

This item is held in Loughborough University's Institutional Repository (<https://dspace.lboro.ac.uk/>) and was harvested from the British Library's EThOS service (<http://www.ethos.bl.uk/>). It is made available under the following Creative Commons Licence conditions.



creative
commons

C O M M O N S D E E D

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **BY:** **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

**The Diffusion of Process Innovation in the UK Financial
Sector: An Empirical Analysis of Automated Teller Machine
(ATM) Diffusion**

BY

Adrian Robert Gourlay

**A Doctoral Thesis
Submitted in partial fulfilment
of the requirements for the award of
Doctor of Philosophy**

**Department of Economics
Loughborough University
April 1999**

© Adrian Robert Gourlay

Abstract

Recent policy initiatives have identified that the diffusion of innovation constitutes an important component in technical change and progress and is the impetus behind changes in firm productivity. To date, however, the main emphasis of economists has been on the diffusion of process innovations in the industrial sector with diffusion in the financial sector either ignored or, at best, summarised by a number of stylised facts relating to the spread of information.

The objective of this thesis is to explore the inter-firm determinants of ATM adoption and diffusion in the UK financial sector and identify firm-specific and market factors in the diffusion process. The empirical analysis draws on duration analysis which represents the current state-of-art modelling approach to inter-firm diffusion. This approach conceptualises inter-firm diffusion as a cross-section of durations of non-adoption from which, most importantly, hypothesised factors (or ‘covariates’) can be examined by their significance or otherwise on the conditional probability of adoption.

The main findings of this thesis support the stylised fact often made in the diffusion literature that the inter-firm diffusion curve is sigmoid and characterised by a non-monotonic hazard function. Furthermore the empirical analysis supports the hypothesis that firm-specific characteristics and expectations have played a crucial role in the inter-firm diffusion of ATMs. In addition, the results indicate that the diffusion of ATMs in the UK has been characterised by the existence of positive network externalities. The results are also shown to be robust across a number of model specifications and assumptions concerning the time-path of covariates.

Acknowledgements

I would like to express my gratitude to Eric J Pentecost for his guidance and support and Thomas Weyman-Jones for advice.

I would also like to acknowledge the financial support from the Department of Economics at Loughborough.

I am also particularly grateful to my colleagues at Loughborough University for their moral support.

Contents

	<i>Page</i>
Chapter 1: Introduction	1.1
Chapter 2: The Economics of Process Innovation Diffusion	2.1
2.1 Introduction	2.1
2.2 Preliminary Definitions	2.3
2.2.1 The Concept of Innovation	2.3
2.2.2 The Dimensions of Diffusion	2.9
2.3 Formal Theoretical Models of Process Innovation Diffusion	2.11
2.3.1 The Early Literature	2.12
2.3.1.1 Epidemic and Information Based Models	2.13
2.3.1.2 Vintage Capital Models	2.20
2.3.2 Contemporary Economic Literature	2.22
2.3.2.1 Probit or Rank Models	2.23
2.3.2.2 Stock Effect Models	2.43
2.3.2.3 Order Effect Models	2.46
2.3.2.4 Network Effect Models	2.49
2.3.2.5 New Investment Theory	2.51
2.3.2.6 Evolutionary Models	2.53
2.4 Concluding Remarks	2.55
Chapter 3: Empirical Diffusion Models: A Literature Review	3.1
3.1 Introduction	3.1
3.2 The Literature	3.1
3.2.1 Early Contributions	3.2
3.2.1.1 Two-Step Models	3.2
3.2.1.2 Time-to-Adoption Models	3.11
3.2.2 Contemporary Economic Literature	3.14
3.2.2.1 Discrete Choice and Count Data Models	3.14
3.2.2.2 Duration Models	3.23

3.3	Concluding Remarks	3.41
Chapter 4: The Market for ATMs		4.1
4.1	Introduction	4.1
4.2	The Nature of ATMs	4.2
4.3	The Adoption and Diffusion of ATMs in the UK	4.8
4.3.1	ATM Reciprocal Networks	4.13
4.4	ATM Diffusion and Technical Change	4.17
4.4.1	Factor Bias and ATM Technology	4.22
4.4.2	Scale and Scope Economies and ATM Technology	4.31
4.4.3	ATMs, Competitive Strategy and Regulatory Change	4.33
4.5	Consequences of ATM Adoption for the Financial Sector	4.37
4.5.1	Barriers to Entry	4.37
4.5.2	Money Demand Stability	4.39
4.6	Concluding Remarks	4.41
A4.1	Appendix One: The Set of Potential Adopters	4.43
A4.2	Appendix Two: ATMs Operated by Financial Institution	4.45
Chapter 5: A Profile of ATM Diffusion in the UK Financial Sector -		
A Duration Analysis		5.1
5.1	Introduction	5.1
5.2	Data and Methodology	5.2
5.2.1	Non-Parametric Estimation	5.9
5.2.2	Parametric Estimation	5.12
5.3	Results	5.16
5.3.1	Non-Parametric Estimates	5.16
5.3.2	Parametric Estimates	5.20
5.4	Goodness-of-Fit Tests for Parametric Models	5.29
5.4.1	A Non-Parametric test of Duration Dependence	5.30
5.4.2	Nested Tests Based on the Likelihood Ratio Test	5.31
5.4.3	Tests based on the Empirical Distribution Function (EDF)	5.32
5.5	Concluding Remarks	5.35
A5.1	Appendix One: The Effects of Sample Attrition	5.37

A5.1.1	Introduction	5.37
A5.1.2	Empirical Analysis	5.38
A5.1.3	Concluding Remarks	5.39
A5.2	Appendix Two: Parametric Duration Models	5.40
A5.3	Appendix Three: Economic Theory and Duration Dependence	5.43
A5.3.1	Introduction	5.43
A5.3.2	The Epidemic Model	5.43
A5.3.3	Rank Effects	5.45
Chapter 6: The Determinants of ATM Adoption: Evidence from the UK Financial Sector		6.1
6.1	Introduction	6.1
6.2	Theoretical and Methodological Considerations	6.2
6.3	A Theoretical Model of ATM Adoption	6.5
6.3.1	An Empirical Representation of the Theoretical Model	6.12
6.4	The Data Set	6.16
6.4.1	Institution-Specific Characteristics and Rank Effects	6.19
6.5	Econometric Specification and Estimation	6.22
6.6	Estimation Results	6.26
6.6.1	Mis-Specification Tests	6.35
6.6.6.1	Unobserved Heterogeneity	6.35
6.6.6.2	Robustness of the Estimates	6.38
6.6.6.3	Collinearity between the Covariates	6.42
6.6.2	Extensions to the Model	6.46
6.7	Concluding Remarks	6.48
A6.1	Appendix One: The Mathematics of Duration Models	6.49
A6.1.1	Introduction	6.49
A6.1.2	The Proportional Hazards Model	6.51
A6.1.3	The Accelerated Lifetime Model	6.53
A6.2	Appendix Two: Descriptive Statistics	6.56
A6.3	Appendix Three: Data Sources	6.59
A6.4	Appendix Four: Estimating the Cumulative Number of Adopters	6.60
A6.5	Appendix Five: A Mis-Specification Test	6.61

Chapter 7: ATM Adoption and Network Externalities: An Empirical Investigation	7.1
7.1 Introduction	7.1
7.2 Network Externalities	7.2
7.2.1 Network Externalities and the Diffusion of New Technology	7.5
7.2.1.1 The Micro Approach	7.7
7.2.1.2 The Evolutionary Approach	7.14
7.3 ATM Technology and Network Externalities	7.16
7.4 A Model of ATM Adoption with Network Externalities	7.20
7.4.1 Empirical Representation of the Theoretical Model	7.31
7.4.1.1 Econometric Implementation	7.34
7.4.1.2 Measurement of ATM Network Externalities	7.39
7.4.1.3 Measurement of Other Covariates	7.41
7.5 The Data Set	7.44
7.6 Estimation Results	7.44
7.6.1 Estimation Results for the Period 1972 to 1986	7.46
7.6.2 Estimation Results for the Period 1972 to 1992	7.55
7.7 Concluding Remarks	7.62
A7.1 Appendix One: The Demand for Technology in the Presence of Network Externalities	7.64
Chapter 8: Conclusion	8.1

References

List of Figures

Chapter 2

Figure 2.1:	Schumpeter's Linear Model of Technical Change	2.5
Figure 2.2	The Epidemic Inter-Firm Diffusion Curve	2.15
Figure 2.3	The Probit Model Mechanism of Diffusion	2.24

Chapter 4

Figure 4.1:	Total ATMs Operated in the UK Financial Sector	4.11
Figure 4.2:	Diffusion of ATMs in the UK Financial Sector	4.11
Figure 4.3:	Ratio of Branch Staff to Branches in the UK	4.30
Figure 4.4:	Ratio of ATMs to Branches in the UK	4.30

Chapter 5

Figure 5.1:	Possible Life Histories of Financial Institutions	5.7
Figure 5.2:	Estimated Non-parametric Hazard Function - 1972 to 1992	5.20
Figure 5.3:	Estimated Non-parametric Survivor Function - 1972 to 1992	5.20
Figure 5.4	Estimated Non-parametric Integrated Hazard Function - 1972 to 1992	5.21
Figure 5.5	Estimated Exponential and Weibull Hazard Functions - 1972 to 1992	5.26
Figure 5.6	Estimated Lognormal and Log-logistic Hazard Functions - 1972 to 1992	5.26
Figure 5.7	Estimated Exponential and Weibull Survivor Functions - 1972 to 1992	5.27
Figure 5.8	Estimated Lognormal and Log-logistic Survivor Functions - 1972 to 1992	5.27
Figure 5.9	Estimated Exponential and Weibull Integrated Hazard Functions - 1972 to 1992	5.28
Figure 5.10	Estimated Lognormal and Log-logistic Integrated hazard Functions - 1972 to 1992	5.28

Chapter 7

Figure 7.1	A Simple Star Network	7.4
Figure 7.2	A Sequential Adoption Game	7.10

Figure 7.3	The ATM 'Forming System'	7.17
Figure 7.4	A Simplified ATM 'One-Way' Network	7.19
Figure 7.5	Separating Network Externalities and Cost-Side Effects	7.30
Figure A7.1.1	The Fulfilled Expectations Demand Curve	7.66
Conclusion		
Figure 8.1	UK Total Employment in Service Industries	8.4
Figure 8.2	UK Total Employment in Banking, Finance and Insurance	8.4

List of Tables

Chapter 4

Table 4.1:	The Diffusion of ATMs in the UK - end 1972 to end 1996	4.9
Table 4.2:	Location of ATMs in the UK	4.12
Table 4.3:	Reciprocal ATM Network Sizes in the UK – end 1996	4.14
Table 4.4	ATM Withdrawals in the UK	4.33
Table 4.5	Payment of Wages and Salaries in the UK	4.41
Table A4.1.1	Retail Banks - as at end of 1992	4.44
Table A4.1.2	BSA Members - as at end of 1992	4.44
Table A4.2.1	Total Number of ATMs Operated by Retail Banks - end 1972 to end 1992	4.46
Table A4.2.2	Total Number of ATMs Operated by Building Societies - end 1972 to end 1992	4.47

Chapter 5

Table 5.1:	Non-Parametric Hazard, Integrated Hazard and Survivor Functions Estimates - 1972 to 1992	5.19
Table 5.2:	Maximum Likelihood Estimates of Parametric Models - 1972 to 1992	5.22
Table 5.3:	Summary of Kolmogrov EDF Statistics for Estimated Parametric Models	5.35
Table A5.1.1	Relative Frequency of Exit Times Conditional on Adoption Times - 1992 to 1996	5.38
Table A5.2.1	Summary of Parametric Duration Models	5.42
Table A5.3.2	Non-Linear Least Squares Estimation of the Diffusion Curve - 1972 to 1992	5.45

Chapter 6

Table 6.1:	Definitions of Explanatory Variables	6.25
Table 6.2:	Expected Sign of Explanatory Variables	6.26
Table 6.3:	Summary of Models Estimated	6.28
Table 6.4	Maximum Likelihood Estimates of the Fully Specified Lognormal and Log-logistic models	6.32

Table 6.5	Maximum Likelihood Estimates of Restricted Stock Lognormal and Log-logistic models	6.33
Table 6.6	Maximum Likelihood Estimates of Restricted Myopic Lognormal and Log-logistic models	6.34
Table 6.7	Summary of Generalised Error Specification Tests for Estimated Accelerated Lifetime Models	6.37
Table 6.8	Maximum Likelihood Estimates of the Fully Specified Proportional Hazards Model	6.41
Table 6.9	Maximum Likelihood Estimates of the Restricted Stock and Myopic Proportional Hazards Model	6.42
Table 6.10	Maximum Likelihood Estimates of Restricted Size Lognormal and Log-logistic Models	6.45
Table A6.2.1	Descriptive Statistics of the Explanatory Variables	6.57
Table A6.2.2	Measures of Dispersion of the Explanatory Variables	6.58
Table A6.4.1	Autoregressive Estimates for K_{t-1}	6.59
Chapter 7		
Table 7.1	Definition and Expected Sign of Covariates	7.43
Table 7.2	Maximum Likelihood Estimates of the Restricted Branch Weibull and Log-logistic Models – end 1972 to end 1986	7.53
Table 7.3	Maximum Likelihood Estimates of the Fully Specified Weibull and Log-logistic Models – end 1972 to end 1986	7.54
Table 7.4	Maximum Likelihood Estimates of the Restricted and Fully Specified Exponential Model – end 1972 to end 1986	7.55
Table 7.5	Maximum Likelihood Estimates of the Restricted Branch Weibull and Log-logistic Models – end 1972 to end 1992	7.60
Table 7.6	Maximum Likelihood Estimates of the Fully Specified Weibull and Log-logistic Models – end 1972 to end 1992	7.61
Table 7.7	Summary of the Network Effect on the One-Period Conditional Probability of Adoption	7.62

List of Acronyms

APACS	Association for Payment Clearing Services
ATM(s):	Automated Teller Machine(s)
BBA:	British Bankers' Association
BSA:	Building Societies Association
CDF:	Cumulative Density Function
CN:	Computer Numeric
CNC:	Computer Numerically Controlled
DTI	Department of Trade and Industry
EDF:	Empirical Distribution Function
EFT:	Electronic Funds Transfer
HM Treasury	Her Majesty's Treasury
IP	Intellectual Property
OECD:	Organisation for Economic Co-operation and Development
OST	Office of Science and Technology
PDF:	Probability Density Function
SIC	Standard Industrial Classification
UK:	United Kingdom
US:	United States
USA:	United States of America

CHAPTER 1

INTRODUCTION

It is only through the wider use (diffusion) of new processes and products that their effects and benefits will be realised. Thus, the diffusion of innovation constitutes a significant stage for both economic development and growth and can be viewed as the last phase in Schumpeter's (1934) trilogy of technical change, that is: invention, innovation and diffusion. It can be argued that in order to understand the process of how new technology generates, for example, higher economic growth, social welfare and firm productivity that it is necessary to understand the process of innovation diffusion [Karshenas and Stoneman (1995)].

The importance of innovation diffusion has recently been expressed in policy orientated documents such as the OECD (1992), Papaconstantinou (1996), the UK Government's White Paper 'Realising Our Potential' [Cabinet Office (1993)] and Department of Trade and Industry (1998), and at a European level by the European Commission (1995). Papaconstantinou (1996) has emphasised the importance of research and development (R&D) spillovers as a mechanism of technology diffusion between countries, particularly for the diffusion of information technology. The OECD has emphasised the need for policy-makers to direct industrial policy towards the dissemination of information ('research spillovers') pertaining to new processes, designs and technologies and to improve the capacity of firms to absorb new technologies. The Cabinet Office (1993), European Commission (1995) and Department of Trade and Industry (1998) have all recognised, *inter alia*, that technological diffusion can be facilitated most effectively through the transfer of people (as a means of transferring 'Know-How') particularly between the science base and industry. As noted by Greenaway (1994) and Stoneman and Diederer (1994), these policy initiatives reflect a gradual reorientation in the focus of policy programmes in OECD countries away from the invention and innovation stages of technological change towards a definite and formal diffusion policy.

In addition to these shifts in industrial policy there has also been a growing recognition in the economics literature in the last fifteen years of the importance of innovation diffusion vis-à-vis the more observable aspects of technical change such as R&D [Greenaway (1994) and Karshenas and Stoneman (1995)]. Early contributions from economists [see, for example, Griliches (1957), Bain (1964) and Bass (1969)] were largely influenced by previous non-economic research traditions such as sociology and anthropology, exemplified by the work of Rogers (1962), where the study of innovation (within defined societies and social groups as opposed to economies) had already been established. Lissoni and Metcalfe (1994) have argued that the inheritance of an already existing framework of analysis, designed for disciplines other than economics, severely constrained the scope of early research (essentially pre-1980) in the area of innovation diffusion by economists in two ways.

First, early contributions inevitably inherited a framework of analysis, which was not designed for an economic analysis of innovation diffusion. Consequently, there tended to be heavy reliance on so-called 'epidemic' explanations of diffusion in which the diffusion of an innovation was made analogous to the spread of a disease within a defined (usually constant) population (perceived as potential adopters). Adoption of innovation was then conceived as taking place via the receipt of a piece of information (either pertaining to the existence of the innovation or a piece of tacit information reflecting the attributes of the innovation) and diffusion would thus proceed by a process of imitation as the information spread through 'contact' (word-of-mouth). Hence, there was a distinct lack of any rigorous economic analysis and, in particular, a deficiency of any formal choice-theoretic framework [Stoneman (1983, 1986)].

Second, early contributions [see, for example, Romeo (1975) and Davies (1979)] had concerned themselves mainly with confirming certain empirical regularities or so-called 'stylised facts'. In particular is the empirical observation that plotting usage or ownership against time yields an S-shaped (or sigmoid) curve. Moreover, it was found in these early contributions that the shape (and hence the speed of diffusion) of the diffusion curve differed across innovations, industries, regions and economies. These differences were often explained by a few key variables, such as adoption profitability and average firm size. Such studies have subsequently been criticised on the grounds

that they are characterised by an absence of any theoretical justification for the inclusion of certain explanatory variables and in relation to their reliance on the epidemic model as an explanation of adoption [Stoneman (1983) and Karshenas and Stoneman (1995)].

Arguably it is only within the last fifteen years that economists have really started to analyse the diffusion process within a definite choice-theoretic framework, rather than one dominated by imitation contained in early approaches [Lissoni and Metcalfe (1994)]. Moreover, the contemporary theoretical literature and contemporary empirical approaches to innovation diffusion have been increasingly concerned with such issues as the compatibility, interrelatedness, co-development of technological diffusion and the related issue of network externalities [see, for example, Ayres and Ezechoyer (1991), Stoneman and Kwon (1994), Colombo and Mosconi (1995) and Saloner and Shepard (1995)].

Despite all these developments there has been relatively little attention paid in the economics literature to the determinants of innovation diffusion in the financial sector and those that do exist tend to examine exclusively the US financial sector [see, for example, Hannan and McDowell (1984a, 1984b, 1987), Saloner and Shepard (1995) and Molyneux and Shamroukh (1996)]. This is in stark contrast to the more rigorous (and growing) theoretical and empirical work examining innovation diffusion in the industrial sector. The latter is often perceived as having greater impact on economic development and growth and therefore of more significance from a policy standpoint. Although this point is not being directly addressed in this thesis, there is the contention that symmetry exists - that is, to fully understand the implications of innovations in the financial sector an understanding of the diffusion process is essential. As Podolski argues:

Just as the mechanics of the diffusion of technical innovation is potentially relevant to the study of industrial policy, the mechanics of the diffusion of financial innovation may be relevant to the study of financial and monetary policies. Yet few systematic enquiries into this aspect have been conducted. Only fragmentary illustrations can be offered. [Podolski (1986), p. 110].

This argument is even more pertinent given that financial sectors in developed economies, including the UK, have experienced rapid innovation and change in the last

two decades, characterised by the development of new financial instruments and techniques by financial institutions [Bank of England (1983, 1986), Kirkman (1987), Pawley (1993), Llewellyn (1992, 1997)]. Innovation in the UK financial sector can be seen as a three-fold phenomenon, although the three aspects are not mutually exclusive [Adam (1986), Spencer (1986) and Llewellyn (1992)]. The first and second aspects of innovation in the financial sector concern the 'legislative' and the 'responsive' aspects. These relate to the development of parallel markets (the Eurocurrency and interbank market for example) and new financial instruments (the introduction of high interest current accounts by banks and building societies for example). Both are concomitant to changes in legislation. They can be viewed as either the direct responses of profit maximising financial intermediaries to changes in legislation or as changes in behaviour in response to the removal of these legislative constraints.¹ The third aspect of innovation concerns technological innovations and is the focus of this thesis.

The payments system of the UK financial sector has presented an abundance of opportunities for the wider use of technology - information technology in particular - due to its main functions being centred around the availability, storage, retrieval and transmission of information. These four functions tend to be extremely labour intensive and involve large transfers of paper [Akhtar (1983), Podolski (1986) and Kirkman (1987)]. Given this and the effects of legislative changes during the 1980s affecting the provision of financial services to the non-bank private sector (NBPS)², retail banks in particular rapidly adopted available information technology. Institutions adopting this technology had as their twin aims: reducing costs of intermediation for both themselves (in terms of bank staff, paper and possibly branches) and for customers (through greater convenience); and increasing their market share for NBPS deposits [Kirkman (1987), Pawley (1993) and Vesala (1994)]. One such technology that met the requirements of these twin aims is the automated teller machine (ATM) which has been widely adopted by retail banks and building societies since their commercialisation in 1972. An ATM is essentially an automated cash dispenser (with additional services) which allows approved holders of an appropriate cash card (more commonly a debit or credit card) to access their bank accounts in order to withdraw and deposit cash [Kirkman (1987)].

¹ This view of innovation causation invokes the hypotheses of 'constraint-induced innovation' developed by Silber (1975, 1983) and 'circumventive' innovation discussed by Kane (1981).

² Legislative changes have included, for example, the Building Society Act (1986).

The ATM can be classified as both a process and product innovation. Firstly, it has a process innovation interpretation because it has provided a new and novel means of producing traditional demand deposit services. Secondly, it has a product innovation interpretation because although it has not provided any additional services that financial institutions have provided in the past (although institutions can theoretically now provide a subset of services 24 hours a day) it has, arguably, acted as an additional characteristic in an institutions spectrum of services [Vesala (1994)].

The response of economists to these developments has been threefold. Firstly, there has been a focus on the consequences for the structure and performance of the financial sector [see Goodhart (1986) and references therein]. This research agenda has included such related issues as whether new technology is making financial markets more contestable and providing opportunities for economies of scale [Revell (1986), Drake (1989) and Ferguson and McKillop (1994)]. Secondly, there is the question of whether innovations in the financial sector are increasing its efficiency [see Akçaoglu (1996) and references therein]. Thirdly, and more widely discussed, there has been concern for the implications for the conduct and performance of monetary policy and the related issue of money demand stability resulting from the perceived reduction in transactions costs and rise in 'money substitutes' [Judd and Scadding (1982), Goodhart (1986), Adam (1987) and Hall *et al* (1989)]. Despite these research agendas, there has been no formal examination of the diffusion of new technology in the UK financial sector. As Podolski states:

... accounts of the diffusion of modern technology in the financial industry, of its determinants, processes and influence on financial innovations tend to be impressionistic rather than systematic [Podolski (1986) p. 168].

There are arguably two reasons for this. Firstly, there may be a perception amongst policy-makers that innovations in the industrial sector are more important than innovations in the service industries from the perspective of economic welfare. This perception may be misleading given that in mature economies the production of services and information is becoming a significant part of their output [Gourlay *et al* (1997a) and Keely and Quah (1998)]. Secondly, empirical studies of diffusion have suffered from a paucity of appropriate data. To empirically investigate diffusion at the

inter-firm level a panel data set is a necessary requirement. This requirement implies a set of individual adoption histories from the time of commercialisation of the new technology together with measurement of individual adopter characteristics. The costs of collecting such data are demanding, both in pecuniary terms and time. A major contribution of this thesis is the compiling of such a data set for ATM diffusion in the UK.

In addition to the paucity of empirical work examining the diffusion of new technology in the UK financial sector an additional three weaknesses in the diffusion literature may be identified. First, there is a distinct shortage of models that explicitly examine the role that prices and price expectations may have on the diffusion process. Although the seminal work of Rosenberg (1976a, 1982) and Ireland and Stoneman (1986) have indicated the possibility that price expectations formed on the price of new technology may delay adoption, and consequently slow diffusion, the empirical literature has largely been bereft of such considerations. Thus, there is a contention in these arguments that previous studies, by assuming myopic expectations, are mis-specified. A second weakness of the empirical literature concerns the lack of rigour and frequent *ad hoc* approaches that have been employed to model selection and comparison. As noted by Karshenas and Stoneman (1995), this deficiency provides a major research opportunity. Third, although network effects have been shown to have important implications for a variety of activities of importance to firm strategy, including technology adoption [see Katz and Shapiro (1985, 1986)], there has been relatively few attempts to test econometrically for these effects [a notably exception being a recent paper by Saloner and Shepard (1995)].

The aim of this thesis is, therefore, to extend the empirical modelling of the analysis of innovation diffusion. In general, the objective of this thesis is to identify the determinants of ATM adoption at the inter-firm level and explore the role of price expectations and network externalities on ATM adoption. Furthermore, the empirical analysis is explicitly set within a duration model framework [Kiefer (1988), Lancaster (1990) and Neumann (1997)] in order to acknowledge the role of time and time-varying characteristics in the diffusion process. Thus, duration models are an integral component in modelling inter-firm diffusion in this thesis with the main focus of

attention being on those firm-specific and market-specific factors affecting the conditional probability of adoption.

There are perhaps four main contributions to the diffusion literature made by this thesis. First, the thesis compiles the first ever extensive panel data set of annual adoption histories for a set of UK potential ATM adopters (measured as a stock of retail banks and building societies at the end of 1992) from the date of ATM commercialisation in 1972 to the end of 1992. This involved extensive fieldwork during 1993 and 1994 in which retail banks and building societies were contacted in order to ascertain individual adoption dates (or not in the case of non-adopters), number of units subsequently adopted for each proceeding year and various measures of institution-specific characteristics such as size and growth. The final data set contains ninety-eight institutions in total, of which thirty-five had adopted ATMs by the end of 1992. It is found that the resulting inter-firm diffusion curve is sigmoid in shape, confirming the stylised fact often made in the diffusion literature.

Second, the thesis modifies previous empirical ATM studies [Hannan and McDowell (1984a, 1984b, 1987)] which have ignored the role of technology price and expectations formed on this price in the factors determining the path of diffusion. The empirical contribution makes the distinction between the *profitability* (adoption yields positive profits) and *arbitrage* (the net benefit from adopting is not increasing over time) criteria for determining the optimal date of adoption analogous to that made by Karshenas and Stoneman (1993). The quality-adjusted, or 'hedonic', price of ATM technology provided by the Office of National Statistics (ONS) is included in regressions examining the institution-specific determinants (rank effects) of ATM adoption. It is reassuring that in this thesis price expectations formed on the price of ATMs is found to have a positive and significant effect on the conditional probability of adoption. In general, the importance of rank effects is confirmed as important determinants of innovation adoption. A positive and significant role is found for institution deposit growth, size, and a measure of technological opportunities facing institutions. In addition, institutional factors are found to have played a significant role in the diffusion process. Moreover, given the inter-firm nature of the study a stronger test of the

relationship between firm-heterogeneity and the adoption of new technology than has previously been conducted has been possible.

A third contribution is that the existence of network externalities in ATM adoption decisions are explored, thus bridging the gap between theoretical advances in the study of network technology [see Economides (1996) and references therein] and empirical verification of these advances. It is argued that the network externality present in ATM technology arises from the increased number of locations from which individual customers can access their accounts. Benefits to both the deposit customer and financial institution rise when the number of locations increases. This positive network externality ultimately arises from the complementarity between ATM hardware and the debit or credit card software. This compatibility produces a two-way ATM network [Economides (1996)]. In the empirical contribution, a distinction is made between the pure network effect and the scale economies effect, the latter arising from the cost-side effects of an increase in the number of depositors. The former effect is approximated by the number of branches operated by the institution, while the number of depositors approximates the latter effect. The empirical results indicate that the diffusion of ATMs in the UK has been characterised by the existence of positive network externalities.

The thesis also addresses the research agenda put forward by Karshenas and Stoneman (1995) who argue that a major task of empirical diffusion research should be aimed at testing a variety of empirical models to any given data set. This agenda is met by examining the robustness of all empirical results by comparing results obtained from estimating non-parametric proportional hazard [Cox (1972)] and parametric accelerated lifetime models [Kalbfleisch and Prentice (1980) and Neumann (1997)]. A number of formal specification tests are employed in order to test the robustness of the underlying assumptions made by the various empirical models. In general, empirical results are found to be consistent and robust across different model formulations. This approach improves on the previous empirical literature that is characterised by either an absence of formal testing or testing of an *ad hoc* nature.

The thesis is structured as follows. Chapter 2 considers the dimensions of innovation diffusion and the theoretical basis for inter-firm diffusion. Chapter 3 provides a review

and critique of the main approaches that have been employed in the literature to model inter-firm diffusion and presents the main results of past research, with a particular emphasis on ATM studies. The technical development and technical attributes (in particular factor bias and scale and scope economies) of ATM technology are discussed in Chapter 4 together with an exploration of the consequences of greater ATM adoption for the financial sector. In addition, Chapter 4 presents the set of UK potential ATM adopters used throughout this thesis. The main focus of Chapter 5 is on the estimation of non-parametric and parametric estimates of the survivor, hazard and integrated hazard functions for the set of ATM adopters assuming a homogenous population. The consequences for economic theory of the subsequent results are also explored. Chapter 6 relaxes the assumption of a homogenous population of ATM adopters and examines the role of institution-specific characteristics ('covariates'), price and price expectations in the adoption and diffusion of ATMs. In addition, the results from a set of rival empirical models are compared for robustness for both time-invariant and time-varying covariates. Chapter 7 examines the significance or otherwise of positive network effects in the diffusion of ATMs. Finally, Chapter 8 offers some concluding remarks relating to the empirical analysis of ATM diffusion and highlights policy implications and avenues for future research.

To summarise, this thesis presents a number of original empirical results relating to the diffusion of ATMs in the UK financial sector. The application of a methodology utilising duration analysis serves to enhance the empirical analysis vis-à-vis past methodologies. Moreover, the estimation of competing models addresses an agenda that has, to date, tended to be overlooked by economists. Furthermore, the empirical analysis yields results which are broadly in line with those from other studies as well as being capable of interpretation within the richer research environment.

CHAPTER 2

THE ECONOMICS OF PROCESS INNOVATION DIFFUSION

2.1 Introduction

A distinguishing feature of the literature on the economics of technical change in the last two decades has been the gradual recognition amongst economists (and policy makers) of the importance of innovation diffusion in the wider process of economic growth and change. These relatively contemporaneous changes are somewhat surprising given that the seminal contributions on the linkages between innovation and economic development made by Schumpeter (1934, 1939) are conventionally perceived as differentiating between three distinct and sequential stages in the process of technical change at the economy-wide level, that is: invention, innovation and diffusion [Nelson (1996)]. Moreover, this so-called 'linear model' of technical change [OECD (1992)] initially dominated the research agenda amongst economists and became an integral component in the formulation of science and technology policy in developed countries during the post-war period [Stoneman (1987), Metcalfe (1993, 1995), Gourlay *et al* (1997a) and Gourlay (1998a)].

Proceeding Schumpeter's work, initial attention by economists centred on two aspects of the invention and innovation stages of technical change: the sources of inventive activity and the economic determinants of innovative activity [Thirtle and Ruttan (1987)]. Although this is certainly a continuing (and ever growing) theme within the literature [see Bosworth *et al* (1996) and Keely (1996)] it was not until the early 1970's that economists started to more formally examine the last stage of the Schumpeter trilogy of technical change. This is not to imply that economists were unaware of the importance of innovation diffusion, but rather that they had not considered it a stage that required a formal modelling approach. Such concerns were initially addressed by the early contributions of Griliches (1957) and Bain (1964). These contributions tended,

however, to be purely empirical based and were arguably characterised by a paucity of formal economic analysis. It was not until the seminal work of Mansfield (1968, 1969) that the economics of innovation diffusion with its own distinctive theories and empirical contributions came to be recognised within the economics profession. Subsequently, the economics of innovation diffusion has now become an established and prominent element in the economics of technical change [see, for example, David (1991), Lissoni and Metcalfe (1994) and Karshenas and Stoneman (1995)].

