of investment quantity			
Overall	0.754	(0.049)*	-0.029	(0.016)*	0.542	(0.130)*	0.022	(0.018)	-0.029	(0.130)		
1984-1989	0.807	(0.059)*	0.209	(0.134)	0.657	(0.152)*	0.068	(0.029)*	-0.038	(0.028)		
1990-1993	0.676	(0.049)*	-0.115	(0.050)*	0.498	(0.115)*	0.067	(0.019)*	-0.016	(0.021)		
1994-1997	0.764	(0.047)*	-0.067	(0.115)	0.477	(0.141)*	0.066	(0.015)*	0.034	(0.016)		

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Significant, positive estimates of the partial elasticities of cost with respect to the price of labour and deposits are recorded for all time periods for the intermediation model in accordance with established cost and production theory. Negative, but statistically insignificant estimates are found for the partial elasticity of cost with respect to the price of capital price. Both partial elasticities of cost with respect to output quantities are positive in accordance with expectation both overall and for all

time periods. For the production model, both partial elasticities with respect to input prices are positive in accordance with expectation, overall and across time periods. The partial elasticities with respect to output quantity are also significantly positive for deposits, as expected, but are significantly negative for loans. The presence of negative partial elasticities of cost with respect to input prices or output quantities, is important to consider as it provides an indication of specification error.

Table 7.4 Economies of scale, economies of scope and cost complementarities

Economies of scale	Intermediation		Production	
	Coefficient	Standard Error	Coefficient	Standard Error
Overall	0.776	(0.044) [†]	0.384	(0.054) [†]
1984-1989	0.935	(0.046)	0.410	(0.056) [†]
1990-1993	0.704	(0.044) [†]	0.369	(0.053) [†]
1994-1997	0.700	(0.045) [†]	0.376	(0.055) [†]
	Economies of scope		Cost complementarities	
Intermediation	Estimate	Standard Error	Estimate	Standard Error
Overall	-1.577	(0.151)*	-0.007	(0.056)
1984-1989	-2.412	(0.136)*		
1990-1993	-2.377	(0.139)*		
1994-1997	-2.675	(0.408)*		
Production	Estimate	Standard Error	Estimate	Standard Error
Overall	7.923	(26.151)	0.931	(0.050)*
1984-1989	9.374	(9.458)		
1990-1993	6.324	(10.444)		
1994-1997	31.489	(240.89)		

*Note: Standard errors in brackets. † = significantly different from 1 at 10%. * = significantly different from 0 at 10%.*

Overall, statistically significant economies of scale are found in both models (see Table 7.4). An average value of 0.776 is recorded for the intermediation model and an average value of 0.384 is recorded for the production model. The value of economies

of scale for the production model is very low, indicating that a 100 per cent rise in output would only lead to 38 per cent rise in cost (overall). This finding is considered to be less plausible than the economies of scale estimate obtained with the intermediation model. An increase in the degree of the economies of scale is observed over time for both models. For the intermediation model, economies of scale increase from 0.935 in the 1984-89 period to 0.700 in the 1994-97 period. Economies of scale increase from 0.410 to 0.376, over the same period in the production model.

Economies of scope estimates are shown for both model forms, overall and over time periods, in Table 7.6. The intermediation model with two outputs considers the cost advantage of producing loans and investments jointly as opposed to separately. The production model with three outputs (loans, investments and deposits) considers the cost advantage of producing these outputs jointly as opposed to separately. Cost complementarities are also reported for both models.

For the intermediation model, substantial diseconomies of scope are recorded suggesting that there would be a potential cost advantage of producing loans and investments separately. The magnitude of these, significant, estimates also increased over time from -2.412 in 1984-89 to -2.675 in 1994-1997. An insignificant, albeit contradictory estimate of cost complementarity is reported for this model, reducing the degree of confidence in this finding. This may represent a degree of specification or approximation error within the model form.

For the production model insignificant³ economies of scope are reported. Significant and contradictory estimates of cost complementarity are also reported.

In Table 7.5, firm-specific distribution-free cost efficiency (see equation 7.7) estimates are reported for both production and intermediation models of bank production. An average efficiency of 81.9 per cent is recorded for the production model and an average of 98.8 per cent is provided for the intermediation model. This indicates that present levels of output could be provided with 18.1 and 1.2 per cent reduction in inputs for the production and intermediation models respectively. A Pearson correlation coefficient of 0.82 is found 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 differ in magnitude between the model forms with standard deviations of 9.95 and 0.66 estimated for the production and intermediation models respectively. The smaller retail banks appear to be relatively more efficient than their larger counterparts.

Table 7.5 Firm-specific estimates of cost efficiency

	Production	Intermediation		Production	Intermediation
Barclays	72.478	98.782	Royal Bank of Scotland	79.433	98.185
Co-operative	93.503	99.464	Yorkshire	90.939	99.769
Clydesdale	100	100	Bank of Scotland	82.618	98.185
Lloyds	71.678	98.259	TSB	77.267	98.408
Midland	74.285	98.259	Abbey National	85.036	98.632
Natwest	73.434	98.707	All banks	81.879	98.786

7.6 Conclusions

In this chapter, economies of scale, economies of scope and cost efficiency in the UK retail-banking sector are estimated. This analysis differs from previous studies by considering the UK retail-banking sector in isolation from other institutions and over

time. Two models of bank production are estimated, the production and the intermediation models. A panel data fixed-effects model is employed to consider cost efficiency over a number of years, removing any potential bias from disequilibrium effects that may have influenced estimation made by previous cross-sectional studies. A translog specification of productive technology is employed. Cost efficiency is estimated using distribution-free methods removing potential bias that may occur when assumptions concerning the distribution of efficiency are imposed.

Overall, significant and substantial economies of scale are observed for UK retail banks. A preliminary conclusion that may be suggested is that expansion of retail banks, either through internal growth or merger, has some justification in terms of cost efficiency. 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, as indicated by the intermediation model, although the potential cost advantage to be gained from separate production may be trivial. The presence of statistically insignificant economies of scope for the production model provides limited empirical support for the joint provision of loans, investments and deposits. The confidence that may be placed in all the economies of scope estimates is limited due to the provision of conflicting cost complementarity estimates. A low level of firm-specific cost efficiency dispersion exists indicating that only limited cost efficiency improvements may be made within this industry, particularly by the larger retail banks. This result is similar to that reported by Altunbas *et al* (1995)⁴ despite the very different econometric techniques employed, the different model form and the use of a panel data sample in this study.

It may be stated that the estimation of economies of scale may be affected by the model form, the small sample size, the techniques applied and the different definition of costs used in the two model forms applied. Such variance in results due to the model employed, whilst indicating the fallibility of analysis, enables a broader view of the problem to be considered. The use of dissimilar definitions of cost is considered to be a primary source of such a difference. 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 economies 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 in the production of loans. Additionally the intermediation approach appears to display a lower degree of approximation error than the production model. This finding suggests that the production approach whilst 'fitting' the data does not concur fully to what would be expected.

In conclusion, the analysis of UK retail banks over the sample period 1984-1997 indicates a variety of results distinct from the findings made in previous studies of this sector. Substantial economies of scale appear to be present in UK retail banks. A variety of economies of scope estimates are produced with the two model forms. These estimates were mostly inconclusive. A low level of efficiency dispersion is observed, with smaller retail banks appearing to experience the highest levels of relative cost efficiency. Lastly, it is decided that due to the substantial level of

approximation error and possible specification error in the production model, the intermediation model will be employed in the remaining studies.

¹ As cited in Molyneux *et al*, (1996).

² The size of these variables are different to those employed in chapter 6, to reduce the degree of variance between factor prices, costs and output variables. Through a reduction in the dispersion of variable sizes, it is hoped bias in standard errors and other residuals used in this analysis will be minimised.

³ Very large standard errors are reported for the production model indicating a high level of approximation error and possibly specification error.

⁴ As cited in Molyneux *et al*, (1996).

Chapter 8 *Total factor productivity growth and characteristics of production technology in the UK retail banking industry.*

8.1. *Introduction*

This chapter provides estimates of total factor productivity growth, technical change and the components of technical change in the UK retail banking sector between 1985-97. A time-trend approach with an intermediation model and a translog functional form is used. To the author's knowledge, this is the first study to consider technical change and total factor productivity growth for UK retail banks.

Total factor productivity growth is the improvement in productive technology generated from changes in the efficiency of production and the state of technology. In chapter 2, it was stated that the UK retail banking industry has been restructured and invested substantially in new technology during the last decade. Amongst such changes are the rapid advances in the computing and communication technology. Innovations in these areas have led to the development of telephone and internet banking and an expansion in the use of a range of financial instruments. This anecdotal evidence of changes in the technology underlying the service provision of retail banks (i.e. technical change) would be expected to have a significant influence on the productive technology and in turn the total factor productivity growth of UK retail banks.

Following the recommendations made in chapter 5, econometric cost efficiency techniques are employed in this chapter. A time-trend model is estimated in this study

due to the relatively small numbers of observations in the UK retail-banking sector. The chapter is structured as follows. A brief outline of relevant literature is provided in section 8.2. Section 8.3 outlines the model specification employed in the study. Data and results are outlined in section 8.4 and a brief summary and conclusions are presented in section 8.5.

8.2 *Relevant Literature*

A number of studies have estimated total factor productivity growth and/or technical change in banking. Most previous studies have considered US commercial banks, although other depository institutions analysed have included Australian building societies (Esho and Sharpe, 1995), German co-operative banks (Lang and Welzel, 1996) and US savings and loan associations (Stiroh, 1997). Most of these studies have modelled production either by means of econometric estimation techniques or non-parametric methods based on index number approaches. Most of the econometric studies have used either a cost or production function to represent productive technology, with a time-trend variable included in the model. Implicit within this approach is the assumption that productive technology is known. It is assumed that variations in the levels of productive technology are captured by the time-trend variable and are representative of technical change.

As mentioned above, previous studies may be broadly delineated into studies that use econometric estimation techniques and studies that employ non-parametric index number methods. Previous econometric studies in banking and other financial services sectors, have fairly consistently found very low levels of positive or negative

technical change and similarly low positive or negative levels of total factor productivity growth.

Esho and Sharpe (1995) estimated the total factor productivity growth and technical change of Australian permanent building societies between 1974 and 1990. A consistent decline in total factor productivity (-2 per cent per annum) and technical change were reported. Hunter and Timme (1986) reported significant levels of positive technical change for 91 large US banks between 1972-82. Particular emphasis was placed on the importance of economies of scale in determining the levels of technical change. This study has been criticised for using only a single output, constructed as the sum of total assets and total deposits.

Humphrey (1993) estimated the technical change of US commercial banks between 1977 and 1988, reporting a number of findings. Between 1977-80, positive technical change was found in the US commercial banking sector. This positive change was reversed as negative technical change was reported for the 1980-82 period. Between 1983 and 1988 minimal technical change was reported. This study also considered the technical change of banks with different asset sizes. Smaller banks, defined as those between \$200m and \$300m in assets, experienced substantial negative technical change over the sample period relative to larger banks.

Stiroh (1997) used two samples of saving and loan associations (S & L's) for the sample period, 1990-95 to estimate technical change and other economic characteristics. The two samples included a complete set of S & L's (i.e. including observations of S & L's which had merged or discontinued operations during the

sample period) and a set of the S & L's that operated continually throughout the sample period (i.e. excluding S & L's which had merged or discontinued operations during the sample period). Stiroh (1997) reported that S & L's that operated throughout the sample period (i.e. the S & L's that remained viable and did not merge with other institutions) had a higher level of positive technical change than the complete set of S & L's. This broadly suggested that a lower level of technical change was prevalent in S & L's that had merged or ceased to trade. Overall, low levels of positive technical change were reported for the sample period, with the level of positive technical change stable across asset sizes and falling over time. Lang and Welzel (1996) examined technical change for German co-operative banks between 1989 and 1992 using a panel data cost function model. Small, yet significant, negative technical change was found.

A number of other studies have measured total factor productivity growth and technical change with alternative techniques. Elyasiani and Median (1990) employed a DEA non-parametric analysis of large US commercial banks with assets in excess of \$300m. A significant level of positive technical change was observed in the sample period 1980-85 for 191 large commercial banks. Employing a DEA non-parametric analysis of productivity growth, technical change and technical efficiency, Wheelock and Wilson (1996) found that some US commercial banks experienced both technological improvement and productivity gains between 1984 and 1993. In this study, relatively smaller banks, with assets less than \$300 million, had lower levels of technical change. Daniel and Tirtiroglu (1998) quantified total factor productivity growth for US commercial banks between 1935 and 1991. They initially calculated total factor productivity growth by employing an index number approach, using a

non-parametric Tornquist index, and then attempted to determine what components of total factor productivity growth may be attributed to trend and cyclical components by employing a Kalman filter technique. A declining trend of total factor productivity growth was observed over the sample period, averaging 2.27 per cent per annum. Bukh *et al* (1995) assessed total factor productivity growth for a sample of 159 Norwegian banks between 1980-89, employing non-parametric, Malmquist indices. A fall in total factor productivity growth between 1981–83 and a rise in productivity growth after 1987 were reported.

To summarise, fairly mixed results are reported for total factor productivity growth and technical change in banks and other depository institutions. These conflicting results may have occurred due to the broad range of data samples, the analysis of institutions operating under distinct regulatory conditions and the different techniques employed. Employing econometric techniques, it could be stated that overall most studies have found that low or negative levels of total factor productivity growth and technical change have been prevalent, particularly in the late 1980s and early 1990s. This overall decline in productivity and technical change is not isolated to just banking. Wolff (1991), for example, recorded a decline in total factor productivity growth for the US insurance industry between 1948 and 1986.

8.3 *Model definition*

Following the discussion in chapters 3 and 6, the intermediation model is employed. This approach to modelling production in retail banking (see Alhadeff, 1957, Sealey and Lindley, 1977 and the discussion in chapter 3) views the production process as a

transformation of three inputs (e.g. capital, labour and deposits; X_1, X_2, X_3) into outputs (e.g. loans (Y_1) and investments (Y_2)). Cost is the sum of both interest and operational costs. Banks are assumed to minimise costs. A cost function of such a correspondence may be presented as:

$$C = g(Y_1, Y_2; P_1, P_2, P_3) \quad (8.1)$$

Outputs are quantified by their values at the end of the financial year. The price of labour (P_1) is measured 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 quantity of physical capital, multiplied by 1000, providing a measure of capital cost for every £1000 of physical capital. The price of deposits is represented by total interest payable divided by the quantity of deposits multiplied by 1000. This measure is used to quantify the interest cost of every £1000 of deposits. Following the recommendation made in chapter 3, aspects of risk are incorporated in the model. Risk in production is proxied by the value of provisions for bad and doubtful loans (denoted *PROV*) at the end of the financial year. For further discussion of this variable specifically, and variable definition and model specifications generally, see chapters 3 and 6.

A non-decreasing relationship between inputs and outputs, concavity of the cost function and homogeneity of degree one of input prices are assumed. The expansion of the problem into a second-order translog cost model allows measurement of potential benefits of multi-product production. A time-trend translog cost function (Humphrey, 1993, Baltagi and Griffin, 1988) is used to represent average technical

change over the sample period. The level of technical change for a multi-product framework may be represented by the partial derivative of the logarithm of total cost with respect to time. A production relation could be presented as:

$$y = f(L, K, t), \quad \text{i.e.} \quad y(t) = f(L(t), K(t), t), \quad (8.2)$$

where L represents labour, K represents capital and t represents time. Time is incorporated in a time trend model through the use of a time specific dummy; *i.e.* t for 0, 1, 2, ..., T years. This approach has been developed to incorporate both quadratic time terms and to incorporate the relationships between both time and input prices. The intermediation time-trend cost model (for simplicity the t subscripts are omitted) may be written as:

$$\begin{aligned} \text{Ln}C = & \alpha_0 + \sum_i \beta_i \text{Ln}Y_i + \sum_j \alpha_j \text{Ln}P_j + \eta t + 1/2 \varphi(t)^2 + \\ & 1/2 \left(\sum_i \sum_l \gamma_{il} \text{Ln}Y_i \text{Ln}Y_l \right) + 1/2 \left(\sum_j \sum_k \omega_{jk} \text{Ln}P_j \text{Ln}P_k \right) + \\ & \sum_i (\kappa_i t \text{Ln}Y_i) + \sum_j (\lambda_j t \text{Ln}P_j) + \sum_i \sum_j \delta_{ij} \text{Ln}Y_i \text{Ln}P_j + \\ & \psi \text{Ln}PROV + 1/2 \xi (\text{Ln}PROV)^2 + \varepsilon \end{aligned} \quad (8.3)$$

for $i, l = 1, 2$ and $j, k = 1, 2, 3$.

Following Shephard's Lemma¹, cost share equations are obtained and are used to form the system to be estimated, where:

$$S_j = \partial \text{Ln}C / \partial \text{Ln}P_j = \alpha_j + \sum_k \omega_{jk} \text{Ln}P_k + \lambda_j T + \sum_i \delta_{ij} \text{Ln}Y_i + e_j \quad (8.4)$$

Following Greene (1993), the deposit share equation (S_3) is dropped from the estimation. This step is taken to make the model operational and to “... solve the problem of singularity of the disturbance covariance matrix of the share equations” (p.505). In accordance with established cost and production theory, the following restrictions are imposed:

$$\sum_j a_j = 1, \quad \sum_j \delta_{ij} = 0, \quad \sum_j \lambda_j, \quad \sum_j \omega_{jk} = 0, \quad (k = 1, 2, 3) \quad (8.5)$$

where C represents total costs, Y represents outputs, P represents input prices, t represent time, $PROV$ represents provisions for bad and doubtful loans and $\alpha, \beta, \gamma, \delta, \lambda, \kappa, \varphi, \eta$ and ω are parameters to be estimated. ε and e represents the error terms, which are assumed to have zero means and constant variances and may vary across the three equations. A seeming unrelated regression (SUR) system of three equations is used to estimate a restricted time trend cost model. This estimation technique is discussed in more detail by Greene (1993, 1995).

Average technical change is viewed as the first derivative of the logarithm of cost with respect to time. The average technical change from the translog cost model, may be written as:

$$-[\eta + \varphi t + \sum_i \kappa_i \text{Ln}Y_i + \sum_j \lambda_j \text{Ln}P_j] \quad (8.6)$$

Average technical change can be decomposed into radial (or neutral) technical change. Radial change ($-\left[\eta + \varphi t\right]$) is equivalent to Hick's neutral technical change where the marginal substitution between factors is unchanged. Radial technical change quantifies the shifting of the cost function geometrically towards the origin. Non-neutral or dis-embodied technical change ($-\left[\sum_j \lambda_j \text{Ln} P_j\right]$) represents the change in the efficiency or quality of the factor inputs in the production process. This would represent a shift towards the cost function or homogenisation within the sample. Scale augmenting technical change ($-\left[\sum_i \kappa_i \text{Ln} Y_i\right]$) quantifies the technical change linked to changes in the scale of institutions in the sample.

An associated measure which may also be derived from a time-trend model is the 'bias of technical progress' (see Stiroh, 1997). This measure quantifies how technical change influences the cost shares of inputs in the production process. The 'bias of technical progress' may be written as:

$$\text{Bias of technical progress} = \partial S_j / \partial t = \partial^2 \text{Ln} C / \partial \text{Ln} P_j \partial t \quad (8.7)$$

Where S_j is the cost share of input j for $j = 1, \dots, 3$. The value of the bias of technical progress indicates that as positive technical change occurs more or less of an input may be employed in the production process when the value is greater or less than zero respectively.

From the time-trend model, the percentage change in total factor productivity growth may be viewed as the negative technical change plus one minus the partial elasticity

of cost with respect to outputs multiplied by the change in outputs. This may be written as:

$$\text{Total Factor Productivity growth} = -\text{Technical change} + (1 - \epsilon_{\text{scale}})\dot{Y} \quad (8.8)$$

This measure is an approximation to a total factor productivity index. Economies of scale are estimated using the *OES* formula (see equation 5.1) following the approach outlined in chapter 5 and employed in chapters 7 and 10.

8.4 *Data and Results*

As in chapter 7, the sample was constructed with data from the Annual Reports and Accounts of 12 UK retail banks from 1985 to 1995. All of the 12 banks are recorded in the Annual Abstract of Banking Statistics produced by the British Bankers Association. The data are deflated to 1985 prices using the Retail Price Index. The banks included in the sample are the Royal Bank of Scotland, Standard Chartered, TSB, Barclays, Clydesdale, the Co-operative, Lloyds, Midland, Natwest, Bank of Scotland, Abbey National and Yorkshire banks. A discussion of the descriptive statistics is contained in chapter 7.

Parameter estimates and diagnostic statistics are shown in Table 8.1. Estimates of economies of scale and the partial derivatives of the logarithm of cost with respect to the logarithm of input prices and output quantities are displayed in Table 8.2. Results are presented for total factor productivity growth, technical change and the components of technical change in Table 8.3.

Table 8.1 *Parameter estimates of the time trend model*

	Coeff.	Standard Error		Coeff.	Standard Error		Coeff.	Standard Error
Constant	-3.764	(0.985)*	ω_{22}	0.049	(0.006)*	δ_{23}	-0.036	(0.024)
α_1	-0.176	(0.069)*	ω_{33}	-0.082	(0.018)*	η	-0.032	(0.063)
α_2	-0.177	(0.035)*	ω_{12}	-0.005	(0.001)*	ϕ	-0.001	(0.002)
α_3	1.001	(0.076)*	ω_{13}	0.024	(0.005)*	κ_1	0.007	(0.005)
β_1	0.934	(0.217)*	ω_{23}	-0.016	(0.001)*	κ_2	-0.007	(0.005)
β_3	0.129	(0.148)	δ_{11}	0.002	(0.002)	λ_1	0.019	(0.006)*
γ_{11}	-0.028	(0.029)	δ_{12}	0.002	(0.000)*	λ_2	-0.001	(0.000)*
γ_{22}	-0.026	(0.004)*	δ_{13}	0.030	(0.024)	λ_3	0.009	(0.009)
γ_{12}	-0.015	(0.011)	δ_{21}	0.004	(0.001)*	ψ	-0.023	(0.037)
ω_{11}	0.031	(0.017)*	δ_{22}	-0.002	(0.000)*	ξ	0.012	(0.009)

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Diagnostic Statistics

F statistic for the model = 269.68, prob. = 0.0000, adj. R² = 0.983, Log-Likelihood = 533.17

F statistic for the restrictions = 230.35, prob. = 0.0000, Durbin-Watson = 1.973,

The diagnostic statistics measuring model fit (the adjusted R² and F statistic) display a high level of model fit. A majority of the coefficients are statistically significant at the 10 per cent level, as indicated by the t statistics. The restrictions placed on the model (equation 8.5) are tested using an F statistic and are not rejected. The degree of autocorrelation is assessed by the Durbin-Watson statistic. Autocorrelation (used here as a test of mis-specification) within the sample data is rejected. The log-likelihood statistic allows linearity in the model to be rejected. Positive partial elasticities of cost with respect to input prices are recorded in accordance with established cost and

production theory. The partial elasticities of cost with respect to loan quantity are significant and positive for all time periods in accordance with theory. However, the partial elasticities of cost with respect to investment quantity are negative, although not significant, for all time periods.

Table 8.2 *Economies of scale and partial derivatives of the logarithm of cost with respect to the logarithm of output quantities and input prices*

	Ray scale economies (OES)		Partial derivative of the logarithm of cost with respect to the logarithm of loan quantity		Partial derivative of the logarithm of cost with respect to the logarithm of investment quantity	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Overall	0.896	(0.016) [†]	0.830	(0.076)*	-0.116	(0.103)
1985-1989	0.893	(0.022) [†]	0.810	(0.088)*	-0.097	(0.096)
1990-1993	0.896	(0.015) [†]	0.840	(0.072)*	-0.121	(0.103)
1994-1997	0.898	(0.016) [†]	0.840	(0.071)*	-0.127	(0.111)
	Partial derivative of the logarithm of cost with respect to the logarithm of labour price		Partial derivative of the logarithm of cost with respect to the logarithm of capital price		Partial derivative of the logarithm of cost with respect to the logarithm of deposit price	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Overall	0.176	(0.021)*	0.058	(0.003)*	0.606	(0.033)*
1985-1989	0.055	(0.081)	0.052	(0.003)*	0.545	(0.055)*
1990-1993	0.175	(0.022)*	0.059	(0.003)*	0.599	(0.033)*
1994-1997	0.174	(0.021)*	0.065	(0.003)*	0.673	(0.057)*

Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance;

† = Significantly different from one at 10% significance.

As shown in Table 8.3, low levels of total factor productivity growth (overall, an increase of 0.803 per cent per annum) and only minor shifts in technical change (overall -0.002 per cent per annum) are observed. The constituents of technical

change indicate a small degree of improvement in the productive capabilities of the industry as a whole or radial technical change (overall, 0.037 per cent per annum) and a limited decline in the quality or efficiency of factor inputs (-0.016 per cent per annum). Scale augmenting technical change, the improvement in technical change from economies of scale present in the industry, displays a slight decline overall (-0.023 per cent per annum).

Table 8.3 *Total factor productivity and its components*

	Radial technical change		Non neutral technical change		Scale augmenting technical change		Technical change		Total factor productivity growth	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Overall	0.037	(0.049)	-0.016	(0.027)	-0.023	(0.038)	-0.002	(0.004)	0.803	(0.125)*
1985-1989	0.034	(0.055)	-0.018	(0.028)	-0.026	(0.042)	-0.010	(0.006)	0.508	(0.099)*
1990-1993	0.037	(0.048)	-0.016	(0.026)	-0.024	(0.039)	-0.003	(0.005)	0.211	(0.030)*
1994-1997	0.039	(0.043)	-0.014	(0.027)	-0.003	(0.003)	0.022	(0.033)	1.547	(0.264)*

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Across time, total factor productivity growth increases, rising from 0.508 per cent (per annum) in the 1985-89 period to 1.547 per cent (per annum) in the 1994-97 period. Total factor productivity growth reached a low point in the 1990-93 period, when 0.211 per cent per annum growth was reported. The components of total factor productivity growth display a range of features over time. Radial technical change rises slightly over the sample period from 0.034 per cent per annum in 1985-89 to 0.039 per cent per annum in 1994-97. Non-neutral technical change is slight and negative across all time periods rising from -0.018 per cent per annum between 1985-89 to -0.014 per cent per annum in the 1994-97 period. Scale augmenting technical

change improves over the sample period, increasing from -0.023 per cent in the 1985-89 period to -0.003 per cent in the 1994-97 period.

Bias in technical progress (estimated using equation 8.7) is only recorded for the model overall. A negative estimate is found for the deposit cost shares (-0.072 , with a standard error of 0.012^*)². Positive estimates are found for the capital cost share (0.027 , with a standard error of 0.006^*) and the labour cost shares (0.058 , with a standard error of 0.017^*). The general conclusion that may be drawn from these findings is that labour and capital costs would experience a relatively greater share of overall costs with continued positive technical change. Conversely, deposit costs would both display a relative decline in their share of overall costs with continuing positive technical change. This result broadly concurs with what would be expected in an environment of rising capital investment in a labour intensive service industry. Further explanation of the bias of technical progress measure is presented in chapter 5.

The estimates of economies of scale, shown in Table 8.2, are slightly lower than the economies of scale estimates reported in chapter 7. This difference in the magnitude of economies of scale is attributed to bias resulting from the inclusion of the time-trend 'dummy' in the model³. All estimates of economies of scale indicate positive economies of scale and are significantly different from one (with 10 per cent significance). These estimates therefore may be interpreted as an indication of slight positive economies of scale.

8.5 *Conclusions*

In this chapter, total factor productivity growth and its components are assessed. An econometric time-trend model employing an intermediation specification of productive technology with a translog functional form is used. Overall, very low levels of total factor productivity growth and technical change are observed. Notably the technological potential of the industry appears to be nearly static over this sample period.

Previous evidence from testing the levels of average technical change and productivity growth also indicates either static or negative technical change in the 1980s and 1990s. The presence of very low levels of technical change and total factor productivity growth for such a large and prominent commercial sector is of great concern. Whenever a substantial quantity of resources is employed in a low productivity area, the potential productivity growth of society is adversely affected. The reduction of productivity growth throughout society may, in turn, have a broader influence on future levels of output, the level of sustainable wage increases and inflation.

Explaining the low levels of technical change and total factor productivity growth in banking is therefore of some importance. The results appear to conflict with the anecdotal evidence (for example, the high levels of investment in new technology) forwarded in chapter 2 and mentioned at the beginning of the chapter. This change, therefore, may provide a good starting point for any explanation. All change in a

company may lead to a degree of disorganisation limiting future productivity improvement in the short to medium term. Equally, the introduction of new technology requires substantial efforts from both the organisation and employees to adapt and learn how to use such technology in an efficient manner. It may be hypothesised that we may be observing a short- to medium-term fall in technical change and total factor productivity growth as the retail banks undergo substantial change.

In a similar vein, the introduction of new technology and changes in the organisation of service delivery may have altered the form of the final service provided and the nature of work performed by the employee. If this has occurred, then the explanation for low technical change and total factor productivity growth may be mis-measurement. The possibility of mis-measurement would in turn suggest that we are observing a far broader phenomenon than we initially suggested. These issues will be considered in greater depth in chapter 11.

¹ For further discussion, see Shephard (1970)

² * denotes significance at the 10 per cent level.

³ Esho and Sharpe (1995) report a similar influence in their model, when the inclusion of a time-trend 'dummy' reduced the level of positive economies of scale.

9.1. *Introduction*

In this chapter, firm-specific cost efficiency in the UK building society industry between 1990-96 is estimated. It is important to estimate the relative performance of building societies because of their importance in the functioning of the UK economy. Building societies, in common with most financial firms in the UK, are highly regulated. The analysis of the efficiency of production in building societies should provide a clearer understanding of the ramifications of this regulation in terms of performance, where “ ... the motivation for such regulation demands an understanding of the behaviour of these firms” (Hancock, 1991, p.2). Empirical evidence is also of particular importance when attempting to assess developments, such as the continued external growth of building societies through acquisition and merger and the future regulation of building societies.

Following the recommendations made in chapter 5, this chapter differs from previous studies of efficiency of UK building societies and extends the literature in three ways. First, efficiency is estimated from a data panel, analysing efficiency over time, and so differs from previous cross-sectional approaches. Cross-sectional approaches have been criticised as reflecting a disequilibrium position through only considering a ‘snapshot’ of the industry at a single year. The data panel used in this chapter examines the average efficiency over a seven-year period, which should limit the degree of bias that may be attributed to disequilibria in the sample period. Secondly, the cross-sectional estimation of firm-specific efficiency with techniques such as

econometric frontiers which requires the use of probability distributions. The assumptions that underpin the use of probability distributions may lead to possible estimation bias (see chapter 5 for further discussion). To free firm-specific estimates of efficiency from such assumptions, distribution-free measures of cost efficiency are used. Lastly, productive technology is represented by a flexible Fourier functional form. This functional form should estimate without the specification error that may have adversely influenced previous studies, which used, for example, the translog functional form. To the author's knowledge this is the first study to consider cost efficiency in building societies using a flexible Fourier functional form.

The chapter is organised as follows. Section 9.2 outlines the model specification used in the study. Section 9.3 presents the measure of cost efficiency to be estimated. Section 9.4 outlines data used in the study. Section 9.5 discusses the choice of functional form and considers the estimates derived from the initial cost model and considers potential problems in estimation. Section 9.6 reports the results from the re-estimated models. Section 9.7 provides a summary of the chapter and offers some concluding remarks.

9.2. Model specification employed

Building societies create services as opposed to recognisable goods. Both physical and non-monetary inputs (physical capital and labour) and monetary inputs (funds from a number of sources) are employed to produce services. These services consist of non-monetary services, such as the manipulation and administration of customers'

accounts, and monetary services, such as loans (Hancock, 1985). Following recommendations made in chapters 3 and 7, the intermediation model is employed.

An intermediation model of building society production assumes that building societies aim to minimise costs and employ labour, capital and deposits to produce advances. The intermediation approach in the dual cost function would be written as:

$$C=g(Y_1, Y_2; P_1, P_2, P_3) \quad (9.1)$$

where outputs are quantified by their values at the end of the financial year; Y_1 represents mortgage loans (Class 1 advances) and loans secured on property (Class 2 advances) and Y_2 denotes advances made without security (Class 3 advances). The outputs therefore represent the services that all building societies were allowed to provide (Class 1 and Class 2 advances) and services that only some qualifying building societies were allowed to provide after the re-regulation (Building Societies Act, 1986) of the sector (Class 3 advances). The price of capital, P_1 , is proxied by the aggregation of the annual cost of property and equipment rentals and depreciation divided by the book-value of physical capital. The price of deposits, P_2 , is defined as the value of the total interest payable divided by the value of all deposits. The price of labour, P_3 , is proxied by the total wage bill divided by the number of full-time equivalent employees. C represents the total cost of production for the building society, including all administration expenses, depreciation and interest costs.

To estimate a total cost function over a data panel, a one component fixed-effects model is used. The model is estimated by the two-step least squares dummy variable

(LSDV) model. The procedure for estimation is set out in detail in chapter 7, Greene (1993, 1995) and Baltagi (1995).

9.3. Long-run distribution-free cost efficiency

Distribution-free cost efficiency is a measure of long-run, firm-specific cost efficiency. In all studies that consider firm-specific efficiency the primary technical difficulty (and the probable cause of the differing results reported in most studies) is that of distinguishing random error from firm-specific inefficiency. Most econometric frontier studies have used probability distributions to represent both random error and inefficiency. Long-run, distribution-free measures of firm-specific cost efficiency do not employ probability distributions to estimate firm specific efficiency, avoiding these possible problems. The estimation of distribution-free, firm-specific efficiency is based on the “... identifying assumption ... that cost differences owing to X-efficiency are persistent, while random errors tend to average over time. That is, good management maximises long-run profits by keeping costs relatively low over long periods of time, although costs may fluctuate from trend because of luck and measurement error” (Berger 1993, p. 263). Firm-specific efficiency is derived directly from the individual effects produced by the fixed effects model. Distribution-free efficiency may be written as:

$$\text{Efficiency}_i = \exp(\min[\ln v] - \ln v) = \min[v]/v \quad (9.2)$$

for the i^{th} building society, where $\min[v]$ is assumed to be the individual effect of the most efficient building society in the sample. The measure is bounded by [0, 1] where

1 represents 100 per cent relative cost efficiency (for further discussion of this measure see DeYoung (1997) and chapter 7).

9.4 Data employed

The entire sample has been constructed using data from the Annual Reports and Accounts for all UK building societies operating between 1990 to 1996. The data are deflated using the Retail Price Index for 1990. Overall, outputs, input prices and total costs all display high levels of dispersion both between individual building societies and over time. This high level of dispersion is clearly seen by contrasting the size of the largest and smallest societies in the sample. The smallest building society in the sample is the Londonderry Provident with approximately £1 – 2 million in total assets, whilst the largest society in the sample, is the Halifax with in excess of £50 billion in total assets.

Descriptive statistics are presented in Tables 9.1 and 9.2 for a range of variables both over time (between 1990-96) and over the two asset groups. Due to the low number of incumbents in each sample (1990 has the largest number of building societies with 99 incumbents) only two groups are defined to leave a reasonable number of societies in individual asset groups and to reduce the problem of bias imposed by individual institutions¹. Group 1, defined as 'small', contains all building societies with less than £100m in total assets. Group 2, the 'large' asset group, contains all building societies with more than £100m in total assets. The two groups contained 38 and 61 building societies respectively. The two asset groups are so defined to ascertain if any difference, in terms of cost efficiency, exists between building societies that have a 'qualifying asset holding'² (societies with greater than £100m in total assets) and

those building societies that do not (building societies with less than £100m in total assets). Building societies with a 'qualifying asset holding' may engage in a range of activities sanctioned by the Building Societies Act (1986) including borrowing money from the wholesale money markets and the power to hold class 3 assets³. The data set is unbalanced, yet contiguous, as some building societies have merged or demutualised over the sample period.

Table 9.1 Descriptive statistics: Overall and over asset size.

	Overall	N=597	Small	N=164	Large	N=433
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Class 1 assets (£m)	1706.54	5251.26	19.01	10.68	2728.16	6454.01
Class 2 assets (£m)	62.75	156.02	1.73	1.38	99.28	188.87
Class 3 assets (£m)	61.39	256.14	0.02	0.09	98.51	319.39
Retail deposits (£m)	1625.00	5668.65	23.46	12.77	2593.21	6236.98
Non-retail deposits (£m)	441.18	1237.76	1.13	1.72	708.31	1509.56
Total costs (£m)	177.41	532.11	2.39	1.67	283.37	652.94
Labour price (£)	11264.28	2763.91	11284.33	3833.25	11156.55	2081.91
Capital price (£)	548.71	439.12	768.41	679.26	452.87	219.76
Deposit price (£)	56.89	23.95	58.10	24.41	56.62	23.83

The level of class 1 assets, including mortgage loans for house purchase, rises substantially between asset groups and displays a high level of dispersion in the two asset groups. The average level of class 2 assets also varies between the asset groups with the large building societies holding the largest average amounts. Only a small amount of Class 3 assets, including insurance and estate agency business, is held by the 'small' building societies. This result is expected, as some of the 'small' societies have been legally limited in their ability to issue these assets for much of the sample period.

The labour price varies only slightly across asset groups with the lowest labour price recorded for the 'large' group of building societies (£11,156). The level of capital price appears to decline as the size of building society increases and the deposit price remains stable.

Table 9.2 Descriptive statistics: Overall and over time.

	1990	N=99	1991	N=92	1992	N=87	1993	N=83
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Class 1 assets (£m)	1357.55	4241.51	1549.22	4596.32	1694.23	4902.60	1829.31	5175.89
Class 2 assets (£m)	48.01	126.71	54.24	135.83	57.58	135.24	65.08	151.10
Class 3 assets (£m)	25.07	110.41	32.09	120.11	38.89	137.76	48.82	159.60
Retail deposits (£m)	1313.09	4131.94	1476.79	4477.34	1600.11	4723.64	1714.99	4890.87
Non-retail deposits (£m)	296.00	863.35	368.58	973.67	424.43	1106.59	487.05	1270.57
Total costs (£m)	218.97	686.17	214.58	628.19	196.48	551.67	158.05	435.32
Labour price (£)	10845.2	2370.93	10358.04	2531.59	10870.1	2649.55	11426.08	2777.16
Capital price (£)	649.91	537.29	625.33	649.69	528.84	378.41	490.67	264.51
Deposit price (£)	95.89	7.36	79.15	6.19	61.54	4.36	42.44	4.88
		1994	N=81	1995	N=79	1996	N=76	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	
Class 1 assets (£m)		1923.16	5395.99	1838.54	6280.21	1863.48	6384.79	
Class 2 assets (£m)		68.26	160.60	71.81	178.91	80.34	206.16	
Class 3 assets (£m)		67.61	209.16	100.33	363.03	136.60	493.61	
Retail Deposits (£m)		1792.80	5019.18	1777.46	6142.74	1803.64	6278.74	
Non-retail deposits (£m)		535.20	1407.08	499.37	1496.54	526.54	1543.87	
Total costs (£m)		149.92	408.79	149.33	469.90	136.59	428.68	
Labour price (£)		11646.05	2801.02	11849.34	2984.00	12166.97	2946.41	
Capital price (£)		490.054	290.55	510.13	353.79	512.87	388.05	
Deposit price (£)		36.73	2.318	36.98	1.89	31.58	2.49	

Some of the variables employed in the model also vary substantially over the seven years of the sample period. Class 1, class 2 and class 3 assets have all risen substantially over the sample period. The average level of total cost has consistently declined over the sample period falling from £219m in 1990 to £137m in 1996. The average prices of labour and capital have changed little over the sample period,

slightly increasing and decreasing, respectively. However, the average deposit price varies greatly over the sample period, mirroring the average level of interest paid, which in turn is strongly associated with the prevailing interest rate.

9.5 *The choice of functional form and the cost efficiency model*

In previous studies of the UK building society sector, the translog functional form has been mainly used. This Diewert flexible functional form provides a cost function that could accommodate a second-order differential approximation to an arbitrary, twice continuously differentiable cost function. Such a cost function will satisfy the linear homogeneity in input prices property at any point only in certain parameters or an 'admissible domain' (Diewert, 1971). This representation of productive technology may therefore only be quantified or estimated consistently within a certain range of observations or 'specified domain', leaving the possibility of specification error in estimation.