There are two aims to this chapter. The first is to consider some of the definitional aspects that have arisen in the study of innovation diffusion. The second, and most important, is to trace the development of formal economic theories of innovation diffusion from the early contributions to the contemporary literature. Throughout this chapter the focus is on inter-firm diffusion models of process innovation diffusion (i.e. the diffusion of new technology embodied in capital goods within a defined industry) as this is consistent with the focus of the thesis. Moreover, the main emphasis will be on the mainstream economics literature largely to the exclusion of the marketing and geography literature.¹

The rest of the chapter is set out as follows. Section 2.2 presents some of the definitional aspects involved in the economics of innovation diffusion and examines them within the wider context of the economics of technical change. In particular, distinctions will be made between product and process innovations, disembodied and embodied technology and between intra-firm, inter-firm and economy-wide diffusion. Section 2.3 reviews the main theoretical literature and distinguishes between early contributions and the contemporary literature and emphasises, where appropriate, possible linkages with diffusion of ATM technology. Concluding remarks are collected in Section 2.4.

¹ For a discussion of the contribution of geography to the study of diffusion see Lissoni and Metcalfe (1994) and for a more extensive one on the marketing contribution see Meade (1984), Mahajan and Wind (1986), Mahajan *et al* (1990), Cooke and Mayes (1998) and Satchell (1998). The former discipline focuses on the spatial nature of diffusion, whilst the latter tends to focus primarily on the diffusion of new consumer products and the role of communication channels, such as advertising, in transmitting information pertaining to the innovation.

2.2 Preliminary Definitions

The diffusion of innovation is the process by which innovations (be they new products, new processes or new management methods) spread within and across economies [Stoneman (1986) and Karshenas and Stoneman (1995)]. As in many other research sub-fields in economics, a number of well-established definitions have emerged from the literature on the diffusion of innovation. Knowledge of these definitions is an essential pre-requisite for fully understanding the theoretical and empirical literature. Thus, this section outlines and reviews those definitions that will be used extensively in the thesis.

2.2.1 The Concept of Innovation

There exists an immense literature that has examined the broad concept of innovation and which has encompassed many research elements within economics [see, for example, Akçaoğlu (1996) and references therein]. In the context of the diffusion literature, with its emphasis on diffusion within the industrial sector, it is possible, however, to identify Schumpeter's (1934, 1939) contributions on technical change and economic development as an initial starting point in the discussion of innovation. Indeed, it is convention in the literature [David (1991) and Lissoni and Metcalfe (1994)] to conceptualise the diffusion of innovation as the last stage in Schumpeter's (1939) trilogy of technical change.²

Schumpeter's theory of technical change developed from two distinct, although certainly interrelated, strands in his work. The first, contained in Schumpeter (1934), was concerned with identifying those endogenous (or 'internal') factors underlying economic development. The second, subsumed in Schumpeter (1939), concerned those

² The term 'technical change' is used here in its broadest sense to mean the process by which economies change overtime in respect to the products they produce and the processes used to produce them [Stoneman (1983, 1987)]. This is consistent with Schumpeter's (1934) concept of economic development that argues that such changes are endogenous to the economic system. A more succinct concept could embrace, for example, a total factor productivity interpretation [see Gourlay *et al* (1997b)].

factors underlying the existence of business cycles. In both works Schumpeter identifies and evaluates three main endogenous factors: changes in consumer tastes, growth in productive resources (population, savings and accumulation) and the creation of 'new combinations' (i.e. innovation). The first was dismissed as an insignificant source of endogenous change and, instead, it was argued that consumer tastes were generated largely by the initiatives of producers. The second was argued to be essentially an outcome of the third and that the economy adjusted quickly and smoothly to such changes. Schumpeter, however, identified the third, as being the main cause of economic development and *the* quintessential feature of the capitalist economy.³

In both strands of his work, Schumpeter identifies four broad cases pertaining to the concept of innovation. These are: the introduction of new products ('product innovation'); the introduction of a new method of production to produce existing products ('process innovation'); the conquest of a new source of supply of raw materials and the carrying out of the new organisation of any industry.⁴ These four cases are subsumed under the general set-up of a new production function. The significant aspect of Schumpeter's theory of technical change is that the invention, innovation and diffusion stages of technical change are separate and distinct. The essential elements of his model are summarised in the schema contained in Figure 2.1 below.

Referring to the schema contained in Figure 2.1, technical change is conceptualised by Schumpeter as being a sequential, unidirectional and time-intensive process with invention being antecedent to innovation and innovation being antecedent to diffusion [Stoneman (1987)]. Moreover, invention is not synonymous with innovation. The first stage, invention, represents the generation of new ideas. The second stage represents the development of new ideas into marketable products and processes and the third represents the diffusion of these products or processes in and across economies.

³ In his later work, Schumpeter (1942) refers to the changes brought about by innovation and the response to them as 'economic evolution' and argues that innovation is the main tool of competition in capitalist economies.

⁴ According to Schumpeter this includes, for example, the breaking-up of a monopoly position or establishment of a monopoly position.

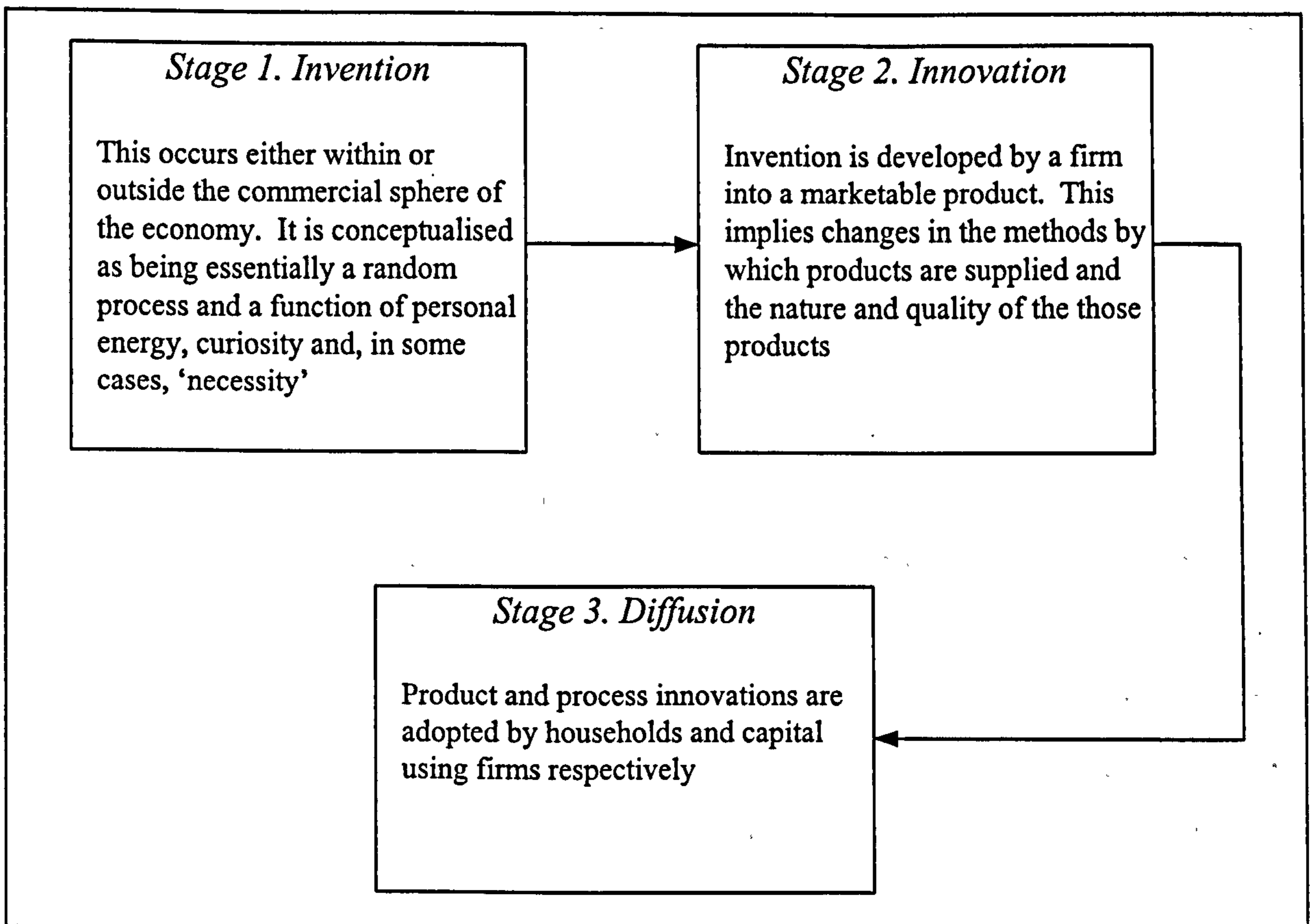


Figure 2.1. Schumpeter's Linear Model of Technical Change

An important aspect of this model of technical change is that the origins of innovation are not the 'entrepreneur'. The role of the 'entrepreneurial function' is, rather, to turn the invention into an innovation. Arguably more significant is that the distinction between invention and innovation leads Schumpeter to conclude that invention *per se* has no real economic relevance and that innovation does not necessarily require the invention stage. As Schumpeter (1938) argues:

Innovation is possible without anything we should identify as invention, but produces of itself no economically relevant effect at all. The economic phenomena which we do observe in the special case in which innovation and invention coincide do not differ from those we observe in cases in which pre-existing knowledge is made use of. [Schumpeter (1939), p. 59].

As noted by Nelson (1996), however, in his later work Schumpeter (1942) puts less stress on the sharp distinction between the invention and innovation stages contained in

Figure 2.1 and argues that the venue of invention and innovation is relatively large firms with R&D facilities.⁵

The approach to technical change outlined by Schumpeter is certainly has its critics and, arguably, there are two main themes to these critiques. The first centres on the invention stage and raises two objections to Schumpeter's view that invention is a random process initiated outside the domain of the entrepreneurial function. The first objection is contained in the historical and sociological perspectives of Usher (1954) who argues that inventions emerge or originate from the cumulative synthesis of relatively simple inventions, each of that requires an individual 'act of insight'.⁶ Thus, for Usher a major invention represents the cumulative synthesis of many individual inventions. This process is broad enough to encompass both the invention and innovation stages [Thirtle and Ruttan (1987)]. These themes have been carried forward in the contemporary marketing literature which has split the invention and innovation stages into many sub-stages based mainly on case study evidence [see, for example, recent contributions by von Hippel (1988) and Cooke and Mayes (1998)].

The second objection to Schumpeter's conception of invention is contained in the mainstream economics literature and centres on whether there exist economic incentives in the production of inventions. This objection has arisen with the paralleled growth of institutionalised research and development (R&D) in both the private and public sectors and empirical evidence that suggests that an increasing proportion of inventions are derived from formal R&D units [see Freeman (1974), Cohen (1995), Mowery (1995) and references therein]. It is the seminal work of Arrow (1962) that initially investigated the possible incentives in the production of inventions. Arrow gives a broad definition to the concept of invention as the production of (new) knowledge. Arrow then considers the incentives to invest in a cost reducing invention (embodied in

⁵ This re-formulation of his initial model lead to a growth in literature empirically testing the so-called 'Schumpeterian hypothesis' concerning the links between firm size and innovative activity [see, for example, Kamien and Schwartz (1982) and Thirtle and Ruttan (1987)]. Stoneman (1983) and Nelson (1996) have argued that this literature has misinterpreted Schumpeter's original work.

⁶ The work of Usher arguably forms a borderline between economists view of innovation, which centres on innovation as an *objective* economic phenomena, and that of sociologists who focus on innovation as a

a capital good) that could be monopolised by the holding of a patent by the inventor under three differing product market conditions⁷, these being: competitive, monopolistic and socially managed. Arrow then assumes those firms in the competitive and monopolistic market conditions are profit-maximisers and that the socially managed firm maximises social welfare. These assumptions allow Arrow to show that the incentives to invent are greatest in the case of a competitive market vis-à-vis the monopolistic, but that both yield less social benefits than the socially managed economy. The work of Arrow stimulated a set of empirical studies that have examined the possible relationship between innovative output (usually proxied by patent counts), inputs (R&D) and demand and supply side influences [see, for example, Scherer (1965), Schmookler (1966), Bosworth and Wilson (1980), Wyatt (1986), Hausmann, Hall and Griliches (1984), Ernst (1997) and Hall (1998)]. The general conclusions from these studies gives tentative support to the role of past R&D inputs and demand and supply side influences, although the studies are frequently derived from reduced form models, are beset with inadequate proxies for inventive activity and are largely industry-specific [Cohen (1995) and Keely (1996)].

The second main objection to Schumpeter's model of technical change focuses on the sequential and unidirectional conception of technical change embodied in Figure 2.1. This conceptualisation has been most rigorously criticised by, *inter alia*, Rosenberg (1982), OECD (1992), Metcalfe (1993, 1995), Rosegger (1996) and Nelson (1996). The alternative approach put forward by these critics, in particular by Rosenberg (1982) is that innovation involves an implicit learning-by-using mechanism that constitutes a 'feedback loop' between innovation and invention.⁸ This learning-by-using mechanism

social process and an individuals *perception* of what is new [see, for example, the work of Rogers (1962)].

⁷ It is important to note that Arrow distinguishes between the market for knowledge and the market for the (tangible) product resulting from inventive activity. For Arrow the production of knowledge is characterised by market failure and sub-optimal investment in inventive activity under competitive market conditions. These arise from the inherent economic characteristics of knowledge, these being indivisibility, inappropriability (due to its public good characteristics of non-rivalry and non-excludability) and uncertainties in outcomes. Geroski (1995) has argued that the quintessential problem of inventive activity surrounds the issue of appropriability deriving from the public good characteristics of knowledge. He argues, however, that Arrow's view of invention is too simplistic and that foreknowledge and first mover advantages lessens the public good consequently.

⁸ As noted by Gourlay *et al* (1997a) there are many such feedback loops (such as the movement of people) especially when a formal research sector (such as universities) is included in the analysis.

is generated as a result of subsequent use of the new technology and is, thus, a different concept to learning-by-doing invoked by Arrow (1962) which is internal to the production process [Rosenberg (1982)]. Contemporary policy-makers have openly embraced these so-called 'feedback' models of technical change as a basis of science and technology policy [see Mowery (1995), Gourlay *et al* (1997a), Gourlay (1998c) and Department of Trade and Industry (1998)].

Despite the challenges and extensions to Schumpeter's model of technical change, the contemporary theoretical and empirical diffusion has, arguably, kept many of its essential elements (the exception are evolutionary theories of diffusion - see Section 2.3.2.6). Firstly, the distinction between product and process innovation remains. According to Stoneman (1986, 1987) and Karshenas and Stoneman (1995) these are distinguishable by which agents are the ultimate buyers of the innovation. In respect to new product technologies the ultimate buyers are assumed to be capital-using firms (public or private) who make decisions on whether or not to install the technology. In respect to new products, the potential buyers are assumed to be households. Moreover, it is convention in the literature [OECD (1992) and Karshenas and Stoneman (1995)] to distinguish between embodied and disembodied technology. The former is embodied in capital vintages and diffuses predominately through market channels. The latter has two interpretations. The first involves shifts in firms' isoquant that are independent of factor proportions and is an integral component of neo-classical theories of economic growth [Burmeister and Dobell (1970) and Keely and Quah (1998)]. The second pertains to those research 'spillovers' (or externalities such as 'know-how') which cannot be fully appropriated by firms doing independent R&D and which are transmitted mainly by non-market means [OECD (1992), Faulkner and Senker (1995) and Geroski (1995)]. The diffusion literature tends to focus mainly on the diffusion of embodied technologies. The second element of the Schumpeter model that has been retained is the distinction between the invention stage and the innovation stage. Moreover, the definitions used by Schumpeter to differentiate between invention and innovation currently remain in use. It is convention in the diffusion literature that intellectual property law does not legally protect the invention in order to distinguish the modelling of the diffusion process from the related issue of licensing and patent race models

[Beath *et al* (1995)]. Thus, diffusion is assumed to begin at the first date at which the invention is commercialised [Lissoni and Metcalfe (1994)]. Models that have attempted to integrate the demand and supply (i.e. R&D as an invention and innovation creating activity have however, recently challenged this latter distinction) sides of diffusion [see, for example, Stoneman and Ireland (1984) and Metcalfe (1995)]. Such models are discussed in Section 2.3.2.

It is convention in the literature to distinguish between ‘adoption’ and ‘diffusion’ studies [Thirtle and Ruttan (1987) and Lissoni and Metcalfe (1994)]. The former refers to those studies that examine the firm in isolation and focuses attention on those economic factors that determines why some firms are early adopters and others are late. The latter is conventionally viewed as those studies that examine the aggregate behaviour of a sample of firms, without necessarily relying on an explicit microeconomic modelling of single firms’ decision processes. This distinction is, however, somewhat controversial and arguably ambiguous. In neo-classical economics there is an explicit assumption that modelling the adoption decision is a pre-requisite to understanding and deriving predictions of the time-path of diffusion [Stoneman (1983) and Karshenas and Stoneman (1995)]. More recent evolutionary economics contributions have, however, challenged this neo-classical perspective and have stressed that ordered patterns of diffusion may emerge from apparently non-maximising decision-making at the firm level. These themes will be more fully explored in the review of the theoretical literature in Section 2.3

2.2.2 The Dimensions of Diffusion

The spread of a new capital-embodied technology can occur in a number of different dimensions [Stoneman (1986)]. If the focus of attention is restricted to the diffusion of process innovations then there are three possible levels of aggregation: intra-firm, inter-firm and economy-wide diffusion.⁹ Each of these ‘dimensions’ can be formally defined

⁹ The following definitions can also be used for the diffusion of product innovations. See Stoneman (1983) for subsequent definitions.

as follows [Stoneman (1983) and Karshenas and Stoneman (1995)]. Consider a firm i ($i = 1, \dots, n_{jt}$) in industry j at time t , producing output Y_{ijt} , of which an amount X_{ijt} is produced by the new technology. Definitions for the three levels of aggregation follow directly from the following expression in (2.1) below:

$$Z_{ijt} \equiv X_{ijt}/Y_{ijt} \quad (2.1)$$

Note that the proportions in (2.1) are defined as flow concepts rather than as stocks. In each case the diffusion process is assumed to begin at the first date of commercialisation. Firstly, the analysis of the time path of Z_{ijt} , up to a point where Z_{ijt} is at a maximum¹⁰, is labelled *intra-firm* diffusion. Thus, intra-firm diffusion orientated studies examine the speed at which an innovation reaches given levels within a single firm.

Inter-firm diffusion studies, however, ignore the possible gradualism of internal adoption and focus their attention on the number of firms using new technologies. Defining some base level of use of a new technology as Z^* , a firm is defined as a user of a new technology at time t if $Z_{ijt} \geq Z^*$. It is convention to set Z^* equal to unity but this is not strictly necessary. Assuming that m_{jt} and n_{jt} are the number of adopters of the new technology and the total number of firms (or potential adopters) in industry j at time t respectively, then the analysis of inter-firm diffusion then concerns the time path of m_{jt}/n_{jt} ($\equiv M_{jt}$). The value of M_{jt} is bounded from below and above to be $0 \leq M_{jt} \leq 1$. It is usual in the theoretical and empirical literature to assume that n_{jt} is constant throughout the diffusion process ($= n$) and has generally been set equal to the stock of potential adopters at the first date of commercialisation of the innovation.¹¹ After the maximum level of diffusion has been reached, M_{jt} may well decline as other superior technologies appear. To date, however, the study of declining use of the new technology has not been part of the remit of the literature.

¹⁰ This may either be a satiation or equilibrium measure. See Stoneman (1983) for further details.

¹¹ This assumption has been heavily criticised by Gold (1981) and Gort and Klepper (1982).

An alternative measure of inter-firm diffusion is to consider the time path of X_{ijt}/Y_{ijt} as it approaches $(X_j/Y_j)^*$, the maximum level or post-diffusion ratio of X_{jt}/Y_{jt} . This is purely an output measure and can be combined with the intra-firm measure to yield an output measure of diffusion at the industry level [Karshenas and Stoneman (1995)]. The usual measure in the theoretical and empirical literature is, however, to take the number of adopting firms as the appropriate measure of the diffusion.

One further level of aggregation is possible, that of economy-wide or 'inter-industry' diffusion. This is measured by summing X_{jt} over the different industries j to obtain an economy-wide measure. There is a problem, however, with this approach - that of summing heterogeneous outputs. An alternative measure put forward is therefore the composition of capital stock [Stoneman (1983)]. This type of measure is also a possible alternative definition for inter-firm diffusion. In this case Davies (1979) has labelled such a measure as the 'overall' rate of diffusion as it cancels out information about single firms' adoption. This measure may, however, provide a more accurate measure of the level of diffusion within an industry or country depending whether the focus of attention is total output or the number of adopting firms.

Inter-firm studies are the most frequent, both at the theoretical and empirical level. Given the focus of this thesis, the rest of the chapter is devoted to a discussion of formal theoretical models of inter-firm diffusion.

2.3 Formal Theoretical Models of Process Innovation Diffusion

This section surveys the theoretical literature that has examined the inter-firm diffusion of process innovations. The literature has arguably been concerned with providing answers to four main questions: first, what factors will determine the post-diffusion level of use or ownership; second, why are some firms early adopters and others late; third, what will be the subsequent time path of diffusion and fourth, what firm-specific, market-specific and technology-specific characteristics will be key factors in

determining the shape of that time path? All the models surveyed in this section have attempted, to varying degrees, to provide answers to these four questions - although, arguably, the second question has been the focus of much more attention in the literature.

The section falls into two parts. Section 2.3.1 surveys the early literature (essentially pre-1970), whilst Section 2.3.2 surveys the contemporary literature (essentially post-1970). The dividing line between the early and contemporary literature is not simply a function of time, but also has an economic rationale. The early literature is dominated by information spreading mechanisms and a paucity of a choice-theoretic framework. The contemporary literature, in contrast, is dominated by a thoroughly neo-classical framework in which firms are assumed to be profit-maximising and have perfect information. The exception to this is the evolutionary approach (see Section 2.3.2.6) in which firms are characterised by bounded rationality and is, to date, the most recent strand in the literature. In both sections the initial starting point of the model is that a new, superior, technology has appeared and that there is no legal impediment to it being adopted (although financial ones may or may not exist). Moreover, possible linkages with the diffusion of ATM technology will be made throughout the remaining sections where appropriate.

2.3.1 The Early Literature

Although the seminal work of Schumpeter (1934, 1939) is conventionally viewed as laying the foundations for economists' study of innovation diffusion, it should be recognised that many of the initial analytical tools that were employed by early contributors to the study of diffusion were to a large extent derived or borrowed from other social sciences such as sociology, rural sociology and anthropology where the study of innovation diffusion had already been well established and had focused on the social characteristics of early and late adopters.¹² This aspect has had a significant

¹² See Rosegger (1996) for an extended discussion of these research traditions.

impact on the development of diffusion research in the economics literature. The early contributions of economists were concerned with verifying (theoretically) two stylised facts that previous research traditions had derived. First, that diffusion is a time-intensive process [Mansfield (1961, 1968) and Nasbeth and Ray (1974)]. Second, that plotting the number of users against time (M_{jt} in Section 2.2.2) yields a sigmoid or S-shaped diffusion curve [see, for example, the empirical work of Griliches (1957)].¹³ Early contributions thus relied heavily on the so-called ‘epidemic’ and information spreading explanations of innovation diffusion that conveniently yielded a logistic curve. An exception to this is the vintage capital model of Salter (1966) which is derived implicitly from his work on productivity, contains no information spreading explanation of diffusion and was the first in the neo-classical paradigm of profit-maximising firms with perfect information.

2.3.1.1 Epidemic and Information Based Models

The starting point of economists theoretical contributions to the study of innovation diffusion is conventionally taken [Karshenas and Stoneman (1995)] to be the work of Griliches (1957), Bain (1964) and Bass (1969) who all developed (and applied) epidemic theories of diffusion in which the spread of an innovation is made analogous to the spread of a disease.

The basic assumptions of the epidemic model are that there exists a fixed population of potential adopters (firms), n , of whom there already exists a number of adopters at time t , m_t . It is further assumed that there exists a constant rate of adoption (or being ‘infected’) and that each non-adopter has a constant and equal probability of learning (or ‘catching’ the disease) about the attributes of the innovation under investigation (as reflected in β) from informal contact with adopters. Finally, it is assumed that the proportion of the population who has already adopted (assuming homogenous mixing)

¹³ This stylised fact was made as early as Tarde (1903) from a sociological perspective.

will determine the number of such contacts.¹⁴ This set of assumptions leads to the following expression concerning the change in the number of adopters [Davies (1979)]:¹⁵

$$m_{t+1} - m_t = \beta(n - m_t)m_t/n, \beta > 0 \quad (2.2)$$

From the expression contained in (2.2), the number of firms adopting between times t and $t +$ is proportionate to the product of the number of non-adopting firms and the proportion of the firms that have already adopted, both at time t . If the period t to $t +$ is very small (2.2) may be alternatively be stated as:

$$\frac{dm_t}{dt} \left(\frac{1}{n - m_t} \right) = \beta \frac{m_t}{n} \quad (2.3)$$

The differential equation in (2.3) has the following solution [Davies (1979)]:

$$m_t/n = [1 + \exp(-\alpha - \beta t)]^{-1} \quad (2.4)$$

where α is the constant of integration. The expression in (2.4) is a logistic curve and can be interpreted as the cumulative density function (CDF) of the logistic distribution [Bailey (1957) and Friedman (1974)]. This curve predicts that the proportion of firms having adopted the innovation will increase at an accelerating rate until 50% adoption has been attained at time $t = -(\alpha/\beta)$. Thereafter, adoption increases at a decelerating rate and 100% adoption is approached asymptotically [Davies (1979)]. Two logistic curves pertaining to equation (2.4) are sketched in Figure 2.2 below for the parameter values $\alpha = 5.0$ (common to both), $\beta = 0.30$ and $\beta = 0.50$.

¹⁴ Stoneman (1983) has argued that the parameter β can be conceptualised as the probability of adopting the innovation after receipt of the information.

¹⁵ Alternatively, the expression could refer to the proportion of output produced in the industry by the new technology.

As can be observed from Figure 2.2, diffusion is more rapid the higher the value of β and as a consequence this variable is frequently referred to as the *speed of diffusion* and will be a function of the frequency of contact and the efficiency of information channels [Davies (1979)].¹⁶

The epidemic model was extremely popular in early studies mainly because of its analytical convenience. Firstly, the log transformation of (2.2) leads to an estimable equation and, secondly, β can be used in cross-sectional empirical studies to observe what factors determine the speed of diffusion (see Chapter 3, Section 3.2.1.1 for more discussion).

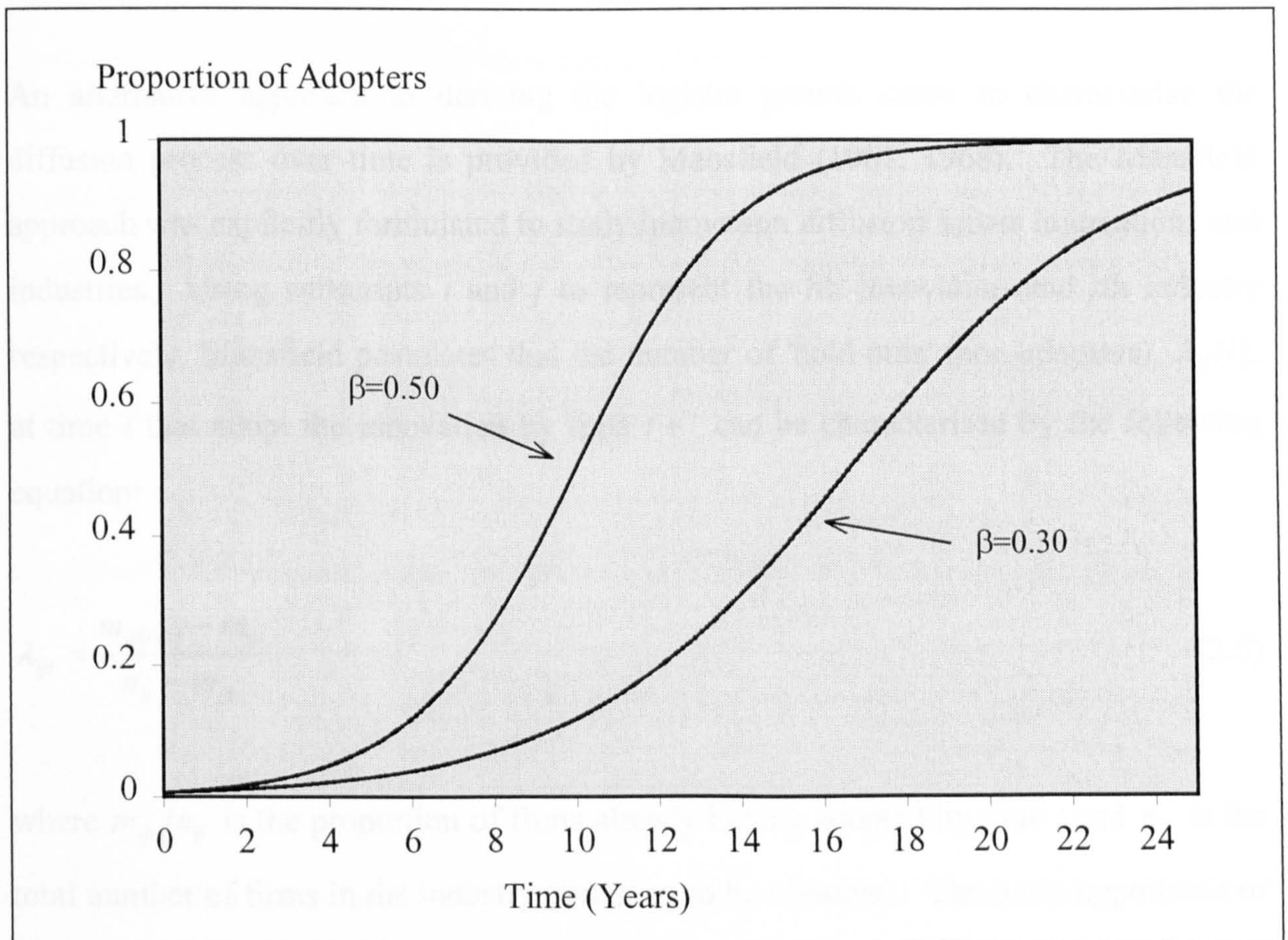


Figure 2.2 The Epidemic Inter-Firm Diffusion Curve

¹⁶ The speed of diffusion, β , should not be confused with the rate of growth of diffusion, which is given by $(dm_t / dt) / m_t$, and which falls continuously over time [Davies (1979)].

A contemporary interpretation of the epidemic model [see Metcalfe (1995)] is that what is being spread is not so much 'awareness' of the innovation or information pertaining to its 'existence', but rather a sub-set of knowledge that can be used to effectively evaluate the new technology. In this case the knowledge that will spread will predominately be tacit in nature [Faulkner and Senker (1995)] and will, thus, be transmitted by demonstration and largely by non-market means. Moreover, as noted by Thirtle and Ruttan (1987) and Karshenas and Stoneman (1995) the essence of the model is essentially disequilibrium one. That is, there is an end point level of use m_i/n and the diffusion process is interpreted as the movement towards this end-point. Adjustment to the new equilibrium is not instantaneous because of the asymmetric distribution of information.