To amend for this possible difficulty, the flexible Fourier functional form (see Gallant, 1981, 1984) is employed. The relatively large data set employed in this study (overall 597 observations) makes the use of this function possible. The flexible Fourier functional form is a second-order polynomial in the explanatory variables together with a combination of sine and cosine functions in the explanatory variables. This form is Sobolev flexible (see Gallant, 1981) and therefore can estimate elasticities consistently and has negative prediction bias, thus removing the potential for specification bias in the representation of productive technology. The flexible Fourier form includes the translog form nested inside it and should provide a similar

interpretation of productive technology to the translog functional form, whilst reducing the potential for specification error.

Symmetry is imposed on the translog portion of the model. The trigonometric vectors within the model are chosen *a priori*, following the approach adopted by Mitchell and Onvural (1996). The vectors include single output prices $\text{Cos}(Z_j)$ and $\text{Sin}(Z_j)$ and single input prices $\text{Cos}(l_r)$ and $\text{Sin}(l_r)$. Pairs of outputs [$\text{Cos}(Z_j + Z_s)$, $\text{Sin}(Z_j + Z_s)$, $\text{Cos}(Z_j - Z_s)$, $\text{Sin}(Z_j - Z_s)$, $j \neq s$], and input prices [$\text{Cos}(l_r + l_q)$, $\text{Sin}(l_r + l_q)$, $\text{Cos}(l_r - l_q)$, $\text{Sin}(l_r - l_q)$, $r \neq q$], are also employed.

Linear homogeneity, the assumption that an increase in all input prices leads to a similar increase in costs, has been imposed on the flexible Fourier forms in a number of ways in earlier studies. Berger and DeYoung (1997), for example, imposed standard homogeneity restrictions on the translog portion of the functional form.

Developing this approach, Mitchell and Onvural (1996) imposed both standard homogeneity restrictions on the translog portion of the functional form and restricted the sum of the coefficients of all price variables in the trigonometric terms to zero.

Berger and Mester (1997), however, suggest that the normalisation of all input prices, including the trigonometric terms, by one of the input prices, is “... the only way to impose homogeneity on Fourier-flexible specification, since unlike the translog terms, the Fourier terms are not multiplicative” (p.918). The approach adopted by Berger and Mester (1997) is employed. The price of labour, P_3 , is used to normalise other input price terms. Monotonicity and quasi-concavity in input prices are not imposed due to the semi non-parametric technique underlying the flexible Fourier functional form.

The flexible Fourier functional form may be written as follows, where for simplicity the i and t subscripts are omitted:

$$\begin{aligned}
 \text{Ln}(C/(P_3TA)) = & \sum_j \alpha_j \text{Ln}(Y_j/TA) + \sum_r \beta_r \text{Ln}(P_r/P_3) + \\
 & \frac{1}{2} \sum_j \sum_s \chi_{js} \text{Ln}(Y_j/TA) \text{Ln}(Y_s/TA) + \frac{1}{2} \sum_r \sum_q \omega_{rq} \text{Ln}(P_r/P_3) \text{Ln}(P_q/P_3) + \\
 & \sum_j \sum_r \delta_{jr} \text{Ln}(Y_j/TA) \text{Ln}(P_r/P_3) + \sum_j \varphi_j (\text{Cos } Z_j + \text{Sin } Z_j) + \sum_r \gamma_r (\text{Cos } l_r + \text{Sin } l_r) + \\
 & \sum_{js} \varphi_{js} [\text{Cos}(Z_j + Z_s) + \text{Sin}(Z_j + Z_s)] + \sum_{js} \varpi_{js} [\text{Cos}(Z_j - Z_s) + \text{Sin}(Z_j - Z_s)] + \\
 & \sum_{rq} \gamma_{rq} [\text{Cos}(l_r + l_q) + \text{Sin}(l_r + l_q)] + \sum_{rq} \psi_{rq} [\text{Cos}(l_r - l_q) + \text{Sin}(l_r - l_q)] + \\
 & \sum_{rq} \kappa_{rqj} [\text{Cos}(l_r - l_q + Z_j) + \text{Sin}(l_r - l_q + Z_j)] + \\
 & \sum_{rq} \theta_{rqj} [\text{Cos}(l_r - l_q - Z_j) + \text{Sin}(l_r - l_q - Z_j)] + v + v
 \end{aligned} \tag{9.3}$$

for $j, s = 1, 2$ and $r, q = 1, 2, 3$, where $\alpha, \beta, \chi, \omega, \delta, \varphi, \kappa, \varpi, \theta, \gamma$ and ψ are coefficients to be estimated. $v + v$ denotes non-random disturbance of the individual building societies and random error respectively. It is assumed that random error is independent of non-random error. Values used in the construction of the trigonometric terms are contained in Table 9.3.

Table 9.3 Scale Factors and Scaled Variables for the Flexible Fourier

Functional Form

P_r^{min} = sample minimum for the r^{th} input price P_r^{max} = sample maximum for the r^{th} input price Y_j^{min} = sample minimum for the j^{th} output quantity Y_j^{max} = sample maximum for the j^{th} output quantity $Wp_r = 0.00001 - Ln p_r^{min}$ $Wy_i = 0.00001 - Ln y_j^{min}$ $M = Ln p_r^{max} + Wp_r$ $\lambda = 6/M$ $\mu = 6/([Ln y_j^{max} + Wy_j] * \lambda)$ Input price $l = \lambda [Ln p_r + Wp_r]$ Output quantity $Z = \mu [Ln y_j + Wy_j]$					
Overall					
Wp_1^{Min}	2.2609	Wy_1^{Min}	0.6801	λ	1.3476
Wp_2^{Min}	3.6893	Wy_2^{Min}	7.4642	U_1	0.5106
				U_2	0.3290
	Minimum	Maximum		Minimum	Maximum
Z_1	0.000	5.8643	L_1	0.000	5.9999
Z_2	0.000	5.0554	L_2	0.000	5.8166
Large					
Wp_1^{Min}	2.2609	Wy_1^{Min}	1.0520	λ	1.3919
Wp_2^{Min}	0.6801	Wy_2^{Min}	7.4642	U_1	0.5093
				U_2	0.3229
	Minimum	Maximum		Minimum	Maximum
Z_1	0.000	5.8166	L_1	0.000	5.9999
Z_2	0.000	5.0554	L_2	0.000	5.8166
Small					
Wp_1^{Min}	2.0288	Wy_1^{Min}	0.6702	λ	1.2307
Wp_2^{Min}	3.6151	Wy_2^{Min}	4.6315	U_1	1.2257
				U_2	0.9639
	Minimum	Maximum		Minimum	Maximum
Z_1	0.000	5.8105	L_1	0.000	4.2484
Z_2	0.000	4.9702	L_2	0.000	4.4473

As shown in section 9.4, the sample used in this study is characterised by substantial variation. This substantial difference in the scale of institution could lead to inefficient estimates. The spread of costs and output quantities across the sample may produce estimates where random errors are widely dispersed in terms of their variance (i.e. larger societies would be expected to display higher variance in random errors than smaller societies). To amend for this possible heteroscedasticity, all costs and output quantities are normalised by total assets as outlined in equation (9.3). This

amendment produces variables, which should not vary substantially in magnitude. Berger and Mester (1997) suggested that this “ ... is particularly important because inefficiency terms ... are derived from the composite residuals, which might make the variance of the efficiencies dependent on bank size in the absence of normalization” (p.918).

9.6 Preliminary results

The cost model was estimated first using all observations. Parameter estimates from this model are displayed in Table 9.4. A high proportion of the estimates in the estimated cost model were statistically significant for this model and the diagnostic statistics indicate an acceptable level of model fit (an adjusted R^2 value of 0.999 was recorded). Partial elasticities of cost with respect to input prices and output quantities were found to be positive in accordance with established cost and production theory. A likelihood ratio test is used to test⁴ if the flexible Fourier functional form provides a better ‘fit’ for the data than the nested translog functional form. The test statistic may be written as:

$$LR = \text{likelihood ratio test} = -2(L_{TL} - L_{FF}) \quad (9.5)$$

Where L_{FF} and L_{TL} are the maximised values of the log-likelihood functions of the flexible Fourier and translog models. Following Hardwick (1990) the likelihood ratio test is a χ^2 statistic “ ... with degrees of freedom equal to the number of parameter restrictions imposed” (p.455). The likelihood ratio test was significant with 10

degrees of freedom suggesting that the flexible Fourier model provides a better 'fit' for the data than the translog model⁵.

Table 9.4 *The efficiency model: parameter estimates, partial derivatives and diagnostic statistics*

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
β_1	-0.2111	(0.0216)*	ω_{22}	-0.0540	(0.0254)*	φ_{12}	-0.0304	(0.0148)*
β_2	0.0131	(0.0652)	ω_{12}	-0.2432	(0.0210)*	$\bar{\omega}_{12}$	-0.1103	(0.0491)*
α_1	0.4855	(0.0678)*	δ_{11}	-0.0265	(0.0065)*	γ_1	-0.0225	(0.0128)*
α_2	0.0308	(0.0479)	δ_{12}	0.0280	(0.0055)*	γ_2	-0.0981	(0.0193)*
χ_{11}	0.0522	(0.0119)*	δ_{21}	-0.0078	(0.0049)	γ_{12}	0.0129	(0.0095)
χ_{22}	0.0150	(0.0068)*	δ_{22}	0.0150	(0.0073)*	ψ_{12}	-0.0570	(0.0086)*
χ_{12}	0.0018	(0.0094)	φ_1	-0.2652	(0.0354)*	κ_{121}	-0.0518	(0.0080)*
ω_{11}	-0.0328	(0.0057)*	φ_2	-0.0559	(0.0270)*	φ_{122}	-0.0011	(0.0072)
Partial derivative of the logarithm of cost with respect to the logarithm of capital price							0.3261	(0.0622)*
Partial derivative of the logarithm of cost with respect to the logarithm of deposit price							0.0291	(0.1448)
Partial derivative of the logarithm of cost with respect to the logarithm of class 1 and 2 asset quantity							0.8300	(0.0740)*
Partial derivative of the logarithm of cost with respect to the logarithm of class 3 asset quantity							0.1656	(0.0307)*
Model size	Observations = 597		Parameters = 123		Deg.Fr. = 474			
Model Fit	R-squared = 0.9998		Adjusted R-squared = 0.9998					
Log-likelihood	$L_{FF} = 1133.3154$		$L_{TL} = 877.1932$		$LM = 512.2444*$			
F test values are not reported as they were too large to be estimated by the software employed (Limdep7)								

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Table 9.5 *Distribution free cost efficiency estimates from the cost model:*

Society	Efficiency	Society	Efficiency	Society	Efficiency
Alliance & Leicester	67.992	Hanley Economic	47.764	Norwich and Peterborough	70.296
Barnsley	59.728	Harpenden	60.972	Nottingham	59.606
Bath Investments	66.281	Haywards Heath	72.085	Nottingham Imperial	61.270
Bedford Crown	88.360	Heart Of England	58.638	Penrith	79.891
Beverley	82.622	Hendon	53.810	Portman	82.571
Bexhill	92.746	Hinckley & Rugby	62.924	Portsmouth	67.349
Birmingham Midshires	68.881	Holmesdale	65.866	Principality	57.954
Bradford & Bingley	67.939	Ilkeston Permanent	79.794	Progressive	67.643
Bristol & West	62.402	Ipswich	75.646	Saffron Walden	69.646
Britannia	69.960	Kent Reliance	56.378	Herts & Essex	
Buckinghamshire	82.390	Lambeth	60.371	Scarborough	59.714
Cambridge	57.050	Lancastrian	63.546	Scottish	63.500
Catholic	87.132	Leamington Spa	55.935	Shepshed	58.429
Chelsea	74.072	Leeds	62.909	Skipton	74.872
Cheltenham & Gloucester	69.738	Leeds & Holbeck	65.053	Southdown	66.734
Chesham	67.712	Leek United	74.135	St Pancras	68.827
Cheshire	72.323	Loughborough	63.592	Stafford Railway	61.442
Cheshunt	66.448	Londonderry Provident	100	Staffordshire	70.072
Chorley And District	74.449	Manchester	66.768	Standard	70.883
City & Metropolitan	63.165	Mansfield	60.344	Stroud and Swindon	100.000
Clay Cross	76.215	Market Harborough	64.307	Surrey	60.316
Coventry	67.957	Marsden	66.148	Swansea	53.526
Cumberland	59.271	Melton Mowbray	53.094	Teachers	79.313
Darlington	59.204	Mercantile	62.108	Tipton & Coseley	67.852
Derbyshire	63.439	Mid-Sussex	56.621	Town & Country	71.247
Dudley	62.909	Monmouthshire	70.959	Tynemouth	68.274
Dunfermline	64.861	Mornington	69.334	Universal	69.079
Earl Shilton	70.806	National & Provincial	71.576	Vernon	62.255
Ecology	100	National Counties	63.105	West Bromwich	57.286
Furness	57.713	Nationwide Anglia	96.838	West Cumbria	67.006
Gainsborough	84.858	Newbury	59.378	Woolwich	58.913
Greenwich	61.802	Newcastle	60.916	Yorkshire	64.480
Halifax	78.650	North of England	65.166	Average	67.518
Hampshire	52.355	Northern Rock	67.452	Stdev	9.287
				Min	52.355

The distribution-free firm-specific cost efficiency estimates are reported in Table 9.5.

Both the Londonderry Provident and the Ecology building societies are treated as extreme outliers. The estimates indicate an average level of cost efficiency for UK building societies between 1990-1996 was of 67.5 per cent. The magnitude of efficiency dispersion is also considerable with a standard deviation of 9 per cent and a minimum efficiency of 52 per cent. These results appear to be skewed with the some

of the smallest building societies operating with higher levels of efficiency than larger building societies. On average the 'large' building societies have an average efficiency of 66.765 per cent whilst the 'small' societies have an average efficiency of 70.746 per cent. Small societies, however, have a greater dispersion of efficiency with a standard deviation of 12.588 per cent, a value larger than the standard deviation of 'large' building societies (9.312 per cent).

It is possible that the differential treatment in terms of regulation and legislation of 'large' and 'small' building societies may have resulted in some societies operating with a distinct productive technology. This difference may have resulted in an exaggeration of the 'relative' efficiency of building societies leading to an underestimate of efficiency. To amend for this possibility, the cost model is re-estimated separately for the 'large' and 'small' societies.

9.7 The re-estimated efficiency models

The cost function (equation 9.3) is re-estimated first using a smaller sample of 61 'large' building societies with more than £100m in total assets, and secondly using the sample of 38 'small' building societies with less than £100m in total assets. As before, the data sets are deflated by the Retail Price Index to 1990 prices and are contiguous and unbalanced. The amended scale factors and scaled variables for the flexible Fourier functional form model are displayed in Table 9.3. The parameter estimates and diagnostic statistics for the re-estimated 'large' building society cost model are shown in Table 9.6. The parameter estimates and diagnostic statistics for the re-estimated 'small' building society cost model are shown in Table 9.7. The re-

estimated distribution-free cost efficiency estimates for 'large' building societies are reported in the Table 9.8, whilst the re-estimated distribution-free cost efficiency estimates for 'small' building societies are reported in the Table 9.9.

Table 9.6 *The efficiency model: parameter estimates, partial derivatives and diagnostic statistics: Large building societies*

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
β_1	-0.1728	(0.0391)*	ω_{22}	-0.0948	(0.0399)*	φ_{12}	-0.0170	(0.0130)
β_2	-0.3435	(0.1323)*	ω_{12}	-0.3372	(0.0384)*	$\bar{\omega}_{12}$	-0.0427	(0.0770)
α_1	0.2928	(0.1210)*	δ_{11}	-0.0021	(0.0140)	γ_1	0.0007	(0.0163)
α_2	0.0686	(0.0667)	δ_{12}	0.0207	(0.0059)*	γ_2	-0.0933	(0.0274)*
χ_{11}	0.1044	(0.0276)*	δ_{21}	0.0080	(0.0053)	γ_{12}	0.0145	(0.0111)
χ_{22}	0.0183	(0.0085)*	δ_{22}	-0.0049	(0.0106)	ψ_{12}	-0.0392	(0.0102)*
χ_{12}	-0.0142	(0.0155)	φ_1	-0.3450	(0.0507)*	κ_{121}	-0.0257	(0.0080)*
ω_{11}	-0.0103	(0.0157)	φ_2	0.0262	(0.0344)	φ_{122}	-0.0028	(0.0067)
Partial derivative of the logarithm of cost with respect to the logarithm of capital price							0.6841	(0.1520)*
Partial derivative of the logarithm of cost with respect to the logarithm of deposit price							0.4893	(0.2727)*
Partial derivative of the logarithm of cost with respect to the logarithm of class 1 and 2 asset quantity							1.3013	(0.1581)*
Partial derivative of the logarithm of cost with respect to the logarithm of class 3 asset quantity							0.0239	(0.0152)
Model size	Observations = 433		Parameters = 94		Deg.Fr. = 339			
Model Fit	R-squared = 0.9998		Adjusted R-squared = 0.9998					
Log-likelihood	$L_{FF} = 917.1894$		$L_{TL} = 786.6675$		$LM = 261.0438^*$			
F test values are not reported as they were too large to be estimated by the software employed (Limdep7)								

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Relative to the overall model, a lower proportion of the parameters in the estimated cost model (50 per cent) were statistically significant for the 'large' model, as would be expected with fewer observations. However, the diagnostic statistics indicate an acceptable level of model fit (an adjusted R^2 value of 0.999 was recorded) and partial elasticities of cost with respect to input prices and output quantities were found to be

positive in accordance with established cost and production theory. A significant likelihood ratio test was recorded indicating that the flexible Fourier functional form provides a better 'fit' for the data than the nested translog functional form.

Table 9.7 *The efficiency model: parameter estimates, partial derivatives and diagnostic statistics: Small building societies*

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
β_1	0.0358	(0.0603)	ω_{22}	0.0342	(0.0562)	φ_{12}	-0.0786	(0.0864)
β_2	2.1329	(0.7963)*	ω_{12}	0.4730	(0.2602)*	ϖ_{12}	0.0365	(0.0896)
α_1	0.1541	(0.4279)	δ_{11}	-0.0021	(0.0308)	γ_1	0.0382	(0.0438)
α_2	2.0066	(0.5924)*	δ_{12}	-0.0089	(0.0180)	γ_2	0.1510	(0.1355)
χ_{11}	0.1615	(0.1533)	δ_{21}	-0.0205	(0.0430)	γ_{12}	-0.0199	(0.0251)
χ_{22}	0.4839	(0.1396)*	δ_{22}	0.5627	(0.1725)*	ψ_{12}	-0.0054	(0.0224)
χ_{12}	-0.0412	(0.1402)	φ_1	-0.1571	(0.1080)	κ_{121}	-0.0191	(0.0179)
ω_{11}	-0.0391	(0.0191)*	φ_2	0.5371	(0.1486)*	φ_{122}	0.0352	(0.0437)
Partial derivative of the logarithm of cost with respect to the logarithm of capital price							-1.1059	(0.6166)*
Partial derivative of the logarithm of cost with respect to the logarithm of deposit price							1.3385	(0.6531)*
Partial derivative of the logarithm of cost with respect to the logarithm of class 1 and 2 asset quantity							0.9286	(0.5764)
Partial derivative of the logarithm of cost with respect to the logarithm of class 3 asset quantity							2.0066	(0.6831)*
Model size	Observations = 164		Parameters = 54		Deg.Fr. = 110			
Model Fit	R-squared = 0.9978		Adjusted R-squared = 0.9968					
Log-likelihood	$L_{FF} = 331.3232$		$L_{TL} = 314.0146$		LM = 34.6172*			
Model test	F = 946.29		Prob. Value = 0.0000					

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

A low proportion of the parameters in the estimated cost model were statistically significant for the 'small' model. The diagnostic statistics indicate an acceptable level of model fit (an adjusted R^2 value of 0.998 and a statistically significant F test were recorded). Whilst most partial elasticities of cost with respect to input prices and output quantities were found to be positive in accordance with established cost and

production theory, the partial elasticity of cost with respect to the price of capital was significant and negative. This may be an indication of approximation error associated with the relatively small sample used. A significant likelihood ratio test was recorded indicating that the flexible Fourier functional form provides a better 'fit' for the data than the nested translog functional form.