An alternative approach to deriving the logistic growth curve to characterise the diffusion process over time is provided by Mansfield (1961, 1968). The Mansfield approach was explicitly formulated to study innovation diffusion across innovations and industries. Using subscripts i and j to represent the i th innovation and j th industry respectively, Mansfield postulates that the number of 'hold-outs' (non-adopters), $\lambda_{ij}(t)$, at time t that adopt the innovation by time $t+$ can be characterised by the following equation:

$$\lambda_{ijt} = \frac{m_{ij(t+1)} - m_{ijt}}{n_{ij} - m_{ijt}} \quad (2.5)$$

where m_{ijt}/n_{ij} is the proportion of firms already having adopted by time t and n_{ij} is the total number of firms in the industry (assumed to be constant). The basic hypothesis of the Mansfield model is then contained in the following relationship:

$$\lambda_{ijt} = f_i \left(\frac{m_{ijt}}{n_{ij}}, \pi_{ij}, S_{ij} \right) \quad (2.6)$$

where π_{ij} is the profitability of installing this innovation relative to that of alternative investments, and S_{ij} is the size of the investment outlay required to install the innovation as a proportion of the average total assets of firms in the industry. Mansfield postulates that the larger the proportion of firms that have already adopted the greater the probable competitive pressures on non-adopters and the more likely they are to accept that the innovation is profitable and relatively risk-free. Therefore, λ_{ijt} should be higher the larger is m_{ijt}/n_{ij} . In addition, Mansfield argues that λ_{ijt} should be higher the larger π_{ij} because the latter will increase the probability that a firm's estimate of the profitability will be high enough to 'compensate' for whatever risks are involved in adopting the new technology. Finally, Mansfield assumes that λ_{ijt} will be inversely related to S_{ij} as S_{ij} will indicate the extent of caution and financing problems (due to liquidity and funding constraints) associated with potential adoption.

Mansfield then develops the above model assuming that the general function $f_i(\cdot)$ in equation (2.5) can be approximated by a Taylor series expansion that omits all third and higher order terms. In addition, Mansfield makes the explicit assumption that the coefficient of $(m_{ijt}/n_{ij})^2$ in the Taylor expansion is set equal to zero because λ_{ijt} is not highly correlated with $(m_{ijt}/n_{ij})^2$ for the innovations in his sample.¹⁶ By assuming that the period t to $t+1$ is very short the equation in (2.5) can be written as:

$$\frac{dm_{ijt}}{dt} \frac{1}{(n_{ij} - m_{ijt})} = A_{ij} + \beta_{ij} \frac{m_{ijt}}{n_{ij}} \quad (2.7)$$

where $\beta_{ij} = a_{i1} + a_{i2}\pi_{ij} + a_{i3}S_{ij}$, and A_{ij} is the sum of all remaining terms from the expansion which do not contain (m_{ijt}/n_{ij}) . Finally, Mansfield (1961) imposes the limit condition that as time tends to zero the number of firms having introduced the innovation must also tend to zero. That is:

$$\lim_{t \rightarrow -\infty} m_{ijt} = 0 \quad (2.8)$$

The differential equation in (2.6) has the following solution that is the familiar logistic curve previously given in equation (2.4):

$$m_{ijt}/n_j = [1 + \exp(-\alpha_{ij} - \beta_{ij}t)]^{-1} \quad (2.9)$$

where α_{ij} is the constant of integration. The limit condition specified in equation (2.8) constrains A_{ij} to be equal to zero.¹⁷ Given the use of this limit condition, the equation (2.8) also coincides (ignoring the subscripts) with (2.4). Stoneman (1983) has argued that this similarity with the epidemic model is also reflected in the learning mechanism that is essentially epidemic (by infection). The firm is assumed to adopt the new technology if profitability is high enough or uncertainty sufficiently low. Uncertainty is reduced in proportion to the number of firms already using the new technology. Thus, in the Mansfield model, adoption (and subsequent diffusion) is not immediate because uncertainty prevents this and only as use extends will uncertainty be reduced.

Despite the wide use of the epidemic model (and Mansfield's re-interpretation) in early studies of the diffusion process, it has subsequently come under severe criticism by, *inter alia*, Lekvall and Wahlbin (1973), Davies (1979), Stoneman (1983), Lissoni and Metcalfe (1994) and Karshenas and Stoneman (1995). There are two sources of these critiques. First, those assumptions underlying the use of the logistic curve to characterise the growth in innovation usage over time. Second, those assumptions made in the derivation of Mansfield's model contained in (2.9).

Critiques of the epidemic model have focused on those assumptions used in deriving (2.4) as a representation of inter-firm diffusion across time. There are two themes to these critiques. The first is that the derivation of the logistic curve depends on β remaining constant over time, for all firms, and that all firms must have an equal

¹⁶ This is based on empirical evidence presented in Mansfield (1961).

¹⁷ Davies (1979) notes that since it is explicitly assumed by Mansfield that π and S are constant over time, this implies that A must be zero at all times, and not only at the limit. This is clearly a restrictive assumption of the Mansfield model since there seems to be no obvious reason why π and S should remain constant over time.

probability of adoption. As noted by Davies (1979), if these assumptions are dropped the resulting diffusion curve will not be a logistic, but rather will be skewed. The assumption that all firms have an equal probability of adoption certainly appears highly restrictive as it ignores firm-specific characteristics (such as firm size) and has certainly been severely undermined by recent empirical evidence (see Chapter 3). Moreover, the model cannot predict which firms will be relatively early and which will be relatively late adopters. The second line of criticism concerns the sources of information in the model. As shown by Lekvall and Wahlbin (1973), the epidemic model relies exclusively on internal sources of information (or 'endogenous learning') and, thus, excludes the possibility of external sources such as advertising. Lekvall and Wahlbin show that if such external sources are included in the model the resulting diffusion path will be positively skewed, with the degree of skewness being greater for more heavily advertised innovations.¹⁸

Critics of Mansfield's (1961) model have focused their attention on those assumptions underlying the derivation of (2.8). The model has been criticised on three counts [Davies (1979), Stoneman (1983), Karshenas and Stoneman (1995)]. First, there is no theoretical explanation of why the firm's adoption decision should depend on the risk and profitability of the innovation and why risk attached to the innovation is reduced by usage. Second, there is an internal contradiction contained in the model. 'Risk' in the Mansfield model is the uncertainty attached to the profitability of the innovation. Although this uncertainty reduces over time as more firms adopt the innovation there is the implicit assumption made in equation (2.8) that the firm's estimate over time remains unchanged. The firm must therefore be learning that its estimate of the profitability of the innovation must be the correct one. This may be a very restrictive assumption to make, especially when it comes to the empirical implementation of the model which necessitates an arbitrary date to be chosen for the 'correct' measure of profitability [Stoneman (1983)]. Third, the Mansfield model only considers *ex post* profitability as a determinant of the speed of diffusion. As will be emphasised in Section 2.3.2, however, the *expected* profitability (which can change over time) of an

¹⁸ The marketing literature has arguably developed this approach much further than the economics one. See, for example, Mahajan *et al* (1990).

innovation has been identified in contemporary theoretical literature as being more important for the adoption decision than *ex post* profitability.

2.3.1.2 Vintage Capital Models

The first approach to new technology diffusion that includes the neo-classical elements of profit maximisation and perfect information is Salter's (1966) vintage capital model. Moreover, Salter's model has arguably contributed a great deal to the development of the contemporary theoretical literature. Indeed, Salter was arguably one of the first economists to stress the importance of diffusion in economic analysis. As Salter (1966) states:

Quite obviously this delay in the use of new methods is extremely important in productivity analysis: *it cannot be neglected, or even relegated to a minor role*. An understanding of productivity movements must include an analysis of the reasons for this delay in the utilisation of new techniques, and an appreciation of the forces which determine the rate at which new methods displace the old. [Salter (1966), p. 49].¹⁹

Salter's theoretical model of diffusion is developed implicitly from his seminal work on the linkages between labour productivity and technical change and can be applied at both an industry level of aggregation or economy-wide one. The former level is considered here. The pivotal assumption of the model is that new technology is completely embodied in new capital equipment so that *gross investment* is made the impetus behind the diffusion process.²⁰ No allowance is made in the model for learning-by-doing or learning-by-using effects. A further three assumptions are made which are crucial to the development of the model. First, existing plants embody the best-practice technologies at the date of their construction, cannot be adopted to any other technique and are infinitely lived. Second, there is perfect competition in product and factor

¹⁹ The author has added italics.

²⁰ In this regard the Salter model is antecedent to the macro-models of diffusion such as Chow (1967) and Stoneman (1976) where investment in time t is assumed to be some proportion of the difference between desired and actual capital stock. As shown by Stoneman (1983) these 'stock adjustment' models can be derived from the existence of adjustment cost functions facing profit maximising firms.

markets, and third, plants work at normal capacity and labour and managerial efficiency is equal in all plants.

Diffusion of new technologies proceeds in this model by the interaction of declines in industry product prices brought about by the adoption of capacity-increasing new technologies (or so-called 'best-practice techniques') and plant scrapping and investment criteria (or rules). At any moment in time the capital stock of the industry consists of a distribution of different vintages with different productive potentials (measured by labour productivity). The oldest vintages are embodied in those plants that are 'marginal' in that they only just cover their operating costs. These marginal plants are scrapped when they fail to yield a surplus greater than operating costs (i.e. when they yield zero quasi-rents). Indeed, the range of technologies in existence is defined by this condition.²¹

Industry price for the homogenous product is assumed to be composed of operating costs plus capital costs of best-practice plants (i.e. those adopted in the current period). Next period a new best-practice technology emerges which yields super-normal profits. Capacity in the industry increases with paralleled investment in new plants and prices fall to a point at which best-practice techniques yield just normal profits. At these lower prices, marginal plants are no longer covering their operating costs. A continuous series of new techniques will, over time, lead to a gradual scrapping of old plants and their replacement by better ones.

Thus, there are two main factors in the vintage capital model that determines the speed of diffusion: the time path of the price of the innovation and the vintage distribution of existing technologies [Davies (1979)]. A sigmoid inter-firm diffusion curve is not guaranteed in this model as it is in the epidemic model.

The model has two limitations. First, it assumes that new investments are made exclusively in best practice technologies. This assumption does not confer with the

empirical evidence [see Goodacre and Tonks (1995) and references therein]. Secondly, the model assumes perfect information, which may be an overly restrictive one to make in the case of new technology, whose economic benefits is inherently uncertain [Arrow (1962a)]. These critiques are, arguably, too harsh and Salter's model should be viewed within the context of the historical development of diffusion theory [Gourlay (1998a)]. Salter's model represents the first step in developing formal theoretical models of diffusion. More importantly, it established an economic basis (rather than an sociological or exclusively informational-based one) of why it was rational for firms to delay adoption when a superior technology had emerged. Moreover, it can explain why old technologies co-exist with new ones.

2.3.2 The Contemporary Literature

There are two main elements that distinguish the contemporary literature from the early contributions outlined in Section 2.3.1. First, there is the dismissal of information dissemination as the key explanatory factor of innovation diffusion. Instead, in the majority of contemporary theoretical models information is assumed to be perfect. In so far as information imperfections do impinge on the diffusion process it is within a definite Bayesian learning framework which allows firms to adjust their prior perceptions of the nature of the new technology to the true nature. Second, it is assumed that firms behave optimally in that they maximise (discounted) profits. Thus a self-imposed methodology is that at any point in time all those firms for which the technology is profitable will have adopted. If some potential adopters have not yet done so then it is not due to being 'ill informed', but rather that they are waiting for the optimal time for adoption to arrive.

Consequently, the majority of theoretical models outlined in this section are thoroughly neo-classical in character. The exception to this is the most recent strand in the

²¹ Salter shows that this is simultaneously determined by the investment criteria - that is, investment in new plants occurs when operating costs equal total costs (including amortisation and interest) of a new plant. Salter shows that this investment criteria and the scrapping criteria yield the same scrapping date.

literature that of the evolutionary approach inspired by the seminal work of Nelson and Winter (1982). This approach assumes that firms are characterised by bounded rationality and that a diffusion process may emerge from the apparently irrational behaviour of individual firms.

2.3.2.1 Probit or Rank Models

In this class of models the impetus behind the diffusion process is that firm heterogeneity and changes in factor prices or learning-by-doing over time, combine via a specified adoption rule, to create different preferred adoption dates. Diffusion in these models is a time-intensive process because it is not profitable for all firms to adopt immediately. The diffusion literature [Davies (1979), Stoneman (1983) and Karshenas and Stoneman (1995)] has labelled models in this mould as 'Probit Models' or 'Rank Models' because they closely resemble econometric probit (and logit) models which are applicable to situations where the decision of an economic agent is typically discrete (for example, the decision to adopt or not to adopt). In such cases the subsequent dependent (or exogenous) variable under interest assumes discrete values and thus requires specialised econometric models to deal with this aspect [see Section 3.2.2.1 of Chapter 3 for more details].

The origins of these models can be found in the early literature pertaining to the diffusion of product innovations [see, for example, Cramer (1969), Bain (1964) and Bonus (1973)] in which a consumer is assumed to adopt the new product at time t if and only if their income exceeds some critical level. In these models the probability of adoption is typically related to income by a so-called 'Quasi-Engel' curve [Bonus (1973)] which is usually assumed to have a cumulative lognormal distribution. Diffusion then proceeds as income grows overtime and is quicker the smaller the variance in population income [Davies (1979)]. By changing 'consumers' into 'firms', 'income' into 'size', and 'product innovation' into 'capital-embodied innovation' the model can then be adapted to the inter-firm diffusion of process innovations.

The main elements of inter-firm probit models can be illustrated by means of the following, highly simplified example. It is assumed that information is perfect and that potential adopters differ from each other by a crucial inherent characteristic (such as firm size for example), z . Moreover, gross benefits from adoption are assumed not to decline as diffusion proceeds. The probability density function (PDF) and cumulative density function (CDF) of z are assumed to be $f(z)$ and $F(z)$ respectively. At time t a potential adopter, i ($i = 1, \dots, n$), will be an adopter of the technology if its characteristic level z_i is greater or equal to some critical level of the characteristic \bar{z}_t . The proportion of the population who have then adopted at time t will thus be given by $1 - F(\bar{z}_t)$. These elements are illustrated in Figure 2.3 below, where diffusion proceeds over time as either $f(z)$ shifts and/or \bar{z}_t falls over time. The resulting diffusion path will, in this highly simplified version of the probit model, reflect the shape of $f(z)$ and the movement of \bar{z}_t over time [Stoneman (1986)].

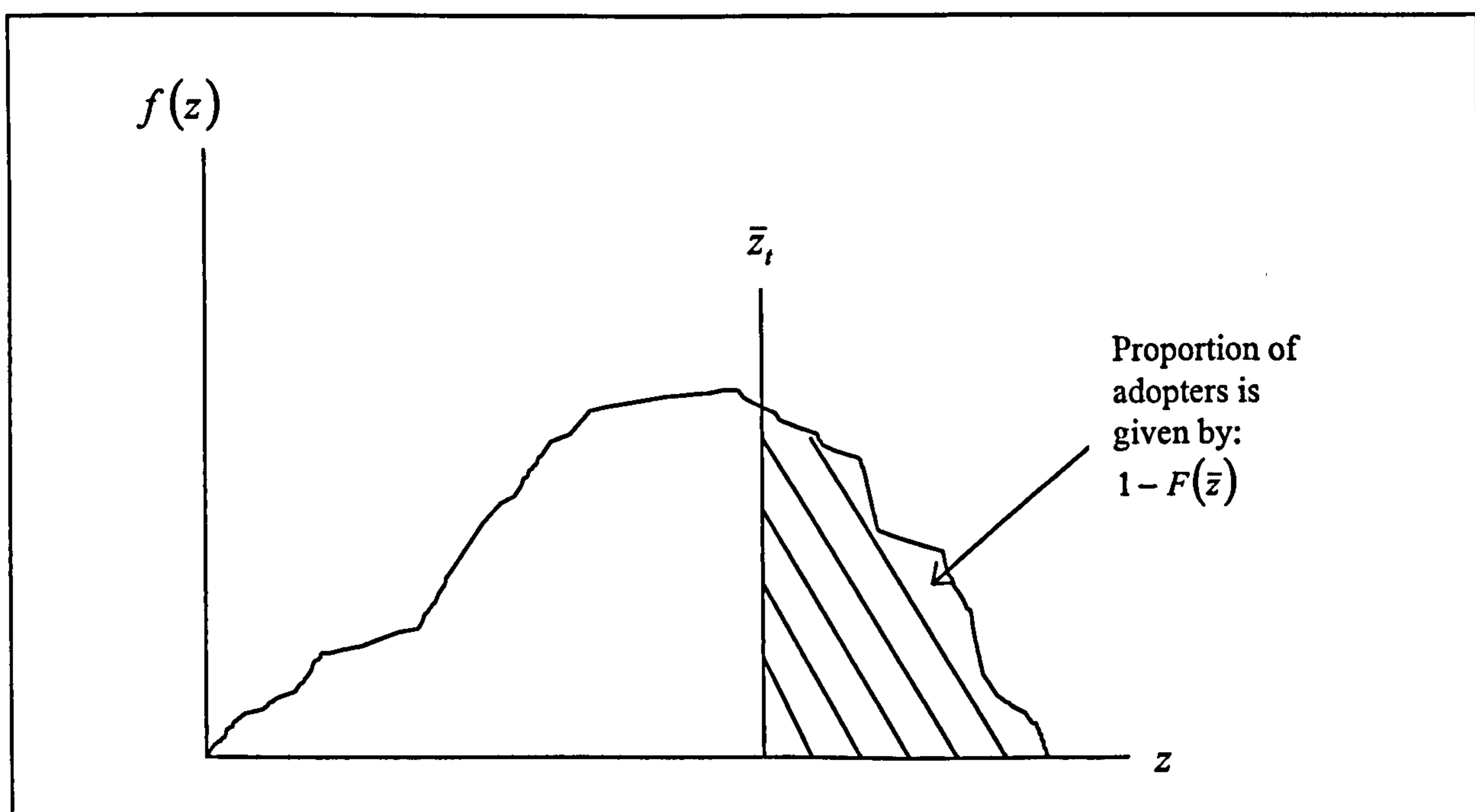


Figure 2.3 The Probit Model Mechanism of Diffusion

Formal inter-firm versions of these models closely resemble this simplified account of product innovation diffusion. Potential adopters (firms) of a new technology are assumed to have different inherent characteristics (firm size is conventionally taken to

be the critical characteristic) and as a result obtain different gross returns resulting from the adoption of the new technology. These different returns then generate different preferred adoption dates. Models are then generally made operational by ranking potential adopters in terms of their returns from adoption (from highest to lowest), thereby generating a benefit distribution across potential adopters. An acquisition rule relating benefits to the cost of acquisition enables the derivation of a distribution of reservation acquisition costs from the benefit distribution. Firms adopt the new technology as acquisition costs fall below reservation acquisition costs. Acquisition costs are then assumed to fall over time through a learning-by-doing mechanism on the supply side (i.e. the manufactures of the new technology). As acquisition costs fall, the cumulative benefit distribution is mapped out as a diffusion path, with those firms achieving high returns adopting early and those firms achieving low returns adopting late [Karshenas and Stoneman (1993)].

Early models in this mould are those of David (1969, 1975) and Davies (1979). The model of David explicitly considers the diffusion of a specific capital-embodied innovation - that of mechanical reapers in the US agricultural sectors during the last century - and identifies firm size as the critical firm characteristic that determines the adoption decision. The identification of firm size as the critical variable is derived from David's arguments concerning the inherent factor bias of mechanical reapers. The model assumes that the reaper involves fixed costs above and variable costs below those of the replaced technique. The model then has the following elements. The purchase cost of the new equipment for a firm i at time t is given by C_t , with an imputed rental value of R_t and c_t is defined as $c_t = C_t R_t$. It is further assumed by David that C_t is zero for the replaced technique and that the main incentive underlying the adoption of reapers is that the technique saves labour inputs relative to that of the old technique. Labour savings are such that for each unit produced the labour saving is denoted by L_{sit} , where i denotes the i th firm and s denotes 'saving'. Both the old and new technology is assumed to have constant economies of scale. Letting w_t to be the wage rate, David argues that there will be some level of scale (output) at which labour savings just

compensate for the increased capital costs. This level of size defines the critical value of firm size \bar{X}_{it} . Thus, \bar{X}_{it} is defined by David as that value of X_{it} where:

$$\bar{X}_{it} w_t L_{sit} = C_t R_t \quad (2.10)$$

The expression in (2.10) states that it is profitable to adopt the new technology when total (monetary) labour saving is equal to the cost of the new technology. Defining $\omega_t = w_t / C_t R_t$ (i.e. relative prices of inputs), then (2.10) can be written as:

$$\bar{X}_{it} = \frac{1}{L_{sit}} \frac{C_t R_t}{w_t} \equiv \frac{1}{\omega_t} \frac{1}{L_{sit}} \quad (2.11)$$

The expression in (2.11) gives the essence of the probit approach to inter-firm diffusion. To generate a time-intensive path from (2.11) requires that either \bar{X}_{it} or the distribution of X_{it} to change over time [Stoneman (1983)]. Given that (2.11) does not hold for all i at the date when reapers were first available, diffusion will not be instantaneous but will proceed in this model if and only if: actual firm size increases over time (X_{it} increases); the cost of reapers declines relative to wages (ω_t falls); and reapers save more labour inputs over time (L_{sit} falls). In the simplest version of this model David assumes that $f(X_{it})$ remains constant over time and argues that \bar{X}_{it} will change over time as wages rise relative to capital costs.²² David then argues that \bar{X}_{it} evolves over time according to the following relationship:

$$\frac{d\bar{X}_{it}}{dt} = -\bar{X}_{it} \frac{d\omega_t}{dt} \frac{1}{\omega_t} \quad (2.12)$$

and that $(d\omega_t/dt)(1/\omega_t) = \lambda$, a constant independent of time. This implies that relative factor prices, ω_t , follows an exponential time trend [Stoneman (1983)]. Defining D_t as the proportion of potential adopters using the technique in time t , this can be defined as:

²² The argument that wages increase relative to capital costs is a stylised fact.

$$D_t = \Pr(X_{it} = \bar{X}_{it}) = \int_{\bar{X}_{it}}^{\infty} f(X_{it}) dX_{it} \quad (2.13)$$

Differentiating (2.13) with respect to critical firm size obtains:

$$\frac{dD_t}{d\bar{X}_{it}} = -f(\bar{X}_{it}) \quad (2.14)$$

and

$$\frac{dD_t}{dt} = \frac{dD_t}{d\bar{X}_{it}} \frac{d\bar{X}_{it}}{dt} = f(\bar{X}_{it}) \lambda \bar{X}_{it} \quad (2.15)$$

David shows that D_t , as defined in equation (2.13), given $f(X_{it})$ is lognormal and λ is constant, will trace out the standard cumulative normal curve when plotted against a positive linear transformation of the time variable, whose curve is sigmoid. It is significant to note that the crucial role of firm size in generating a diffusion path in this model reflects the inherent ‘lumpiness’ of capital (i.e. all firms have to pay identical rental costs) rather than any scale economies advantages of being a larger firm. Under these circumstances, large firms must always be at an advantage in this model because the cost of new technology can be ‘spread’ over a larger scale of operations [Davies (1979)].

Davies’s (1979) model of process innovation diffusion is similar to that of David’s outlined above, but assumes that the existence of uncertainty and imperfect information pertaining to the innovation implies that adoption decisions are characterised by behavioural (or ‘satisficing’) rules rather than explicit profit maximising ones. More specifically, in Davies’s model a firm i will adopt the new technology if the expected pay-off period (the inverse of profitability) from its use, $ER_{it} \leq \bar{R}_{it}$, some critical or ‘target’ pay-off period. The expected pay-off period is then made a function of firm size X_{it} , other firm-specific technological characteristics (Y_{it}) and time.

These assumptions lead to the following expression for the expected pay-off period:

$$ER_{it} = \theta_{it} X_{it}^{\beta} \varepsilon_{1it} \quad (2.16)$$

with $\beta \geq 0$

where:

$$\varepsilon_{it} = \prod_{j=1}^r Y_{ijt}^{v(j)} > 0 \quad (2.17)$$

where θ is a time-varying parameter characterising learning-by-using and by-doing effects both in the innovation using and producing industry (as reflected by declining labour inputs) and improvements in the quality of information over time (through both greater dissemination and search activity by firms). Davies assumes that $\theta_{it} > 0$ and $d\theta_{it}/dt(1/\theta_{it}) < 0 \forall t$. These two assumptions imply that ER_{it} declines monotonically over time for all i .

Thus, from equation (2.16) the expected pay-off period resulting from adoption at time t is assumed to be a multiplicative function of firm size, X_{it} , and r other characteristics of the firm. Davies argues that $\beta > 0$, that is, larger firms have a higher probability of adoption *ceteris paribus*. This hypothesis is based on five main tenets: firstly, that large firms are likely to use technology more intensively and thus benefit from any inherent economies of scale existing in the technology; secondly, that they are more likely to employ skilled management and staff; thirdly, they will have large resources and therefore will be less affected by disturbances to output that accompanies adoption; fourthly, they replace capital equipment on a more frequent basis, and; fifthly, they will have higher capacity and therefore be able to better optimise their technology mix when adopting the new technology.

The critical pay-off period is \bar{R}_{it} and is related to firm size and other firm-specific financial characteristics, Z_{ijt} , by the following expression:

$$\bar{R}_{it} = \theta_{2t} X_{it}^{\beta_2} \varepsilon_{2it} \quad (2.18)$$

with $\beta_2 \geq 0$

where

$$\varepsilon_{2it} = \prod_{j=1}^u Z_{ijt}^{r(j)} > 0 \quad (2.19)$$

with $\theta_{2t} > 0$ and $(d\theta_{2t}/dt)(1/\theta_{2t}) > 0 \forall t$. These two assumptions imply that \bar{R}_{it} increases monotonically over time.

A firm will become an adopter when its assessment of the profitability of adoption is sufficiently favourable to suggest that the initial outlay required will be recouped within an acceptable time period. This implies that the condition $ER_{it} \leq \bar{R}_{it}$ is a necessary and sufficient one.²³ Using the expressions in (2.16) and (2.18) and assuming $\beta > 0$ this adoption rule implies that the following condition must hold for adoption:

$$\frac{\theta_{1t} X_{it}^{\beta_1} \varepsilon_{1it}}{\theta_{2t} X_{it}^{\beta_2}} \leq 1 \quad (2.20)$$

defining $\theta_t = \theta_{1t}/\theta_{2t}$, $\varepsilon_{it} = \varepsilon_{1it}/\varepsilon_{2it}$ and $\beta = \beta_1 - \beta_2$, then (2.20) may be re-written as:

$$\theta_t X_{it}^{\beta} \varepsilon_{it} \leq 1 \quad (2.21)$$

Thus a firm will adopt the new technology if and only if:

²³ Davies (1979) ignores the possibility of dis-adoption in the construction of the theoretical model.

$$X_{it}^{\beta} \leq (\theta_t \varepsilon_{it})^{-1} \quad (2.22)$$

and critical firm size will be defined as that size such that the following equality holds:

$$\bar{X}_{it} = (\theta_t \varepsilon_{it})^{-1/\beta} \quad (2.23)$$

with $\beta > 0$, $\theta_t > 0$ and $(d\theta/dt)/\theta > 0 \forall t$. From equation (2.23), critical firm size, \bar{X}_{it} , falls monotonically over time then probability of adoption increases over time.²⁴ Moreover, from (2.23), critical firm size varies across firms reflected by ε_{it} and as this term is the ratio of two error terms which are assumed independent of firm size Davies shows that ε_{it} will therefore be lognormally distributed across firms. Moreover, given that β and θ_t are constant across firms then critical firm size, \bar{X}_{it} , is also lognormally distributed with mean $-\log(\theta_t/\beta)$ and variance $(\sigma_t/\beta)^2$ [Davies (1979)]. The model, outlined to this point, has specified a distribution for firm sizes and an adoption rule. To generate a diffusion path over time still requires specification of how critical firm size in (2.23) declines over time. In this respect, Davies considers two types of innovations - Group A and Group B - distinguished by the nature of their respective learning mechanisms over time. These are specified as follows:

$$\text{Group A Innovations: for which } \theta_t = \alpha t^{\psi}; \quad \alpha > 0 \quad (2.24)$$

$$\text{Group B Innovations: for which } \theta_t = \alpha e^{\psi t}; \quad 0 < \psi < 1 \quad (2.25)$$

where α is the initial value of θ_t (at $t = 0$ for Group B and $t = 1$ for Group A) and ψ is a growth parameter in θ_t , reflecting the strength of the learning effect, and improvements in the quality information [Davies (1979)]. Group A innovations are relatively cheap, technologically simple innovations and are largely produced off-site. The learning effects for these innovations are initially quite large but soon fall away. Moreover, information about their economic attributes is likely to be quite accurate early

²⁴ For the case of $\beta < 0$, then as critical firm size increases over time so does the probability of adoption.

on in their commercialisation [Davies (1979)]. In contrast, Group B innovations are characterised as being relatively expensive, technologically sophisticated and are subject to sustained post-innovation improvements. The learning effects for this group of innovations is initially slow, but more sustained and more substantial proceeding their commercialisation. Davies shows that for Group A innovations the diffusion curve will follow a cumulative lognormal time path and a cumulative normal time path for Group B innovations. In variations of the basic model, Davies (1979) considers the role played by industry growth and cyclical factors, but these do not change the substance of the above results.

The models of David (1969, 1975) and Davies (1979) can be subjected to two main critiques that, together, conveniently lead to a discussion of recent developments in the theoretical literature. Firstly, diffusion in these two models is driven by exogenous changes in factor prices and is characterised by an absence of any attempt to endogenise such changes. The essence of the David and Davies models is that at a moment in time, firms hold that stock of technology that is appropriate with their estimates of the returns to that technology. Only as factor prices change does the actual profitability of adoption change (increase) and, thus, diffusion proceeds. Any modelling of the supply side is, however, absent in their approaches. Secondly, learning is given only a minor *ad hoc* role in these models. Given that the economic and technical attributes of the new technology are unlikely to be known with a high degree of certainty at the initial point of adoption [Arrow (1962a) and Rosenberg (1982)] it can be argued that models which ignore learning effects are particularly weakened.

These weaknesses in early probit models have lead to two major advances in the theoretical literature. The first advance centres on the introduction of more sophisticated and formal learning mechanisms, while the second strand has attempted to model the supply side in order to endogenise price movements. In both cases the essential elements of the probit approach have been retained - that is, firm heterogeneity is the force behind diffusion.

Bayesian Learning Models

The incorporation of learning effects within a distinct probit approach was initiated by the seminal work of Stoneman (1980, 1981) and has subsequently lead to a number of similar contributions conventionally labelled under the generic name of the ‘Bayesian learning approach’ [Lissoni and Metcalfe (1994)]. This approach amalgamates both intra- and inter-firm dimensions of diffusion within a single model. Firm heterogeneity remains at the forefront of the diffusion process, but rather than firm size being the critical variable, the attitude and perceptions of the firm’s management (or ‘entrepreneurs’) toward the innovation is made the only source of difference amongst firms.

In Stoneman (1980, 1981) the individual entrepreneur is faced with the choice of using two technologies: the new technology and the old technology. These technologies have anticipated returns normally distributed as $N(\mu_{nt}, \sigma_{nt}^2)$ and $N(\mu_{ot}, \sigma_{ot}^2)$ respectively with a covariance of returns being $\rho\sigma_{nt}\sigma_{ot}$, where ρ is the correlation between new and old technologies ($\rho > -1$). At time t , the new and old technologies are used in the proportions $\alpha_t:1-\alpha_t$, respectively, where α_t is defined as the proportion of (fixed) output produced on the new technology. The entrepreneur is assumed to choose those proportions in order to maximise the following utility function subject to an adjustment cost constraint:

$$H(\mu_t, \sigma_t^2) = \mu_t - \frac{1}{2}b\sigma_t^2 \quad (2.26)$$

where μ_t and σ_t^2 are the parameters of the joint distribution of returns and b represents the ‘risk coefficient’ of the firm.²⁵ Assuming that there are no adjustment costs, the optimal value of α_t is shown by Stoneman (1981) to be given as the following:

²⁵ If $b > 0$, then the firm is risk averse. The implications of the term containing b is that if the variance of the new technique is less than that of the old and/or if the old and new techniques are less than perfectly correlated, then diffusion of the new technique will reduce risk and will be faster for a risk averse firm than for one that is risk neutral [Thirtle and Ruttan (1987)].

$$\alpha_t^* = \frac{(1/b)(\mu_{nt} - \mu_{ot}) + \sigma_{ot}^2 - \rho\sigma_{nt}\sigma_{ot}}{\sigma_{nt}^2 + \sigma_{ot}^2 - 2\rho\sigma_{nt}\sigma_{ot}} \quad (2.27)$$

with $0 \leq \alpha_t^* \leq 1$. The incorporation of learning effects into the model centre around the anticipated values of the means and variances of the new technology. Stoneman assumes that from past experience the entrepreneur knows with certainty the returns to the old technology. These are assumed to be fixed and have the distribution $N(\mu_o, \sigma_o^2)$. In contrast, the returns to the new technology are not known with certainty. More succinctly, Stoneman assumes that the entrepreneur learns about the characteristics of the new technology in a Bayesian manner [Lindley (1965)]. Explicitly, the true returns to the new technology are assumed to be $N(\bar{\mu}_n, \bar{\sigma}_n^2)$ with $\bar{\sigma}_n^2$ known but $\bar{\mu}_n$ is not known with certainty. The essence of the Bayesian approach is that at time t the entrepreneur holds a prior distribution on the difference in mean profitability to be gained from using the new rather than the old technology, that is, $N(\mu_{nt} - \mu_o, z_t)$. As time proceeds, the entrepreneur monitors the performance of this initial batch of new technology and then adjusts his prior distribution of the returns to that pertaining to the new technology in a Bayesian manner. As perceptions of the parameters of the distribution of returns to the new technology change so does the mix in which he wishes to use the two technologies. This will, under certain conditions, lead to α_t rising above some reservation proportion k after which the firm is defined as a user. The time path of α_t towards α_t^* (the 'ceiling level') sketches out the intra-firm diffusion curve. Stoneman shows that the greater the entrepreneurs initial estimate of the mean profitability, the greater its true mean returns and the lower the mean return to the old technology, the earlier will the firm be expected to use the technology. Low risk aversion is also associated with early use. Since firms differ in the initial conceptions of the returns to the new technology, the date at which α_t will become greater than k will differ across firms. If an PDF of priors is specified a inter-firm diffusion curve may thus be obtained in the conventional probit approach [Stoneman (1983, 1986)]. Stoneman (1983) illustrates that if the mean and variance of returns to the innovation are lognormally distributed then a cumulative lognormal diffusion curve results.

The interesting aspect of the Stoneman model is that there are no exogenous changes in the economic environment - diffusion results purely from learning, although learning is not by infection as it was in the epidemic model but, rather learning is from experience. It is rational for firms not to immediately adopt because they will not correctly perceive the true returns and risk associated with the technology.

Jensen (1982, 1983, 1988a, 1988b) has provided variations on Stoneman's learning model. In Jensen (1982), the adoption of new technology is viewed as a problem of decision making under uncertainty in which learning can occur. The firm is confronted by an exogenously developed innovation which, if adopted, can either increase or decrease the firm's expected present value. Information about the innovation's existence is costless and is derived purely from external sources to the firm. These sources provide information at discrete intervals so long as the firm remains a non-adopter. These 'pieces' of information are assumed to be represented by a Bernoulli random variable Z which takes the value of unity if the information is 'favourable' or zero otherwise. The distribution of the Z_i 's are assumed independent and distributed with unknown parameter $\theta = \Pr(Z_i = 1) \in (0,1)$. It is additionally assumed by Jensen that θ can only take two values, θ_1 and θ_2 , where $0 < \theta_2 < \theta_1 < 1$. Moreover, $\theta = \theta_1$ if the innovation is profitable and unprofitable if $\theta = \theta_2$. Associated with these parameters are discounted revenues R_1 (if $\theta = \theta_1$) and R_0 (if $\theta = \theta_2$), for which $R_1 > R_0$. The cost of adoption is assumed to be constant.

The firm is assumed to use Bayesian updating rules to learn about the distribution of θ and at each moment in time can make three decisions: adopt, not adopt, or wait for further information before making a decision. The decision to adopt is assumed to be irreversible and decision-makers are assumed to be risk neutral, acting to maximise expected profits. The decision problem is shown by Jensen (1982) to be an optimal stopping problem [see Lambert (1990)] in which the stopping value is the expected returns from adoption and the value from optimal continuation is the discounted expected value of the next piece of information. Jensen then shows that adoption will occur at any date at which the firm's current prior of the innovation's profitability $p \geq p^*$, some reservation level. Moreover, Jensen shows that delay in adoption is

optimal for the firm the more optimistic the firm is, the more favourable the information received, the higher the discount rate or period adoption returns, and the lower the cost of adoption. The model is converted to a probit inter-firm model by assuming a (uniform) distribution of prior beliefs. At the inter-firm level, Jensen shows that diffusion is faster the lower the costs of adoption, the higher the per period adoption returns and the higher the discount rate. In Jensen (1983), the model is extended to a choice between competing innovations. In both cases it is shown that the model can predict a sigmoid inter-firm diffusion curve. A similar model is contained in Feder (1982a, 1982b) and Just and Zilberman (1983) for the case of innovation in the agricultural sector.

In Jensen (1984), the basic learning model is extended to allow for more than one information message per period. Most interesting, however are the extensions provided in Jensen (1988a, 1988b) where information pertaining to the innovation contains costless and costly elements (i.e. the firm has to pay for part of the information). The decision problem remains threefold, as in Jensen (1982), but are now redefined as: to adopt, wait and receive a piece of costless information or wait and buy some information in addition to the costless information. The problem is set within a two-period time dimension. Jensen shows that if learning is costly then immediate adoption becomes more likely and eventual adoption of a profitable innovation is not certain. Under certain circumstances the reservation policy in Jensen (1982) becomes sub-optimal. The major contribution in Jensen (1988a, 1988b) is, however, to show that if information is costly some firms may never learn about an innovation that is profitable for them because they are unwilling to pay for the information. The welfare implications of adoption under costly information are explored in Jensen (1992).

The intra-firm model of Stoneman (1981) has been most recently extended by Tsur *et al* (1990) for the explicit case of the agricultural sector. The model includes both dynamic elements and learning effects. Two types of learning are considered: learning-by-doing and Bayesian learning effects. The model predicts that risk aversion positively affects the probability of adoption because the more cautious producer has a greater appreciation of future declines in risk resulting from learning.

Demand and Supply Models

The second recent advance in the literature has been the incorporation of the supply-side into the modelling framework. The explicit modelling of the supply-side has been designed to generate three essential elements [Karshenas and Stoneman (1995)]. First, a supply curve relating for each time period the quantity at each price that the (capital producing) industry is willing to supply. Second, a time path for technological improvements, and third, a resolution of any conflict between quantities supplied and demanded.

Although exact specifications of demand and supply diffusion models differ, they do share a common framework labelled by Stoneman (1987) as the 'basic type' of technical change. This framework characterises the diffusion process as involving a capital producing industry that invents, produces and markets an embodied innovation. For the capital-producing industry the new technology is a product innovation. This innovation is then adopted by a capital-using industry to which the new technology represents a process innovation. Subsequent price reductions and improvements in the technology may occur through a process of learning-by-doing and this effect leads, through interaction with a specified adoption rule, to greater diffusion. In extending the basic probit models outlined above, economists have been concerned with formalising this 'basic type' by being more precise about the time structure of the supply industry's costs, the number of firms in the supply industry and the nature of their capacity constraints and the price and quantity setting behaviour of firms [Lissoni and Metcalfe (1994)].

An early attempt to incorporate a supply-side was made by Glaister (1974) who considered the intersection of a Mansfield (1961, 1968) type demand structure with a monopolised supply sector and shows that Mansfield's demand-based results are considerably modified. Since the Mansfield model suffers from a number of considerable weaknesses (see Section 2.3.2.1) this early attempt at incorporating a supply-side has been somewhat discredited by contemporary commentators [Stoneman (1986)]. The first paper to incorporate a supply-side in combination with a probit-type

demand side is that of Stoneman and Ireland (1983). The demand-side is modelled in a similar fashion to that proposed in David (1969, 1975) and Davies (1979) in that a firm will be an adopter of a new technology in time t if and only if its size S_{it} exceeds some critical level S_t^c . Critical firm size is assumed to be defined by the following ratio of input prices:

$$S_t^c \equiv hp_t/w_t \quad (2.28)$$

where $w_t = w_0\gamma(t)$, h is a constant, p_t is the price of the new technology and w_t is an index of cost of other inputs in the using industry both measured in time t . Assuming that each firm buys only one unit of the capital good (with no replacements) and that there are a fixed number of potential adopters, then specifying a (continuous) distribution for S_{it} then enables the derivation of the inverse stock demand curve faced by suppliers. This relates the stock of the new capital good demanded, x , to the current price of that capital good (assumed to be declining in x). If $\gamma(t) > 0$, then the new technology becomes more attractive overtime as costs of other inputs rise yielding the prospects of greater cost savings from increased efficiency. In this case, the inverse demand function will shift upwards over time.

The suppliers of capital goods are assumed to be fixed in number, identical and aim to maximise their discounted stream of profits subject to: the industry demand function (known to all suppliers); the initial condition $\dot{x}(0) = 0$; the condition that $\dot{x} = q$, where \dot{x} is the derivative of x with respect to time and q is the current output of capital good suppliers; the reaction of rival producers and cost conditions. Learning effects within the capital producing industry are subsumed within the cost function of producers in the capital goods industry. For producer i with rivals j , Stoneman and Ireland (1983) specify the general cost function, C , as follows:

$$C_i = C(x_i, x_j, q_i, t) \quad (2.29)$$

where x_i and x_j are cumulative outputs of the firm and its rivals respectively and $x = x_i + x_j$. Learning effects are then modelled in a similar fashion to Arrow (1962b) in that unit costs fall with cumulative output. From (2.28) this implies that $C_{xi} \leq 0$. It is further assumed that there are no learning 'spillovers' so that firms can appropriate their own learning effects. This implies that $C_{ji} = 0$. Finally, it is assumed that costs fall over time, $C_t \leq 0$, and that costs increase with current output, $C_{qi} > 0$.²⁵

Using this framework outlined above, Stoneman and Ireland consider the diffusion process under two different market structures in the supply industry: monopoly and oligopoly. The diffusion path in both cases traces out the movement of q over time towards that value, q^* , where profits become zero and beyond which q becomes zero.

Under monopoly conditions it is assumed that $C_t = 0$ and $\gamma(t) = 0$ so that the impetus behind diffusion are purely endogenous factors (i.e. costs are time-invariant). Along this path firms are assumed to be maximising profits over an infinite time horizon. Stoneman and Ireland show that the nature of the diffusion path thus defined depends crucially on assumptions pertaining to the value of the discount rate and the shape of the profit function. In their model, if the discount rate is set equal to zero for all t (i.e. there is no 'impatience') then the optimal path involves production at minimum average cost for all t and no sigmoid diffusion curve obtains. If the discount rate is positive, however, the optimal path will be sigmoid if the profit function is non-monotonic for a given q - at first increasing and then decreasing and obtaining a maximum for some positive x .

In contrast, under oligopolistic supply conditions it is assumed that demand is growing over time such that $\gamma(t) = e^{\lambda t}$ and that costs are related to time as well as to accumulated output. Firms in the supply industry are assumed to learn only from their own accumulated output and that costs are proportional to output. This implies the following cost function for firm i :

²⁵ In the case of the cost function the time subscript indicates the (partial) derivative of C with respect to

$$C_i = C(x_i, x_j, q_i, t) \equiv c(x_i, t)q_i \quad (2.30)$$

As shown by Stoneman and Ireland, the form of the cost function expressed in (2.30) implies that only 'steady-state' solutions [see Dixit (1990) and Lambert (1990)] can be obtained to the optimisation problem. Assumptions pertaining to price and quantity setting follow Spence (1981) and employ the open-loop equilibrium concept whereby each firm selects its optimal path given the paths of its competitors which are also optimal and correctly anticipated by the individual firm. A number of steady state solutions are shown to exist depending on assumptions concerning the value of the discount rate, the growth in demand parameter, λ , and the shape of costs over time. For the case of $c_t = c_{xi} = 0$, $\lambda = 0$ and $r > 0$, for example, the price along the steady state path is shown to be constant and that adoption is greater for all t the larger the number of producers or users, and prices and price cost margins are lower the larger the number of producers. The speed of diffusion is, however, shown to be invariant to the number of producers and buyers. Thus, the model predicts that monopolisation of the supply side will reduce the level of use of a new technology for all t .

Ireland and Stoneman (1986) retain a similar probit-type demand structure as contained in Stoneman and Ireland (1983), but extends the analysis by considering the role of price and technology expectations as outlined in Rosenberg (1976a, 1982). Their paper shows that if buyers of the new technology are risk neutral and profit maximising then the criteria for adoption in time t depends on two necessary and sufficient conditions being satisfied [of which only the former is considered in Stoneman and Ireland (1983)]: the profitability and arbitrage conditions. The former stipulates that for adoption to occur in time t that the new technology is profitable for the x th ranked buyer. The second stipulates that its profitability is not expected to increase over time in the sense that it is believed that the purchase price will fall by less than the flow of benefits. In addition, both buyers and sellers are assumed to hold expectations on the date of obsolescence of the new technology, each agent (buyer or seller) holding the same subjective probability. These expectations are incorporated into the model by augmenting the discount rate by a term adjusted for the hazard [Kalbfleisch and Prentice

time.

(1980)] of obsolescence. The supply industry is assumed to be an n -firm symmetric oligopoly (of which $n = 1$ is a monopoly and $n \rightarrow \infty$ is perfect competition are the special cases). Firms are assumed to be quantity setters with Cournot conjectures [see, for example, Kreps (1990)] who know the demand regime and maximise expected profits given the behaviour of other firms. Unlike Stoneman and Ireland (1983), costs of production on the supply-side are assumed to fall exogenously over time and no formal learning effects are introduced. Diffusion then proceeds by firms reducing prices over time that increases use by movements down the reservation price distribution over potential users.

The main aim of Stoneman and Ireland (1986) is to explore how the nature of the buyer's expectation regime will affect the diffusion path. Their results indicate four main conclusions. First, for a given number of suppliers (for $2 \leq n \leq \infty$) diffusion will be faster if buyers have perfect foresight on prices rather than hold myopic expectations.²⁶ Second, with a given expectations regime the greater the number of suppliers the faster is diffusion. Third, perfect competition on the supply-side with buyers having perfect foresight yields the same diffusion curve as a monopolist supplier combined with buyers who are myopic. Fourth, under both myopic and perfect foresight a higher expectation of obsolescence reduces usage over the entire diffusion path. Stoneman and Ireland also explores the welfare implications of different expectations regimes and most importantly demonstrates that if the industry supply of the technology is competitive the slower diffusion implied by perfect foresight is the optimal path. Thus in this model faster diffusion is not necessarily desirable from a welfare perspective.

David and Olsen (1986) develop a diffusion model that has also attempted to incorporate expectations within a demand and supply framework. The model assumes that both the capital-producing and capital-using sectors are perfectly competitive and that expectations are based upon perfect foresight held by profit maximising firms who understand the true structure of the economy. The structure of the demand side is a

²⁶ Adaptive expectations are shown to imply non-binding constraints on the diffusion process if the price of technology falls monotonically over time.

probit-type one and firms are assumed to have access to costless information pertaining to the new technology. The model extends earlier models in two main ways. First, it allows the distribution of the critical variable that determines the proportion of adopters to change through the process of diffusion.²⁷ This is done by making the stock of 'experience' (measured by the stock of cumulated output) proportional to the critical variable. Second, it introduces the possibility of capacity constraints in the supply industry. Diffusion is shown to follow a so-called 'perfect foresight equilibrium diffusion path' on which decisions predicated upon them must validate the expectations by issuing a consistent set of outcomes. The subsequent shape of this diffusion path is shown to be dependent on the nature of distribution function for the critical variable and the profit and learning functions. The main conclusion from David and Olsen is that the diffusion of a new technology may not take-off because of market failure. This market failure occurs because agents cannot completely appropriate learning-by-doing effects. If rental costs are above gross profit increments for small values of the critical variable then no equilibrium path can be defined. David and Olsen use this outcome to argue that in the case of an infant industry, for example, policy-makers need to ensure that a certain level of firms need to adopt before the diffusion process is self-sustaining. More recent developments in the literature have attempted to be more explicit in modelling improvements in technology overtime. These improvements are assumed to originate on the supply-side. Stoneman (1989) and Gruber (1990), for example, model the diffusion process incorporating vertical product differentiation. In Stoneman (1986), the link back to the R&D spending of capital goods suppliers is explored as an alternative way of addressing the same problem.

Another recent aspect is the consideration of horizontal product differentiation. Stoneman (1990b) has argued that as the variety of products incorporating new technology increases the number of firms acquiring that technology increases. By incorporating a supply side, Stoneman shows that it is possible to make the extent of product variety endogenous. The next step from discussions of product variety and product quality is product compatibility and standards. There is now a growing literature on the economics of standardisation [see David and Greenstein (1990) for a

²⁷ Although the exact definition of the critical variable is left unspecified.

recent review], although as yet this literature has had little impact on the analysis of diffusion. As explored by Stoneman (1990b, 1991) standardisation and compatibility issues will particularly be relevant for those technologies that have joint hardware and software inputs. These issues are also relevant for so-called 'network technologies' [Economides (1996)], such as ATMs, although discussion of these is left until Section 2.3.2.4.

Overall, the theoretical literature indicates that the inclusion of the supply-side (unsurprisingly) changes the nature of the diffusion process and that its incorporation is a pre-requisite for any discussion of the welfare optimal path. The models discussed above do, however, suffer from a number of limitations. First, they assume that markets always clear. There is empirical evidence to suggest [Stoneman (1976) and Freeman and Soete (1997)] that the price of new technology will not always clear the market and orders of inventories may build up. Second, except for the paper by Glaister (1974), the possible use of advertising and other forms of non-price competition by new technology suppliers tends to be ignored. Third, all the models constructed to date have assumed that over the diffusion period the number of suppliers is constant. The empirical work presented in Gort and Klepper (1982) suggests that this is not a reflection of reality. Finally, as stated by Ireland and Stoneman (1986) and David and Olsen (1986) as soon as the assumption of a given output price is relaxed, firm's strategic behaviour needs to be arbitrarily ruled out by an *ad hoc* hypothesis, otherwise firms which are ranked at the low levels of the critical variables (small firms, for example) could be induced to anticipate their adoption dates.²⁸ The economics literature has not really addressed the second and third of these critiques and as noted by Karshenas and Stoneman (1995) this paucity in the literature provides an area for future research. The first critique has to some extent been met by evolutionary models (see Section 2.3.2.6) and the latter critique by so-called stock and rank effect models. The discussion turns to the latter of these models.

²⁸ This point is particularly relevant for the nature of learning in the models of David and Olsen (1986) and Ireland and Stoneman (1986). In these models all firms in the capital-producing sector are assumed to benefit equally from their collective experience. If learning can be appropriated completely, however, then this would arguably require the admission of the possibilities of strategic behaviour amongst suppliers who would seek to pre-empt other firms' opportunities for learning.

2.3.2.2 Stock Effect Models

The defining characteristic of this class of diffusion models is that strategic pre-emptive behaviour is brought to the forefront of the diffusion process by incorporating the reaction of rival firms behaviour within a firms' adoption decision rule. This feature represents a major dichotomy between these models and the probit-type ones discussed earlier, the latter of which assume that the flow of benefits to adoption are independent of the adoption decisions of other potential adopters or rivals.²⁹ Stock effect models acquire their name from the assumption that the benefits to the marginal adopter from acquisition decreases as the number of previous adopters increase [Karshenas and Stoneman (1993)].

As stated by Beath *et al* (1995), as soon as the likely reactions of rival firms are deemed significant then two main effects come into play.³⁰ First, there is the so-called 'rent-grabbing' effect associated with being a first-mover. If a firm is first to adopt a new cost reducing technology its profits may increase because it will obtain a greater market share. This may cause firms to adopt sooner than would be optimal from a welfare perspective [Tirole (1988)].³¹ This effect, however, has to be traded-off against realising possible declines in the cost of adoption due to learning-by-doing effects in the capital-producing industry. Second, if there exists short information lags a firm can observe the actions of its rival and react to this. This gives rise to the possibility of strategic pre-emptive adoption. This implies that firms will generally adopt sooner than they would if their rivals' adoption dates were fixed (as determined by the distribution of the critical variable in probit-type models for example). The motivation for adoption is, therefore, not simply to realise the cost-reducing benefits of adoption *per se* but to prevent or delay adoption by rival firms.

²⁹ As noted by Tirole (1988), these strategic models also represent a break with Arrow's (1962a) model of the incentive mechanisms initiating innovative activity under different market structures which ignores the possibility of such behaviour.

³⁰ In the case of a monopoly buyer these following effects disappear as the firm will wait until the increase in the profit flow just equalled the marginal cost of adopting [Fudenberg and Tirole (1985)].

³¹ This implication of 'rent-grabbing' behaviour is particularly pertinent in patent race models. See Crampes (1991) for further discussion.

Given the prominence to the possibility of strategic behaviour to occur when the assumption of independent benefits is relaxed, it is unsurprising that the consequences for the diffusion process has been examined by economists using exclusively game-theoretic approaches. Such an approach has been most rigorously formulated in the seminal work of Reinganum (1981a, 1981b, 1989).

In Reinganum (1981a), diffusion is modelled as a non-cooperative duopoly game between *ex ante* identical firms under the assumption of complete certainty and who pre-commit themselves to adoption dates [see Kreps and Wilson (1982)].³² The major contribution of this paper is to show that even in the case of identical firms and complete certainty, adoption is not simultaneous in the sense that the Nash equilibrium in pure strategies is asymmetric. The model has the following elements. At time zero a new cost-reducing technology becomes available. If a firm has not adopted it when m others have, its profit flow is given by $\pi_0(m)$. If, however, the firm is one of the m to have adopted it, its profit flow is $\pi_1(m)$. Furthermore, it is assumed that T_i is firm i 's adoption date and $c(t)$ the present value cost of implementing the adoption of the new technology over the adjustment period. For a duopoly, the pay-off functions are specified as follows, where $V_i(T_i, T_j)$ represents the present value of profits from adopting at T_i when the other firm had adopted at T_j :

$${}_1V(T_1, T_2) = \int_{T_0}^{T_1} \pi_0(0)e^{-rt} dt + \int_{T_1}^{T_2} \pi_1(1)e^{-rt} dt + \int_{T_2}^{\infty} \pi_1 e^{-rt} - c(T_1) \quad (2.31)$$

$${}_2V(T_2, T_1) = \int_{T_0}^{T_1} \pi_0(0)e^{-rt} dt + \int_{T_1}^{T_2} \pi_0(1)e^{-rt} dt + \int_{T_2}^{\infty} \pi_1(2)e^{-rt} - c(T_2) \quad (2.32)$$

There are two assumptions concerning the nature of the pay-off functions contained in expression (2.31) and (2.32) that are crucial to the model. Firstly, the $\pi(m)$'s are decreasing in m and the profit gain from adoption, $(\pi(m) - \pi(m-1))$, is also decreasing in m . Secondly, the present value of the cost of adoption will decline at first but it is

³² As noted by Fudenberg and Tirole (1985) this assumption is equivalent to assuming infinite information lags.

assumed that $\lim_{t \rightarrow \infty} c'(t) > 0$ so that delaying adoption eventually leads to increased costs. Without this assumption firms may postpone adoption forever.

Reinganum shows that this basic framework yields a pre-commitment equilibrium of adoption times, $\{T^*\}$, in which one firm adopts relatively early with the other adopting at a relatively late date so long as the net value of being first is positive. Moreover, in this equilibrium the pay-offs decline monotonically with the order of adoption.

In Reinganum (1981b) the basic model of Reinganum (1981a) is extended for the n firm case. It is shown that technically there is $n!$ possible Nash equilibria, each firm adopting in sequence. It is also shown that under the assumption of linear demand functions and constant marginal cost that increases in the number of potential adopter causes most firms to delay adoption as each firm captures less of the post-adoption market share and therefore has less incentive to adopt. Thus, the model predicts that the greater concentration in the user industry increases the speed of diffusion.

Quirnbach (1986) extends the model of Reinganum (1981b) and shows that the result that diffusion proceeds through sequential adoption does not depend on the existence of strategic behaviour by firms. What does matter, however, is the nature of the patterns of incremental benefits and adoption costs over time. In essence, Quirnbach shows that if there are decreasing incremental benefits and decreasing adoption costs in the number of adopters in a non-cooperative game then strategic behaviour is not necessary for adoption to be sequential. Quirnbach also compares the diffusion path in the Reinganum model with three different regimes: a monopoly seller who holds a patent on the technology; a joint venture on the buyer's side who act to maximise industry profits and a social welfare maximising path. In general, the results indicate that diffusion in a non-cooperative environment is faster than the social optimum and that seller-side market power accelerates early adoption but that buyer-side market power retards diffusion. The last outcome is the reverse of the outcome of the Reinganum model. The reason for this difference is that in a more concentrated user- industry, where firms have larger market shares, a given cost reduction is worth more. In Reinganum's framework, however, firms continue to choose adoption dates in a non-co-operative fashion. Thus, greater concentration increases a firm's incentive to adopt, while

bringing no greater recognition of the harm the adoption imposes on rivals. A joint venture, however, internalises these externalities whilst leaving market shares unchanged. Ultimately, this reduces incentives to adopt. The main contribution of Quirnbach's paper is, however, to show that even within a game-theoretic approach the incorporation of a supply-side alters the diffusion path vis-à-vis a framework that ignores it.

The models of Reinganum (1981a, 1981b) have, however, been severely criticised by Fudenberg and Tirole (1985) who argue that her models rule out strategic behaviour by assumption of pre-commitment adoption dates which are assumed to be exogenously determined. Fudenberg and Tirole offer an alternative game-theoretic model of technology diffusion that does not rely on the existence of pre-commitment dates. This paper is discussed in Section 2.3.2.3.

2.3.2.3 Order Effect Models

This set of models assume that the gross returns to a firm adopting a new technology depends upon its position in the order of adoption, with high-order adopters achieving a greater return than lower-order adopters. In general, order effect models are made operational by arguing that the firm's adoption decision will take into account how waiting and thus moving down the adoption order will affect its profits. For any given cost of acquisition it will be profitable only for firms down some point in the order of adoption to actually adopt. This mechanism is then assumed to determine the number of adopters. In addition, the cost of acquisition is assumed to fall over time and as it does so the number of adopters increases. This then maps out the diffusion path [Karshenans and Stoneman (1993, 1995)].

The order effect has been rationalised on two grounds. First, at a general level in the paper by Ireland and Stoneman (1985) who argue that early adopters may, for example, obtain prime geographic sites or pre-empt the pool of skilled labour. Second, on a more succinct level by the game-theoretic approach in Fudenberg and Tirole (1985).

The paper by Fudenberg and Tirole (1985) extends the models of Reinganum (1981a, 1981b) by relaxing the assumption that firms' adoption dates are pre-committed and instead, assuming that firms can respond immediately to their rivals' decisions; this is, adoption is perfectly observable and instantaneous.³³ This allows for the possibility of pre-emptive behaviour in which one firm, observing the adoption of its rival, will attempt to adopt at an earlier time and thus gain a higher pay-off. For the case of a duopoly, Fudenberg and Tirole show that a perfect equilibrium [see, for example, Kreps (1990) and Shy (1990)] will result in one firm (the 'leader') adopting earlier than another (the 'follower'), but that both achieve equal pay-offs. This outcome is derived by considering the continuation of the game at time t where one firm has already adopted the new technology and the problem for the rival firm is to choose the date $t' \geq t$ that maximises $V_2(t', t)$. Fudenberg and Tirole then consider the case for which the pay-off from adopting first, $L(T_1^*) > M(T)$ is greater than joint adoption ('Case A'), where '*' signifies a pre-commitment equilibrium time. Both firms know that $L(T_1^*)$ is the maximum pay-off and, thus, each will want to pre-empt its rival at some $(T_1^* - \varepsilon)$, where ε is a small positive integer. It can then be shown that a unique equilibrium will exist in which one firm adopts at T_1 and the other at T_2^* , with $T_1 > T_2^*$. As adoption is staggered so a diffusion path is obtained. Moreover, it is shown that rents are equalised for both firms at these adoption dates.

Fudenberg and Tirole extend the basic duopoly model for the case of an oligopoly. In this case it is shown that equilibrium profits are not necessarily equalised as in the case of a duopoly. This occurs because in cases where there are more than two firm's pay-offs additionally have to be equated along subgames and not only along the equilibrium path [Shy (1996)]. With unequal continuation pay-offs, different firms will then have different gains from moving first and so the threat of pre-emption need not equate the equilibrium pay-offs.

³³ As noted by Quirmbach (1986) this can be interpreted as an open-loop equilibrium.

As noted by Sadanand (1989) and Mariotti (1989, 1992), however, diffusion may occur within a game-theoretic approach without resorting to a stock or order effect mechanism. In the paper by Sadanand, a two-period duopoly model of adoption is developed in which symmetric, risk-neutral firms are confronted with an exogenously produced process technology that, if adopted, produces an uncertain reduction in marginal costs. Firms are assumed to make production and adoption decisions simultaneously. Strategic behaviour occurs in this model because both firms have to trade-off the advantages of becoming a Stackelberg leader [Tirole (1988)] if they adopt first with the chance that the new technology, acquired at a cost, is unsuccessful. It is shown that both mixed strategy and pure strategy equilibria result in staggered adoption (i.e. a diffusion process exists) with the possibility that a 'bad' innovation can be adopted. In the papers by Mariotti the new technology is either 'good' or 'bad' and the expected returns to adoption are measured by a von Neumann-Morgenstern expected utility function. The true nature of the technology is only revealed after one firm has adopted. It is shown that if there exists a degree of informational externalities then strategic behaviour by firms can imply that the technology is not diffused at all. This occurs when the firms' balance exactly the incentive to wait provided by the possibility of information disclosure and the incentive to adopt provided by the discount factor and by the undesirability of never adopting the technology.

Papers by Vickers (1986) and Riordan and Salant (1994) have indicated that under certain conditions a game-theoretic approach can produce a new technology adoption pattern which displays 'increasing dominance' [Beath *et al* (1995)]. This implies that a single firm makes all technology adoptions' as the game proceeds. The paper by Riordan and Salant explores the perfect subgame equilibrium of a continuous duopoly game in which firms choose different vintages of a non-drastic technology³⁴ (assumed to be improving over time) at a constant cost. Pay-off functions for firms are assumed to be increasing in its own vintages, but decreasing in its rivals. It is shown that if the product market is characterised by pure Bertrand competition that an equilibrium adoption pattern has the same firm making all of the adoptions at dates such that the cost of each adoption exactly dissipates the rents from the new technology. This is

because the leading firm has more to lose from the follower adopting than the lagging firm has to gain. Therefore, the leader will always pre-empt the follower. In Vickers (1986) a similar outcome is obtained in a sequential bidding model.

A recent paper by Saracho (1997) has explored the strategy of a monopolist (patentee) who produces a durable capital innovation at some cost to buyers who act in a strategic manner. The paper shows that adoption levels are greatest when the new technology is sold by means of an auction.

The models of Vickers (1986) and Saracho (1997) are only a short-step away from game-theoretic 'waiting models' of R&D rivalry. Discussion of these models lie outside the ambit of this chapter and reviews of this literature can be found in Dasgupta (1988), Reinganum (1984, 1985a, 1989), Bridges *et al* (1991) and in Beath *et al* (1995).

Although bringing the importance of strategic elements to the fore of the diffusion literature, both stock and order models may be criticised on two grounds. First, they do not provide clear, testable empirical implications because they analyse behaviour in highly stylised and counterfactual settings. Second, the results of these models depend upon typically unverifiable assumptions concerning the identity of decision variables and sequence of moves [Cohen (1995)]. As will be shown in Chapter 3, the possible existence of stock and order effects in the diffusion of new technology has been not subjected to the same degree of empirical testing as rank effects have been.

2.3.2.4 Network Effect Models

An important development in the industrial economics literature in the last decade has been the study of so-called network industries, such as telecommunications, transport and money transmission services and their implications for, *inter alia*, firm strategy, technology adoption, predatory pricing and product pre-announcements [Economides (1996)]. The theoretical study of these industries by economists started with the

³⁴ As defined in Arrow (1962a).

seminal work of Katz and Shapiro (1985, 1986) and Farrell and Saloner (1986). The principal economic characteristic of network industries is that the utility of a user or adopter of the network derives from the consumption of the good or services that the network produces *increases* with the number of other agents consuming the good or service [Katz and Shapiro (1985)]. This characteristic reverses the traditional view of technology adoption (contained in the stock and order effects models for example) that returns to adoption are decreasing. Rather, for network industries returns are increasing and, thus, display positive consumption externalities. These positive externalities may be direct (resulting from a direct physical effect of the number of users on the quality of the good) or indirect (resulting from, for example, increases in the amount and variety of software and post-purchase services)

The basic source for these positive externalities is the complementarity and compatibility between the links and nodes of a network [Economides (1996)]. In the game-theoretic models of Katz and Shapiro (1986) and Farrell and Saloner (1986), the implications for technology adoption arise in a framework in which two or more technologies compete. In these models, certain technologies can get locked-in and dominate others. Moreover, there may exist 'excess inertia' in adoption decisions in that the move to a new, superior technology is not forthcoming because a first-mover cannot fully appropriate the returns from the network technology. Papers in this mould are those of Farrell and Saloner (1988, 1992). Lissoni and Metcalfe have interpreted these developments in the literature as representing a substitution of a narrow definition of innovation as a piece of embodied machinery with a much broader definition of technology as a set of interrelated hardware pieces, software packages and human skills. Thus, problems of compatibility and standardisation are emerging as key themes in the networks literature [see, for example, Farrell and Saloner (1992), Economides (1993), Economides and Lehr (1995) and Economides (1996)].