The average level of efficiency is somewhat higher when only 'large' building societies are considered relative to the overall sample. An average cost efficiency of 76.8 per cent is recorded, indicating that on average 'large' building societies could operate with 23 per cent fewer costs. The lowest efficiency estimate recorded was 65 per cent, whilst the standard deviation of efficiency is 5.8 per cent. This shift in the magnitude of efficiency is assumed to be a product of considering a more homogenous sample⁶.

The average level of efficiency for building societies with less than £100m in total assets is far greater than the average efficiency of larger building societies. An average cost efficiency of 91.7 per cent is recorded, with the lowest efficiency estimate of 82.7 per cent. The standard deviation of efficiency is lower than that recorded for the large and overall groups (5.8 per cent). The higher levels of efficiency recorded by smaller societies may have resulted from a number of causes. The sample is smaller which may have led to greater approximation error. Additionally, the smaller sample may have provided a more homogenous sample reducing the relative dispersion of efficiency. Equally, the smaller societies, which have operated under a different regulatory environment, are more efficient than larger societies which have become more diversified in their operations⁷.

Table 9.8 *Distribution-free cost efficiency estimates from the re-estimated cost model: Large building societies*

Society	Efficiency	Society	Efficiency	Society	Efficiency
Alliance & Leicester	75.196	Ipswich	69.823	Nottingham	74.711
Barnsley	73.690	Kent Reliance	72.670	Portman	74.288
Birmingham Midshires	79.959	Lambeth	75.081	Portsmouth	69.076
Bradford & Bingley	77.050	Lancastrian	66.927	Principality	76.067
Bristol & West	71.379	Leamington Spa	72.805	Progressive	82.161
Britannia	79.060	Leeds & Holbeck	86.039	Saffron Walden	73.129
Cambridge	69.003	Leeds Permanent	75.891	Herts & Essex	
Chelsea	83.698	Leek United	76.023	Scarborough	75.470
Cheltenham & Gloucester	78.949	Loughborough	81.431	Scottish	74.190
Chesham	86.417	Manchester	72.191	Skipton	78.823
Cheshire	83.894	Mansfield	78.713	Southdown	79.636
Cheshunt	77.231	Market Harborough	78.634	Staffordshire	82.161
City & Metropolitan	80.008	Marsden	66.152	Stroud and Swindon	71.587
Coventry	81.549	Melton Mowbray	75.542	Teachers	79.491
Cumberland	71.289	Mercantile	71.006	Tipton & Coseley	86.152
Darlington	72.470	Monmouthshire	82.575	Town & Country	78.870
Derbyshire	76.377	Mornington	81.837	Universal	74.739
Dunfermline	77.504	National & Provincial	74.485	Vernon	72.751
Furness	69.637	National Counties	100.000	West Bromwich	78.011
Greenwich	75.397	Nationwide Anglia	71.405	Woolwich	74.910
Halifax	92.333	Newbury	74.400	Yorkshire	81.263
Hanley Economic	73.718	Newcastle	78.011	Average	76.763
Haywards Heath	73.265	North of England	79.797	Standard Deviation	5.804
Heart Of England	65.037	Northern Rock	79.204	Minimum	65.037
Hinckley & Rugby	78.399	Norwich and Peterborough	72.737		

Table 9.9 *Distribution free cost efficiency estimates from the re-estimated cost model: Small building societies*

Society	Efficiency	Society	Efficiency	Society	Efficiency
Bath Investments	89.792	Gainsborough	94.298	St Pancras	84.357
Bedford Crown	94.226	Hampshire	82.726	Stafford Railway	91.335
Beverley	93.165	Harpenden	91.268	Standard	96.474
Bexhill	95.069	Hendon	91.023	Surrey	84.568
Buckinghamshire	90.845	Holmesdale	90.316	Swansea	93.141
Catholic	93.375	Ilkeston Permanent	94.346	Tynemouth	90.031
Chorley And District	91.067	Londonderry Provident	100.000	West Cumbria	91.000
Clay Cross	94.370	Mid-Sussex	91.650		
Dudley	87.925	Nottingham Imperial	92.770	Average	91.666
Earl Shilton	89.663	Penrith	91.179	Standard Deviation	3.635
Ecology	95.680	Shepshed	92.654	Minimum	82.726

9.8 Conclusions

In this chapter, the cost efficiency of UK building societies between 1990 and 1996 was investigated. This analysis differs from previous studies in three central respects. Firstly, the model is estimated using a panel of building society observations over a seven-year period, reducing the possibility of reporting on what may be a time of distinct operating conditions. Secondly, distribution-free cost efficiency is estimated. This measure provides an indication of average relative efficiency, without the imposition of distributional assumptions employed in previous studies. Thirdly, a flexible Fourier functional form is employed to represent productive technology. This relatively under-used functional form, whilst inclined to provide higher levels of approximation error, is statistically superior in reducing the levels of specification error than other functional forms used in the literature.

Both differences and similarities are identified with previous studies that have considered efficiency in UK building societies. The efficiency estimates made in this

chapter differ from the econometric estimates produced by Drake and Weyman-Jones (1996) in that a substantial dispersion of efficiency appears to exist in the UK building society sector. The dispersion of efficiency reported in this study appears to support many of the conclusions made by the non-parametric DEA studies. For example, Piesse and Townsend (1995) found that substantial efficiency dispersion, which may be systematically related to scale factors, existed in the building society sector.

To conclude, differential levels of cost efficiency are observed for the UK building society sector between 1990 and 1996. Building societies of different sizes appear to experience distinct levels of efficiency. This finding is particularly important in policy terms due to the differential treatment of building societies of different sizes in terms of the activities the societies may undertake. An average cost efficiency of 76.8 per cent was estimated for the building societies deemed to be 'qualifying asset holding' societies under the 1986 Building Societies Act. This qualification allowed a far greater degree of diversification in terms of form of operations undertaken and the services offered by these societies. The 'small' group of societies, which were not 'qualifying asset holding' societies and were denied many of these opportunities for diversification, have a greater average efficiency of 91 per cent. This finding is of significance in policy terms, as it seems that the on-going processes of re-regulation and diversification appear to have adversely affected the cost efficiency of this sector. Equally, it may be stated that despite previous empirical evidence (see chapter 4) that large depository institutions are generally more efficient than small depository institutions, this does not appear to be the case for UK building societies. The

different opportunities available to large and small societies could be an explanation of such a result.

The estimation of a broader range of economic characteristics and tests of economies of scope for the joint production of class 1 and 2 with class 3 assets, sanctioned by the Building Societies Act (1986), are considered in more detail in chapter 10.

¹ This problem prevented the use of individual asset groups for the retail bank data set and is discussed in chapter 7.

² Section 118 Building Societies Act (1986) –“ (1) This section has effect for determining for the purposes of this Act whether, in any financial year, a building society has a “qualifying asset holding”. (2) A building society has a qualifying asset holding in any financial year, if, and only if, the aggregate value of its total commercial assets, as shown in its annual accounts for the previous year, is not less than £100 million or such other amount as may be substituted for it under subsection (3) below.”

³ Class 3 assets include money lent to individuals, with or without security, whether or not at interest and whether or not they are members of the society. Advances fully secured on land do not constitute class 3 loans except under certain circumstances (Building Societies Act (1986) 16- (1-2).).

⁴ This test is devised for use with maximum likelihood estimators. It is used here to provide a broad guide as to whether the flexible Fourier functional form provides a better ‘fit’ for the data than the translog functional form (see Greene, 1993 p.380).

⁵ 10 degrees of freedom are assumed for the LM test resulting in a critical value of 15.9871.

⁶ Consideration of possible heteroscedasticity in the sample is made in chapter 10.

⁷ Further investigation of the existence of differing productive technologies is suggested. A possible approach for considering distinct productive technologies was presented by Beard *et al*, (1991). In this study, a finite mixture analysis was used to test for the existence of multiple technologies in a sample of US S&L’s. The presence of multiple technologies was not rejected.

Chapter 10 Economies of scope, economies of scale, technical change and total factor productivity growth in UK building societies

10.1 Introduction

In this chapter, a number of cost characteristics of the UK building society sector are calculated from cost functions which are estimated for every year of the sample period, 1990-96. The cost characteristics considered include economies of scope, economies of scale, cost complementarities, technical change and total factor productivity growth. In chapter 2, it was suggested that the UK building society sector had undergone substantial re-structuring and change during the 1980s and early 1990s. A study of total factor productivity growth and technical change, in association with other cost characteristics, should help to establish the effects of these changes in terms of the cost structure of the industry. To the author's knowledge, this is the first study to consider total factor productivity growth and technical change in the UK building society sector.

Productivity measures are used to quantify the change over time in the 'efficiency' of production. An analysis of this kind may allow a greater understanding of the implications of the regulatory, organisational and market-related changes in terms of technical change and total factor productivity growth. Measures of total factor productivity growth incorporate aspects of technical change, scale efficiency, technical efficiency and allocative efficiency. A broader discussion of productivity was presented in chapter 5. A review of studies considering the productivity and technical change of depository institutions is contained in chapters 5 and 8.

This chapter is structured as follows. Section 10.2 outlines the model specification used in this chapter and section 10.3 considers the methods of measuring total factor productivity growth and technical change. Section 10.4 reports the results of the study. A summary of the study and conclusions are presented in section 10.5.

10.2 Model specification

Following the recommendations made in chapters 3 and 7, an intermediation model of depository institution production is employed. As in chapter 9, building societies are assumed to produce outputs in the form of class 1 and class 2 assets (Y_1) and class 3 assets (Y_2) using labour (X_1), capital (X_2) and deposits (X_3). The cost (C) of production is therefore an amalgam of operational and interest cost. A cost function may be represented as:

$$C = g(Y_1, Y_2, P_1, P_2, P_3) \quad (10.1)$$

Output quantities are measured by their values at the year-end. Input quantities are denoted by the total full-time equivalent staff numbers, tangible fixed assets and the value of both retail and non-retail deposits at the year end. As in chapter 9, the price of labour is proxied by staff costs divided by the number of full time equivalent staff (P_1). The price of capital (P_2) is defined as the sum of property and equipment rental costs and depreciation costs, divided by the level of total fixed assets, multiplied by 1000, to provide a measure of capital cost for every £1000 of physical capital. The

price of deposits (P_3) is obtained by dividing the value of total interest paid by the value of total deposits, multiplied by 1000, to provide a measure of the interest cost of every £1000 of deposits. Further discussion of the variable definitions and model specifications employed in this chapter can be found in chapters 3 and 5.

10.3 Approaches to measuring productivity and characteristics of productive technology

Total factor productivity growth and technical change are quantified using a cost function shift technique from Humphrey (1993). The cost function shift approach examines the change in average costs for each of the sample years. A translog specification is employed to represent productive technology. The translog functional form is employed due to the relatively small samples that are available on a yearly basis (the largest sample from a single year is 99 observations in 1990; see chapters 5 and 9 for further discussion of this issue).

The translog cost function model may be written as:

$$\begin{aligned} \ln C = & \alpha_0 + \sum_j \alpha_j \ln Y_j + \sum_r \beta_r \ln P_r + \frac{1}{2} \sum_j \sum_s \chi_{js} \ln Y_j \ln Y_s + \\ & \frac{1}{2} \sum_r \sum_q \omega_{rq} \ln P_r \ln P_q + \sum_j \sum_r \gamma_{jr} \ln Y_j \ln P_r + \varepsilon \end{aligned} \quad (10.2)$$

for $j, s = 1, 2$ and $r, q = 1, 2, 3$.

Following the approach used in chapter 8, cost shares are derived from Shephard's Lemma and the deposit share equation (S_3) is dropped to make the model operational.

The cost share equations may be written as:

$$S_r = \partial \ln TC / \partial \ln P_r = \beta_r + \sum_q \omega_{rq} \ln P_q + \sum_j \gamma_{jr} \ln Y_j + e \quad (10.3)$$

Following established cost and production theory, the following restrictions are imposed:

$$\sum_r \beta_r = 1, \quad \sum_r \omega_{rq} = 0, \quad \sum_j \gamma_{jr} = 0 \quad (10.4)$$

Estimation of total factor productivity may be derived from estimates of technical change and scale elasticity. Following Humphrey, (1993) technical change may be written as:

$$Tech = -(AC_{t+1} - AC_t) / AC_t \quad (10.5)$$

Where *Tech* represents technical change, AC_t represents average costs in period t and AC_{t+1} represents average costs in time period $t+1$. Average costs are (fitted) total costs divided by the value of total assets of the building societies. For each year, the

logarithm of total costs are 'predicted' by multiplying the parameter estimates by the relevant building society input price and output values.

Total factor productivity growth may be viewed as a measure of the change in cost efficiency over time. Following Baltagi and Griffin (1988), total factor productivity growth may be written as:

$$TFP = -Tech + (1 - \epsilon_{scale})\dot{Y}$$

where TFP = change in Total Factor Productivity Growth

$Tech$ = Technical change

ϵ_{scale} = Economies of scale (OES)

and \dot{Y} = change in output (10.6)

Average change in output, \dot{Y} is calculated as the percentage change in the average level of output. The change in output is recorded for all outputs (including Class 1, 2, and 3 assets) and for all years and assets groups.

Overall, ray economies of scale (*OES*) are estimated for all years and asset groups. Ray economies of scale, which are cost savings resulting from proportional increases in the quantities of all outputs produced, are used to represent scale economies. Thus, for a building society producing j outputs, we can write :

$$OES = \sum_j (\partial \ln C / \partial \ln Y_j) \quad (10.7)$$

for $j = 1, 2$.

According to this radial measure, there are overall economies of scale (and therefore decreasing ray average costs) if $OES < 1$, and overall diseconomies if scale (increasing ray average costs) if $OES > 1$ and constant ray average costs if $OES = 1$.

Economies of scope and cost complementarities are also assessed for individual years. Economies of scope used in this study may be defined as cost savings achievable from the joint production of two or more goods or services within a single enterprise, compared with their separate production by specialised firms. As in chapter 7, for a firm producing two outputs, Y_1 and Y_2 , the magnitude of scope economies (*ESC*) may be measured by:

$$ESC = [C(Y_1, 0) + C(0, Y_2) - C(Y_1, Y_2)] / C(Y_1, Y_2) \quad (10.8)$$

ESC measures the relative increase in cost that would result from a separation of the firm into two specialist firms. There are economies of scope if $ESC > 0$ and diseconomies of scope if $ESC < 0$. As previously mentioned, an alternative way of

investigating the existence of local economies of scope, used in this study, is to test for *cost complementarities* by checking whether the second derivative of cost with respect to output ($\partial^2 C / \partial Y_1 \partial Y_2$) is significantly less than zero. If the value is less than zero, then the marginal cost of one good would decrease as the level of output of the other good increases. When this is the case, cost complementarities and therefore (local) economies of scope are deemed to exist. Further discussion of economies of scope and cost complementarities can be found in chapter 5.

10.4 *Data and results*

A sample of UK building societies for the period 1990-96 was used. As in chapter 9, data on costs, prices and quantities were taken from the Annual Reports and Accounts. Positive levels of prices and quantities were recorded for all societies in all years. The data set is contiguous and unbalanced and the data are deflated by the Retail Price Index for 1990, where appropriate. Summary statistics for this data set are outlined and discussed in chapter 9. A generalised least squares (GLS), iterative, seemingly unrelated regression (SUR) estimator is used for all estimations, following a procedure elaborated in Greene (1993). The parameter estimates are displayed in Tables 10.1 and 10.2 with diagnostic statistics shown in Table 10.3.

Table 10.1 Parameter estimates: 1990-93

	1990		1991		1992		1993	
α_0	0.101	(0.003)*	0.104	(0.003)*	0.127	(0.004)*	0.147	(0.005)*
β_1	-1.037	(0.038)*	-1.104	(0.038)*	-0.930	(0.051)*	-0.405	(0.056)*
β_2	0.102	(0.006)*	0.082	(0.006)*	0.060	(0.007)*	0.067	(0.008)*
β_3	1.934	(0.041)*	2.022	(0.040)*	1.871	(0.054)*	1.338	(0.060)*
α_1	2.850	(0.068)*	3.194	(0.085)*	3.617	(0.107)*	3.570	(0.071)*
α_2	-1.769	(0.069)*	-2.106	(0.088)*	-2.538	(0.110)*	-2.573	(0.073)*
χ_{11}	-0.001	(0.004)	-0.022	(0.004)*	-0.019	(0.005)*	-0.014	(0.004)*
χ_{22}	0.008	(0.006)	0.008	(0.007)	0.003	(0.006)	0.001	(0.005)
χ_{21}	-0.005	(0.005)	0.005	(0.005)	0.003	(0.005)	0.004	(0.004)
γ_{11}	-0.342	(0.010)*	-0.370	(0.010)*	-0.310	(0.015)*	-0.131	(0.018)*
γ_{12}	-0.057	(0.003)*	-0.037	(0.003)*	-0.058	(0.005)*	-0.128	(0.007)*
γ_{13}	0.341	(0.007)*	0.341	(0.006)*	0.310	(0.009)*	0.214	(0.009)*
γ_{22}	-0.005	(0.000)*	-0.003	(0.000)*	0.000	(0.001)	0.001	(0.001)
γ_{23}	0.068	(0.002)*	0.072	(0.002)*	0.058	(0.003)*	0.023	(0.003)*
γ_{33}	-0.004	(0.001)*	-0.002	(0.001)*	0.001	(0.001)	0.022	(0.002)*
ω_{11}	-0.001	(0.000)*	-0.001	(0.000)*	0.001	(0.000)*	0.001	(0.000)*
ω_{12}	-0.015	(0.001)*	-0.009	(0.001)*	-0.005	(0.002)*	0.007	(0.003)*
ω_{13}	0.266	(0.010)*	0.293	(0.013)*	0.342	(0.015)*	0.324	(0.010)*
ω_{21}	0.001	(0.000)*	0.001	(0.000)*	0.000	(0.000)	0.000	(0.000)
ω_{22}	0.009	(0.001)*	0.007	(0.002)*	0.006	(0.002)*	-0.002	(0.003)
ω_{23}	-0.260	(0.010)*	-0.292	(0.013)*	-0.343	(0.015)*	-0.331	(0.010)*

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

In all years, the majority of parameter estimates are significant at the 10 per cent level. The adjusted R^2 and F statistics both provide a measure of model fit with all years experiencing high and statistically significant levels of model fit. The F test for restrictions provides a test of the restrictions imposed on the translog model. In all cases, the restrictions cannot be rejected. The Breusch-Pagan test is used to test for conditions of homoscedasticity in the sample; the presence of homoscedasticity is rejected for 1990 and 1991 indicating that heteroscedasticity cannot be rejected for these years. This finding means that the estimates may be inefficient in these years. The log likelihood statistic is significant for all time periods, indicating that a linear form of the cost models is rejected.

Table 10.2 Parameter estimates: 1994-1996

	1994		1995		1996	
α_0	0.152	(0.006)*	0.152	(0.004)*	0.163	(0.007)*
α_1	-0.863	(0.070)*	-1.080	(0.067)*	-0.934	(0.060)*
α_2	0.043	(0.007)*	0.020	(0.008)*	0.041	(0.010)*
α_3	1.820	(0.074)*	2.061	(0.071)*	1.893	(0.064)*
β_1	3.669	(0.118)*	3.726	(0.153)*	3.936	(0.125)*
β_2	-2.701	(0.120)*	-2.776	(0.155)*	-2.911	(0.127)*
χ_{11}	0.001	(0.004)	0.008	(0.004)*	0.002	(0.005)
χ_{22}	0.012	(0.004)*	0.018	(0.003)*	0.034	(0.004)*
χ_{21}	-0.007	(0.003)*	-0.013	(0.003)*	-0.022	(0.003)*
γ_{11}	-0.279	(0.023)*	-0.351	(0.021)*	-0.310	(0.018)*
γ_{12}	-0.072	(0.009)*	-0.054	(0.008)*	-0.047	(0.007)*
γ_{13}	0.289	(0.011)*	0.333	(0.011)*	0.297	(0.010)*
γ_{22}	0.000	(0.001)	0.001	(0.001)	-0.002	(0.001)*
γ_{23}	0.049	(0.004)*	0.062	(0.004)*	0.054	(0.003)*
γ_{33}	0.012	(0.002)*	0.009	(0.002)*	0.007	(0.002)*
ω_{11}	0.001	(0.000)*	0.001	(0.000)*	0.001	(0.000)*
ω_{12}	0.007	(0.003)*	0.008	(0.003)*	0.006	(0.003)*
ω_{13}	0.340	(0.015)*	0.354	(0.020)*	0.368	(0.016)*
ω_{21}	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
ω_{22}	-0.002	(0.002)	-0.002	(0.002)	0.000	(0.002)
ω_{23}	-0.347	(0.015)*	-0.361	(0.020)*	-0.375	(0.016)*

Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.

Partial derivatives of the logarithm of cost with respect to the logarithm of input prices are displayed in Table 10.4. Positive partial elasticities of cost with respect to input prices are found, in accordance with expectation. Estimates of the cost characteristics are reported for both individual years and asset groups. Following the approach adopted in chapter 9, two asset groups are defined, including 'small' societies with less than £100m in total assets and 'large' societies, which have in excess of a £100m in total assets.

Table 10.3 Diagnostic statistics

	1990	1991	1992	1993	1994	1995	1996
Observations	99	92	87	83	81	79	76
Iterations	4	4	5	5	6	6	6
Degrees of freedom	78	71	66	62	60	58	55
R ²	0.998	0.998	0.999	0.999	0.999	0.999	0.999
Adj. R ²	0.997	0.998	0.999	0.999	0.999	0.999	0.998
F test	1644.79*	2463.45*	3253.66*	4326.45*	5435.61*	3210.60*	2062.57*
F test for restrictions	1422.68*	815.379*	313.231*	183.485*	98.547*	147.917*	123.917*
Breusch-Pagan test	33.2379*	90.9551*	25.9807	20.1799	11.4237	16.8033	22.1879
Log-Likelihood Test	554.609	543.142	549.529	529.668	536.345	502.442	468.619

Note: * = Significantly different from zero at 10% significance.

Table 10.4 Partial derivatives of the logarithm of cost with respect to the logarithm of input prices

	Partial derivative of the logarithm of cost with respect to the logarithm of labour price		Partial derivative of the logarithm of cost with respect to the logarithm of capital price		Partial derivative of the logarithm of cost with respect to the logarithm of deposit price	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Overall						
1990	2.011	(0.057)*	0.128	(0.008)*	1.187	(0.044)*
1991	2.159	(0.056)*	0.117	(0.008)*	1.209	(0.051)*
1992	1.735	(0.081)*	0.128	(0.011)*	1.342	(0.062)*
1993	0.709	(0.095)*	0.455	(0.025)*	1.216	(0.035)*
1994	1.481	(0.121)*	0.159	(0.017)*	1.069	(0.057)*
1995	1.854	(0.110)*	0.118	(0.016)*	0.952	(0.075)*
1996	1.637	(0.093)*	0.130	(0.016)*	0.955	(0.059)*
'Large'						
1990	1.895	(0.054)*	0.266	(0.017)*	1.336	(0.049)*
1991	2.117	(0.055)*	0.122	(0.008)*	1.346	(0.057)*
1992	1.734	(0.081)*	0.139	(0.012)*	1.493	(0.068)*
1993	0.709	(0.095)*	0.257	(0.017)*	1.331	(0.038)*
1994	1.473	(0.120)*	0.169	(0.018)*	1.173	(0.062)*
1995	1.848	(0.109)*	0.128	(0.017)*	1.035	(0.079)*
1996	1.654	(0.094)*	0.142	(0.018)*	1.061	(0.062)*
'Small'						
1990	1.972	(0.056)*	0.117	(0.007)*	0.824	(0.037)*
1991	2.123	(0.055)*	0.100	(0.007)*	0.875	(0.037)*
1992	1.740	(0.082)*	0.105	(0.009)*	0.167	(0.046)*
1993	0.708	(0.096)*	0.208	(0.013)*	0.899	(0.081)*
1994	1.503	(0.123)*	0.133	(0.015)*	0.773	(0.044)*
1995	1.871	(0.111)*	0.089	(0.014)*	0.723	(0.063)*
1996	1.603	(0.092)*	0.096	(0.014)*	0.809	(0.051)*

Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.

Economies of scope and cost complementarities are reported in Table 10.5. Economies of scale are reported in Table 10.6. Technical change and total factor productivity growth are reported in Table 10.7.

Estimates for economies of scope are greater than zero and statistically significant for all years in the overall and the large asset classes. This indicates the presence of economies of scope for the joint production of class 1 and class 2 assets with class 3 assets within large building societies. Diseconomies of scope are recorded for most years for small societies suggesting economies of joint production are not present in building societies with less than £100m in total assets. The estimates of cost complementarities are mixed and contradictory for most years. The cost complementary estimates therefore reduce the level of confidence that may be placed in the economies of scope estimates.

Table 10.5 Economies of scope and cost complementarities

Overall	Economies of scope		Cost complementarities		
	Coeff.	S. E.	Coeff.	S. E.	
1990	1.280	(0.075)*	1990	0.0137	(0.0046)*
1991	2.127	(0.157)*	1991	-0.0090	(0.0051)
1992	2.922	(0.243)*	1992	-0.0133	(0.0054)*
1993	3.901	(0.231)*	1993	-0.0093	(0.0048)*
1994	5.297	(0.526)*	1994	0.0050	(0.0040)
1995	7.287	(0.981)*	1995	0.0127	(0.0040)*
1996	10.177	(1.169)*	1996	0.0140	(0.0400)
'Large'	Economies of Scope		'Small'	Economies of Scope	
1990	2.610	(0.187)*	1990	-0.235	(0.046)*
1991	4.662	(0.441)*	1991	-0.332	(0.012)*
1992	7.272	(0.806)*	1992	-0.390	(0.013)*
1993	9.229	(0.763)*	1993	-0.384	(0.009)*
1994	13.559	(1.782)*	1994	-0.418	(0.014)*
1995	21.613	(3.949)*	1995	0.081	(0.037)*
1996	32.360	(5.079)*	1996	-0.454	(0.014)*

*Note: Standard errors are in brackets. * = Significantly different from zero at 10% significance.*

Table 10.6 Economies of scale (OES)

Overall							
Overall	1990	1991	1992	1993	1994	1995	1996
Coeff.	1.039	0.902	0.903	0.887	0.967	1.006	1.012
S.E	(0.035)	(0.026) [†]	(0.030) [†]	(0.029) [†]	(0.025)	(0.026)	(0.029)
'Large'							
Coeff.	1.038	0.876	0.871	0.868	0.970	1.022	1.026
S.E	(0.027)	(0.031) [†]	(0.031) [†]	(0.034) [†]	(0.029)	(0.031)	(0.035)
'Small'							
Coeff.	1.045	0.968	0.974	0.937	0.958	0.963	0.973
S.E	(0.013) [†]	(0.014) [†]	(0.016)	(0.015) [†]	(0.014) [†]	(0.014) [†]	(0.016) [†]

Note: Standard errors are in brackets. † = Significantly different from one at 10% significance.

A mixture of economies of scale results are recorded. Overall, economies of scale are reported for the period 1991-93, with constant returns to scale reported in all other years. Economies of scale are recorded for the small asset group in the 1993-96 period. Constant returns to scale are found for large societies for the period 1994-1996.

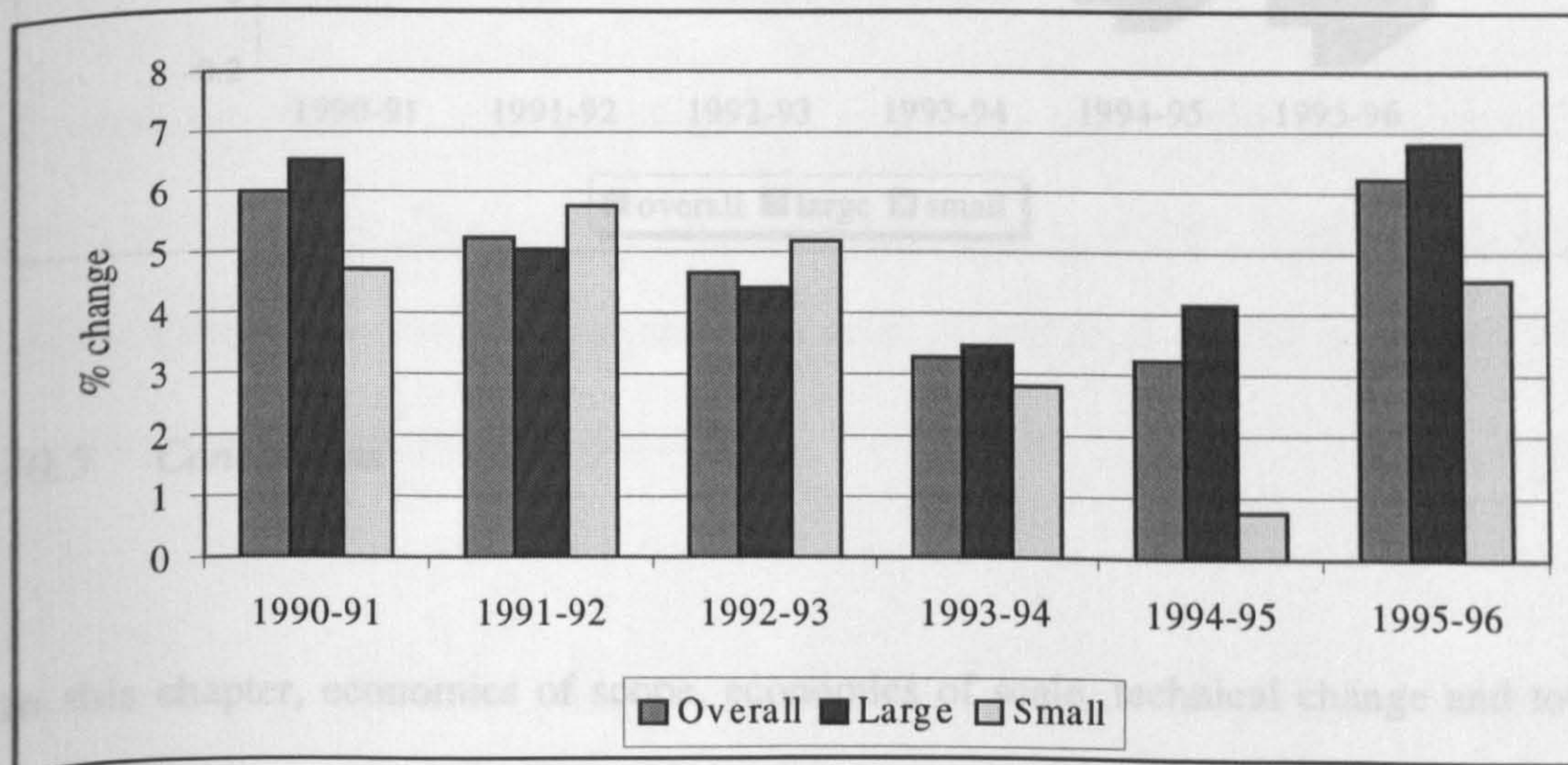
Table 10.7 Total factor productivity growth, output growth and technical change (percentage change)

Overall	Overall			'Large'			'Small'		
	Change in output	Technical change	Total factor productivity growth	Change in output	Technical change	Total factor productivity growth	Change in output	Technical change	Total factor productivity growth
1990-91	6.02	0.0461	0.5891	6.52	-0.0476	0.807425	4.71	-0.1095	0.1492
1991-92	5.26	-0.0976	0.5090	5.06	-0.0791	0.653342	5.78	-0.1258	0.1502
1992-93	4.65	-0.1493	0.5257	4.45	-0.1523	0.587489	5.21	-0.1700	0.3283
1993-94	3.30	-0.0706	0.1089	3.47	-0.0680	0.102931	2.83	-0.0849	0.1187
1994-95	3.24	-0.0954	-0.0195	4.14	-0.0889	-0.09119	0.76	-0.1038	0.0281
1995-96	6.22	-0.0900	-0.0747	6.82	-0.0810	-0.17803	4.55	-0.1156	0.1227
Average	4.78	-0.0761	0.2731	5.08	-0.0862	0.3137	3.97	-0.1182	0.1495

Technical change only exists only at low levels over the sample period. The rate of technical change varies only slightly falling slightly between 1990-91 (0.046 per cent) and 1995-96 (-0.09 per cent). Similarly low rates of technical change are observed in both asset groups during the sample period.

The percentage change in the output of the building society sector is seen to fall from a initial high level of 6.02 per cent growth in the 1990-91 period to a low of 3.3 per cent in the 1993-94 period. A recovery in the rates of output growth is seen in the second half of the sample period as output growth rises to 6.22 per cent in the 1995-96 period. The rate of output growth is also seen to differ between asset groups. 'Large' building societies experience a far greater growth in outputs throughout the sample period with an average growth of 5.08 per cent (overall an average rate of output growth of 4.78 per cent was observed). The growth in outputs of the 'small' building societies is lower with an average growth of 3.97 per cent. The percentage change in output of the UK building society sector is displayed in Figure 10.1.

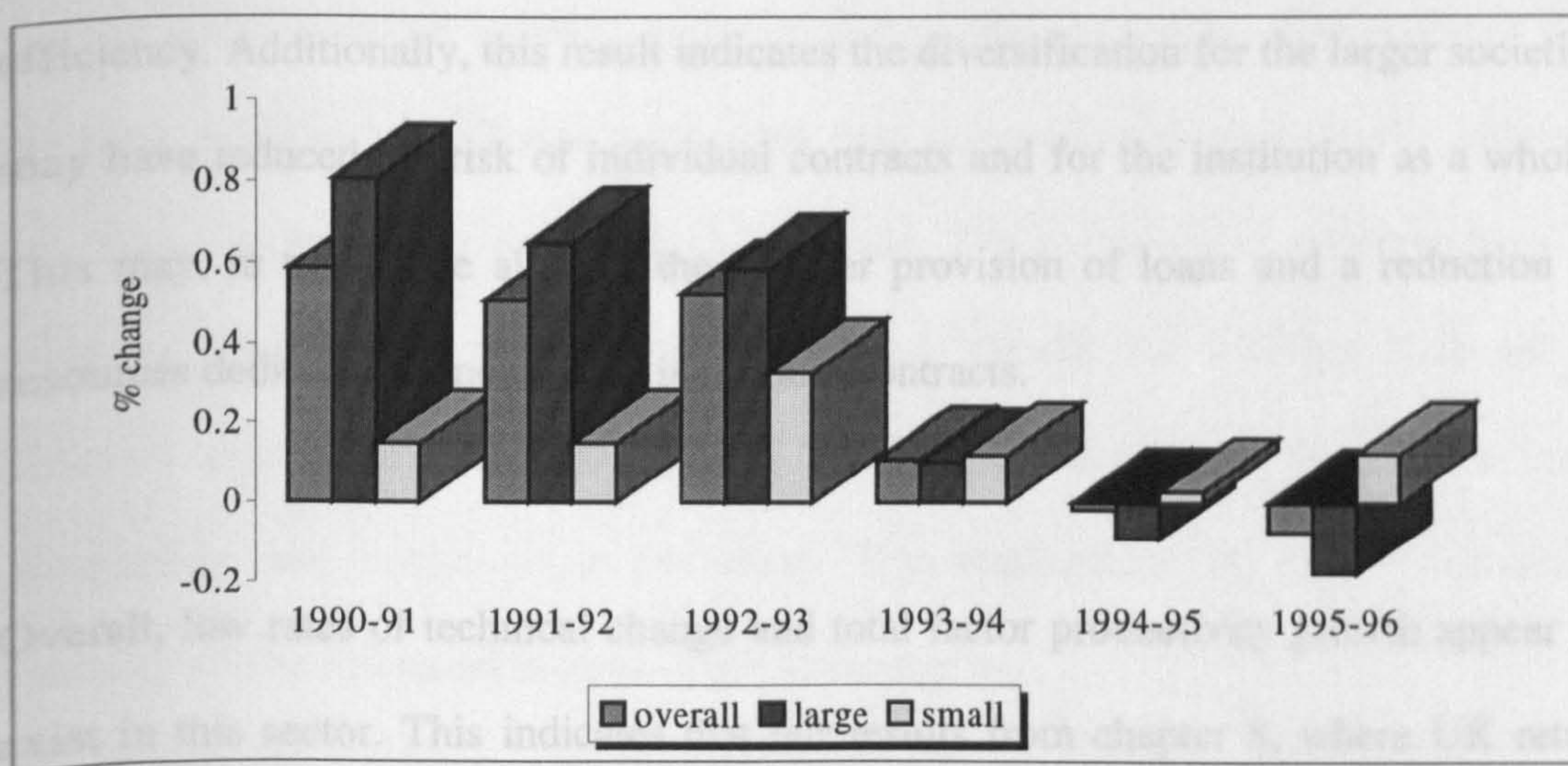
Figure 10.1 Change in output in the UK building society sector: 1990-96



between 1990 and 1996. Cost functions are estimated for every year and the estimated

The rate of total factor productivity growth is low in all years. After an initial rise in total factor productivity growth in 1990-93 period (average 0.541 per cent) a slight decrease in the level of total factor productivity growth is observed in the rest of the sample period (overall total factor productivity growth of 0.273 per cent is recorded). 'Large' building societies experience a slightly higher rate of total factor productivity growth over the sample period of 0.31 per cent on average ('small' building societies have an average total factor productivity growth of 0.15 per cent). Total factor productivity growth is displayed in Figure 10.2

Figure 10.2 Total factor productivity growth in the UK building society sector: 1990-96



10.5 Conclusions

In this chapter, economies of scope, economies of scale, technical change and total factor productivity growth have been estimated for the UK building society sector

between 1990 and 1996. Cost functions are estimated for every year and the estimated co-efficients are initially used to provide estimates of average costs, economies of scale, economies of scope and cost complementarities. The level of technical change and total factor productivity growth are then estimated, using these estimates and the average level of change in the rate of output, following a cost function shift approach. Estimates are made for all years and for two asset groups.

A slight change in the level of economies of scale is observed over the sample period. The estimates of economies of scope indicate substantial economies of the joint production of class 1 and class 2 assets with the production of class 3 assets for large societies. This finding indicates that the re-regulation of the building society sector, which allowed the joint provision of these new services, is justified in terms of cost efficiency. Additionally, this result indicates the diversification for the larger societies may have reduced the risk of individual contracts and for the institution as a whole. This may, in turn, have allowed the cheaper provision of loans and a reduction in resources dedicated to monitoring individual contracts.

Overall, low rates of technical change and total factor productivity growth appear to exist in this sector. This indicates that our results from chapter 8, where UK retail banks were also found to experience low rates of technical change and total factor productivity growth, were not isolated.

In chapter 8, a number of possible causes of the low rates of technical change and total factor productivity growth were introduced. The possible causes included the

possibility of mis-measurement of productivity and technical change and the level of dis-organisation and difficulties experienced in adapting to new technologies. As the building society sector, in common with the banking sector, has also experienced a decade of substantial investment in new technologies, it is proposed that the causes of the similar low level of technical change and total factor productivity growth may have resulted from similar sources.

Differences between the two sectors do exist and may provide an appropriate starting point for further consideration. The building society sector contains a far greater number of incumbents than the retail-banking sector. This larger sample allows the consideration of the differential performance of building societies of different asset sizes. It was reported that the rate of technical change and total factor productivity growth varied between building societies of different asset sizes. This suggests that building societies of different asset sizes may have adopted qualitatively different productive technologies and working practices. The existence of different rates of technical change and total factor productivity growth for building societies of different asset sizes may also indicate that differential rates of diffusion of new innovations and inventions is occurring. The implications of these preliminary findings will be discussed in more depth in chapter 11.

Chapter 11 Conclusions

11.1 Introduction

The conclusions attempt to clarify and assess the results of the thesis. First, a broad outline of the thesis content is provided. Then the findings from the empirical studies are considered in more detail and are subdivided into an assessment of the estimates of economies of scale, economies of scope, cost complementarities, efficiency, technical change and total factor productivity growth. Overall conclusions are suggested and combined with suggestions for further research.