Carlton and Kalmer (1983), Vesela (1994), Matutes and Padilla (1994) and Saloner and Shepard (1995) indicate that network externalities may particularly be relevant for the adoption of ATM technology by financial institutions. These externalities principally arise in the case of ATMs from the increased benefit (increased convenience) an ATM

user obtains from increases in the number of ATMs from which the cardholder can access their account. Thus, the value of the ATM network increases in the number of locations it includes. In the paper by Saloner and Shepard (1995) a model of ATM adoption is developed in which a bank is assumed to be able to increase its network size by adding more ATM locations. The implication of this assumption is that banks expecting to have a larger number of locations in equilibrium adopt sooner. The results from their empirical study of adoption diffusion in the US banking sector are outlined in Chapter 3.

A key focus on this thesis is to explore whether there has been significant network effects in the diffusion of ATMs for the UK experience. Thus, an extensive discussion of network effects models and their implications for new technology adoption and diffusion are delayed until Chapter 6 where a formal theoretical model of ATM adoption is presented and tested for the case of the UK financial sector.

2.3.2.5 New Investment Theory

The new investment theory centres around the seminal work of Dixit (1992, 1993), Pindyck (1993) and Dixit and Pindyck (1994) who develop an options-type approach to investment decisions as an alternative to the conventional net present value rule [see, for example, Primrose (1991)].³⁵ As noted by Dixit and Pindyck (1994) most real world investments by firms do not satisfy the implicit assumptions of the standard net present value rule; namely that either the investment is reversible (that is, the investment can be somehow undone and investments recovered), or if irreversible, the decision has to be taken immediately and no allowance for waiting is included. According to Dixit and Pindyck most real world investments by firms do not conform to these strict assumptions and, alternatively, argue that most investments are in fact irreversible. Their alternative investment rule makes the opportunity to invest analogous to holding a financial call option. This alternative rule implies that the firm has the right but *not* the obligation to buy an asset at some future time of its choosing. Thus, when a firm makes

an irreversible investment, it exercises its option to invest. It is this last option value that is ignored in the standard net present rule and, as argued by Dixit and Pindyck, an element that must be included as part of the true 'cost' of investment. In general, this option value can be shown to increase with the sunk costs of investment and with the degree of uncertainty over the future price of the technology. Optimal investment timing and investment (or adoption) decisions can then be conceptualised as optimal sequential decisions under uncertainty and are solved by dynamic programming methods [see Dixit (1990)].

To date, however, economists have not incorporated these developments within a model of innovation diffusion. A recent paper by Farzin *et al* (1998), however, indicates the possible future research agenda. In their basic model, Farzin *et al*, develop a model in which a single, profit maximising competitive firm is faced with an exogenous stochastic innovation process and is only allowed to 'switch' to the new technology once. Thus, adoption is irreversible. Uncertainty in the investment decision is assumed pertain to both the arrival date and improvements (assumed exogenous) in the new technology over time. The problem facing the firm is shown to be an optimal stopping one and the solution to this problem is solved within a dynamic optimisation framework with an infinite horizon.

Their results indicate that the alternative view of investment proposed by Dixit and Pindyck (1994) leads to a slower pace of adoption vis-à-vis the standard net present value rule. Moreover, the comparative statics of the model indicate that a firm's optimal timing is quicker the higher the discount rate (in contradiction of the net present value rule) and is slower for firms who already have a high degree of technological efficiency. Although this analysis is not extended to the consideration of the diffusion process their results suggest that the diffusion paths of conventional diffusion models could be altered using their approach. This represents a potential opportunity for future research.

³⁵ This is the profitability condition in Ireland and Stoneman (1986).

2.3.2.6 Evolutionary Models

This is the most recent development in the theoretical diffusion literature and although there are a number of different and distinct evolutionary approaches [see the debate in Dosi (1988), Witt (1991), Nelson (1996), Andersen (1997), Hodgson (1993, 1998), and Gourlay *et al* (1998d)] they arguably all have their origins in the early work of Schumpeter (1947), Alchain (1950) and Salter (1966). The defining characteristic of the evolutionary approach to diffusion is its rejection of the neo-classical representation of technology as a choice set facing profit maximising firms who have perfect information and its view of diffusion as a *selection process* [Amendola and Gaffard (1988)]. In contrast to the neo-classical paradigm, firms are assumed to be constrained by bounded rationality [Sugden (1991)] and act in a behaviouralist manner in which decisions are routine-based which accumulate through experience rather than any foresight of the future [Dosi and Orsenigo (1988)]. Consequently, theories that firms' hold about their environment and the process by which these theories are generated and revised are crucial to explaining the way in which firms innovate [Swann (1992) and Metcalfe and Boden (1992)]. This explicit behaviouralist approach emphasises the organisational aspects of the firm and the links between the propensity to innovate and adopt and the productive and marketing activities of the firm [Metcalfe (1995)].

The evolutionary approach to diffusion also rejects the conventional view of technology as a single piece of embodied machinery. Instead, it argued that a specific technology is a 'design configuration' with its own distinct framework of concepts, ideas and relationships that develop over time. This design configuration follows an identifiable 'paradigm' or logic that defines the possible path of technological improvement. Freeman (1984) and Perez (1985) have labelled this logical path of technological improvement as a 'technology trajectory'. Thus, diffusion is viewed an evolving set of design configurations. Incremental innovations are conceptualised as changes within a definite architecture, while radical innovations are conceptualised as involving changes to that architecture and the creation of new knowledge bases in science [Metcalfe

(1995)].³⁶ Diffusion is then modelled as a selection process involving inter-firm variety that arises from different managerial and organisational structures.

There are two formal types of models that conform to the evolutionary approach. The first approach centres on the papers by Nelson and Winter (1982), Silverberg (1988, 1991) and Silverberg *et al* (1988). These papers conceptualise the diffusion process as involving the co-existence of firms with different innovative capabilities and capabilities to invest. In these models, however, analytical solutions are computationally difficult to obtain and therefore computer simulations are frequently employed to portray the diffusion process. In Silverberg *et al* (1988), for example, a set of firms are assumed to exist at the beginning of the process, each firm being characterised by different capabilities. The innovation being diffused generates public spillovers (productivity gains for both adopters and non-adopters) and private productivity gains for adopters. The private and public productivity gains are assumed to change over time and the price of the technology is assumed to decline. By changing the parameters of the model, different diffusion paths can be achieved.

The second approach centres on the papers by Soete and Turner (1984), Metcalfe (1981, 1995) and Cameron and Metcalfe (1988). Lissoni and Metcalfe (1994) have given this approach. In the model of Metcalfe (1988), for example, diffusion of a new technology is the result of competition with an old technology. Two situations are then possible: either the old technology is completely eliminated, or the two technologies share the market for the productive service.³⁷ The outcome depends on the demand curve for the productive service, the price elasticity of the two supply curves and the qualitative superiority of the new technology. The price of the new technology adjusts to keep equilibrium between the growth in industry productive capacity (assumed to be proportional to the profitability of new technology) and the demand for the output of the new technology. The resulting diffusion curve (derived for the total output obtained from the new technology) is a logistic curve, although this result does depend crucially upon the learning mechanism of consumers.

³⁶ See Chapter 4 of this thesis for an application of this evolutionary framework to the analysis of changes in ATM technology and the distinction between first and second-generation technology.

³⁷ Given that technology in this model is modelled as a productivity distribution, Gourlay *et al* (1998a) has given this model an interpretation on the lines of Salter's (1966) model.

Although the evolutionary approach arguably offers a novel view of the diffusion process as a process of competitive change and is able to explain the co-existence of old and new technologies, the major limitation of this approach is that it arguably represents a *post hoc* explanation of diffusion. These models are therefore extremely difficult to verify and test at an empirical level.

2.4 Concluding Remarks

There were two main aims to this chapter. The first was to outline the main definitional issues that have emerged in the diffusion literature and distinguish between the various dimensions of innovation diffusion. The second, and most important, was to trace the development of formal economic theories of innovation diffusion from the early contributions to the contemporary literature.

It was found that the theoretical literature has attempted to provide answers to two main questions. Firstly, why some firms are early adopters and others late and, secondly, what are the key firm-specific, market-specific and technology-specific characteristics that determine the resulting time path of diffusion. A distinction was made between neo-classical and evolutionary approaches to providing answers to these questions. The former approaches have, in the main, taken as their starting point the micro modelling of adoption decisions. In contrast, the latter approach has stressed how ordered patterns of diffusion may emerge from apparently irrational individual firm behaviour. It was noted, however, that the evolutionary approach is weakened by its *post hoc* nature and the difficulty in verifying and testing its main implications for the diffusion process.

Finally, it was noted that a potentially significant aspect of ATM technology is its inherent network characteristics and, consequently, any serious empirical investigation should consider the implications of these effects for the diffusion process.

CHAPTER 3

EMPIRICAL DIFFUSION MODELS: A LITERATURE REVIEW

3.1 Introduction

There are three main aims of this chapter. First, to provide a review and critique of the main approaches that have been used to empirically model the inter-firm diffusion of process innovations. Second, to report the main research findings of these studies. Third, to highlight, where appropriate the contribution to the analysis of inter-firm diffusion made by this thesis. The main emphasis throughout this chapter is on inter-firm studies of ATM diffusion, but given the paucity of research in this area the main findings from early aggregate diffusion models and results from industrial innovation studies are also provided for completeness, although in a more abridged format. Economy-wide models of process and product innovation diffusion are not discussed in this chapter¹ and summaries of these models can be found in Stoneman (1983), Mahajan *et al* (1990) and Karshenas and Stoneman (1995).

In order to trace the progress and to include the latest developments in empirical inter-diffusion models, the chapter will take the following form. Early contributions to the literature will be discussed first in Section 3.2.1, which is then followed by Section 3.2.2, which discusses contemporary models. Section 3.3 summarises the principal findings.

3.2 The Literature

This section begins with a survey of the early empirical literature that has attempted to model the diffusion of a capital-embodied process innovation. This is then followed by a survey of the contemporary literature emphasising, in particular, the recent and

growing literature that has applied duration models to inter-firm diffusion. The aim of this section is to identify and critically appraise the main modelling approaches and techniques that have been employed in the literature and to summarise the extent to which the predictions of the theoretical models outlined in Chapter 2 have been empirically verified.

3.2.1 Early Contributions

This section reviews the early empirical literature, which is dominated by two main approaches, the two-step procedure and the time-to-adoption model. The former approach reflects the dominance of epidemic theories of innovation diffusion employed in early explanations of innovation diffusion by economists. This dominance can, arguably, be interpreted as a tacit acceptance of other disciplines (such as psychology and sociology) given the absence of any rigorous economic models. The second approach does not rely on the epidemic model, but does however, suffer from a number of internal weaknesses which mitigate against its use in contemporary work. The main focus of these early studies was on the perceived Schumpeterian hypotheses [Kamien and Schwartz (1982)] concerning the effects of firm size and market structure on adoption decisions.

3.2.1.1 Two-Step Models

The term 'two-step' is used here to refer to a class of diffusion models where the estimation procedure involves two distinct stages of estimation. The first stage typically involves the fitting of a logistic-type growth curve to diffusion data on a number of different innovations in various industries and then, in a second stage, the estimated slope parameters (usually referred to as the 'speed of diffusion') are used as the

¹ This includes the stock adjustment models of Chow (1967), Stoneman (1976) and Labson and Gooday (1994).

dependent variable to be explained in terms of the characteristics of the industries and innovations concerned.²

The two-step model was used extensively in early studies of inter-industry and inter-sectoral diffusion where the main objective was to examine what innovation-specific characteristics explained how quickly they diffused. In addition, because the two-step method necessitated cross-sectional data the model was used extensively to test theories of the nexus between market structure and the extent and speed of the diffusion process. The two-step modelling approach makes explicit use of the logistic growth curve to characterise the growth in the proportion of adopters across time. The justification for using the logistic curve comes from three theoretical approaches to the innovation diffusion process. First, from the epidemic model as outlined in Chapter 2. Second, from Mansfield's (1961, 1968) model that derives a logistic diffusion curve as the outcome of a Taylor series expansion of a hypothesised functional relationship between the proportion of 'hold-outs' (non-adopters) and a set of determining economic attributes of the innovation and the proportion of firms that have already adopted. Third, from the distinction between Group A and Group B innovations made by Davies (1979).

As indicated in Chapter 2, the justification of using the epidemic model to characterise the diffusion process is made on two grounds. Firstly, empirical research into innovation diffusion has often found that plotting the proportion of adopters against time obtains an *S*-shaped (or sigmoid) curve that can conveniently be approximated (and estimated) by the logistic curve [Griliches (1957), Romeo (1975) and Karshenas and Stoneman (1993)]. Secondly, there is an analogy between the spread of the (tacit) knowledge pertaining to the economic attributes of the innovation through informal contact and demonstration and that of a disease [Metcalf (1995)].³

² This 'two-step' estimation procedure should not be confused with the more common terminology used in the econometrics literature to refer to estimation methods for heteroscedastic regression models [see Greene (1993) for more details of this procedure].

³ There are also the behavioural and sociological justifications for using the epidemic model to represent the diffusion process. These justifications are most eloquently outlined in Ryan and Gross (1943) and Rogers (1962).

To illustrate how the two-step model is implemented empirically, recall that the simple epidemic models [as outlined by Bailey (1957) and Bain (1964)], predicts that the proportion of the population having adopted the innovation at time t will evolve according to the following relationship:

$$m_t/n = [1 + \exp(-\alpha - \beta t)]^{-1} \quad (3.1)$$

with $\alpha < 0$ and $\beta > 0$ and where α is the constant of integration. Sketching equation (3.1) against time then results in an inter-firm diffusion curve that is logistic and therefore *S*-shaped. Moreover, the logistic diffusion curve is symmetrical with a point of inflexion (implying a maximum value for the rate of increase in adoption) at the midpoint where half the population has adopted and the maximum level of adoption, n , is approached asymptotically from below.

The application of the above model to an empirical analysis of the diffusion process then proceeds through a log-transformation of equation (3.1) leading to the following estimable equation [Romeo (1977)]:⁴

$$\ln[m_t/(n - m_t)] = \alpha + \beta t \quad (3.2)$$

where the parameter β in equation (3.2) is often defined in empirical research as the 'speed of diffusion' or 'rate of imitation' [Mansfield (1968) and Romeo (1975, 1977)] and reflects the pace of the diffusion process across time.⁵ By estimating either equation (3.1) or (3.2) across innovations and industries the estimated speed of diffusion, $\hat{\beta}$, can then be used in a secondary weighted-least-squares (WLS) regression⁶

⁴ Unless specified otherwise the log-transformation of the logistic growth curve (and any other equation) discussed in this chapter (and throughout the thesis) always refers to the natural log-transformation.

⁵ The speed of diffusion, β , should not be confused with the rate of growth of diffusion, which is given by $(dm_t/dt)/m_t$, and which falls continuously over time [Davies (1979)].

⁶ This is to eliminate possible heteroscedasticity and involves weighing each observation by the inverse of the estimated standard error of the dependent variable [Saxonhouse (1976)].

on innovation and industry specific characteristics.⁷ Hence the approach is labelled 'two-step.'

As shown in Section 2.3.1.1 of Chapter 2, an alternative approach to deriving the logistic growth curve to characterise the diffusion process over time is provided by Mansfield (1961, 1968). This approach leads also leads to the familiar logistic curve previously given in equation (3.1).

The seminal and pioneering study of hybrid corn diffusion across US states and crop reporting districts for the period 1932 to 1956 by Griliches (1957) was the first application of the two-step model by an economist. Griliches's chooses this model for ease of estimation rather than any economic rationale. Additionally the employment of the logistic curve gives a convenient summary of the diffusion process in terms of the ceiling or maximum level of hybrid corn use [given by n in equation (3.1)], the rate of imitation, β and the date of first use [approximated by α in equation (3.2)]. The results obtained by Griliches indicate that the dominant variables in explaining variations in the speed of diffusion are variations in the profitability of corn across states. The profitability of hybrid corn is defined as the superiority of hybrids over open pollinated corn and is measured as the average increase in yield in bushels per acre and the long-run average pre-hybrid yield of corn. Both measures are found to have a positive and significant effect on the speed of diffusion at the state level and crop reporting level.⁸ In a follow-up study of hybrid corn diffusion, Dixon (1980) argues that given the skewed nature of the diffusion process over time the Gompertz curve is a better representation (as measured by an R^2 measure of fit) of the data than the logistic.

Mansfield (1961) examines the diffusion of twelve industrial process innovations across the bituminous coal, iron and steel, brewing, and rail industries in the US. In this study, profitability of the innovation is measured by the ratio of the average actual pay-off

⁷ The estimation of (3.2) is, however, not straight forward. As noted by Kamenta (1986) the approach used in the initial estimation of (3.2) is complicated by the specification given to the error structure. If the error term enters within the bracketed left hand term of (3.2) then simple application of OLS is appropriate. If, however, the error term enters outside this bracketed term then maximum likelihood estimation should be used. Secondly, if n is unknown (the maximum number of adopters) then Oliver (1964) has suggested estimation by iterative least squares.

period for the innovation to the average pay-out period required by firms in the industry.⁹ The risk and financial commitment attached to the innovation is measured by the ratio of the initial investment in the innovation to the average total assets in the industry for the relevant period. Mansfield finds that the profitability of the innovation has a positive and significant effect on diffusion speed while that of the financial investment required in the innovation has a negative and significant effect. In a more recent study, Mansfield (1989) analyses the diffusion of industrial robots in the US and Japan using similar measures of the profitability and financial commitment associated to the innovation used in Mansfield (1961). Results obtained are identical to those in Mansfield (1961) in terms of the signs and significance attached to these variables.

Romeo (1977) studies the diffusion of numerically controlled machine tools (NCMs) across twelve different industries in the US. Results indicate that inter-industry differences in diffusion speeds are explainable in part by inter-industry characteristics. In particular he finds a positive and significant role for the year that the industry first adopts the new technology. In addition, Romeo proxies investment risk as the ratio of purchase price of the new technology to average total assets of firms. This is found to have a negative and statistically significant effect on diffusion speed. Romeo also examines the effect of industry 'competitiveness' on the speed of diffusion. Further results also suggest that more competitive markets lead to higher diffusion speeds. In addition, Romeo (1977) employs the average rate of return, measured by the reciprocal of the pay-out period, from the investment in NC machine tools in each of the twelve industries under investigation. He finds that this profitability measure has a positive and significant coefficient when included in the regression for the diffusion speed.

The two-step model has been applied to the diffusion of ATMs in the US banking sector by Hannan and McDowell (1984a) who examine the effects of market-specific characteristics on the speed of diffusion across different banking markets. As competition between US commercial banks occurs primarily within geographically limited markets [see Rhoades (1980)], this allows Hannan and McDowell to explore the

⁸ Unless otherwise specified, the significance of an estimated coefficient in this chapter denotes its statistical significance at the 10% level at least.

⁹ Mansfield defines the pay-out period as the capital cost of the innovation divided by the annual savings-annual cost [for more details of this measure see Primrose (1991)].

effects of market-specific characteristics on the diffusion process within the same industry. The Standard Metropolitan Statistical Area (SMSA) defines local banking markets.¹⁰ Hannan and McDowell estimate the speed of ATM diffusion across eighty-nine banking markets for the period 1971 to 1979 using the log-transform of (3.1) and dropping the subscript j obtain:

$$\ln\left[\frac{P_{it}}{1 - P_{it}}\right] = \alpha_i + \beta_i t \quad (3.3)$$

where i is the i th local banking market ($i = 1, \dots, 89$), t is time ($t = 1971, \dots, 1979$), P_{it} is the proportion of banks using ATMs in the i th banking market at time t and β is the speed of diffusion. In that initial year(s) where ATMs had not been adopted, P is assumed equal to 0.01.

Equation (3.3) is estimated by ordinary least squares (OLS) to obtain the estimated speed of diffusion (found to have an average value of 0.698). Hannan and McDowell then use this estimate in a secondary regression on market-specific characteristics.¹¹ Their results indicate that the coefficients on a three-firm market concentration variable, the number of years taken for ATMs to be first introduced from 1971, and a market growth variable (measured as the average growth in market deposits) are all positive and statistically different from zero. Although not explicitly including a variable measuring the profitability of the innovation as previous two-step models have done, they do include the average wage of bank employees in the market to capture the incentives of adopting a potentiality 'labour-saving' ATM (see Chapter 4 for further consideration). They find that this has a positive but insignificant effect on the speed of diffusion, thus providing only partial support to their hypothesis that there is more rapid ATM diffusion (as a result of more expeditious factor substitution between capital and labour) in those markets with relatively higher wage levels. They explore the effects of market legislation on the diffusion of ATMs by including a dummy variable taking a value of unity if the state has legal restrictions on the establishment of branches, and

¹⁰ The Standard Metropolitan Statistical Area (SMSA) is a local labour market definition and was introduced into the US census in 1950 to describe a metropolitan labour market in which 'commuting' connected cities and the surrounding areas [Rhoades (1980)].

zero otherwise. This variable is also found to be positive but insignificant. In addition, they include a measure of investment risk similar to that as Romeo (1977) defined as the ratio of purchase price of ATMs to average total assets of the banking firm. This is found to have a negative but insignificant coefficient.

The results obtained by Hannan and McDowell (1984a) suggest that in those markets experiencing ATM introduction by 1975 their diffusion has occurred more rapidly in the relatively more concentrated and growing markets, *ceteris paribus*. In addition, Hannan and McDowell find a dominant role played by the variable measuring the number of years (taken from 1971) that it takes for the first ATM to appear in the market in explaining variations in the speed of diffusion. One possible explanation of this finding is that the tacit knowledge relating to the operation of ATM technology accumulated in one market spills-over into other markets by informal contacts and the transfer of bank staff (involved in banking technology) across states and between banks.¹² This mechanism is analogous to the 'learning-by-using' concept put forward by Rosenberg (1982). Alternatively, the significance of this variable may reflect the technical improvements (and possible lower quality adjusted price) in ATM vintages over time which may have made ATMs more profitable for those banks that had delayed adoption. As shown by the rank effects models in Chapter 2, the validity of the last argument depends crucially on the subsequent movement of prices over time and how benefits are related to firm-specific characteristics in particular markets.

Hannan and McDowell's study, however, excludes those SMSA's where banks had not adopted ATMs by 1979 the regression results obtained may be biased. This occurs as a direct result of using the two-step model, which requires knowledge of the speed of diffusion and which, can only be obtained if the defined market has adopted the innovation. In addition, their use of the average wage level of bank employees as a measure of the incentives to substitute capital (ATMs) for labour is, arguably, picking up wider state level income effects [Vesala (1994)]. Spatial models of branch and ATM

¹¹ From equation (3.3) there is the implicit assumption made by Hannan and McDowell that all banks in the sample will eventually adopt ATMs. From their paper it is unclear whether this actually occurs or is simply an assumption concerning the eventually nature of the diffusion process.

¹² Rosenberg (1976b, 1982) has forcefully pointed out that many improvements in inventions and innovations take place *during* the process of diffusion itself, both as a result of learning-by-using and as a result of competition between suppliers.

location-choice [see Tirole (1988) and Shy (1996)] predict that a rise in bank customer incomes, *ceteris paribus*, leads to a rise in the opportunity cost of time (through increased transactions costs) and greater incentives for banks to provide ATMs [Salop (1979) and Vesala (1994)].

Davies's (1979) two-step model is arguably less *ad hoc* than the above models and is derived from an underlying behaviouralist approach to the adoption decision. As outlined in Chapter 2, Davies distinguishes between two groups of innovation: Group A and Group B (see Chapter 2 for the basis of this distinction). By assuming that critical-firm size is lognormally distributed, Davies is then able to show that the inter-firm diffusion curve of Group A innovations is a cumulative lognormal curve whilst that of Group B innovations is a cumulative normal curve. Both distributions are specified as follows:

$$D_t = N(\log t | \mu_D, \sigma_D^2) \quad \text{- Group A innovations (cumulative lognormal)} \quad (3.4)$$

$$D_t = N(t | \mu_D, \sigma_D^2) \quad \text{- Group B innovations (cumulative normal)} \quad (3.5)$$

where D_t is the percentage of firms adopting the innovation at time t , μ_D is the mean of the diffusion curve and σ_D^2 its variance.¹³

The Davies model is operationalised empirically by using the linear transform $D_t = N(z_t | 0,1)$, where z_t is the normal equivalent deviate of the level of diffusion (m_t/n) in year t [D'Agostino and Stephens (1986)]. This allows z_t to be written as $z_t = (t - \mu_D)/\sigma_D$. Substituting this term into (3.4) and (3.5), and assuming no cyclical factors, the following estimable equations are obtained respectively:

$$z_t = a_1 + b_1 \log t \quad \text{- Group A innovations} \quad (3.6)$$

$$z_t = a_2 + b_2 t \quad \text{- Group B innovations} \quad (3.7)$$

where $a = -\mu_D/\sigma_D$ and $b = 1/\sigma_D$ and b can be interpreted as a measure of diffusion speed.

The model is empirically applied in two steps. In the first step, Group A and Group B innovations are distinguished by the relative performance (based on a R^2 measure of fit and a Durbin-Watson measure of serial autocorrelation) of (3.6) and (3.7). In the second step, the time series estimates of the b 's obtained in step one are used in cross-sectional WLS regressions across industries on a number of industry- and innovation-specific characteristics.

Results obtained by Davies for twenty industries in the UK indicate that diffusion speed is positively and significantly related to the innovation's profitability (measured as the pay-off period obtained from trade journals), labour intensity of the adopting industry (measured by the share of value added allocated to wages) and industry growth in output. In addition, Davies finds that the diffusion speed is negatively and significantly related to the number of potential adopters (n) and the variance of the log of firm size (measured by employment).

All variables are assumed to be constant over the diffusion period and are measured at the time where half of potential adopters have adopted. Overall, the results confirm the underlying theoretical assumptions of his model (see Chapter 2) which predicts the speed of diffusion is quicker in those industries experiencing faster growth, have smaller inter-firm differences in firm-size and a smaller number of firms.

The Davies (1979) approach has been most recently been applied by Alderman and Davies (1990) and by Gruber (1998). Alderman and Davies study inter-regional differences in the diffusion of four industrial innovations across nine regional areas in the UK. Results obtained by Alderman and Davis support those obtained by Davies (1979) that diffusion speed (at the regional level) is negatively related to the variance of firm size and, in addition, are positively related to the percentage of firms in the fastest adopting industries. The paper by Gruber augments the basic model of Davies to a

¹³ The variance term in (3.4) and (3.5) is a structural parameter in the Davies model and captures the inter-firm differences in the expected pay-off and target pay-off after allowing for differences in firm-

panel data representation. This is then applied to a study of the diffusion of shuttleless looms across twelve industrial countries over the period 1977 to 1992. The results are quite poor with only real wage levels and trade liberalisation dummy variables being statistically significant.

3.2.1.2 Time-to-Adoption Models

This class of models have been developed in order to explain differences between firms in the *time* it takes for them to adopt the same innovation from the date of the first adoption. The time-to-adoption model was originally developed by Mansfield (1963) who specifies the following relationship to explain differences between firms in their 'time-to-adoption':

$$d_{ij} = Q_i H_{ij}^{a_{i2}} S_{ij}^{a_{i3}} e^{\varepsilon_{ij}} \quad (3.8)$$

where d_{ij} measures the number of years the j th firm takes to adopt the i th innovation (taken from an arbitrary date), S_{ij} is the size of the firm, H_{ij} is a measure of the profitability of its investment in the innovation and ε_{ij} is the random error which is assumed to be $NID(0, \sigma_\varepsilon)$. By taking the log-transform of equation (3.8) it is then possible to estimate the coefficients a_{i2} and a_{i3} by OLS from:

$$\ln d_{ij} = \ln Q_i + a_{i2} \ln H_{ij} + a_{i3} \ln S_{ij} + \ln \varepsilon_{ij} \quad (3.9)$$

The coefficients a_{i2} and a_{i3} in equation (3.9) can then be interpreted as the elasticities of adoption delay with respect to firm profitability and firm size respectively.¹⁴

Mansfield argues in particular that d_{ij} will be inversely related to S_{ij} (i.e. a_{i3} is negative) for three reasons. First, the costs and risks of early adoption are borne more easily by large firms. Second, because of their size, there is greater probability of large firms

size.

¹⁴ The elasticity of delay with respect to firm size, for example, is given by $(dd_{ij} / dS_{ij}) / (S_{ij} / d_{ij}) = a_{i3}$.

needing to replace old equipment at any point in time. Thus, if the innovation is embodied in new capital-equipment, large firms have greater opportunities to adopt relatively early on average. Third, because of their size, large firms by definition are likely to encompass a wider range of operating conditions than smaller firms operate in. This aspect of being 'large' may increase technological opportunities for these firms. Mansfield also argues that d_{ij} will be inversely related to S_{ij} (i.e. a_{12} be negative) for identical reasons outlined in his formulation of the two-step model.

Mansfield (1963) fits equation (3.9) to data pertaining to a cross section of firms adopting fourteen different innovations in the bituminous coal, iron and steel, brewing and railroad industries in the US. The number of years that a firm takes to adopt the innovation, d_{ij} , is measured by the date at which it adopts minus the date of the first adopter in that particular industry. Firm size, S_{ij} , is measured by physical output in the coal and brewing industries, ingot capacity in the steel industry, and freight ton-miles in the railroad industry.

Although Mansfield cannot obtain direct estimates of the innovations' profitability, H_{ij} , for each firm so he uses firm-specific characteristics that are conjectured to be highly correlated with potential profitability. For instance, Mansfield argues that profitability from adopting a continuous mining machine will be strongly influenced by the proportion of the adopter's output derived from high coal seams. This measure is used as a proxy for the profitability of the innovation. This measure, arguably, captures many other factors besides profitability and thus restricts the explanatory power of this variable in practice.

The results obtained from estimating equation (3.9) are mixed. The profitability measure, H_{ij} , although has the correct sign based on *a priori* expectations (i.e. negative) is found to be statistically significant in only two of the fourteen innovations. On the other hand, firm size, S_{ij} , is found to be consistently significant and negative. Depending on the exact specification of the equation, a_{13} varies between -0.03 and -1.53, with Mansfield's preferred equation yielding an estimate of -0.4. All other variables are found to be statistically insignificant and in the case of the firm's

profitability, output growth rate, the age of president and profit trend have the incorrect sign as expected by *a priori* expectations. This suggests that there may exist a problem of multicollinearity in the specification of equation (3.9) but Mansfield does not address this issue.

The model may be criticised because Mansfield constrains a_{i3} to be equal for all innovations while allowing the coefficient on H_{ij} (a_{i2}) to vary across innovations.¹⁵ As indicated by Davies (1979) this makes interpretation of a_{i3} extremely problematic because it makes distinguishing between the effects of inter-firm, rather than inter-industry, size effects impossible. Davies shows that this may lead to a spurious correlation between d_{ij} and S_{ij} . Consequently, the results obtained by Mansfield, in particular those relating to the relationship between firm size and d_{ij} , should be treated with caution.

Smith (1974) examines the inter-firm diffusion of shuttleless looms across five countries (Italy, Sweden, UK, USA and West Germany) and employs similar explanatory variables to that of Mansfield (1963). His results are very discouraging. He finds R^2 's typically of only 0.10 and the only significant variable is an arbitrary index reflecting the extent of firms' vertical integration.