11.2 The thesis outline

In the chapters comprising Section One, a number of recommendations are forwarded. Chapter 2 outlines the predominant changes in the conduct and structure of the industry. These are categorised as the re-organisation of market structure, the growth of new products, the development of technology in service provision, the transformation of labour markets, the macro-economic environment and the alteration of the regulatory and supervisory framework. Chapter 3 considers the model specifications and variable definitions that have been employed in previous studies of depository institutions. Both the production and intermediation models are outlined and assessed. No substantial theoretical flaws are observed in either model form. It is therefore suggested that in one empirical chapter both these models should be employed. From this study (chapter 7),

recommendations on the performance of the models in terms of estimation error and plausibility of the estimates are used to determine which model form should be employed in subsequent chapters. Additionally, a number of variable definitions are considered with recommendations made as to which should be used in the empirical chapters.

Chapter 4 provides a comprehensive international review of depository institution studies of productive characteristics and efficiency. This chapter indicates that the differing methods and approaches employed in previous studies, such as the efficiency concept, estimation technique, model definition and functional form influence the estimation of cost and production characteristics. Chapter 5 considers the definition and estimation of efficiency and productivity. Here it is suggested that an econometric, cost efficiency panel data approach, employing either a translog or flexible Fourier functional form, should be used in the empirical studies of efficiency in chapters 7 and 9. Recommendations are also made as to what measures of efficiency, total factor productivity growth and technical change should be quantified in the empirical chapters.

Section Two comprises the empirical chapters in which cost and production characteristics of the UK building society and retail banking sectors are estimated. These chapters draw heavily on Section One for their principal methods and assumptions. Chapter 6 suggests that monopolistically competitive conduct was present in the UK building society mortgage market during the 1990s. Both conditions of monopoly and perfect competition are rejected. These findings are similar to those estimated for the UK retail bank loan market by Molyneux *et al*, (1994). The level of market contestability

appears to increase marginally over the sample period. Disequilibrium conditions are reported for the 1990-92 period and stability of market shares cannot be rejected for any year.

Chapter 7 provides estimates of economies of scale, economies of scope and distribution-free cost efficiency of the UK retail-banking sector between 1984 and 1995. The translog functional form is employed to estimate two distinct models of bank production: the intermediation and production models. A high level of cost efficiency dispersion and very large economies of scale are found in the production model, while a relatively low level of efficiency dispersion and a moderate degree of economies of scale are reported with the intermediation model. Diseconomies of scope are reported for the joint production of loans and investments, and loans, deposits and investments. The intermediation model form is shown to have lower approximation and specification error than the production model. Based on the limited criteria of the plausibility of results and the level of estimation error, the intermediation model is recommended for use in the subsequent empirical studies.

Chapter 8 quantifies the rates of total factor productivity growth and technical change in the UK retail-banking sector during the sample period 1984-1995. A translog time-trend model is employed with an intermediation model of bank production. Low levels of technical change and total factor productivity growth are observed for the sector overall. The bias of technical change indicates that deposit inputs would be increasingly substituted for capital and labour inputs as technical change continued.

Chapter 9 considers cost efficiency in the UK building society industry between 1990 and 1996. A flexible Fourier functional form with an intermediation model is employed as the representation of productive technology. A fixed effects panel data model is used for the estimation of economies of scale, economies of scope and distribution-free cost efficiency. Overall, low levels of efficiency are observed. Higher levels of efficiency and lower levels of efficiency dispersion were recorded when the model was re-estimated separately using data for societies greater or smaller than £100m in total assets. The smaller building societies proved to have substantially higher levels of efficiency. From this analysis, it is suggested that; (a), smaller building societies may have a productive technology which is distinct from all other building societies; and (b), the Building Societies Act (1986) which sanctioned the greater diversification of larger building societies (those societies with excess of £100m in total assets) may have had a negative influence on the cost efficiency of these societies.

Chapter 10 employs a cost function shift approach to estimate the levels of economies of scale, economies of scope, cost complementarities, technical change and total factor productivity growth in the UK building society sector between 1990 and 1996. A translog functional form and an intermediation specification of bank production are employed. Negative rates of technical change and low rates of total factor productivity growth are reported. It is also observed that building societies of different sizes have distinct experiences of total factor productivity growth over the sample period. Economies of scope are reported for the joint production of class 1 and class 2 assets with class 3 assets

for 'large' societies and diseconomies of scope are reported for 'small' societies. Mixed and contradictory cost complementarity estimates are also recorded reducing the level of confidence placed in the economies of scope estimates. Constant returns to scale and slight economies of scale are reported for all asset classes.

11.3 Economies of scale and product-mix

In chapter 7, the UK retail banks are seen to have experienced positive economies of scale. The empirical study considering economies of scale in the UK building society sector (chapter 10) indicates that constant returns to scale and slight positive economies of scale are present for building societies at different times during the sample period.

Positive economies of scale imply that a larger scale of operation may provide a lower average cost. It could therefore be stated that larger-scale production in a market with positive economies of scale would be in the public interest. Such a position would in turn support the continuing trend towards mergers of banks and building societies. In practice, such an outcome of public benefit may not be automatic as banks or building societies may not operate at minimum average costs (as reported in chapter 6). This problem is compounded by the 'local' form of scale economies, i.e. the scale economies estimated in the empirical chapters only measure the effect of a marginal change in the level of output and do not consider the 'global' form of scale efficiency (where changes in scale efficiency are based on the movement of output that would lead to minimum average costs). This approach therefore assumes that in the level of cost inefficiency is minimal

and does not differ substantially between depository institutions. A study considering the sources of economies of scale and the measurement of a 'global' measure of scale efficiency would be an appropriate direction for future research.

The economies of scope results for the retail-banking and building society sectors differ in terms of the products considered. Diseconomies of scope for the retail-banking sector are recorded for both the joint production of all loans, aggregate investments and total deposits with the production model and for the joint production of investments and loans with the intermediation model. However, cost complementarity estimates reported for both these cases are not supportive, reducing the level of confidence in the plausibility of the estimates.

Economies of scope, of producing both class 1 and class 2 assets with class 3 assets, are reported for the 'large' building societies and diseconomies of scope are recorded for 'small' building societies in chapter 10. The main policy implication of this result is that the legislation enabling building societies to diversify into a broader range of activities appears to be justified in terms of cost efficiency. The results recorded for retail banks and building societies provide support for the provision of building society class 1 and class 2 assets separately from class 3 assets for the smallest building societies, the joint provision of class 1 and 2 assets with class 3 assets for 'large' building societies, and the separate provision of retail bank loans, deposits and investments.

11.4 A descriptive analysis of efficiency

It is important to assess the efficiency of depository institutions for a number of reasons. The level of efficiency of a depository institution provides an indication of its relative 'success' in production terms both at the level of the individual depository institution and at the level of the industry as a whole. This 'success' deserves attention as institutions in competitive markets will be better able to maintain and develop their businesses when they operate efficiently. In addition, the efficient depository institution will be able to avoid the high costs of failure for itself, the industry and the broader economy.

In chapter 7, firm-specific distribution-free cost efficiency estimates are made for both production and intermediation specifications of UK retail banks. A correlation coefficient of 0.815 is calculated for the two sets of efficiency results, indicating a high degree of association between the two model forms. The dispersion of cost efficiency, however, differs substantially between the two model specifications.

In chapter 9, the distribution-free cost efficiency estimates indicate that only a moderate degree of cost efficiency is prevalent in the building society sector. The distribution-free cost efficiency estimates suggest that the average level of cost efficiency for large UK building societies between 1990 and 1996 was 77 per cent and the average level of efficiency for small societies was 91 per cent.

The causes of the differing efficiency results may be attributed to a wide range of sources. These include (a) the use of differing estimation methods and models, (b) the growth of the depository institutions over the sample period, (c) a distinct productive technology employed by small depository institutions and (d) the prevailing level of competition within depository institution core markets.

a) The use of different estimation methods and models

In chapter 7, two distinct although theoretically sound models of bank production were estimated. The results from these models, whilst associated, differ greatly in terms of magnitude, suggesting that the model form is influential in the determination of economic characteristics. Equally, the estimates of economies of scale in the retail-banking sector differ between chapter 8, where a time trend model is used and chapter 7, where a fixed effects panel data model is estimated. The differences between these two studies, including the inclusion of a time-trend dummy and different estimation techniques, led to similar estimates, which differed in terms of magnitude. Thus, both model form and the estimation technique may have an influence on at least the magnitude of efficiency estimates.

b) The growth of depository institutions over the sample period

The growth of depository institutions may result in control and allocation difficulties. The larger the institution, the broader and less transparent are the links between the

employees, the management and the customers. This position may have enabled allocative inefficiencies to form. The process of expansion of institutions may engender inefficiency similar to that indicated by Leibenstein (1966), who suggested that expansion to a larger scale of production provides greater opportunities for costs to expand and may allow individual workers and institutions to perform sub-optimally. Whilst this explanation is not proposed as a rejection of the potential cost efficiency benefits of depository institution expansion from sources such as economies of scale, the movement towards larger depository institutions through both internal expansion and external acquisition and merger may in itself result in technical and allocative inefficiency.

c) A distinct productive technology employed by small institutions

In chapter 9, it was reported that cost efficiency in small building societies appeared to be different from the cost efficiency of larger building societies. This finding may suggest that the smaller building societies have a different productive technology from larger societies. It is assumed that different productive technologies would result in greater specification error and possibly an overestimation of the dispersion of cost efficiency.

d) The level of competition

The level of competition may influence the efficiency of a depository institution in a number of ways. The impact of a low cost entrant on a depository institution's core

market may be substantial. A new entrant, with relatively low costs, would be expected to cause a degree of disturbance in the stability of market shares, the long-run equilibrium of the market and the subsequent profitability of incumbents. This would occur as it can be assumed that low cost new entrants would increase the levels of competition where “... competition causes efficient organisations to prosper at the expense of inefficient ones” (Vickers, 1995, p.1). Such a disturbance may have been observed in the earlier part of the sample period (1990-1992) in the building society mortgage market where disequilibrium conditions were observed (see chapter 6). The degree to which these new entrants influenced the building society mortgage market, however, may be limited due to the relative stability of market shares over the sample period.

The degree of competition may also influence the efficiency of depository institutions through the ‘sharpening’ of incentives and monitoring of managers to ensure that the objectives of cost minimisation and efficiency maximisation are followed. Although a limited amount of empirical and anecdotal evidence exists supporting such a process (see Vickers, 1995 and Nickell, 1996, for a discussion of the literature), the methods of transmission of information about the differential performance of individual managers are less than clear. It has even been suggested that the level of competition has very little influence on the efficiency and productivity of the individual firm, where the existence of competition may be influential in providing a framework or environment in which many differing innovations, approaches or ideas may be employed. Nickell (1996) exploring this issue, suggested that “... if there are lots of ways of doing things, competition allows many to be tried and then selects the best, something a monopoly finds hard to

replicate” (Nickell, 1996, p.741). While this eventuality is not considered to be implausible, in chapter 6 evidence was put forward to show that both building societies and retail banks operate within long-run Chamberlinian equilibrium for at least some of the sample period. Under such conditions, these depository institutions will operate where average cost is tangential with the demand curve, a point not necessarily at minimum average costs. This position provides managers with an effective cost ‘cushion’ (Berger and Hannan, 1998) enabling managers to follow objectives other than cost minimisation and efficiency maximisation and limiting the possible influence of competition in reducing ‘slack’ in this industry.

Further discussion of the existence of distinct productive technologies in the building society sector is contained in chapter 9. The potential sources of efficiency (see for example Berger and Mester, 1997 and Hermalin and Wallace, 1994) and the association between efficiency and the level of competition (see for example Vickers, 1995 and Hay and Liu, 1997) are areas where further research activity would be beneficial.

11.5 A survey and assessment of the estimates of total factor productivity growth and technical change.

It has been stated in both the literature review (chapter 4) and the relevant empirical chapters (chapter 8 and 10) that previous empirical studies concerned with estimating total factor productivity growth and technical change in depository institutions have often produced small or static estimates for the 1980s and 1990s. The empirical chapters (chapters 8 and 10) estimate the levels of technical change and total factor productivity

growth for UK retail banks and building societies and also find very low rates of change and growth. An econometric time-trend model (chapter 8) and an econometric cost function shift model (chapter 10) are used in these studies.

The time-trend model and cross-sectional shift model both provide similar estimates. The general finding that technical change appears to be minimal and total factor productivity growth occurs only at low levels is reinforced. It was also shown that building societies with differing sizes had distinct experiences of technical change and total factor productivity growth during the sample period.

The implications of low productivity growth for such a large commercial sector are both substantial and negative. The potential output of the national economy is reduced as more of depository institutions' resources are dedicated to low productivity activities. It follows that as potential output within the entire economy falls, a decline in the level of sustainable wage increases is experienced. This effect, may in turn, transmit into higher rates of inflation. Why minimal levels of technical change and total factor productivity growth have occurred in UK depository institutions during the 1980s and 1990s is therefore a question demanding greater discussion and explanation.

The presence of zero or low levels of technical change and total factor productivity growth is, to a degree, at odds with the anecdotal evidence discussed in chapter 2. Banks and building societies have undertaken substantial change in terms of employment practice, the distribution of services and computerisation of a range of banking functions.

A number of potential causes or explanations of low productivity have been forwarded to explain why technical change and total factor productivity growth has been low. Possible sources or explanations that will be discussed include (a) levels of inefficiency in the sector, (b) the adoption of new technology and structural change in the industry, (c) the presence of low levels of competition, (d) the service industry 'cost disease' and (e) mis-measurement.

a) *The level of inefficiency in the sector*

The presence of a static or low rate of technical change indicates either no improvement or only a slight increase in the level of productive technology. Total factor productivity growth is a construct of technical change, allocative, technical and scale economies. The presence of low total factor productivity growth and static technical change together with positive economies of scale (see chapters 7 and chapter 10) suggests that one of the other components of productivity growth may also be changing. This position supports the conclusion that either technical or allocative inefficiency or both are present in the sector. Such inefficiency may bias the results we have gained and disguise possible improvements in technical change or total factor productivity growth¹ that may have emerged over the sample period.

More directly, the presence of cost inefficiency may have limited the level of potential productivity growth. Total factor productivity growth is dependent on an improvement in the efficiency of production. Any improvement from an inefficient position may therefore

be a 'catching up' process, as inefficient firms become more efficient. This could provide a misleading interpretation of productivity growth and technical change. Such a difficulty was identified in US commercial banking by Berger and Humphrey (1992b) where "... it is also difficult to measure technical change and productivity growth because of the confounding effects of changes in inefficiency over time and the deregulation of the deposit side of banking. If inefficiency is not taken into account, then measures of technical change or productivity growth may confuse shifts in the minimum-cost technology with changes in the deviations from that technology" (p.275).

(b) The adoption of new technology and structural change

The restructuring of UK depository institutions during the 1980s and 1990s has witnessed the introduction of a wide range of new technology. This investment has been used to assist production in several diverse areas, including distribution and information processing, in addition to reducing transaction costs. It can be initially stated that many upheavals, such as staff retraining and widespread disruption, must have been incurred over the sample period. Secondly, staff employed by depository institutions may have had to spend a considerable amount of time learning how to use new technology efficiently. This process of learning may be mis-interpreted as a decline in productivity. Thirdly, the management of new technology may also be problematic. Maclean (1997) suggested that "... initially, technology is used largely to improve the way in which traditional tasks are performed. It takes greater familiarity with the potential of new technology before more fundamental changes in organisation and the production process

can be implemented to better exploit the technology” (p.21). Lastly, the diffusion of new technology may be unequal across the sector as a whole, due to the high costs of purchasing new technology. This may be illustrated by the differential productivity and technical change estimates for different asset sizes of building society, observed in chapter 10. To summarise, high levels of investment and the associated disruptive activity may have adversely affected the level of total factor productivity growth in depository institutions in the short and medium term. Equally, the low productivity growth we have observed may represent productivity during a time of change and perhaps underestimate the true long-term total factor productivity growth.

c) *The presence of low levels of competition*

The low levels of total factor productivity growth may have resulted from low levels of competition within the sector. In chapter 6, it was suggested that the existence of low levels of competition may result in the management of depository institutions following objectives other than cost minimisation and efficiency maximisation. The moderate levels of market power in both the building society and banking core markets indicate that such behaviour could be a strong possibility in UK retail banks and building societies. The consequent low levels of efficiency, associated with such behaviour would in turn lead to low total factor productivity growth and low levels of technical change. It may be stated that the possible link between the levels of competition, productivity and efficiency is an area that requires greater consideration and inquiry. Further discussion of this and other related issues is provided by Hay and Liu (1997) (also see section 11.4).

(d) The service industry's 'cost disease'

The products of building societies and retail banks, such as mortgages, are composite products, which are 'manufactured' using an amalgam of services and transactions (Baumol, 1991). The inclusion of 'personal services' of customer care within products may reduce the potential for total factor productivity growth. Within such an environment, however much is spent on new technology will still leave the productive technology underpinning the service unchanged. This is assumed to lead to rising input costs, in turn associated with low productivity growth and technical change. This low productivity is therefore viewed as being as much a function of the limited scope of 'improvement' possible with composite products. Whilst this argument certainly broadens understanding of the issue at hand, its conclusions have been considered by some to be premature. Maclean (1997), for example draws attention to Baumol's (1967) example of a horn quintet where " ... any attempt to increase productivity here is likely to be viewed with concern by critics and audience alike" (p. 416). Maclean indicated that this argument may be limited when new technology alters the nature of the service provided, where, " ... advances in sound production, recording and broadcasting allow one performance of the quintet to be heard by millions of people simultaneously, and re-heard over and over again." (Maclean, p. 26). This conclusion suggests that an emphasis on the quality or convenience a product may bring to a customer may differ substantially between products, adopting differing types of new technology.²

(e) Measurement problems

The negligible rates of productivity growth and technical change observed in the empirical studies may result from a more fundamental problem. The presence of a global or western slowdown in productivity (see Griliches, 1994) or a process of international productivity convergence (Baumol *et al*, 1994) as differences in the technological sophistication of nations have declined, has been mooted by many commentators. The results we have observed may have been a part of this hypothesised global phenomenon. Productivity growth throughout white-collar work in the western world may have slowed significantly within the last twenty years. Reasons that have been forwarded to explain this phenomena have included data problems, the possibility of exogenous shocks on the industry and imprecise or abstract model assumptions.

Griliches (1994) suggested that during the 1970s and 1980s the sources of productivity growth had become confused as productivity growth was seen to slow down. The characteristics of the productivity-growth decline were suggested to be concentrated particularly in the areas of construction, finance and services sectors, where output measurement is particularly difficult. Griliches indicated that the areas of agriculture, manufacturing and communication, where output measurement is less of an issue had not experienced such a major slowdown in productivity growth. This finding led to the conclusion that we may not have an adequate understanding of the mechanisms producing growth or adequate data to indicate whether there has been such a shift.

Associated with this conclusion is the possibility of mis-measurement of productivity and technical change of depository institutions. In chapter 3, the modelling of production in depository institution was discussed. It was suggested that measurement of most inputs and outputs has been problematic. Some commentators have emphasised that 'huge' mis-measurement occurs due to the changing levels of quality and the impact of innovations on the form of products and services. For example, Nordhaus (1997) suggested that "... constructing price indices that adequately capture the impact on price of new technologies, especially radically new inventions, is beyond the practical capability of current techniques" (p.1549). Other possible problems that may have negatively influenced the efficiency, productivity and technical change estimates include the use of a general price deflator (the Retail Price Index) in all the studies, as opposed to a price deflator specifically designed for UK depository institutions (to my knowledge, unavailable, although desirable). This general price deflator may have led to a bias in results. Compounding this problem is the large number of intermediate goods in bank production, which may experience bias from, for example, double deflation.

More broadly, the definition of technical change may have omitted the degree of technical change that may be 'embodied' in capital itself (see for example, Hulten, 1992) or the influence of past approaches on future innovations, proposed in the 'path dependence' explanations of technical change (see Rutten, 1997 for a critical assessment of path dependence and other theories relating to the sources of technical change). Embodied capital has been suggested to be a significant and substantial determinant of the overall slowdown in productivity since 1973, with the effect on productivity varying

between nations from a low of 23 per cent in Japan to 69 per cent in France (Wolff, 1996).

11.6 Problems with our conception of the 'firm'

A number of assumptions are made in the construction of all cost and production functions. These assumptions include the objective function of firms, the cost free dissemination of technology, the breadth of the efficiency concept considered, the replication of technology, homogeneity of factors of production and the homogeneity of production technology in relation to distinct products.