Overall, the results from time-to-adoption models are rather discouraging, being characterised by low R^2 's and insignificant variables. In addition, the study by Smith had to face a problem not encountered by Mansfield: not all the firms in their sample had adopted at the time of investigation. Consequently observations on d_{ij} are not available for all firms in the sample considered. Smith attempts to solve this problem by allocating an arbitrary date in the future at which they adopt. This approach has two problems. First, the likely outcome of this approach is biased estimates from estimating equation (3.9) by excluding firms that do not adopt. In general the direction of the bias will depend upon three factors: size of the $\text{var}(\varepsilon_{ij})$ in equation (3.9), the arbitrary choice of the future adoption date and the date at which the explanatory variables are measured

¹⁵ This implies that equation (3.9) is fitted to the pooled data for all innovations and industries with various dummies on H_{ij} , but not for S_{ij} [Judge (1985)].

[Davies (1979)]. Second, given that empirical diffusion studies are retrospective in nature there are no economic criteria available to researchers in their choice of the future date at which firms are supposed to adopt.

3.2.2 Contemporary Economic Literature

The 1980s witnessed a revival of interest amongst empirical economists into the causes and determinants of inter-firm diffusion, at last incorporating and testing the advancements that had been made in the theoretical literature. This section reviews the contemporary empirical literature which, similar the early literature is dominated by two main approaches: discrete choice and count data models and duration models. These two approaches are arguably less *ad hoc* than their early counterparts and are characterised by the dismissal of information dissemination as the key to their derivation. The former approach has its roots in probit and rank models of diffusion where the impetus for diffusion is heterogeneity amongst the set of potential adopters. The latter approach pertains to the current state-of-art approach to inter-firm modelling and attempts to merge adoption and diffusion processes together by introducing time and time-varying variables into the modelling procedure.

3.2.2.1 Discrete Choice and Count Data Models

Discrete choice and count data models are applicable to situations where the outcome of an economic agent's decision is typically discrete and qualitative in nature with the subsequent dependent (or endogenous) variable under interest assuming discrete values [Amemiya (1981), Maddala (1983) and Greene (1993)]. The former models are applicable when the values taken by the dependent variable are used merely as a coding for some qualitative outcome often characterised by a 'yes or no' decision. The latter count data models are appropriate when the dependent variable represents *count data*, which although discrete and non-negative, is not qualitative in nature (for example, the number of innovations a firm adopts). These models have been developed because classical regression models (such as OLS), which are almost exclusively formulated for

cases where the dependent variable is continuous, are inappropriate since the residuals will be non-normal and predicted coefficients (probabilities) can be above unity [Greene (1993)].

Both these models are most applicable to the empirical modelling of inter-firm diffusion and the testing of the rank effects models discussed in Chapter 2. Recall that rank models assume that potential users of a new technology differ from each other in some important characteristic such that some firms obtain a greater gross return from the new technology than others do. In these models the interest lies mainly with firm-specific and market-specific determinants of adoption. The dependent variable in the empirical model, y , can then be assumed to have a discrete outcome: either the firm has adopted the new technology (denoted by $y = 1$) or it has not (denoted by $y = 0$). In these cases discrete choice models are appropriate. Alternatively, the interest may lie in those firm-specific characteristics that determine the number of innovations adopted up to a specified time period. In this case the dependent variable, y , will be a discrete non-negative random variable ($y = 1, 2, 3, \dots$) and the appropriate choice of model is a count data one.

It can be noted that discrete choice models can be categorised into three broad categories, each category being dependent upon the nature of the choice(s) available to the economic agent [Amemiya (1986)]. These are: firstly, univariate binary dependent models (or, more simply, binary choice models) in which a single dependent variable can only take two values; secondly, multinomial models in which a single dependent variable takes more than two discrete values; and thirdly, multivariate models that involve more than one discrete dependent variable. The empirical diffusion literature has, to date, only considered application of the first of these models. The *raison d'être* for this choice of model is that once a technological innovation is commercially available the alternative choices available to a potential adopter are limited to only two: either to adopt the innovation or not to adopt. This is clearly a simplification of the actual adoption process, which is often highly complex, and as observed from Chapter 2 may, arguably, involve specific stages.¹⁶ To a large extent, however, this simplification is inevitable for two reasons. Firstly, isolating specific stages in the adoption process is

extremely difficult given the somewhat ambiguous nature of some of the stages often cited [Karshenas and Stoneman (1995)]. Secondly, it is questionable whether specific stages can be measured accurately unless there is continuous monitoring of the potential adopter's behaviour. Even if this monitoring were feasible it would be prohibitively expensive in both pecuniary and non-pecuniary terms.

In order to relate the various binary choice models directly to the analysis of diffusion, it is useful to explicitly consider the case of a cross-section of a set of n potential adopters ($i = 1, \dots, n$) at a specific point in time. It is assumed that a firm will either be in a state of adoption ($y = 1$) or non-adoption ($y = 0$) in the period in which the diffusion study is undertaken. In an inter-firm study the focus is on a set of K , firm-specific characteristics, such as firm size and profitability, which are gathered in a $1 \times K$ vector, $X_i = (X_{1i}, \dots, X_{Ki})$, which differ across individual firms. The basic inter-firm model can then be specified as [Judge (1985)]:

$$y_i = X_i \beta + u_i \quad \text{with} \quad y_i = \begin{cases} 1 & \text{if the firm adopts the new technology} \\ 0 & \text{otherwise} \end{cases} \quad (3.10)$$

where β is a $K \times 1$ vector of unknown parameters that reflect the impact of changes in X' on adoption probabilities and $E(u_i) = 0$, where 'E' is the expectations operator. It is assumed that the data under investigation are individual in nature, which implies that each observation consists of $[y_i, X_i]$ - the actual response of the firm and the associated regressor vector respectively. Estimation of the parameter vector β in (3.10) can then proceed by employment of three models of binary choice that have been considered in the empirical diffusion literature. These are the linear probability model, the logit model and the probit model. Each of these models will be discussed in this section.

The linear probability model proceeds by estimating (3.10) by OLS, which implies that the conditional expectation $E[y_i | X_i]$ is equal to $X_i \beta$. As noted by Judge (1985) given the Bernoulli character of the random variable y_i , it must be the case that

¹⁶ For a critique on these lines see Gold (1980, 1981).

$E[y_i|X_i] = \Pr[y_i = 1] = X_i\beta$. In the case of the linear probability model this is then interpreted as the probability that the event will occur given X_i . Therefore, for the i th institution the estimated value of y from (3.10), \hat{y}_i , will then be the probability that firm i will adopt the new technology given the particular value of X_i .

The linear probability model was used predominately in early (pre-1980s) inter-firm diffusion studies, which examined the role of firm-specific characteristics in determining the probability of adoption. Globerman (1975) investigated the inter-firm diffusion of numerically controlled machine tools (NC's) for a cross section of firms in the Canadian tool and die industry over the period 1961 to 1972. He finds that firm size (measured by the number of firm employees) and the age of the firm's president (to measure the progressiveness of the firm's management towards new technology) both have a positive and significant effect on the probability of adoption.¹⁷ Interestingly, Globerman also includes a squared term for firm size in the regression, which is found to have a negative and significant effect on the probability of adoption. This suggests a non-linear relationship between adoption probability and firm size. The R^2 s obtained by Globerman from the estimated regressions are typically below 0.4.

Romeo (1975) in his study of NC diffusion in the US obtains similar results as Globerman. Romeo finds that firm size (again measured as the number of firm employees) and the education of the firm's president (measured as the number of years spent in education beyond eighth grade) both have a positive and significant effect on the probability of adoption. In addition, Romeo finds, contrary to Globerman, that the age of the president has a negative and significant effect on adoption.

Despite the ease of estimation and interpretation, the linear probability model suffers from three shortcomings, which severely limits its applications in empirical research.¹⁸

Firstly, and arguably the most severe, is that given $X_i\beta$ is unbounded, estimation of this model can give probabilities outside the $[0, 1]$ probability interval. Secondly, as

¹⁷ Globerman also includes dummy variables for the existence, or otherwise, of R&D expenditure, the college education of the firm's president and whether the majority of the firm's assets are owned Canadian residents or otherwise. All of these variables are found to have a statistically insignificant effect on adoption probabilities.

¹⁸ Although, surprisingly, see a recent application of the linear probability model by Goel and Rich (1997) who examine the diffusion of new technologies in the US aircraft industry.

noted by Maddala (1983), since y_i can only take two values, u_i can also only take two values and, thus, it cannot be assumed that u_i is normally distributed. In fact u_i will follow a binomial distribution, which implies that OLS is not, in general, fully efficient. That is, there exist non-linear procedures that are more efficient than the OLS procedure. Thirdly, with the probability structure given above, u_i is heteroscedastic. This leads to OLS techniques producing inefficient estimators [Maddala (1988)].

An alternative and more recently used approach to the linear probability model, which does constrain the estimated probabilities in the $[0, 1]$ interval, is provided by the *logit* and *probit* models which only differ in their specification of the distribution of the error term. The basis of both these models is the assumption that there exists some underlying response or 'latent' variable, y_i^* , which is not actually observed. This response variable could be, for example, the 'propensity to adopt a new technology' by a firm. It is assumed in the logit and probit analysis that this response variable is defined by the following regression relationship, which has the same structure as the one defined in (3.10) for the case of individual data:

$$y_i^* = X_i\beta + u_i \quad (3.11)$$

with $i = 1, \dots, n$ and where y_i^* is the dependent variable and the assumptions concerning the distribution of u_i depends on whether the model is specified either as a logit or a probit model (discussed below). Although y_i^* is not actually observed, the actual decision of the firm on whether to adopt the new technology or not is observed. This decision is then defined as:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \text{ (the firm adopts the new technology)} \\ 0 & \text{otherwise} \end{cases} \quad (3.12)$$

where the variables are defined earlier. The use of the response variable in the probit and logit models implies that $X_i\beta$ is not equal to $E(y_i|X_i)$, as in the linear probability

model, but will be given as $E(y_i^*|X_i)$. From the relationships in (3.11) and (3.12), the probability that $y_i = 1$ (i.e. the probability that the firm will adopt) will be given by:

$$\Pr(y_i^* > 0) = (u_i > X_i\beta) = 1 - F(-X_i\beta) = F(X_i\beta) \quad (3.13)$$

where $F(\cdot)$ is the cumulative distribution for u_i , which is assumed to be symmetric.

To constrain the estimated probabilities from (3.13) in the interval $[0, 1]$, $F(\cdot)$ is specified as any continuous cumulative distribution function. Moreover, the functional form of $F(\cdot)$ in (3.13) will depend on the assumptions made about the cumulative distribution of the error term, u_i , in (3.11). Two symmetric distributions are most commonly specified in the literature: the *logistic* and the *normal* from which the logit and probit models are derived respectively. In the logit model it is assumed that the cumulative distribution of u_i has a logistic distribution with a zero mean. The probability that a firm will not adopt a new technology, $\Pr(y_i = 0)$, is then given by $1 - \Lambda(\cdot)$, where $\Lambda(\cdot)$ represents the logistic cumulative distribution function. Alternatively, in the probit model the cumulative distribution of u_i is assumed $u_i \sim N(0, 1)$. In this case the probability that a firm will not adopt a new technology is then given by, $1 - \Phi(\cdot)$, where $\Phi(\cdot)$ represents the logistic cumulative distribution function [see Judge (1985) for exact specifications].

Unlike the linear probability model, the estimated coefficients obtained from the logit and probit models cannot be interpreted as the increase in the probability of the event occurring given a one-unit increase in the corresponding independent variable. Instead, the estimated coefficients reflect the effect of a change in the independent variable on $F^{-1}(P_i)$, for the probit model, and on $\ln[P_i/(1 - P_i)]$, for the logit model (the *log-odds ratio*)¹⁹, where P_i is the probability that y_i is equal to one. In both the logit and probit models the increase in the probability will depend upon the original probability and,

¹⁹ For a proof of this see Gujarati (1988) and Judge (1985). Note that this implies that the estimated coefficients can lie outside the $[0, 1]$ interval, even though the estimated probabilities will be constrained in the $[0, 1]$ interval.

therefore, upon the initial values of all the independent variables and their coefficients, and these relationships will be non-linear in X_i and β [Judge (1985)].

Benvignati (1982) has applied the probit model outlined above to the diffusion of process innovations in the US textile industry using a cross section of textile firms. The probit model is applied by specifying the dependent variable as unity if the firm adopts any of the thirty-three innovations under consideration and zero otherwise. Only two firm-specific characteristics are considered to have an effect on the firm's decision to adopt: firm size (measured as the firm's overall employment level) and the firm's labour costs. The latter factor is approximated by a trade union dummy variable, which takes the value of unity if more than ten per cent of the firm's textile workers are represented by a union and zero otherwise. Results indicate that firm size has a positive and significant effect on adoption probability. The coefficient on the trade union dummy is found to be positive indicating that, *ceteris paribus*, those firms with trade union representation over ten per cent of their workforce are more likely to adopt new technology.

The study by Benvignati has two principle limitations. Firstly, there is the implicit assumption made by Benvignati that wage settlements in non-unionised firms are unaffected by those in unionised firms. If non-unionised wage settlements *are* affected significantly by wage settlements in unionised firms then this might explain why the trade union dummy is found to be insignificant.²⁰ Secondly, Benvignati pools thirty-three, arguably, non-comparable innovations together in one single regression. This seems to be an extremely crude method given that the early empirical research has found that innovation-specific characteristics are also significant in the diffusion process.

Oster (1982) has applied the logit model to the diffusion of the basic oxygen furnace (BOF) in the US steel making industry. The study is an interesting one because Oster makes the characteristics of adopting plants rather those firms *per se*, the main determinant of the profitability of the new technology. Oster assumes that the

²⁰ In addition, there may be a multicollinearity problem in the specified regression because, as Elliott (1991) has shown, trade union representation tends to be higher in firms that employ more workers.

probability of adoption will depend upon the size of the firm that operates the plant and the profitability of installing the BOF.

For the measure of innovation profitability, Oster distinguishes between two separate adoption decisions: adopting new technology to replace *existing* capital and adopting to *expand* existing capital stock. In the first decision, the adoption criterion used by the firm is assumed to be based on a comparison between the relevant capacity costs of the BOF and the old technology. Oster by measures this assuming that at each plant the firm will replace its smallest capacity furnace, from which (using cost schedules and the size of the old technology provided by the plant) potential cost savings from adopting the BOF are calculated. In the second decision, the growth in output between 1957-1964 in the state in which the plant is located is used as a measure of profitability.²¹

A logit model is estimated by Oster using these two plant-specific profitability measures, together with a firm size variable (measured as the firm's total output), to account for any separate effects on the probability of adoption. The results obtained show that both profitability measures have a positive effect on the probability of adoption. An interesting result is that firm size is found to have a negative and statistically significant effect on adoption probability. This result is contrary to the majority of empirical research, which finds that larger firms adopt earlier than smaller firms do.²² The main limitation of this study is that Oster only considers two plant- and firm-specific characteristics: plant profitability and firm size. There is, arguably, room for expanding the study to examine other possible characteristics, such as the financial resources available to the firm and any institutional factors that may be important.

In a more recent application of the logit model, O'Farrell and Oakey (1992) examine the determinants of CNC (computer-numerically-controlled) machine tools adoption across three geographical regions in the UK obtained from survey evidence gained in 1989. The results indicate that age of the firm (a proxy for the reputation aspects of raising finance), size of the firm (measured by employment) and percentage of output produced

²¹ Oster interprets this as a demand side variable, so that in expanding markets firms are more likely to adopt the new technology available.

that is subcontracted (to approximate the applicability of CNC) have a positive and significant effect on the probability of adoption. In addition, a regional dummy variable for Wales is found to have a negative and significant effect on the probability of adoption indicating the potential importance of regional-specific characteristics.

To date, there have been no studies that have applied count data models to examine the factors determining the number of innovations adopted at the inter-firm level. Recent papers by Blundell *et al* (1994, 1995) have, however, examined the firm- and market-specific determinants of the number of innovations *commercialised* for a panel of data of manufacturing firms in the UK from 1972 to 1982.

In Blundell *et al* the number of product and process innovations that are commercialised at a given moment in time are assumed to be realisations of a non-negative, discrete random variable. Two count data models are then employed: the Poisson model and the negative binomial model. They show that basic specification of these models do not take into account unobserved heterogeneity²³ that arise from not being able to measure all those firm-specific and market-specific determinants of innovative activity. Blundell *et al* show that by directly modelling unobserved heterogeneity as pre-sample innovation activity for individual firms they are able to take account of these fixed effects. This is approximated in two ways: firstly, by a dummy variable that takes the value of unity if the firm innovated pre-sample and zero otherwise, and secondly, by a measure of the firms 'knowledge stock', approximated by the depreciated sum of past innovations. The results obtained by Blundell *et al* show that both these variables have a positive and significant effect on innovative activity, but are reduced by the inclusion of firm-specific characteristics. Moreover, union density and market share of the firm are found to have a positive and significant effect on a firm's innovative activity. Blundell *et al* also find that a market concentration variable (measured by the market share of the five largest firms) has a negative and significant effect suggesting that market concentration dampens the innovative effects of market dominance.

²² One possible explanation of this, hinted by Oster (1982), is that there are significant economies to scale in the old technology, so that firms with relatively small furnaces (predominately the smaller firms) stand to gain more from scrapping and replacement than do larger, more modern works.

²³ These are so-called fixed effects in panel data models [see Greene (1993) for further discussion].

3.2.2.2 Duration Models

Duration models represent the current state-of-the art approach to empirical inter-firm modelling. They have been developed by economists and econometricians in the last two decades in order to analyse so-called *duration data* which is characterised by a group or groups of economic agents for whom (or which) there exists a principal point at which a particular event ends, often referred to as *failure*.²⁴ The event ends after a length of time called the *failure time* or *duration time* [Greene (1993), Neumann (1997)]. There is a direct analogy to inter-firm diffusion inherent in duration data. The group of economic agents can represent a set of potential adopters of the innovation under investigation and the principal point at which a particular event ends can represent that time at which adoption takes place for an individual firm. The simplest type of duration data is *single spell* data [Lancaster (1990)] in which observations are made on an agent's duration of stay in a single state and this is the most frequent type of data encountered in studies of inter-firm diffusion where dis-adoption (perhaps followed by re-adoption) is rarely observed during sample periods.²⁵

As stated by Cox and Oakes (1984), however, there are three prerequisites for a duration to exist. Firstly, a time origin needs to be defined for each agent, secondly, a scale of measurement is required to measure duration time and, thirdly, the point of failure needs precise definition. In the inter-firm diffusion studies reviewed in this chapter, it is convention to measure these three parameters for an individual firm as the time of commercialisation of the innovation (or, due to a paucity of data, the date at which the first firm of the set of potential adopters adopts), calendar time and the date of adoption respectively. More formally, if T is assumed to be a non-negative random variable representing the duration of non-adoption of an individual firm in a homogenous set of n potential adopters (with respect to systematic factors and regressor variables) then duration data is typically represented by a cross-section of duration (adoption) times t_1, t_2, \dots, t_n [Greene (1993)]. Thus, inter-firm diffusion can be conceptualised as the

²⁴ The use of the term 'failure' has arisen from early engineering applications of this approach used to describe electrical component failures. Its use does not convey a value judgment relating to the adoption decision.

²⁵ When agents - in this case firms - are observed in a sequence of states then this type of data is referred to as *multiple spell* data and models to deal with such data lie outside the ambit of this chapter. See Lancaster (1990) for a review of appropriate models in these cases.

transition from a state of non-adoption to one of adoption. It is convention in inter-firm studies (although not strictly required) that all firms enter the state of non-adoption at time $T = 0$.

From this simple framework three representations of probability distributions of T can be specified and which form the central focus of duration models. These are the *survivor function*, the *hazard function* and the *integrated hazard function*. In order to relate the various duration models to the analysis of inter-firm diffusion, each of these distributions will be defined with explicit reference to a set of potential adopters. If T is assumed absolutely continuous²⁶ and has range $[0, \infty)$ then the survivor function gives the probability that the spell of non-adoption is at least t and is given by:

$$S(t) = \Pr(T \geq t) = 1 - F(t) \quad (3.14)$$

where $F(t)$ is the CDF of duration times. As stated by Heckman and Singer (1984, 1985) $S(t)$ will be a monotonic, non-increasing left continuous function with $S(0) = 1$ and $\lim_{t \rightarrow \infty} S(t) = 0$.

The hazard function, in contrast, specifies the instantaneous rate of adoption at $T = t$ conditional upon non-adoption to time t and is defined as:

$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t)}{dt} \quad (3.15)$$

The numerator in (3.15) gives the probability that in the next short time interval, dt , the firm will adopt given that it has not adopted until time t , where the event that $T \geq t$ is the event that the state is still occupied at t . More succinctly, the term in (3.15) gives the instantaneous rate of leaving the state of non-adoption per unit time period at t . An approximate interpretation of $h(t)$ is that $h(t)dt$ is the probability of exit from the state of non-adoption in the short interval of length dt after t , *conditional* on the state still being occupied at t [Lancaster (1990)]. Ultimately, the hazard function in an inter-firm

²⁶ The following definitions can be derived for T being absolutely discrete and being partly discrete and continuous. See Kalbfleisch and Prentice (1980) for appropriate definitions in these cases.

diffusion study will give the conditional probability of a firm that has not adopted an innovation after, say, five years since its commercialisation, will adopt in the fifth year. In terms of relative frequencies, $h(5)dt$ will give the proportion of firms that have not adopted after five years who do adopt within dt of the fifth year. This concept of conditional probability is in contrast to the discrete choice models introduced in Section 3.2.2.1 where the main focus of attention was on the unconditional probability of adoption (assuming a homogenous population). In the simple case given above, the unconditional probability of adoption will simply be given by the probability of a firm adopting after five years.

Lancaster (1990) and Neumann (1997) have shown that the hazard function in (3.15) can be expressed in terms of the CDF and PDF of the continuous random variable T . This is given by:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (3.16)$$

Thus, the hazard function at time t is the PDF divided by the CDF. Moreover, given the definition of the survivor function in (3.14) then it follows that the hazard function can more conveniently be expressed as the PDF divided by the survivor function. Furthermore, as shown by Heckman and Singer (1984, 1985), knowledge of $h(t)$ determines $F(t)$ because integration of (3.14) obtains the following:

$$\int_0^t h(u)du = -\ln[1 - F(x)] \Big|_0^t + c \quad (3.17)$$

and since $c = 0$, given that $S(0) = 1$, equation (3.17) may be written as:

$$S(t) = \exp\left[-\int_0^t h(u)du\right] \quad (3.18)$$

The term on the left-hand side of (3.17) is referred to as the *integrated hazard function* and is conventionally denoted by $\Lambda(t)$. Taking logs of (3.18), it can also be noted that the integrated hazard function it is equal to $-\ln S(t)$. This function is not strictly a probability and is used frequently in diagnostic tests [Kiefer (1988)]. If the probability

of adoption for the set of potential adopters is assumed certain in the sense that $\lim_{t \rightarrow \infty} S(t) = 0$ (i.e. as time tends to infinity the probability of surviving in a state of non-adoption is zero), then Heckman and Singer (1984, 1985) have shown that $h(t)$ must adhere to the following property:

$$\lim_{t \rightarrow \infty} \int_0^t h(u) du \rightarrow \infty \quad (3.19)$$

Duration data with this quality is termed as being *non-defective* [Lancaster (1990)] and is the conventional assumption to make in inter-firm diffusion studies. Sinha and Chandrashekar (1992), however, have attempted to model inter-firm diffusion for *defective* data in the sense that not all potential adopters are assumed to adopt which implies $\lim_{t \rightarrow 0} S(t) \neq 0$.²⁷

These three functions can be employed as a summary device and as a means of selecting parametric models (see later discussion). Of particular interest in inter-firm diffusion studies are the movement of the hazard function over time and the associated nature of duration dependence which exists if:

$$\frac{dh(t)}{dt} \neq 0 \quad (3.20)$$

Moreover, if $dh(t)/dt > 0$ at $t = t_0$ there is said to be *positive duration dependence* at t_0 . Alternatively, if $dh(t)/dt < 0$ at $t = t_0$ there is said to be *negative duration dependence* at t_0 [Heckman and Borjas (1980)]. Consequently, the shape of the hazard function over time will indicate how the conditional probability of adoption is evolving and, thus, is frequently used as a descriptive tool in empirical studies. Indeed, the shape of the hazard function serves as the central focus of Chapter 5.

Extensive discussion of the methods employed in the estimation of the survivor, hazard and integrated hazard functions is delayed until Chapter 5 where such approaches are

²⁷ The terms 'non-defective' and 'defective' do not convey a value judgment on the quality of the data but, is used in the literature purely as a technical description of the underlying assumption of the integrated hazard function [Heckman and Singer (1984, 1985)].

actually employed for a set of potential ATM adopters. It can be noted here, however, that there are two main approaches employed in the literature: the non-parametric and the parametric. Non-parametric approaches are purely empirical based and make no assumptions as to the particular form of the probability distribution of adoption times and, hence, the subsequent shape of the hazard function. The most commonly employed estimates are those derived from the Kaplan-Meier (1958) method.

In contrast, parametric estimates of the survivor, hazard and integrated hazard functions make explicit assumptions concerning the probability distribution of adoption times and, hence, the subsequent shape of the hazard function. There is a plethora of parametric distributions which can represent duration data each assuming a particular form or forms of duration dependence depending on the subsequent estimated parameters of the model [see Kalbfleisch and Prentice (1980) for a comprehensive review of possible distributions]. The inter-firm diffusion literature has typically considered four distributions that allow a high degree of flexibility of duration dependence: the Weibull, exponential, lognormal and log-logistic. Each of these distributions are employed in Chapter 5 for the set of potential ATM adopters and a summary of their properties is contained in Appendix Two of that chapter.

The ultimate aim of employing duration models in inter-firm studies of diffusion is not simply to provide a descriptive tool as to the nature of the hazard function, but rather, to examine the significance or otherwise of firm-specific and market-specific characteristics in the diffusion process. This aim is implemented by relaxing the assumption made in the above analysis that all firms are drawn from a homogenous population, but instead allowing for firm heterogeneity by including covariates or regressor variables into the analysis.²⁸ Regression models are then typically formulated with the dependent variable being represented by the hazard function and the sign of the estimated coefficient(s) indicating the direction of the effect that the covariate has on the conditional probability of adoption.

The inclusion of covariates is, however, complicated by the existence of two distinct types: *time-invariant covariates* and *time-varying covariates*. The former of these do

²⁸ The terms 'covariate' and 'regressor' are used synonymously. The term 'covariate' has replaced the more usual 'regressor' terminology in the duration literature and is the one used in this thesis [Lancaster (1990)].

not change over time and will include, for example, a dummy variable representing the industry-classification the firm belongs to. In contrast, time-varying covariates do change over time and could include, for example, the profitability of the firm if this is purported to be a significant factor affecting the adoption decision of the firm. Time-varying covariates can further be separated into those that change in a discrete fashion over time and those that change continuously [Lancaster (1990)]. To date, the empirical literature has assumed that both sets of time-varying covariates remain constant for finite sub-periods to enable convenient estimation [and Lancaster (1990) and Petersen (1986a, 1986b)].

An adjustment to the definition of the survivor, hazard and integrated hazard functions in (3.17) to (3.20) must be made for heterogeneity amongst the set of potential adopters [Lancaster (1990)]. In the presence of K time-invariant covariates assembled in a $1 \times K$ vector, X , the hazard function at time t is defined as being conditional on the value of X :

$$h(t|X) = \frac{f(t|X)}{S(t|X)} \quad (3.21)$$

with the survivor and integrated hazard functions similarly defined conditional on X . In contrast, in the presence of K time-varying covariates the $1 \times K$ vector, $X(t)$, contains that path of K covariates from time 0 to time t . In this case the hazard function is defined as in (3.21) but with X being replaced by $X(t)$. Similar definitions for the survivor and integrated hazard functions result, all conditional on $X(t)$. In practice the determinants of innovation diffusion are likely to involve both time-invariant and time-varying factors.

As noted by Lancaster (1990), however, the conditional hazard function in (3.21) can only be derived if X strictly contains only exogenous variables. Exogeneity of a covariate has a distinct interpretation in the duration literature and has been formally defined by Lancaster (1990). Lancaster's definition is similar to that of Granger's (1969) definition of causality for time series data. In Lancaster's definition a covariate is exogenous if and only if the information that a firm has not adopted to time $t + dt$, does not aid prediction of the path of the covariate process from t to $t + dt$ given its

history to t . Using this definition, time-invariant covariates are necessarily exogenous [Lancaster (1990)]. As noted by Heckman and Singer (1984, 1985), however, in practice the distinction between exogenous and endogenous covariates is contentious and problematic. These issues are more fully discussed in Chapter 6 when firm-specific and market-specific determinants of ATM adoption are examined.

The specification of the conditional hazard functions $h(t|X)$ and $h[t|X(t)]$ in principle allows a wide range of possible interactions among durations, t , and covariates, X . Consequently, formal regression models have been developed in order to restrict the possible range of restrictions. There are currently two main empirical models employed in the diffusion literature: the proportional hazards model of Cox (1972) and the accelerated hazard model as outlined in Kalbfleisch and Prentice (1980). The former model restricts the interaction between t and X to a multiplicative one, while the latter rescales time. Discussion and analysis of these models are delayed until Chapters 6 and 7 where these models are applied to the set of ATM adopters.

The sign of an estimated coefficient in both classes of models will indicate the direction of the effect of the covariate on the conditional probability of adoption, which will be inversely related to the duration of non-adoption. The interpretation of its value is, however, dependent on the exact specification of the model, although in general it does not have the same interpretation as a marginal effect as in a linear least squares regression model (see Appendix One of Chapter 6). Both the proportional hazards and accelerated time models are estimated by maximum likelihood techniques.

A distinguishing feature of duration data is that it often contains ties (where two or more observed duration times are identical) and censored (incomplete durations) observations. In inter-firm studies of diffusion censored observations typically arise as a result of sampling design. Empirical studies are often retrospective, utilising a stock of potential adopters at a point in time. Consequently, it is commonly found that not all potential adopters have adopted by the end of the sample period and, without a follow-up survey, the precise duration times of non-adopting firms are not known. Such observations are known as *right-censored* or *singly Type I censored* and occur when the

censoring time is under the control of the researcher and is equal for all firms.²⁹ If T^* is the random adoption time in the absence of censoring and L is the censoring time then what the researcher does observe is the following:

$$T = \min(T^*, L) \quad (3.22)$$

An indicator variable, δ , is then employed to denote whether the observation is censored or not. This takes the value of unity if censored and zero otherwise. Such observations have to be treated differently to uncensored observations in the estimation procedures for non-parametric and parametric functions otherwise the estimate hazard function will be biased upwards [Kiefer (1988)].

To date, most empirical studies that have employed duration models have focused their attention on the rank effect models of diffusion. An early study by Hannan and McDowell (1984b) examined the firm-specific and market-specific characteristics purported to affect a banks decision to adopt ATMs. They employ a similar data set to Hannan and McDowell (1984a) consisting of a panel data set of adoption histories from 1971 to 1979. The empirical model is formulated as reduced-form exponential regression one, specified as:

$$h(t_i) = \exp[X_i(t)\beta] \quad (3.23)$$

where the baseline hazard³⁰ is set to unity, and $X_i(t)$ is a vector of time-varying covariates that differ across banks. They find a positive and significant effect on the conditional probability of adoption (or, alternatively, a negative effect on the duration of non-adoption) for market concentration, wage levels and bank size. These confirm the two-step results of Hannan and McDowell (1984a) and are therefore subject to similar interpretation and critique. The possibility that liquidity constraints facing banks may delay adoption are examined by including a variable measuring the banks profit rate as

²⁹ This is not the only censoring mechanism that empirical economists can employ or encounter. See Chapter 5 for discussion.

³⁰ It is convention in the diffusion literature to assume that the right-hand side of (3.23) is multiplied by a time-varying baseline hazard function, which pertains to the condition $X = 0$. Hannan and McDowell

net income divided by total assets. This, however, is found to have a negative sign (inconsistent with *a priori* expectations) and insignificant coefficient. Differences in the banks product mix are captured by a variable measuring the ratio of each bank's demand deposits to its total assets. This is found to have a positive and significant effect on the conditional probability of adoption and may arguably be capturing differences in technological opportunities available to banks given that cash withdrawals (often from demand deposits) are the most frequent transaction that ATMs perform (see discussion in Chapter 4). Institutional arrangements and regulatory arrangements are also shown to significantly affect the diffusion path. They find that unit banking and branching restrictions increase the conditional probability of adoption and that independent banks are less likely to adopt ATM technology as compared to banks owned by holding companies.

Hannan and McDowell (1987) replicate the results obtained in Hannan and McDowell (1984b) for an identical data set, but also consider the possible effects of rival precedence inherent in the stock effect model of innovation diffusion. Hannan and McDowell argue that rival precedence has two counteracting implications for the diffusion path. Firstly, rival precedence may increase the probability of adoption due to reductions in uncertainty concerning the economic and technical attributes of the innovation analogous to the Mansfield (1961, 1968) model. Secondly, it may reduce the gross benefits to adoption as more firms adopt due to a fall in output prices as industry output expands and increases in factor prices. The last argument is similar to the order effects proposed by Quirnbach (1986). Rival precedence is captured in their model by two variables. The first is a dummy variable taking a value of unity at year t if at least one other bank in the market had adopted ATMs prior to year t and zero otherwise. This captures whether the bank's decision was to adopt as a follower, in which case the dummy takes a value of unity, or that of a leader, in which case it takes a value of zero. The second measures the proportion of banks in the market that had adopted ATMs prior to year t . This is an attempt to capture the extent to which a bank's decision to adopt is dependent on the degree to which it has been preceded. Higher values of this variable indicate a higher degree of being preceded.

assume that this baseline hazard is time-invariant and takes a value of unity. See Appendix One of Chapter 6 for technical details.

Hannan and McDowell's results from an exponential regression model show that both measures of rival precedence have positive and significant coefficients indicating that observed adoption by rivals increases the likelihood that potential adopters will themselves choose to adopt. This result suggests that the uncertainty reducing effects of rival precedence is stronger than the possible effects on output prices and factor prices. They also include two interaction variables in the model which enter as multiplicative terms between the two measures of rival precedence, bank size and market concentration. Their results show that larger firms tend to adopt as initial adopters rather than followers and that firms in more concentrated markets exhibit less tendency to adopt in response to rival precedence than do firms in less concentrated markets. This last result indicates that while market concentration may increase the probability of adoption, this effect is less significant when a substantial proportion of banks has already adopted (i.e. at the latter stages of adoption).

An innovative paper by Sinha and Chandrashekar (1992) formulates a so-called split hazard model [Schmidt and Witte (1988, 1989)] to investigate the diffusion of ATMs from a cross-section of US banks from 1971 to 1979. The innovative aspect of the paper is the relaxing of the assumption that all banks eventually adopt ATMs, which is implicit in the condition $\lim_{t \rightarrow \infty} S(t) = 0$ contained in (3.19). They argue that this assumption is a constraint on the modelling approach because only 20% of their sample banks adopt at the end of their sample period in 1979. Their alternative model draws on the criminal recidivism literature of Schmidt and Witte (1988, 1989) in which (for the case of inter-firm diffusion rather than recidivism) the probability of eventual adoption becomes an additional parameter to be estimated and allows this to take a value less than unity. The basis of their model is that there exists some unobservable variable indicating whether or not a firm adopts analogous to the latent variable discussed in relation to the discrete choice models in Section 3.2.1.1. This unobservable variable takes the following values:

$$A_i = \begin{cases} 1 & \text{if the } i\text{th bank eventually adopts} \\ 0 & \text{otherwise} \end{cases} \quad (3.24)$$

where A_i is the unobservable indicator variable for the i th bank. The probability of observing different values of A_i is then given as:

$$\Pr(A_i) = \gamma_i = \gamma(X_i) \quad \text{- for adopting banks} \quad (3.25)$$

$$\Pr(A_i) = 1 - \gamma_i = 1 - \gamma(X_i) \quad \text{- for non-adopting banks} \quad (3.26)$$

where γ_i is the probability of eventual adoption. Equation (3.25) and (3.26) effectively split the sample between those banks who eventually adopt ATMs and those who do not.

From this simple framework, Sinha and Chandrashekar show that the PDF for the i th adopting bank is:

$$\gamma_i f(t_i | X_i, A_i = 1) \quad (3.27)$$

and that for a censored bank (one that has not adopted by the end of the sample period) is:

$$1 - \gamma_i + \gamma_i [S(\tau | X_i, A_i = 1)] = 1 - \gamma_i + \gamma_i S(\tau) \quad (3.28)$$

where τ is end of sample time. Both the terms in (3.27) and (3.28) enter the likelihood function and this aspect of the modelling allows heterogeneity in both the probability of eventual adoption represented by $\gamma_i = \gamma(X_i)$ and in the duration of non-adoption represented by $f(\cdot)$ in (3.26).

As noted by Schmidt and Witte (1989) the general model outlined above nests three distinct duration models. First, by assuming that the probability of eventual adoption is unity ($\gamma_i = 1$) the conventional class of duration model results. Second, assuming the eventual probability is constant across banks ($\gamma_i = \gamma$) a restricted hazard model is obtained and third, allowing the probability of eventual adoption to vary across banks, results in the flexible hazard model. Sinha and Chandrashekar estimate the first and

third of these models and allow the distribution of adoption times to follow a lognormal distribution (chosen by best-fit of the data) and the distribution of γ_i to follow a logit model of the form:

$$\gamma_i = 1/[1 + e^{x_i\beta}] \quad (3.29)$$

Their results from estimation of the conventional model broadly confirm earlier studies of ATM diffusion found by Hannan and McDowell (1984b, 1987). They find a positive and significant effect on the duration of non-adoption for market concentration, wage level and growth, bank income (as a measure of liquidity constraint) and unit bank restrictions. Interestingly, however, they find a negative and significant effect for bank size. Their results from the flexible split hazard model are, however, discouraging. They find a negative and significant effect on the probability of adoption for market growth, bank income and unit bank restrictions. Although market concentration is found to have a positive but insignificant effect on the probability of adoption it is found to have a negative and significant effect on the duration of non-adoption. A similar result is revealed for the proportion of demand deposits as a measure of the banks product mix. In addition, market growth is found to have a positive, but insignificant effect on the duration of non-adoption, reversing its effect (and significance) found on the probability of adoption. The reverse effect is found for the market wage.

Although the model proposed by Sinha and Chandrashekar is arguably innovative it does suffer from two main weaknesses. First, it gives no underlying economic justification for the use of the split hazard model. The motive being purely dependent on the fact that only 20% of their sample had adopted by the end of sample period. No follow-up evidence is provided to support their hypothesis that not all banks in their sample will eventually adopt. Second, Sinha and Chandrashekar offer no explanation of their results. The sign and significance of some variables change when analysed as either affecting the probability of adoption or the duration of non-adoption, but no explanation of why this occurs is given. Moreover, the paper raises a broader issue in that they use time-invariant covariates in their model which are measured at the beginning of the sample period in 1971. As noted by Heckman and Singer (1984, 1985), *ad hoc* selection of the date at which to measure time-invariant covariates may

lead to divergent results for different model specifications. Furthermore, it does not follow that models including time-varying covariates are necessarily superior. This arguably necessitates more rigorous model testing and selection and these important themes are more fully explored in Chapter 6.

Rose and Joskow (1990) examine the diffusion of new electric generating technologies for a cross-section of generating utilities in the US. They decompose the decision to adopt into two component parts: those that arise only because opportunities for adoption are more frequent, and those that reflect an early decision to employ a new technology. Rose and Joskow argue that the former is more likely to occur for larger firms given that they build new generating units more frequently than do smaller firms as a result of different capital configurations, demand and growth rates, and the lumpiness of capital. Thus, adoption decisions are argued to be 'left-censored' [Kiefer (1988)] in the sense that the decision to adopt is not observed until firms build a new generating unit. They account for this by employing a double-censored hazard model [Greene (1993)] which corrects for this potential bias by including a term in the likelihood that captures the probability that the adoption decision occurs between the date of building the last unit and the new unit.³¹ Their results from parametric and non-parametric specifications of the proportional hazard model indicate that firm size has a positive and significant impact on the conditional probability of adoption, but that this effect is exaggerated in models that do not take account of double-censored observations.³² They also find a positive and significant effect for fuel prices (as a proxy for potential savings from adoption) but no significant role for institutional factors.

These early inter-firm studies have been criticised by Karshenas and Stoneman (1993) for ignoring price expectations and consequently, may be seriously mis-specified. They argue that the optimal time to adopt new technology that time that maximises gross net present profit. The optimal time to adopt ultimately depends on two conditions at the individual firm level. First, the profitability condition which requires that the technology yields positive profits and secondly, the arbitrage condition which requires that the net benefits from adoption is not increasing over time. Karshenas and

³¹ It is additionally assumed that the building decision is independent of the adoption decision.

Stoneman show whilst the former condition determines the set of potential adopters it is the arbitrage condition that determines the optimal adoption time. As soon as this arbitrage decision is introduced into the firm's criteria for adoption the benefits from delaying adoption will depend, *inter alia*, on the expected cost of adoption which ultimately depends on the expected price of the new technology. They show that if and only if price expectations are myopic (i.e. the price of technology is assumed constant next time period) will the arbitrage condition not hold.

In their empirical contribution, Karshenas and Stoneman (1993) investigate the diffusion of CNC machines for a panel data set of UK manufacturing firms across nine industries from 1968 to 1980. They attempt to provide an alternative and superior approach to inter-firm modelling by explicitly subsuming the four main theoretical approaches to inter-firm diffusion - epidemic, rank, stock and order models - and price expectations into one single empirical model. This is achieved by specifying the proportional hazards model developed in Cox (1972) and using firm-specific and industry-specific proxies for these four approaches. Epidemic effects are captured by the time dependency of the baseline hazard. Rank effects are captured by firm-specific characteristics such as size, its ownership status and output growth. Stock and order effects are merged together as the cumulative number of owners of the technology up to and including time t . Arguably, this measure may create an endogeneity bias in the model as the price of new technology usually falls as adoption increases due to interaction between demand and supply factors [see, for example, Ireland and Stoneman (1986)]. This potential problem is recognised by Karshenas and Stoneman, but circumvented by arguing that CNC machines were supplied predominately through imports and, thus, that supply-side factors can be ignored for estimation purposes. Finally, price expectations are modelled quite crudely by the one-period ahead first difference in price at the time of adoption. This relatively simple measure is chosen on the basis that because the price of CNC machines fall monotonically throughout the sample period then more sophisticated approaches (by explicitly modelling adaptive or rational expectations, for example) would yield no greater insights.

³² They find that elasticities of the hazard rate with respect to firm size can be exaggerated up to 50% in conventional models.

Their results from estimating a proportional hazard model with a Weibull baseline hazard show that the measure of price expectations has a positive and significant effect on the conditional probability of adoption. Moreover, their results reject the myopic expectations model using the Likelihood Ratio test when compared against the fully specified model. This result is significant because it suggests that the models of Hannan and McDowell (1984b, 1987) and Sinha and Chandrashekar (1992) may be seriously mis-specified by ignoring the arbitrage condition as a key element in the firms adoption criteria. There is strong evidence found in support of the rank effects model, with firm size (measured by employment) and firm growth (measured by output) found to have a positive and significant effect on the conditional probability of adoption. Market concentration is found to have a positive, but insignificant effect on adoption probability, while firms R&D expenditure is found to have a negative and insignificant effect. Karshenas and Stoneman suggest that the inclusion of this latter variable may cause multicollinearity with firm size, but even after excluding firm size, R&D expenditure is found to have a negative effect on adoption probability, which is inconsistent with *a priori* expectations. The baseline hazard is found to have positive time dependency, a result that Karshenas and Stoneman use to argue that the diffusion path is characterised by positive duration dependency and, thus, supports the contention that there are significant epidemic effects present in the diffusion path.³³ No empirical support is found for the existence of stock and order effects in the diffusion path and their variable measuring this effect has a positive sign, which is inconsistent with *a priori* expectations. They suggest that this result occurs because its existence requires a substantial effect on firm costs and output and may be outweighed by the stronger epidemic effects. Overall, their results are found to be consistent across various specifications of the baseline hazard (exponential and logistic) from allowing the baseline to take an arbitrary value consistent with the original Cox (1972) model.

Recent papers by Stoneman and Kwon (1994) and Colombo and Mosconi (1995) have taken a broader, arguably evolutionary perspective, of technology diffusion. Consequently, these papers have attempted to incorporate the interrelatedness and co-

³³ It is argued in Chapter 5, that if the diffusion curve is sigmoid, which is consistent with the epidemic model, then the resulting duration dependency is non-monotonic. This implies that the assumption of a Weibull baseline hazard excludes the possibility of non-monotonic duration dependency and so Karshenas and Stoneman cannot strictly reach this conclusion based on this result.

development of technologies that the evolutionary perspective invokes, into formal duration models. Stoneman and Kwon (1994) distinguish between technologies that are substitutes and those, which are complements in the production process. For the case of two technologies - A and B - two technologies are defined as being complements if the per annum profit gain (relative to the no adoption baseline) for adopting both A and B is greater than adopting A and B alone. The reverse is true for substitute technologies. By assuming perfect foresight in the formation of firms' price expectations, Stoneman and Kwon are able to show that optimal adoption dates are determined by arbitrage conditions alone, requiring equality between costs and benefits of waiting. From this simple framework they are then able to derive a set of simultaneous deterministic and stochastic adoption decisions in which joint adoption decisions are substituted into single adoption decisions. They estimate a set of simultaneous exponential regressions for NC machine tools and coated carbide tools (CCT) for a panel set of manufacturing firms in the UK. Their results indicate that there are significant cross-technology effects. They find a significant and negative effect for technology price and stock effects (measured by the number of users of NC and CCT at time t) and a positive and significant effect for firm size for the conditional probability of adopting NC given previous adoption of CCT. Similar findings are found for the case of adopting CCT given previous adoption of CCT. Price expectations, epidemic effects (approximated by a time trend) and firm size are all found to have a positive and significant effect on the conditional probability of adopting NC and CCT simultaneously. Order effects (measured as the number of users of NC and CCT in the industry at time t) are found to have a negative and significant effect on the conditional probability of adoption. Overall, their results indicate a more significant role for stock and order effects in a multi-technology framework relative to Karshenas and Stoneman (1993).

The results presented in the paper by Colombo and Mosconi (1995) confirm the importance of interdependency and complementarity in technological diffusion. They examine the diffusion of a cluster of innovations [Dosi (1983)] resulting from flexible automation for a cross-section of manufacturing plants in the Italian metalworking industry using a single equation Weibull proportional hazard model. Their methodology separates the rank effects into three components. First, cumulative learning-by-using effects that reflect the stock of knowledge, capabilities and technical

and managerial skills that a firm develops through the use of previous vintages. Second, interactive effects from the adoption of complementary technology which increases the marginal benefits from adopting the technology under investigation, and third, conventional rank effects (such as firm size and market structure for example). Cumulative learning is approximated in the empirical model by a series of dummy variables indicating earlier adoption of previous vintages and managerial and organisational innovations. Learning-by-using is approximated by a series of dummy variables indicating the adoption of complementary technologies. Their results indicate that both cumulative learning and learning-by-using have a significant and positive effect on the conditional probability of adoption. In addition, firm size and the average education level of the plant's workforce are confirmed also to have a positive effect. No significant role is found for ownership status, R&D intensity, regional economic development, and market concentration or stock effects.

A recent paper by Saloner and Shepard (1995) has attempted to substitute the narrow definition of ATM technology as a single capital-embodied with a broader interpretation of the technology as a set of interrelated hardware pieces displaying positive 'network externalities' as outlined in Katz and Shapiro (1985, 1986). As outlined in Chapter 2, the presence of network externalities arises from the existence of complementarity between the components of the network. The assumption of positive network externalities reverses the theoretical predictions of the stock effects model [Reinganum (1981a, 1981b) and Quirmbach (1986)] that as the number of users of the new technology increase the gross benefit from adoption declines. Saloner and Shepard argue, however, that for the case of ATM technology, benefits to depositors increase as the number of geographically dispersed ATMs from which they can access their accounts increases. In addition, because there will be differences in banks' post-adoption network size this will generate different valuations to their depositors. Thus, the benefits of adoption from an ATM system will be higher for banks expecting to have larger proprietary networks in equilibrium.

The model of adoption developed by Saloner and Shepard is operationalised by assuming that the 'network effect' for an individual bank depositor increases (linearly) in the number of locations from which he is able to access his account using an ATM. This is reflected in the benefits to the bank by assuming that the per-period increase in

revenues to the bank is proportional to the per-period benefits to the depositor. Fixed costs of adoption³⁴ are decomposed into 'system costs' which are independent of the number of ATM locations and 'location costs' which increases (linearly) in the number of locations. Diffusion then proceeds by assuming that fixed costs decline over time through a process of learning-by-using and benefits increase over time. The optimal date of adoption then depends on a comparison between the net present value from adoption at time T with that at time $T+1$. If the former is greater than the latter then adoption takes place at T .

In their empirical contribution, Saloner and Shepard measure the network effect as the number of branches a bank has at the start of the sample period as a proxy measure of the number of locations. Implementation of their model is, however, complicated because their model predicts ATM technology is characterised by scale economies in that location costs per depositor are reduced more the greater the number of depositors a bank has. Thus, separating scale economies and network effects is problematic. They resolve this by allowing the estimation of the extent of network effects to be bounded between an upper and lower band. The upper band is defined for when the ratio of depositors to branches is kept constant, which leads to an overstatement of the networks effect. The lower band is defined for when the number of branches are held constant, which leads to an overstatement of the networks effect. Their results from a cross section of US banks employing a Weibull and log-logistic proportional hazard model indicate that network effects have a positive and significant effect on the conditional probability of adoption, increasing the hazard rate between 6% (lower band) and 11% (upper band) for the average bank. This effect is found to be greater than the scale effect, which increases the hazard rate by 5%. In addition, they find a positive and significant effect for the market wage but a significant and negative effect for bank growth.

³⁴ Variable costs are subsumed into the proportional relationship between bank benefits and depositor benefits and so, strictly, the net present value from adopting ATMs becomes the variable net present

3.3 Concluding Remarks

The early empirical models of diffusion were dominated by their reliance on information dissemination and testing of the so-called Schumpeterian effects of firm size and market structure. Firm size and measures of innovation profitability were found to be significant determinants of the diffusion path.

The contemporary empirical literature on diffusion models has identified a significant role for rank effects and the institutional environment as determinants of the diffusion path. Less support has been given to the existence of stock and order effects, although as noted by Stoneman and Kwon (1994), this may result from research ignoring the complementary and substitute aspects of technologies. More recent empirical research has attempted to incorporate learning effects and network externalities. It is clear, however, that the findings of each empirical diffusion study reflects the uniqueness of the innovation and attributes of the market structure under investigation and that previous approaches and results cannot simply be applied to completely different innovations in different market and regulatory environments.

This chapter indicates advances have been made in the empirical analysis of inter-firm diffusion and this is to be welcomed given that diffusion constitutes a significant component of technical change and progress and which consequently has important implications for both household income and firm productivity. There are, arguably, three main weaknesses in the literature that can be identified for the specific case of ATM diffusion. Firstly, there have been no studies of ATM diffusion in the UK financial sector. Secondly, those studies that have been performed (to date, exclusively for the US experience) have so far ignored the role of prices and price expectations in the diffusion path and consequently may be mis-specified. Thirdly, there is often a lack of rigor and frequent *ad hoc* approaches to model selection and comparison. It is these weaknesses that are addressed in proceeding chapters.

profit [Krepps (1990)].

CHAPTER 4

THE MARKET FOR ATMs IN THE UK

4.1 Introduction

Despite the huge literature examining the wider effects of financial innovation on, for example, the conduct and performance of monetary policy there exists only fragmentary analysis in the economics literature tracing the development of the ATM and its diffusion in the UK. It is necessary, therefore, to provide an understanding of ATM technology, its market environment and to identify possible factors involved in motivating ATM adoption by financial institutions before attempting to construct a theoretical framework to explore its diffusion. Therefore this chapter defines some of the terminology used in association with the technology and to trace the technical developments of the ATM. Secondly, it explores the extent of ATM diffusion in the UK since their commercialisation. Thirdly, it explains the nature of the ATM as a capital-embodied process innovation and to examine the associated nature of factor bias and possible scale and scope economies inherent in the technology, and fourthly, it considers some of the consequences of wider ATM adoption for the financial sector.

The main difficulty with examining ATM technology is that the nature (or 'quality') of the technology has advanced so quickly since commercialisation that at the practical level a detailed survey tracing its development is extremely difficult. This aspect is exacerbated due to a paucity of appropriate literature pertaining to precise technical specifications of different technological generations. This is a frequent problem faced by economists examining the diffusion of new technology [see, for example, Stoneman (1976), Davies (1979) and Colombo and Mosconi (1995)]. This problem is usually resolved in the literature by concentrating on its generic nature, rather than individual generations, and to incorporate technological developments within the modelling procedure [Karshenas and Stoneman (1995)]. In this chapter, however, a clear distinction between first-generation and second-generation ATMs is made as a precursor to subsequent empirical modelling. This distinction is based upon the nature

of the input devices necessary to gain access to an individual deposit account and the subsequent degree of flexibility in the amount of cash that can be withdrawn.

The chapter is set out as follows. Section 4.2 considers the nature of ATMs in terms of their function and technical development. Section 4.3 presents quantitative measures of ATM diffusion in the UK since the commercialisation of ATMs in 1972. Section 4.4 examines the ATM as a capital-embodied innovation, highlighting where applicable the links with the theoretical literature on technical change discussed in Chapter 2. The implications of wider ATM adoption for the financial sector are considered in Section 4.5, where emphasis is placed on the potential for lowering barriers to entry and the effects on money demand. Concluding remarks are collected in Section 4.6.

4.2 The Nature of ATMs

It is widely recognised in the economics literature [see, for example, Akhter (1983), Bank of England (1983), Spencer (1986), Podolski (1986), Pawley (1993), Akçaoglu (1996) and Llewellyn (1992, 1997),] that the financial and monetary sectors of developed economies have experienced rapid product and process innovation, the development of new markets (in some cases enabled by product and process innovation) and fundamental structural change in the last three decades. The UK has certainly not been exempt from these changes and change in the UK can be viewed as a threefold phenomena, although these three aspects should certainly not be viewed as being mutually exclusive [Spencer (1986), Adam (1987) and Llewellyn (1992)]. The first aspect is purely technological and involves, in particular, the application of information technology to the payments system.¹ Such technology includes, *inter alia*, the development of the retail-orientated EFTPOS (Electronic Fund Transfer at Point of Sale), the inter-bank clearing system BACS (Bankers' Automated Clearing Services) and the automated teller machines (ATM). The adoption of ATMs has been concentrated, to date, in the retail banking and building society sectors and, thus, can be viewed as a subset of wider technological change within the financial sector. The second and third aspects are less easy to separate but it is common in the literature

¹ By 'payments system' it is meant the country-specific means by which payments in the retail, industrial and financial sectors are made and debts are settled [Podolski (1986) and APACS (1997)].

[Podolski (1986) and Spencer (1986)] to distinguish between ‘legislative’ and ‘responsive’ aspects. These centre on perceived changes in financial institutions’ behaviour in response to either changes in legislation or to changes in institutions’ behaviour in response to the removal of these legislative constraints (for example, abolition of exchange controls and the removal of the Supplementary Special Deposit Scheme in the UK).²

The focus of this chapter (and thesis) centres on the first of these changes – technological changes - and more succinctly on the diffusion of one particular innovation, the ATM. There is no generally agreed international definition of an ATM but it is broadly accepted in the literature [Jones (1981), Podolski (1986), Kirkman (1987), British Bankers’ Association (1996) and APACS (1997)] that an ATM is a cash dispenser with additional services. At a basic level, the ATM is most significantly a *self-service* device that enables members of the personal sector³ to withdraw cash⁴ from their retail bank, building society accounts (predominately current) and credit card accounts. The nature of the additional services are essentially those traditionally provided by retail deposit taking institutions [Lewis and Davis (1987), Pawley *et al* (1991) and Vesala (1994)] and include *inter alia*:

- withdrawal of cash;
- acceptance of cash and cheque deposits;
- ability to order a cheque book and account statements (and in some cases print mini-statements);
- ability to transfer funds between accounts;
- ability to make bill payments.

The range of these additional services has increased since the first adoption of ATMs in the UK. Moreover, recent developments in the US banking sector have shown that

² A full discussion of these three aspects of change and the issues raised by them lies outside the ambit of this thesis but opening references in this section all contain a comprehensive account.

³ Technically, the personal sector comprises of individuals, unincorporated businesses and non-profit-making bodies serving persons [Bank of England (1997)]. The utilisation of ATMs has, however, been predominately made by individuals.

⁴ Currently, ATMs are only able to issue notes. The minimum withdrawal is £5 and the maximum depends on the nature of the account and the (usually) daily limits imposed by the individual institution. Coins cannot be withdrawn from ATMs.

ATMs can now provide a wide range of financial services such as loan applications and share purchasing [Guglielmo (1996) and The Tower Group (1997)]. The basic function of self-service cash withdrawal remains, however, the defining characteristic of ATMs and accounted for approximately 90% of all ATM transactions in the UK at the end of 1996 [APACS (1997)] and approximately 86% (both measured by volume) in the US at the end of 1991 [Humphrey (1994)].

By tracing the technical development of the ATM it is possible to make the important distinction between first-generation and second-generation machines. The former of these two generations were first introduced in 1967 by Barclays Bank in association with National Cash Register (NCR), with later models developed by International Business Machines (IBM) [Kirkman (1987) and Austin (1992)]. In order to withdraw cash from these machines a customer first had to apply to their relevant bank branch for a book of pre-processed vouchers [Marti and Zeilinger (1982) and Austin (1992)].⁵ These vouchers (known as 'BarclayCash') were at a predetermined value of £10 and only one could be used each day. The voucher had to be inserted into the machine, which would then check its eligibility and, if successful, would issue a single £10 note. These early machines were not actually referred to as 'ATMs', but instead, were known simply as 'bank cash dispensers' as the only service provided was cash dispensing and the customers' account was not directly debited when withdrawing cash from the machine [Kirkman (1987) and APACS (1997)]. Adoption of these machines was restricted to three other clearing banks; Lloyds, Midland and National Westminster, and were eventually phased-out by the late 1970s [British Bankers' Association (1985)]. First-generation machines were beset with poor reliability and their inherent inflexibility in terms of the amounts that could be withdrawn was found to be a major constraint for their acceptance and utilisation by bank customers [Scarborough and Lannon (1988) and Austin (1992)].

In contrast to the inherent inflexibility of first-generation machines, second-generation machines have allowed deposit customers more flexibility in the amount that can be withdrawn from their accounts. These machines were commercialised in 1972 by NCR

⁵ These vouchers were punched cards analogous to those used to communicate data with early generation computers in the 1970s [see Stoneman (1976) and Williams (1985)] and, hence, can be conceptualised as input devices into the ATM for the specific purpose of initialising the withdrawal of cash.

and first adopted by Lloyds Bank in 1973 [British Bankers Association (1985)]. A distinguishing feature of this generation of machines is the nature of the input device required to access the customer's deposit account. This consisted in 1973 (and continues today) of a plastic card (initially an ATM 'cash card' with the single facility of cash withdrawal), with a magnetic strip⁶ attached to the back of the card, issued by the deposit taking institution in association with the deposit account. When the deposit holder required cash the cash card was inserted into the ATM. The holder's personal identification number (PIN) - consisting of a four digit number in the UK - and the amount required was typed via a keyboard into the machine. The cardholder was then able to obtain the required amount, subject to the balance held in their account and the daily maximum imposed by the deposit taking institution. This procedure remains largely unaltered today, except that the nature of the input device has altered. From 1987 the ATM cash card began to be replaced by the multi-functional debit (and credit) card which combined ATM withdrawal facilities with the ability to make retail EFTPOS transactions⁷ together with cheque guarantee facilities.⁸)

When first adopted, second-generation machines were designed operate 'off-line.' Transaction details (such as account number and the amount withdrawn) were recorded on disc or magnetic tape attached to the ATM. After each working day the recorded data was processed at the institutions' head office where the appropriate account would be debited by the value of the cash withdrawal(s). This procedure took twenty-four hours to complete from the time of the original withdrawal and was referred to as 'batch mode processing' [Revell (1986)] as each transaction was dealt with at the end of the working day rather than at the instant they occurred. Consequently, off-line operations were a hindrance to the development of additional services because cash withdrawals

⁶ A magnetic strip is a device that allows pre-recorded information to be stored on the card, such as the owner's personal details and nature of the account [Kirkman (1987)]. The information contained on the strip can be read by the ATM and is a key element in the authorisation and transmission process between an 'on-line' ATM and the central computer belonging to the financial institution which authorises and records the withdrawal.

⁷ EFTPOS consists of a 'on-line' network between participating financial institutions [see APACS (1997) for a comprehensive list of participants] and retail establishments which allows the use of the debit card in retail payments, the card holders account being debited generally two days after the initial transaction is made. For more details of this network see Kirkman (1987) and Howells and Hine (1993).

⁸ The first debit card of this type was introduced by Barclays Bank in 1987 [Lindsey (1994)] and are now issued by other individual financial institutions and by institutions licensed by the card issuers Visa or Switch [Lindsey (1994) and APACS (1997)]. At the end of 1996 there were 56.548 million cards in

were not immediately updated and additionally presented security problems for the financial institution as any fraudulent use of cash cards would not be discovered until twenty-four hours after the withdrawal [Revell (1986) and Kirkman (1987)].

In 1977, however, the Royal Bank of Scotland introduced the first 'on-line' ATMs in the UK in association with IBM and by the mid-1980s on-line ATMs had completely replaced off-line ATMs in the UK [Jones (1981), British Bankers' Association (1996) and APACS (1997)]. A key feature of on-line operations is the telecommunication link between the ATM and the institutions' central computer centre, the latter having the dual role of checking the authenticity of the PIN and corresponding account number and commencing the instantaneous debiting of the appropriate account by the amount withdrawn. This feature of continuous account updating has enabled institutions to extend the range of services provided to balance enquiries and account transfer. Furthermore, it has been a major element in reducing fraud associated with ATM cash withdrawals [Kirkman (1987)].

From the mid-1980s technical developments in ATM technology have focused mainly on hardware⁹ improvements (product innovations) in three key areas [Banking Technology (1986), Kirkman (1987), Austin (1992) and Retail Bank Research (1997)]:¹⁰

- increased reliability and speed of withdrawal through improved size, volatility (ability to retain data once stored) and access time of storage devices at the institutions' head office;
- development of 'packet' switching systems for use in on-line working - this system breaks down each message from the ATM into 'packets' of standard length and routes them to the destination where the full message is reconstituted from the parts. If the destination is off-line temporarily then messages can be stored for delivery;

circulation in the UK that combined these three facilities plus single facility ATM cash cards [APACS (1997)].

⁹ By 'hardware' it is meant that part of the ATM system that receives information, stores it, acts upon it and produces new information [Williams (1985)].

- development of 'switching' technology for use in inter-bank networks
(see Section 4.3),⁷

Unfortunately there is a paucity of literature that has examined the underlying causes of the technological changes that ATMs have experienced since the adoption of first-generation machines. It is, however, possible to identify three likely influences. Firstly, first-generation machines were so primitive and inflexible that deposit holders continued to prefer using counter staff to make cash withdrawals rather than utilise self-service ATMs, thus mitigating one of the main reasons why financial institutions initially adopted ATMs (see Section 4.4). Indeed, evidence for this can be found in Marti and Zeilinger (1982), Scarborough and Lannon (1988) and Austin (1992). Secondly, the nature of the market structure confronting ATM producers may have induced competition to be more orientated towards product differentiation and technological advancement rather than price competition. There is evidence to suggest [Stoneman (1976), Williams (1985) and Podolski (1986)] that the computer producing industry (as a proxy for the ATM producing industry) is (or at least was) characterised by increasing returns to scale. In addition, given that NCR and IBM had captured approximately 83% of the ATM market¹¹ by the end of 1991 [Banking World (1992)] then smaller ATM producers (such as Nixdorf, Olivetti and Dassault) would be forced to compete technologically rather than by price [Tirole (1988)].¹² Thirdly, ATM technology has incorporated ongoing innovations in the electronics components industry, such as integrated circuits, magnetic discs and fibre optics and has enabled, for example, the substitution of the debit card for the simple ATM cash card in the mid-

¹⁰ A recent innovation is the development of alternatives to the PIN. The Nationwide Building Society, for example, has recently explored the use of scanning the customers iris as a means of identification [*Independent*, December 1, 1997]

¹¹ This figure is derived from the ratio of the total number of ATMs in operation supplied by NCR and IBM to the total number of ATMs in operation in the UK at the end of 1991. This figure ignores the differences between the retail banking sector, where NCR and IBM had captured 93% of the market, and the building society market, where, in contrast, both companies captured only 35% of the market in total [Banking World (1992)]. Such a dominance in the retail banking sector may have been derived from NCR's and IBM's involvement in office equipment and mainframe computers (which retail banks started adopting in the 1960s) which provided a 'first mover advantage' in the commercialisation of their ATMs.