The assumption of cost minimisation employed in the thesis may be questioned. Efficiency throughout the thesis relates to the minimisation of input costs in the production of output. Cost minimisation is therefore assumed in all the empirical studies. This conception could be viewed as unnecessarily narrow for a number of reasons. Firms within oligopolistic markets, for example, may seek objectives of sales maximisation (Baumol, 1959). The managers of depository institutions may seek expenses that add to their own utilities, over those which improve firm efficiencies and productivity (see Edwards, 1977, Mester 1989, Drake, 1995 for further explanation). Baumol (1990) broadens this argument by suggesting that all entrepreneurial activity is driven by profit. The sources of profit in any society are defined as the prevalent reward system which may or may not be associated with cost minimisation. Whilst these theories both acknowledge and forward explanations for sub-optimal behaviour, they still emphasise

the firm as a maximiser. This assumption that the firm will maximise may also be cast as a most significant criticism. The relaxation of maximisation within a firm context may emerge from a broad range of sources, such as inertia, insufficiently complete contracts, customs and standard procedures within the workplace (see North, 1990 and Streeck, 1992 for further discussion).

The conception of the firm as a command institution where maximisation is reproduced mechanically may additionally be questioned. Managers may be seen to adapt decision making to both their own and worker individualism, defining progress through collective decisions and as such are not in total control of the firm. Therefore, whilst the consideration of such factors shifts the emphasis of the analysis from the firm towards the individual level and away from the approaches adopted throughout the thesis, limitations when assuming firm maximisation are at least acknowledged. It is suggested that future research work, particularly about the corporate governance of depository institutions, may provide further insights and clarification of these criticisms.

In chapter 5, the efficiency concept employed in the thesis was introduced. The assumptions on which the cost efficiency concept is based have experienced a diversity of criticism. Firms or institutions represented by production and cost functions are assumed to be able to replicate existing states of technology achieved by competing firms. In some cases (such as radial measures), assumptions of the construction of an infinite range of technologies is deemed to be both attainable and cost free. The assumptions have been viewed as assuming a 'book of blueprints' for differing

productive states. It is assumed that such knowledge may be acquired and implemented at no cost to the institution. The homogeneity of factors of production is essential to the existence of a 'book of blueprints' conception. It is acknowledged that this assumption may be inconsistent with many aspects of firm behaviour.

Leibenstein (1966) indicated that in all production functions a certain 'experimental element' will remain, where even if the existing relationship between inputs and outputs is understood, change in the input ratios would create uncertainty and limit our understanding. Attributes of skills and organisational knowledge and learning have been posited as potential exceptions to the suggested homogeneity of factors and may form this 'experimental element'. Teece (1982) expands such a view where "... in the exercise of individual skill, many actions are taken that are not the result of considered choices but rather are automatic responses that constitute aspects of the skill" (p.40). Skill is thus viewed as a function of doing, and experience which may not be derived at zero costs. Organisational knowledge may be viewed as the knowledge that facilitates an organisation to co-ordinate production according to certain limits of technology.

Organisational knowledge, conceptualised within a cost or production function, is achievable for all observed combinations and, for radial measures, for all potential permutations of observed technologies. This assumption applies both locally and globally. Difficulties with this assumption occur, in that, while an individual within an institution may fully understand his or her individual and team responsibilities and the importance and relation of such responsibilities within the entire institution, no individual

or sub-set of individuals within the institution is able to understand all the possible responsibilities and their relationship to each other. Such a property or attribute held by the organisation, by its intangibility, is difficult to quantify and exchange freely with other firms.

Questions of the viability of the assumption of homogeneity of technology in relation to the production of distinct products may be posited as a further potential failing of this form of analysis. Emphasis is placed on institutions with experience in a general area of production that may diversify into the production of products or services with similar production technologies. The movement into the production of products with highly distinct production technologies is viewed as difficult. Teece (1982) suggested that “ ... whereas the neo-classical firm selects, according to factor prices technologies off the shelf to manufacture a given end product” ... the conception of the firm that incorporates fungibility “ ... selects an end product configuration, consistent with its organisation and technology which is defined yet fungible over certain arrays of final products.” (p.42).

11.7 Overall conclusions and suggestions for further research.

In the thesis, a number of central themes have been developed. The UK depository sector has a noticeable dispersion of efficiency. Economies of scale appear to be present in both the retail banking and building society sectors. The rates of technical change and total factor productivity growth have been low to negligible and potentially constrained by a

broad array of factors, including the differential application and diffusion of technology, low levels of competition and the level of efficiency dispersion within the sector.

In qualification of the thesis findings, a number of potential difficulties have been identified. Whilst the consideration of particular components of efficiency and productivity change, both in terms of their consistency of estimation, emphasis and scope, is essential for analysis of this sector, a clear demarcation between such components remains unclear. Additionally, the data used and variables applied in the analysis, demand continued attention both in terms of quality and breath of information provided. The sources of efficiency observed within both the retail banking and building society sectors remain unclear; it may only be assumed that the primary source of this inefficiency is poorly applied entrepreneurial and managerial behaviour. The best indications of possible sources of efficiency seem to emerge through the application of differing models and techniques and the link with the low levels of productivity growth, which may be related to the substantial re-structuring and the low levels of competition in the sector. The precise delineation of such an association remains unanswered. Additionally, the econometric cost function approach used throughout the thesis is limited in the scope of the 'true' productive technology it may incorporate, assuming away potentially essential elements such as skill and differential product quality.

Considering such limitations, there are a number of possible avenues for further investigation. The reliance on Diewert functional forms, such as the translog functional form could have overestimated economies of scale and led to more dispersed efficiency

results than may be present. The amendment for, or at least acknowledgement of, the potential for mis-specification is strongly suggested. The use of different efficiency concepts, such as profit efficiency, may enable the differing objectives of the firm to be assessed. For example, Berger and Humphrey (1997) suggested that both cost (or input) efficiency measures and profit (or output) efficiency measures may be contrasted to provide a greater understanding of the level of service intensity individual firms provide. In support of such a view, Braeutigam and Pauly (1986) employ an alternative cost function approach to quantify possible quality bias. The authors strongly argue that assuming away differential product quality may be “ ... an omission” ... that may “ ... seriously bias estimates of economically important parameters, for example, economies of scale” (p.616). The use of both econometric and programming approaches in tandem may provide further insights into potential specification errors that could occur with either technique. Through the comparison of results, further insight into both the levels (Sudit, 1995, p.450) and sources of efficiency within institutions may be provided.

A noticeable rise in the number of products provided and outlets for the customer to use has occurred over the sample period. An acknowledgement of the diversity of production in institutions is therefore suggested. The associated increase in the level of specialisation, diversity and the ramifications of this change may be a starting point for such analysis. The importance and incorporation of risk taken by individual institutions within production may similarly be posited as another avenue for investigation.

A broader analysis of competitiveness and market power is deemed to be very important, perhaps considering more flexible techniques to incorporate issues such as regulation, market power and the dispersion of prices in core product markets. Empirical work on the form of competition or exercise of market power, considering both the structure-conduct-performance (SCP) (Berger, 1995 and Berger and Hannan, 1998) and the new economic industrial organisation (NEIO) paradigms (Bresnahan, 1989 and Vesala, 1995) provides a starting point for further work in this area.

To conclude, the thesis has considered the competitiveness, efficiency and productivity of UK retail banks and building societies. The measurement of these economic characteristics has been the primary aim of the thesis. The type of analysis employed, previously termed “... the industrial economics of banking” (Molyneux *et al*, 1996), is appropriate for questioning any of the assumptions made, and has, hopefully, provided a broader assessment of this subject area than has previously been available.

¹ This is possible as total factor productivity growth and technical change are estimated as average as opposed to extremal measures.

² This point is associated with the mis-measurement issues also considered in this section.

Bibliography

Alhadeff, D. A. (1954) *Monopoly and Competition in Banking*, University of California Press.

Aigner, D., C. A. Knox Lovell, and P. Schmidt, (1977) "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics*, vol.6, 21-37.

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

Aly, H., R. Grabowski, C. Paskurka, and N. Rangan, (1990) "Technical, scale and allocative efficiencies in U.S. banking", *Review of Economics and Statistics*, vol.72, 211-218.

Bain, J. S. (1951) "Relation of profit rate to industry concentration: American manufacturing, 1936-1940", *Quarterly Journal of Economics*, vol.65, no.3, 293-324.

Baltagi, B. H. and J. M. Griffin, (1988) "A general index of technical change", *Journal of Political Economy*, vol.96, 20-41.

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

Bank of England, (1987) "Building Societies: a changing role", *The Bank of England Quarterly Review*, August, 400-403.

Banker, R. D., A. Charnes, W. W. Cooper, and A. Maindiratta, (1988) "A comparison of DEA and translog estimates of production frontiers using simulated observations from a known technology", from Dogramaci, A. and R. Fare, (eds) *Applications of Modern Production Theory: Efficiency and Productivity*, Kluwer Academic Publishers, Boston.

Baumol, W. J. (1959) "The revenue maximisation hypothesis", in *Business Behaviour, Value and Growth*, Macmillan, New York, 45-53.

Baumol, W. J. (1982) "Contestable markets: An uprising in the theory of industrial structure", *The American Economic Review*, vol.72, no.1, 1-15.

Baumol, W. J. (1990) "Entrepreneurship: productive, unproductive, and destructive", *Journal of Political Economy*, vol.98, no.5, part.1, 893-921.

Baumol, W. J. (1991) "Technological imperatives, productivity and insurance", *The Geneva Papers on Risk and Insurance*, vol.116, 154-165.

Baumol, W. J., R. R. Nelson, and E. N. Wolff, (1994) *The Convergence of Productivity. Cross-National Studies and Historical Evidence*, Oxford University Press, New York.

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

Barnes, P. and M. Ward, (1997) "The consequences of deregulation: A comparison of the experiences of UK building societies with those of US savings & loan associations", presented at the *4th Annual Conference of the Multinational Finance Society*, June 25-29, 1997, Thessaloniki, Greece.

Barnett, W. A. (1978) "The user cost of money", *Economics Letters*, vol.1, 145-149.

Beard, R. T., S. Caudill, and D. M. Gropper, (1991) "Finite mixture estimation of multiproduct cost functions", *Review of Economics and Statistics*, 654-664.

Bell, F. and N. Murphy, (1968) "Economies of scale and division of labor in commercial banking", *Southern Economic Review*, vol.35, October, 131-139.

Benston, G. J. (1965) "Branch banking and economies of scale", *National Banking Review*, vol.2, June, 507-549.

Benston, G. J. (1972) "Economies of scale in financial institutions", *Journal of Money Credit and Banking*, (April), 312-41.

Benston, G. J., G. Hanweck, and D. B. Humphrey, (1982) "Scale economies in banking. A restructuring and reassessment", *Journal of Money, Credit and Banking*, vol.14, November, 435-456.

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, MIT Press, Washington.

Benston, G. J. and C. W. Smith, (1976) "A transactions cost approach to the theory of financial intermediation", *The Journal of Finance*, vol.31, no.2, 215-231.

Berg, S. A., F. R. Førsund, L. Hjalmarsson, and M. Suominen, (1993) "Banking efficiency in the Nordic countries", *Journal of Banking and Finance*, vol.17, 371-388.

Berg, S. A., F. R. Førsund, and E. S. Jansen, (1992) "Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89", *Scandinavian Journal of Economics* (Supplement), 211-228.

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. (1995) "The profit-structure relationship in banking – Tests of market power and efficient-structure hypothesis", *Journal of Money, Credit and Banking*, vol.27, no.2, May, 404-431.

Berger, A. N., D. J. Cummings, and M. A. Weiss, (1997) "The coexistence of multiple distribution systems for financial services: The case of property-liability insurance", *Journal of Business*, vol.70, no. 41, 515-546.

Berger, A. N. and R. DeYoung, (1997) "Problem loans and cost efficiency in commercial banks", *Journal of Banking and Finance*, vol.21, 849-870.

Berger, A. N. and T. H. Hannan, (1989) "The price concentration relationship in banking", *The Review of Economics and Statistics*, vol.71, 291-299.

Berger, A. N. and T. H. Hannan, (1998) "The efficiency cost of market power in the banking industry: a test of the 'Quiet Life' and related hypotheses", *The Review of Economics and Statistics*, vol.80, no.3, 454-465.

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

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

Berger, A. N. and D. B. Humphrey, (1992a) "Measurement and efficiency issues in commercial banking" from *Studies in Income and Wealth*, vol.56, *Output measurement in the service sectors*, (Ed) Griliches, Z., University of Chicago Press, Chicago, 245-300.

Berger, A. N. and D. B. Humphrey, (1992b) "Megamergers in banking and the use of cost efficiency as an antitrust defence", *The Antitrust Bulletin*, Fall, 541-600.

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

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

Bourke, P. (1989) "Concentration and other determinants of bank profitability in Europe, North America and Australia", *Journal of Banking and Finance*, vol.13, 65-79.

Bresnahan, T. (1989) "Empirical studies of industries with market power", from Schmalensee, R. and R. Willig, (Eds)., *Handbook of Industrial Organization*, vol. II, Elsevier Science Publishers B.V., Netherlands.

Braeutigam, R. R. and M. V. Pauly, (1986) "Cost function estimation and quality bias: the regulated automobile insurance industry", *Rand Journal of Economics*, vol.17, no.4, 606-617.

Brigham, E. F. (1964) "Economies of scale in the savings and loan industry", *Western Economic Journal*, vol.3, 7-20.

British Bankers Association, *Abstract of Banking Statistics*, various editions.

Brown, R. S., D. W. Caves, and L. R. Christensen, (1979) "Modeling the structure of cost and production for multiproduct firms", *Southern Economic Review*, vol.46, July, 256-273.

Bukh, P. N. D., S. A. Berg, and F. R. Førsund, (1995) "Banking efficiency in the Nordic countries: A four country Malmquist Index Analysis", *Arbeidsnotat Research Department 1995/7*, Norges Bank, September, Oslo.

Caves, D. W. and L. R. Christensen, (1980) "Global properties of flexible functional forms", *American Economic Review*, vol.70, 422-432.

Chamberlin, E. H. (1933) *The Theory of Monopolistic Competition*, Harvard University Press, Cambridge MA.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau, (1971) "Conjugate duality and the transcendental logarithmic production function", *Econometrica*, vol.39, 255-256.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau, (1973) "Transcendental Logarithmic Production Frontiers", *The Review of Economics and Statistics*, vol.55, 28-45.

Clark, J. A. (1984) "Estimation of economies of scale in banking with a generalised functional form", *Journal of Money, Credit and Banking*, vol.16, 53-68.

Coles, A. (1992) "Concentration and mergers in the building society industry", *Housing Finance*, no.16, November, 6-13.

Colwell, R. J. and E. P. Davis, (1992) "Output and productivity in banking", *Scandinavian Journal of Economics Supplement*, 111-129.

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

Crapp, H. R. (1983) "Scale economies in the N.S.W. permanent building society sector", *Australian Journal of Management*, vol.1, 33-43.

Daniel, K. N. and D. Tirtiroglu (1998) "Total factor productivity growth in U.S. commercial banking for 1935-1991: A latent variable approach using the Kalman filter", *Journal of Financial Services Research*, vol. 13, no.2, 119-135.

Denny, M., M. Fuss, and L. Waverman, (1981) "The measurement and interpretation of total factor productivity in regulated industries with an application to Canadian telecommunication", from *Productivity Measurement in Regulated Industries*, (eds) Cowing, T. G. and R. E. Stevenson, Academic Press, New York.

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

DeYoung, R. (1998) "Management quality and X-inefficiency in national banks", *Journal of Financial Services Research*, vol.13, 4-22.

- DeYoung, R. and I. Hasan, (1998) "The performance of de novo commercial banks: A profit efficiency approach", *Journal of Banking and Finance*, vol.22, 565-587.
- Diamond, D. W. (1984) "Financial intermediation and delegated monitoring", *Review of Economic Studies*, vol.51, 393-414.
- Dietsch, M. (1993) "Economies of scale and scope in French commercial banking industry", *The Journal of Productivity Analysis*, vol.4, 35-50.
- Diewert, W. E. (1971) "An application of the Shephard duality theorem: A generalized Leontief production function", *Journal of Political Economy*, vol.79, 481-507.
- Diewert, W. E. (1992) "The measurement of productivity", *Bulletin of Economic Research*, vol.44, no.3, 163-190.
- Donovan, D. J. (1978) "Modelling the demand for liquid assets: An application to Canada", *IMF Staff Papers*, vol.25, no.4, (December), 676-704.
- Dowling, W. A. and G. C. Philippatos, (1990) "Economies of scale in the US savings and loan industry: 1973-83", *Applied Economics*, vol.22, 237-248.

Drake, L. (1992) "Economies of scale and scope in U.K. building societies: an application of the translog multi-product cost function", *Applied Financial Economics*, vol.2, 211-219.

Drake, L. (1995) "Testing for expense preference behaviour in UK building societies: The implications for measures of scale and scope economies", *The Service Industries Journal*, vol.15, no.1, (January), 50-65.

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.

Drake, L. and T. G. Weyman-Jones, (1992) "Technical and scale efficiency in U.K. building societies", *Applied Financial Economics*, vol.2, 1-9.

Drake, L. and T. G. Weyman-Jones, (1996) "Productive and allocative inefficiencies in UK building societies: A comparison of non-parametric and stochastic frontier techniques", *The Manchester School*, vol.64, no.1, (March), 22- 37.

Eastwood, B. J. and R. A. Gallant, (1991) "Adaptive rules for seminonparametric estimators that achieve asymptotic normality", *Econometric Theory*, vol.7, 307-40.

Edgar, R. J., J. H. Hatch, and M. K. Lewis, (1971) "Economies of scale in Australian banking, 1947-68", *The Economic Record*, vol.117, 17-37.

Edwards, F. R. (1977) "Managerial objectives in regulated industries: Expense preference behaviour in banking", *Journal of Political Economy*, vol.85, no.1, 147-162.

Elyasiani, E., S. Mehdian, and R. Rezvanian, (1994) "An empirical test of association between production and financial performance: the case of the commercial banking industry", *Applied Financial Economics*, vol.4, 55-59.

Elyasiani, E. and S. Mehdian, (1990) "A nonparametric approach to measurement of efficiency and technological change: the case of large U.S. commercial banks", *Journal of Financial Services Research*, vol.4, 157-168.

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

English, M., S. Grosskopf, K. Hayes, and G. Yaisawarng, (1993) "Output allocative and technical efficiency of banks", *Journal of Banking and Finance*, vol.17, 349-366.

Esho, N. and I. G. Sharpe, (1995) "Long-run estimates of technological change and scale economies in a dynamic framework: Australian permanent building societies, 1974-1990", *Journal of Banking and Finance*, vol.19, 1135-1157.

European Economy, (1997), Supplement A, no.12, European Commission.

Evanoff, D. D. and P. R. Israilevich, (1994) "Scale elasticity versus scale efficiency in banking", *Southern Journal of Economics*, 1036-1046.

Favero, C. A. and L. Papi, (1995) "Technical efficiency and scale efficiency in the Italian banking sector: a non-parametric approach", *Applied Economics*, vol.27, 385-395.

Ferrier, G. D., S. Grosskopf, K. J. Haynes, and G. Yaisawarng, G. (1993) "Economies of diversification in the banking industry", *Journal of Monetary Economics*, vol.31, 229-245.

Field, K. (1990) "Production efficiency of British building societies", *Applied Economics*, vol.22, 413-426.

Fixler, D. J. and K. D. Zieschang, (1992) "User costs, shadow prices, and the real output of banks", from *Studies in Income and Wealth (Output measurement in the service sectors)*, vol.56, (Ed). Griliches, Z. University of Chicago Press, Chicago, 219-243.

Flannery, M. J. (1974) *An Economic Evaluation of Credit Unions in the United States*, Research Report, no.54, Federal Reserve Bank of Boston.

Flannery, M. J. (1983) "Correspondent services and cost economies in commercial banking", *Journal of Banking and Finance*, vol.7, 83-99.

Førsund, F. R., C. V. Knox Lovell, and P. Schmidt, (1980) "A survey of frontier production functions and of their relationship to efficiency measurement", *Journal of Econometrics*, vol.13.

Fried, H. O., C. V. Knox Lovell, and P. V. Eeckaut, (1993) "Evaluating the performance of US credit unions", *Journal of Banking and Finance*, vol.17, 251-265.

Fukuyama, H. (1993) "Technical and scale efficiency of Japanese commercial banks: a non-parametric approach", *Applied Economics*, vol.25, 1101-1112.

Gallant, A. R. (1981) "On the bias in flexible functional forms and an essentially unbiased form", *Journal of Econometrics*, vol.15, 211-245.

Gallant, A. R. (1984) "The Fourier flexible form", *American Journal of Agricultural Economics*, May, 201-208.

Gilbert, R. A. (1984) "Bank market structure and competition", *Journal of Money, Credit and Banking*, vol.16, no.4, 617-645.

Gilligan, T., M. Smirlock, and W. Marshall, (1984) "Scale and scope economies in the multi-product banking firm", *Journal of Monetary Economics*, vol.13, 393-405.

Glass, J. C. and D. G. McKillop, (1991) "Efficiency in Irish banking: theory and evidence", *Applied Financial Economics*, vol.1, 235-240.

Glass, J. C. and D. G. McKillop, (1992) "An empirical analysis of scale and scope economies and technical change in an Irish multi-product banking firm", *Journal of Banking and Finance*, vol.16, 423-437.

Goldberg, L. G. and Rai, A (1996) "The structure-performance relationship for European banking", *Journal of Banking and Finance*, vol.20, 745-771.

Gough, T. J. (1979) "Building society mergers and the size-efficiency relationship", *Applied Economics*, vol.11, 185-194.

Guilkey, D. K. and C. A. Knox Lovell, (1980) "On the flexibility of the Translog approximation", *International Economic Review*, vol.21, 137-147.

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.

Grabowski, R., N. Rangan, and R. Rezvanian, (1993) "Organisational forms in banking: An empirical investigation of cost efficiency", *Journal of Banking and Finance*, vol.17, 531-538.

Greene, W. H. (1990) "A gamma distributed stochastic frontier model", *Journal of Econometrics*, vol.46, 141-163.

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.

Griliches, Z. (1994) "Productivity, R&D, and the data constraint", *The American Economic Review*, vol.84, no.1, 1-23.

Gropper, D. M. (1991) "An empirical investigation of changes in scale economies for the commercial banking firm 1979-86", *Journal of Money, Credit and Banking*, 718-727.

- Gropper, D. M. (1995) "Product-line deregulation and cost structure of US savings and loan associations", *Applied Economics*, vol.27, 183-191.
- Hannan, T. H. (1994) "Recent trends in retail fees and services of depository institutions", *Federal Reserve Bulletin*, (September).
- Hancock, D. (1985) "The financial firm: Production with monetary and non monetary goods", *Journal of Political Economy*, vol. 93, no.5, 859-880.
- Hancock, D. (1991) *A Theory of Production for the Financial Firm*, Kluwer Academic Publishers, Boston
- Hardwick, P. (1989) "Economies of scale in building societies", *Applied Economics*, vol.21, 1291-1304.
- Hardwick, P. (1990) "Multi-product cost attributes: A study of U.K. building societies", *Oxford Economic Papers*, vol.42, 446-461.
- Hardwick, P. (1996) "Concentration and Stability in the Mutual Building Society Industry", *The Service Industries Journal*, no.1, vol.16, 47-57.
- Hay, D. A. and G. S. Liu, (1997) "The efficiency of firms: what difference does competition make", *The Economic Journal*, vol.107, (May), 597-617.

Hermalin, B. E. and N. E. Wallace, (1994) "The determinants of efficiency and solvency in savings and loans", *Rand Journal of Economics*, vol.25, no.3, (Autumn), 361-381.

Hicks, J. R. (1939) *Value and Capital. An Enquiry into Some Fundamental Principles of Economic Theory*, Oxford University Press, London.

Hulten, C. R. (1992) "Growth accounting when technical change is embodied in capital", *The American Economic Review*, vol.82, no.4, September, 964-980.

Humphrey, D. B. (1993) "Cost and technical change: effects of bank regulation", *The Journal of Productivity Analysis*, vol.4, 9-34.

Hunter, W. C. and S. G. Timme, (1986) "Technical change, organizational form, and the structure of bank production", *Journal of Money, Credit and Banking*, vol.18, May, 152-166.

Hunter, W. C. and S. G. Timme, (1995) "Core deposits and physical capital: A re-examination of bank scale economies and efficiency with quasi-fixed assets", *Journal of Money, Credit and Banking*, vol.27, no.1, (February), 165-185.

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

Ivaldi, M., N. Ladoux, H. Ossard, and M. Simioni, (1996) "Comparing Fourier and Translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers", *Journal of Applied Econometrics*, vol.11, 649-667.

Jagtiani, J., A. Nathan, and G. Sick, (1995) "Scale economies and cost complementarities in commercial banks: On-and off-balance-sheet activities", *Journal of Banking and Finance*, vol.19, 1175-1189.

Jagtiani, J. and A. Khanthavit, (1996) "Scale and scope economies in large banks: Including off-balance sheet products and regulatory effects (1984-1991)", *Journal of Banking and Finance*, vol.20, 1271-1287.

Jondrow, J., C. A. Knox Lovell, I.S. Materov, and P. Schmidt, (1982) "On the estimation of technical inefficiency in the stochastic frontier production function model", *Journal of Econometrics*, vol.19, 233-238.

Kaparakis, A. G., S. M. Miller, and A. G. Noulas, (1994) "Short-run cost inefficiency of commercial banks: A flexible stochastic frontier approach", *Journal of Money, Credit and Banking*, vol.26, 875-893.

Karafolas, S. and G. Mantakas, (1996) "A note on cost structure and economies of scale in Greek banking", *Journal of Banking and Finance*, vol.20, 377-387.

Kilbride, B. J., B. McDonald, and R. E. Miller, (1986) "A re-examination of economies of scale in banking using a generalised functional form", *Journal of Money, Credit and Banking*, vol.18, 519-526.

Kim, H. Y. (1986) "Economies of scale and economies of scope in multiproduct financial institutions: further evidence from credit unions", *Journal of Money, Credit and Banking*, vol.18, no.2, (May), 220-226.

Kim, M. and J. Weiss, (1989) "Total factor productivity growth in banking: The Israeli banking sector 1979-1982", *The Journal of Productivity Analysis*, vol.1, 139-153.

Kolari, J. and A. Zardkoohi, (1990) "Economies of scale and scope in thrift institutions: the case of Finnish cooperative and savings banks", *Scandinavian Journal of Economics*, vol.92, no.3, 437-451.

Koopmans, T. C. (1951) *Activity analysis of production and allocation: Proceedings of a conference*, John Wiley & Sons, New York.

Koot, R. S. (1979) "On economies of scale in credit unions", *The Journal of Finance*, vol.33, (September), 1087-1094.

Lang, G. and P. Welzel, (1996) "Efficiency and Technical progress in banking. Empirical results for a panel of German Co-operative banks", *Journal of Banking and Finance*, vol.20, 1003-1023.

Lawrence, C. (1989) "Banking costs, generalized functional forms, and estimation of economies of scale and scope", *Journal of Money, Credit and Banking*, vol.21, 368-379.

Le Compte, R. L. B. and S. D. Smith, (1990) "Changes in the cost of intermediation: the case of savings and loans", *The Journal of Finance*, vol.45, no.4, (September), 1337-1346.

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

Lloyd-Williams, D. M. and P. Molyneux, (1994) "Market Structure and performance in Spanish banking", *Journal of Banking and Finance*, vol.18, 433-443.

Lohéac, F. (1991) "Deregulation of financial services and liberalization of international trade in services", *The Geneva Papers on Risk and Insurance*, vol.16, no.16, 406-413.

Longbrake, W. A. and J. A. Haslem, (1975) "Productive Efficiency in commercial banking", *Journal of Money, Credit and Banking*, 317-330.

Lucey, B. M. (1993) "Profits, efficiency and Irish banks", *Report to Statistical and Social Inquiry Society of Ireland*, (November).

Mahajan, A., N. Rangan, and A. Zardkoohi, (1996) "Cost structures in multinational and domestic banking", *Journal of Banking and Finance*, vol.20, 283-306.

Martin, S. (1993) *Advanced Industrial Economics*, Blackwell, Oxford.

McAllister, P. H. and D. McManus, (1993) "Resolving the scale efficient puzzle in banking", *Journal of Banking and Finance*, vol.17, 389-405.

Maclean, D. (1997) "Lagging productivity growth in the service sector: mismeasurement, mismanagement or misinformation?", *Bank of Canada Working paper 97-6*, Bank of Canada, March.

McKillop, D. G. and C. Ferguson, (1994) *Building Societies: Structure Performance and Change*, Graham and Trotman, London.

McKillop, D. G. and C. J. Glass, (1994) "A Cost Model of Building Societies as Producers of Mortgages and Other Financial Products", *Journal of Business, Finance and Accounting*, vol.21, no.7, 1031-1046.

McKillop, D. G., J. C. Glass, and Y. Morikawa, (1994) "The composite cost function and efficiency in Japanese banking", *Working paper A/F 94-3, The Queen's University of Belfast, Accounting & Finance Division*, (May).

McKillop, D. C. Ferguson and D. Nesbitt, (1995) "Paired difference analysis of size economies in UK credit unions", *Applied Economics*, vol. 27, 529-537.

McNulty, J. E. (1982) "Economies of scale: a case study of the Florida savings and loan industry", *Federal Reserve Bank of Atlanta Economic Review*, (November), 22-31.

Meeusen, W. and J. van de Broek, (1977) "Efficiency estimation from Cobb Douglas production functions with composed error", *International Economic Review*, vol.18, no.2, 435-444.

Mester, L. J. (1987) "A multi-product cost study of savings and loans", *The Journal of Finance*, vol.42, 423-445.

Mester, L. J. (1989) "Testing for expense preference behaviour: mutual versus stock savings and loans", *Rand Journal of Economics*, vol.20, no.4, (Winter), 483-498.

Mester, L. J. (1991) "Agency costs among savings and loans", *Journal of Financial Intermediation*, vol.1, no.3, (June), 257-278.

- Mester, L. J. (1993) "Efficiency in the savings and loan industry", *Journal of Banking and Finance*, vol.17, 267-286.
- Mester, L. J. (1996) "A study of bank efficiency taking into account risk-preferences", *Journal of Banking and Finance*, vol.20, 1025-1045.
- Miles, D. K. (1994) "Economic issues in the regulation of mutual financial firms: the case of U.K. building societies", *The Manchester School*, vol.62, no.3, 227-250.
- Miller, S. M. and A. G. Noulas, (1996) "The technical efficiency of large bank production", *Journal of Banking and Finance*, vol.20, 495-509.
- Mitchell, K. and N. M. Onvural, (1996) "Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form", *Journal of Money, Credit and Banking*, vol.28, no.2, (May), 178-199.
- Modigliani, F. and M. H. Miller, (1958) "The cost of capital, corporation finance and the theory of investment", *American Economic Review*, June, 261-297.
- Molyneux, P., D. M. Lloyd-Williams, and J. Thornton, (1994) "Competitive conditions in European banking", *Journal of Banking and Finance*, vol.18, 445-459.

Molyneux, P. and W. Forbes, (1995) "Market structure and performance in European banking", *Applied Economics*, vol.27, 155-159.

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

Muldur, U. and M. Sassenou, (1993) "Economies of scale and scope in French banking and savings institutions", *The Journal of Productivity Analysis*, vol.4, 51-72.

Mullineaux, D. (1978) "Economies of scale and organizational efficiency in banking: A profit function approach", *Journal of Finance*, vol.33, 259-280

Murray, J. D. and R. W. White, (1983) "Economies of scale and economies of scope in multi-product financial institutions: A study of British Columbia credit unions", *Journal of Finance*, vol.38, 887-902.

Nathan, A. and E. H. Neave, (1989) "Competition and contestability in Canada's financial system: empirical results", *Canadian Journal of Economics*, vol.22, no.3, (August), 576-594.

Nathan, A. and E. H. Neave, (1992) "Operating efficiency of Canadian banks", *Journal of Financial Services Research*, vol.6, 265-276.

Nelson, R. A. (1990) "Productivity growth, scale economies and the Schumpeterian hypothesis", *Southern Economic Journal*, vol.57, 521-527.

Nelson, R. W. (1985) "Branching, scale economies and banking costs", *Journal of Banking and Finance*, vol. 9, 177-191.

Newman, J. A. and R. E. Shrieves, (1993) "The multibank holding company effect on cost efficiency in banking", *Journal of Banking and Finance*, vol.17, 709-732.

Nickell, S. J. (1996) "Competition and corporate performance", *Journal of Political Economy*, vol.104, no.4, 724-746.

Nordhaus, W. D. (1997) "Traditional productivity estimates are asleep at the (technological) switch", *The Economic Journal*, vol.107, (September), 1548-1559.

North, D. C. (1990) *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge.

Noulas, A. G., S. C. Ray, and S. N. Miller, (1990) "Returns to scale and input substitution for large U.S. banks", *Journal of Money, Credit and Banking*, vol.22, 94-107.

O'Reilly, J. (1992) "Banking on flexibility: a comparison of the use of flexible employment strategies in the retail banking sector in Britain and France", *The International Journal of Human Resource Management*, vol.3, no.1, (May), 35-58.

Ovey, E. and M. Waters, (1987) *Building Societies Act 1986 with annotations*, Sweet and Maxwell, London.

Panzar, J. C. and R. D. Rosse, (1977) "Economies of scale in multi-product production", *Quarterly Journal of Economics*, vol.91, 481-493.

Panzar, J. C., and R. D. Willig, (1987) "Testing for 'Monopoly' equilibrium", *The Journal of Industrial Economics*, vol.35, no.4, 443-456.

Pavloppolos, P. G. and A. K. Kouelis, (1990) "Cost behaviour in the banking industry: evidence from a Greek commercial bank", *Applied Economics*, vol.22, 94-107.

Perrakis, S. (1991) "Assessing competition in Canada's financial system: a note", *Canadian Journal of Economics*, vol.24, no.3, 727-732.

Pi, L. and S. G. Timme, (1993) "Corporate control and bank efficiency", *Journal of Banking and Finance*, vol.17, 515-530.

Piesse, J. and R. Townsend, (1995) "The measurement of productive efficiency in UK building societies", *Applied Financial Economics*, vol.5, 397-407.

Pulley, L. B. and Y. M. Braunstein, (1992) "A composite cost function for multi-product firms with an application to economies of scope in banking", *The Review of Economics and Statistics*, 221-230.

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

Rossi, P. E. (1985) "Comparison of alternative functional forms in production", *Journal of Econometrics*, vol.30, 345-361.

Ruttan, V. R. (1997) "Induced innovation, evolutionary theory and path dependence: sources of technical change", *The Economic Journal*, vol.107, (September), 1520-1529.

Saunders, A. and C. Ward, (1976) "Regulation, risk and performance of U.K. clearing banks 1965-75", *The Journal of Industrial Economics*, vol.25, no.2, (December), 143-159.

Schimdt, P. and C. A. K. Lovell, (1979) "Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers", *Journal of Econometrics*, vol.9, 342-366.

Sealey, C. W. and J. T. Lindley, (1977) "Inputs, outputs, and a theory of production and cost at depository financial institutions", *The Journal of Finance*, vol.32, no.4, (September), 1252-1253.

Shaffer, S. (1982) "Competition, conduct and demand elasticity", *Economic Letters*, vol.10, 167-171.

Shaffer, S. (1993) "Can megamergers improve bank efficiency?", *Journal of Banking and Finance*, vol.17, 423-436.

Shaffer, S. and D. Edmond, (1991) "Economies of superscale in commercial banking", *Applied Economics*, vol.23, 283-293.

Shephard, R. W. (1970) *Theory of Cost and Production Functions*, Princeton University Press, Princeton.

Siegel, S. and N. J. Castellan, (1988) *Nonparametric Statistics for the Behavioural Sciences*, Second Edition, McGraw-Hill, New York.

Solow, R. (1957) "Technical change and the aggregate production function", *Review of Economics and Statistics*, vol.39, 312-320.

Spiller, P. T. and E. Favaro, (1984) "The Effects of Entry Regulation Oligopolistic Interaction: the Uruguayan banking sector", *Rand Journal of Economics*, vol.15, no.2, 244-254.

Stevenson, R. E. (1980) "Likelihood function for generalized stochastic frontier estimation", *Journal of Econometrics*, vol.13, 57-66.

Stiroh, K. J. (1997) "The evolution of an industry: US thrifts in the 1990s", *Journal of Banking and Finance*, vol.21, 1375-1394.

Streeck, W. (1992) *Social Institutions and Economic Performance. Studies of Industrial Relations in Advanced Capitalist Economies*, Sage Publications, London.

Sudit, E. F. (1995) "Productivity measurement in industrial operations", *European Journal of Operational Research*, vol.85, 435-453.

Suominen, M. (1994) "Measuring Competition in Banking: A Two Product Model", *Scandinavian Journal of Economics*, vol.96, no.1, 95-110.

Tachibanaki, T., K. Mitsui, and H. Kitagawa, (1991) "Economies of scope and shareholding of banks in Japan", *Journal of Japanese and International Economies*, vol.5, 261-281.

Talyor, R. A. (1972) "Economies of scale in large credit unions", *Applied Economics*, vol.4, 33-40.

Teece, D. (1982) "Towards an economic theory of the multi-product firm", *Journal of Economic Behaviour and Organisation*, vol.3, 39-63.

Timmer, C. P. (1971) "Using a probabilistic frontier production function to measure technical efficiency", *Journal of Political Economy*, vol.64, 776-794.

The Financial Services and Markets Bill: A Consultation Document (1998) London Stationery Office.

Varian, H. R. (1985) "Non-parametric analysis of optimizing behaviour with measurement error", *Journal of Econometrics*, vol.30, 445-458.

Vesala, J. (1995) "Testing for competition in banking: behavioural evidence from Finland", *Bank of Finland Studies*, E:1, Bank of Finland, Helsinki.

Vickers, J. (1995) "Concepts of competition", *Oxford Economic Papers*, vol.47, 1-23.

Wales, T. J. (1977) "On the flexibility of flexible functional forms", *Journal of Econometrics*, vol.5, 183-193.

Wheelock, D. C. (1993) "Is the banking industry in decline? Recent trends and future prospects from a historical perspective", *Federal Reserve Bank of St Louis Review*, September/October, 3-22.

Wheelock, D. C. and P. W. Wilson, (1995) "Evaluating the efficiency of commercial banks: Does our view of what banks do matter?", *Federal Reserve Bank of St Louis Review*, July/August, 39-53.

Wheelock, D. C. and P. W. Wilson, (1996) "Technical progress, inefficiency and productivity change in US Banking, 1984-1993", *Federal Reserve Bank of St Louis Research Division Working Papers*, 94-021B, Federal Reserve Bank of St Louis, St Louis.

White, L. J. (1991) *The S & L Debacle. Public policy lessons for bank and thrift regulation*, Oxford University Press, Oxford.

Wilson, J. H. (1981) "A note on scale economies in the saving and loan industry", *Business Economics*, (January), 45-49.

Wolff, E. N. (1991) "Productivity growth, capital intensity, and skill levels in the U.S. Insurance industry, 1948-86", *The Geneva Papers on Risk and Insurance*, vol.16, no.59, 173-190.

Wolff, E. N. (1996) "The productivity slowdown: the culprit at last? Follow up on Hulten and Wolff", *The American Economic Review*, vol.86, no.5, December, 1239-1252.

Wolken, J. D. and F. J. Navratil, (1980) "Economies of scale in credit unions: Further evidence", *The Journal of Finance*, vol.35, no.3, (June), 769-777.

Worthington, A. C. (1998) "Efficiency in Australian building societies: an econometric cost function approach using panel data", *Applied Financial Economics*, vol. 8, 459-467.

Yue, P. (1992) "Data envelopment analysis and commercial bank performance: A primer with applications to Missouri banks", *Federal Reserve Bank of St Louis Review*, vol.74, (January/February), 31-45.

Zardkoohi, A. and J. Kolari. (1994) "Branch office economies of scale and scope: evidence from savings banks in Finland", *Journal of Banking and Finance*, vol.18, 421-432.