¹² Due to innovations in the components industry this may have additionally affected the bias of technical change towards an emphasis on hardware improvements. Moreover, given that NCR and IBM may have built up a considerable amount of reputation in the industry (and given the possibility of increasing returns to scale present at the industry level) there may have existed substantial barriers to entry into the ATM producer industry. Unfortunately there is no evidence to support these conjectures, although Stoneman (1976) finds similar results for the early years of the computer industry in the UK.

1980s. Such innovations have yielded major benefits for the reliability, speed and range of services that the ATM was able to provide [Kirkman (1987)].

4.3 The Adoption and Diffusion of ATMs in the UK¹³

To date the adoption of ATMs in the UK has been confined exclusively to two separate sectors within the financial sector: the retail banking sector and the building society sector [British Bankers' Association (1996) and APACS (1997)]. This is unsurprising given that ATMs have been designed and developed explicitly as a self-service device to enable the personal sector to withdraw cash from deposit accounts and that accounts with these institutions form approximately 50% of the personal sectors total financial assets (and most liquid assets) at the end of 1996 [Financial Statistics (1997)].

Before examining several quantitative measures of ATM adoption and diffusion in the UK contained in Table 4.1, the methodology employed in its construction is discussed since it explicitly relates to the definition of potential ATM adopters utilised throughout this thesis. As indicated in Chapter 3, the empirical analysis of inter-firm diffusion necessitates the definition and selection of a set of potential adopters. In this thesis the set of potential adopters is defined as the *stock* of retail banks and building societies at the end 1992. This date was chosen purely for practical reasons as extensive fieldwork during 1993 and 1994 was required in order to obtain specific dates of ATM adoption and number of ATMs operated by each institution.¹⁴ The criteria used in the selection of relevant retail banks and building societies is given in Appendix A4.1 of this chapter. After excluding those institutions from which appropriate data was not forthcoming a total of 98 potential adopters were identified, of which 12 are retail banks and 86 are building societies. Of these, 35 institutions (12 retail banks and 23 building societies) had adopted ATMs by the end of 1992.¹⁵ Following convention [Karshenas and

¹³ For a recent survey of international ATM use see Retail and Bank Research (1997).

¹⁴ This was achieved by directly contacting the relevant finance, planning and research departments at building societies and retail banks and asking them which year they first adopted second-generation ATMs and the subsequent numbers adopted for proceeding years.

¹⁵ The number of adopters has remained at the same level (and identical adopters) from 1992 to the end of 1996.

Stoneman (1995)] it was assumed that the number of potential adopters was constant throughout the period 1972 to 1996.

Furthermore, in the construction of Table 4.1, figures from 1993 onwards have been collected from APACS (1997) and it should be noted for interpretation purposes that all figures contained in Table 4.1 refer strictly to the set of potential adopters as defined in Appendix A4.1. Specific dates of adoption data pertaining to the number of ATMs operated by each adopting institution from 1972 to 1992 is provided in Appendix A4.2.

Year	Total ATMs operated by Retail Banks (A)	Total ATMs operated by Building Societies (B)	Total ATMs in the UK Financial Sector (A+B)	Change in Total ATMs Operated	Number of Adopting Institutions (C)	Ratio of Actual Adopters to Potential Adopters (C/98)
1972	0	0	0	-	0	0.00
1973	230	0	230	230	1	0.01
1974	337	0	337	107	1	0.01
1975	568	0	568	231	1	0.01
1976	676	0	676	108	3	0.03
1977	875	0	875	199	4	0.04
1978	1000	0	1000	125	4	0.04
1979	1184	0	1184	184	6	0.06
1980	1735	0	1735	551	8	0.08
1981	2805	0	2805	1070	9	0.09
1982	3928	0	3928	1123	9	0.11
1983	5466	17	5483	1555	11	0.15
1984	6534	146	6680	1197	15	0.23
1985	8254	410	8664	1984	23	0.32
1986	9266	766	10032	1368	27	0.33
1987	10615	1134	11749	1717	31	0.36
1988	11743	1599	13342	1593	32	0.36
1989	13172	2063	15235	1893	35	0.36
1990	14089	2641	16730	1495	35	0.36
1991	14444	3022	17466	736	35	0.36
1992	15010	3248	18258	792	35	0.36
1993	15026	3443	18469	211	35	0.36
1994	15637	3588	19225	756	35	0.36
1995	16242	3623	19865	640	35	0.36
1996	16838	4146	20984	1119	35	0.36

Note: Figures from 1993 are taken from APACS (1997).

With these definitions in mind, the number of ATMs operated by retail banks in the UK from the end of 1972 (the date at which second-generation machines were commercialised) to the end of 1996 is given in the second column of Table 4.2. The number of ATMs operated by building societies is given in the third column of Table 4.2. The fourth column is simply the sum of the second column (A) and the third column (B) and represents the total number of ATMs in the UK financial sector, whilst the change in the number of ATMs operated is given in column five. The total number of ATMs is sketched against its corresponding year in Figure 4.1 to graphically illustrate the expansion in the number of ATMs operated over time. The number of adopting institutions over time is given in column six in Table 4.1 and the ratio of adopters to the set of potential adopters (assumed constant at 98) is given in column seven. As indicated in Chapter 2, the movement of this ratio overtime can be employed as a measure of inter-firm diffusion and this is sketched in Figure 4.2 below.¹⁶

From Table 4.1 and Figures 4.1 and 4.2, it can be observed that the growth in the total number of ATMs operated in the UK was most rapid during the period 1982 to 1990, which coincides with the first adoption of ATMs by the building society sector. In addition, the diffusion curve sketched in Figure 4.2 resembles (approximately) a sigmoid curve, characterised by a period of convexity in the first period of diffusion (1972 to 1982), then by concavity during the second half of diffusion (1983 onwards) and then, finally, tending towards a (horizontal) asymptotic level of 0.36 (or 36% of potential adopters). This lends empirical support to the stylised fact frequently made by diffusion theory [Karshenas and Stoneman (1993, 1995)] that the plot of the use of new technology against time yields an S-shaped, or sigmoid, diffusion. Furthermore, it illustrates the time length involved in new technology diffusion. For the banking sector it has taken seven years (see Table A4.2.1 of Appendix A4.2) for half of potential adopters to adopt ATMs since their commercialisation. This is an identical figure to the average diffusion time found by Mansfield (1989) in his study of 14 innovations in the US manufacturing industry. In contrast, the building society sector has taken twenty-one years to reach an adoption level of 27% of potential adopters (see Table A4.2.2 of Appendix A4.2).

¹⁶ Alternatively an output measure could have been employed, although this is not the focus of the thesis and raises the problematic question of how to measure the output of a financial institution (see Section 4.4).

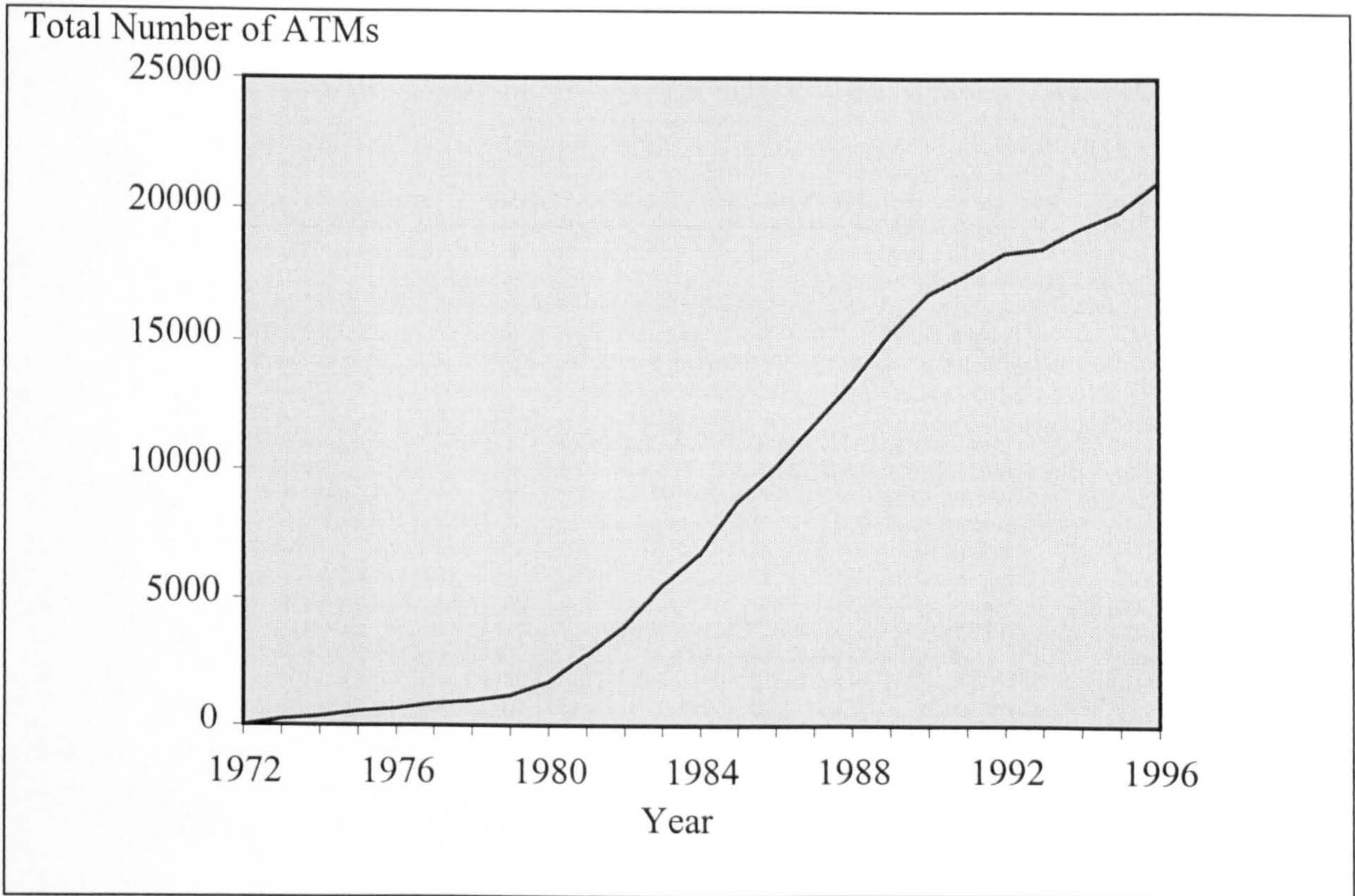


Figure 4.1: Total ATMs Operated in the UK Financial Sector

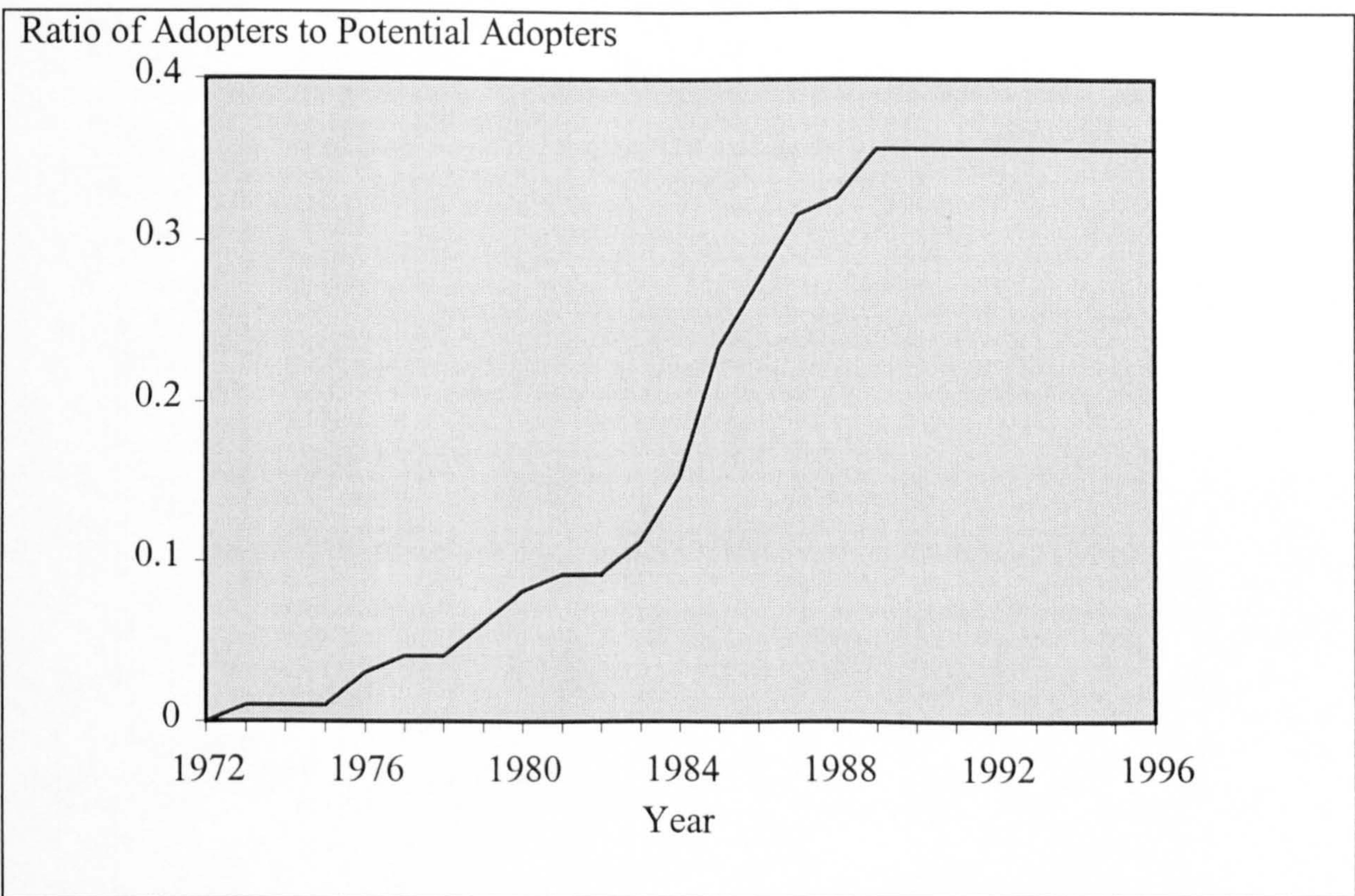


Figure 4.2: Diffusion of ATMs in the UK Financial Sector

The possible reasons for this dichotomy in diffusion speed between these two sectors centre on three main issues [Drake (1989), Pawley *et al* (1991) and Pawley (1993)].

First, the nature of the activities building societies are involved in vis-à-vis retail banks means that ATM technology is less appropriate for providing their services. Second, changes in the regulatory framework in which banks and building societies operated in during the 1980s. The most important of these changes from the viewpoint of ATM diffusion is arguably The Building Societies Act (1986). This Act allowed building societies to participate and compete fully (in non-price and price terms) in the provision of deposit services. Third, building societies may have adopted less quickly than clearing banks because of institution-specific characteristics unique to building societies themselves. This explanation relies on the rank effects model of innovation diffusion. Building societies may have delayed adoption because their gross benefits from adoption were less than those benefits accruing to clearing banks. These issues are more fully explored and examined in Section 4.4.

The figures presented in Table 4.1 do not disaggregate total ATMs by location and this is performed in Table 4.2 which displays the percentage split between three different categories of location. 'Through the wall' ATMs are those embedded into the outside wall of the institution, 'customer area' ATMs are those within an institution's branch and 'remote' ATMs are those that are not physically attached to an institution's branch (for example, outside a supermarket). The figures indicate a move towards greater use of 'remote' ATMs and this may reflect the fact that by 1992 there were more ATMs than branches in the UK (see Figure 4.4).

Location	% of total ATMs at end 1991	% of total ATMs at end 1996
'Through the wall'	64.00	59.00
Customer area	28.00	23.00
Remote	8.00	18.00
Total	100.00	100.00

Source: APACS (1997). Note: figures before 1991 are not available.

The decline in the share of ATMs in customer areas contradicts arguments put forward by McKillop and Ferguson (1993) that financial institutions are attempting to shift the share of ATM locations towards these areas in order to entice customers into the branch. McKillop and Ferguson argue that a shift towards a higher proportion of ATM locations

in branches is aimed at selling customers a wider range of financial services such as insurance and mortgages. This argument does not appear to be borne out by the evidence presented in Table 4.2.

4.3.1 ATM Reciprocal Networks

A major development in the provision of ATM services in the UK since the mid-1980s has been the establishment of reciprocal ATM networks. Before the establishment of these shared networks deposit holders were only able to make transactions at those ATMs operated by the institution that issued their debit (or ATM) card and held their cash deposits. In shared networks each institution issues its own debit (or ATM) cards and has its own proprietary ATM network, but each ATM is connected to other ATMs belonging to the same network [Economides (1995)]. Deposit holders are then able to make transactions at all ATMs belonging to the shared network, although currently these additional transactions are restricted to just cash withdrawals.

The current network structure in the UK is relatively complex and encompasses international as well as domestic linkages. There are currently two distinct networks operating in the UK. The first pertains to those institutions that issue either MasterCard or Visa debit or credit cards.¹⁷ The majority of these cards have ATM cash withdrawal functions. The MasterCard and Visa networks are also linked up with the two large international networks 'Cirrus' and 'Plus' respectively. Current UK membership of these two networks is given in APACS (1997).

The second distinct network pertains to holders of ATM and cash cards that do not have debit or credit facilities. At present, there are three ATM networks of this type in the UK: FOUR BANKS, MINT and LINK. The members of these networks are given in Appendix One. The FOUR BANKS network was the first to be established in 1986 and was set-up by the, then, four largest clearing banks in the UK. The LINK network was established in 1987 from a merger between itself and the MATRIX network [BSA

¹⁷ There are also institutions issuing 'Switch' debit cards that are excluded from these two networks. See APACS (1997) for details.

(1989) and Banking World (1989)]. Both the LINK and MATRIX networks were established in 1986, with the Matrix network originally set-up under the auspices of the Building Societies Association (BSA). The LINK network is currently the only network that building societies participate in [APACS (1997)]. Three clearing banks established the MINT network in 1989 and remain the sole members. The total numbers of ATMs in each of these networks at the end of 1996 are given in Table 4.3 below.

Table 4.3 Reciprocal ATM Network Sizes in the UK - end 1996	
Name of Network	Number of ATMs in Network
MINT	7865
LINK	7762
FOUR BANKS	6941

Source: APACS (1997)

The establishment of shared networks has only been possible by the development of packet-switching technology that allows transactions to be routed to the appropriate account that issued the debit card [The Bankers' Magazine (1982) and Banking World (1984)]. This technology has been developed through joint ventures with vertically related computer and telecommunications firms [Kirkman (1987) and Banking World (1992)]. In particular, the packet-switching technology employed by the LINK network has been developed by International Business Machines (IBM) and British Telecom (BT). The computer software firm Nexus have developed that employed by MINT [Banking World (1989, 1992)]. As noted by Carlton and Klammer (1983) and Economides and White (1993), joint ventures such as these are able to set compatibility standards in ATM software and hardware technology through co-ordination between institutions. Moreover, compatibility between differently owned ATMs has strong welfare benefits for depositors because they are able to access a greater number of geographically dispersed ATMs [Saloner and Shepard (1995)]. Katz and Shapiro (1984, 1985, 1994) and Economides (1996) refer to the positive benefits that deposit holders obtain from greater ATM compatibility as 'network externalities' or 'demand-side externalities.' These externalities have important implications for, *inter alia*, the adoption and diffusion of ATMs and are more fully explored in Chapter 7.

During the early 1990s access to ATMs belonging to rival networks and the subsequent pricing of these transactions became a particularly contentious issue between financial institutions. Holders of Visa debit (and credit) cards issued by the FOUR BANKS network were initially blocked by the MINT network from using their ATMs. From 1995 access was granted only to those depositors who held cards with the Visa insignia, but with the addition of interchange fees (that is, transaction fees charged to depositors for use of ATMs not belonging to the network). Currently, interchange fees are set at a minimum level of £1.50; otherwise the fee is 1.50% of the total cash withdrawal [see Retail Bank Research (1997) for further details of pricing policies]. In general, holders of non-Visa or non-MasterCard cards are denied access to rival ATM networks [Retail Bank Research (1997)]. For the majority of institutions, however, ATM withdrawal fees (that is, a transaction fee charged to depositors of other compatible institutions for use of one's own ATM) remain free at the point of use.¹⁸

There is also a system of inter-institution Visa transaction fees (that is, fees charged to each other institution for use of its ATMs), but the nature of these fees remain largely unpublished [Retail Bank Research (1997)].

Matutes and Padilla (1994) have examined the incentives for ATM compatibility within a two-period locational game based on Salop's (1979) 'circular-city' model of oligopolistic competition [see Tirole (1988)]. In the first period institutions propose a compatibility agreement which states which institutions are to join the shared network. In the second period institutions compete for retail deposits and simultaneously set deposit rates to maximise discounted profits. Matutes and Padilla show that in considering compatibility with other institutions an institution have to trade-off the 'network effect' and the 'substitution effect.' The network effect derives from the willingness of deposit holders to accept lower interest rates on their deposits for a given increase in the size of the ATM network. Deposit holders' value a greater number of geographically dispersed ATMs because they reduce transport costs associated with unexpected cash demand. The network effect necessarily creates incentives for institutions to reach compatibility agreements. In contrast, the substitution effect derives from the credible threat of deposit holders to switch deposits to that institution

¹⁸ The exceptions are the National and Provincial and Nationwide building societies [APACS (1997)].

offering the highest deposit rate post-compatibility. The substitution effect increases price competition between institutions and acts as a disincentive for institutions to share ATM networks. Matutes and Padilla argue that the substitution effect is higher the lower the proportion of transactions that can only be done at the institution the depositor has an account with.¹⁹ Nash equilibrium in this game is shown to result in partial compatibility between institutions. Full compatibility cannot denote Nash equilibrium, as there is always an incentive for an individual institution to deviate and increase profits. Matutes and Padilla show that the imposition of withdrawal fees or interchange fees can limit the extent of the substitution effect of compatibility. The imposition of fees can thus be used as a mechanism by institutions to achieve full compatibility.

Economides and White (1993), Economides (1995) and McAndrews and Rob (1997) have argued that shared ATM networks between vertically integrated institutions although increasing the available network size to depositors may also act as an entry barrier to potential entrants even in the absence of economies of scale. This entry barrier may take the form of simply refusing potential entrant access to a network 'switch'. Alternatively, compatible institutions may impose a fixed sunk cost of joining the network such that the profits gained by the entrant from either joining the network or being incompatible are lower than the value of the fixed cost [Matutes and Padilla (1994)]. Furthermore, compatible institutions may use the network benefits of providing a larger network of ATMs to capture a larger market share of deposits and forcing incompatible institutions out of business. As noted by Matutes and Padilla (1994), this latter outcome depends on the strength of the network benefits that deposit holders gain from the increased network size. In 1994 UK holders of Fidelity's 'GM' card were refused from using the MINT and FOUR BANKS networks. This was viewed by many commentators as being anti-competitive [*The Sunday Times* (July 1994) and *The Guardian* (October 1994)].

¹⁹ Or, conversely, the substitution effect is higher the higher are those transactions that can be performed at compatible ATMs (such as cash withdrawal, for example) as a proportion of total transactions.

4.4 ATM Diffusion and Technical Change

Current ATM technology can be interpreted as the application of information technology (IT) by financial institutions in the provision of deposit services to the personal sector [de Wit (1990) and The Tower Group (1997)]. At the purely technical level, IT represents the merger of computer technology (embodied in the ATM and the institutions central computer) and telecommunications (forming the link between the ATM and the central computer). This interpretation forms a demarcation between first-generation and second-generation machines discussed in Section 4.2.

Given the generic nature of IT, however, it is conceivable that the diffusion of ATMs has been influenced by the diffusion of complementary innovations, such as the electronic component technology underlying the debit card for example. This point is even more pertinent if the ATM is given a 'radical innovation' interpretation in the vein of Freeman (1984, 1988). As indicated by Grübler (1991) and Nakicenovic (1991), the diffusion paths of this class of innovations are either competing with many rival technologies and/or are influenced by the diffusion of complementary innovations. The second influence has arguably had more significance in the case of ATM diffusion. Firstly, because the decision to adopt ATM technology has been viewed by financial institutions as being strictly a choice between an unchanging 'old technology' (i.e. providing deposit services by counter staff) and the 'new technology.' Evidence for this proposition is presented shortly. Moreover, because, as shown in Section 4.2, the technical development of the ATM can be traced to innovations in the electronics components industry.

Diederer *et al* (1990, 1991) have gone a step further in this analysis and argued that IT should be regarded as a new 'techno-economic paradigm' [Dosi (1983), Freeman (1984), Perez (1985) and Metcalfe (1995)] characterised by, *inter alia*, obvious potential for all-pervasive influence in the productive sphere of the economy²⁰ and a generally recognised capacity to reduce production costs and change the quality of capital equipment, labour and products. Such a new paradigm, as illustrated by Freeman

²⁰ See Guy (1987) for sectoral-specific penetration measures of IT in the UK economy and discussion of the subsequent changing nature of the capital stock in the service industry in particular.

(1984), produces 'clusters' of interrelated products, processes, technical and organisational innovations affecting many branches of the economy.²¹ The ATM can, thus, be conceptualised as one of these 'cluster' innovations resulting from technological breakthroughs in electronics and subsequent IT developments.

Furthermore, ATM technology can be analysed within the evolutionary approach to technical change and diffusion [Nelson (1968), Nelson and Winter (1982), Andersen (1994) and Metcalfe (1995)] introduced in Chapter 2. In these models 'technology' is not defined in the conventional manner as a single capital-embodied innovation (or, in general, an 'artefact'), but rather, is distinguished by a unique design configuration and specific architecture at a particular point in time. In this framework, the initial commercialisation of ATMs can be given a 'radical innovation' interpretation which simultaneously creates a new design architecture and new knowledge bases in science and engineering which underpin evolving design configurations [Hendersen and Clarke (1990) and Metcalfe (1995)]. ATM diffusion can then be conceptualised as the diffusion of a sequence of artefacts (differentiated by incremental changes in design configuration) with the architecture virtually unchanged [Bell and Pavitt (1993)]. In addition, the evolutionary approach emphasises positive feedback between users and suppliers of ATMs via 'producing-by-learning' and 'learning-by-using' processes [Rosenberg (1982)] as an important source of improved design configurations. This is, arguably, a close approximation to the technical development of the first-generation ATM as outlined in Section 4.2. Moreover, formal inter-firm diffusion models in this vein [Metcalfe (1988) and Antonelli *et al* (1992)] model the diffusion process as process of selection of competing technology vintages within a non-equilibrium framework. Those firms with the 'better' technologies re-investing their extra profits and increasing their market share, while those with 'inferior' technologies realise losses and either contract or exit. Thus, diffusion in evolutionary models is concomitant with the increase of its adopters' market share [Lissoni and Metcalfe (1994)].

²¹ It has been argued by Diederer *et al* (1990, 1991) that these wider effects are analogous to the 'clustering' of process innovations described by Schumpeter (1939) in his theory of business cycles. Although not denying the existence of such 'clusters', Nelson (1996), however, questions Schumpeter's (1939) hypothesis that such clustering is the main cause of the subsequent 'upswing' in economic activity.

Within the evolutionary framework, therefore, the adoption and diffusion of ATMs by the financial sector is not an isolated act, but rather, is part of a greater process of change in which the adopting institutions change their organisation and culture. Indeed, as shown by Scarborough and Lannon (1988) in their case study of the Royal Bank of Scotland and Bank of Scotland, IT has led to managerial and organisational innovation within banks. Both these banks had set-up new 'automation planning departments' in the mid-1980s as a consequence of the wider use of IT (including ATMs) in their organisations which had the dual aim of:

... acting as a catalyst to stimulate awareness of technology throughout the organisation, at the same time presenting senior management with a coherent view of the long-term implications of new technology for banking operations. [Scarborough and Lannon (1988), p.46].

Similar evidence concerning organisation restructuring as a consequence of ATM adoption can be found in Howells and Hine (1993), Pawley (1993) and Llewellyn (1997), further lending support to the predictions made by evolutionary models.

The evolutionary approach additionally predicts that the 'technology trajectory' of a new paradigm (in this context IT) will be dependent upon the social, cultural and institutional framework of the relevant country under investigation [Dosi (1983) and Nelson (1992)].²² Consequently, there are likely to be inter-country differences in the technological trajectory of IT and, hence, also in the trajectory of ATMs. In this context, Diederer *et al* (1990, 1991) and de Wit (1990) have argued that the technology trajectory for IT in the financial sector is largely dependent on the prevailing payments system in that country. As shown by Kirkman (1987) and Pawley (1993), two broad types of payments systems can be distinguished. These are the giro system and the cheque system. The giro system is the prevailing system in Scandinavian countries, while the latter is the prevailing one in the UK and USA. As shown by Vesala (1994), adoption of ATMs has been earlier and on a larger scale (per capita) for those countries in the former category vis-à-vis those countries in the latter category for the period 1983 to 1990. Diederer *et al* (1990, 1991) have argued that the main reasons for these differences is inter-country heterogeneity in 'technological opportunities' resulting from

²² This hypothesis is linked to the concept of National Systems of Innovation. See Nelson (1996) and Edquist (1997) for a recent comprehensive exposition of this approach and Gourlay (1998a) for a critique.

institutional and regulatory asymmetry. Diederer *et al* argue that in giro-based systems it is more profitable to invest in large mainframe computer systems at the banks' central offices in order to automate giro payments. In contrast, IT related investments in those areas leading to the automation of periodical payments and withdrawals are more profitable in cheque-based systems. As a consequence, not only were ATMs introduced earlier in cheque-based countries but so were credit cards.²³

These issues raise the question of why the financial sector has provided such extensive opportunities for the application of IT and, therefore, for the adoption of ATMs. An initial insight into answering this question can be found in the financial sectors' perception of what IT implies for their industry:

IT can be seen as the use of computers, microelectronics and telecommunications to help produce, store, obtain and send information in the form of pictures, words or numbers more reliably, quickly and cheaply. [Banking Information Services (1982), p. 12].

With this definition of IT in mind, Podolski (1986), Diederer *et al* (1990, 1991) and Llewellyn (1997) have identified three main factors to explain the relatively high rate of IT application in the retail banking and building society sectors:

- financial institutions' role in the storage, retrieval, processing and modification of information;
- operation of an extensive branch network;
- large funds available for IT investment.

The first two of these three factors both relate to the technological opportunities inherent in the nature of the services provided to the personal sector by retail banks and building societies. The first relates to the main functions required in the provision of deposit services.²⁴ Niehans and Llewellyn (1997) have interpreted the wider utilisation of IT by financial institutions as reflecting their desire to reduce the costs of these four functions. Indeed, the literature has identified the reduction of transactions costs, by

²³ Vesala (1994) finds no significant role for macro-economic variables, such as GDP per capita or relative labour costs in explaining inter-country differences in ATM intensity (defined as the number of ATMs in operation divided by the total number of branches).

²⁴ These can be loosely interpreted as being 'transactions costs' [Podolski (1986)].

displacing paper and people, as the main motivation in the adoption of ATMs by financial institutions (see Section 4.4.1). The second factor has two elements to it. The first refers to the necessity for institutions to link the operations of their branch network to their central office for the authorisation and recording of transactions. The development of IT has enabled these requirements to be carried out instantaneously for ATM transactions. The second element is based on the networks literature [Economides (1996)] which suggests that extensive branch networks provided by retail banks and building societies (as a proxy for the 'networks effect') have provided incentives to share ATMs and allowed smaller institutions to realise economies of scale [Matutes and Padilla (1994) and Saloner and Shepard (1995)]. Such incentives, it is argued, do not exist in other sectors. The basis for these arguments is more fully explored in Chapter 7 of this thesis. The last factor is based on the perceived effects on new investment expenditure that may have to be imposed by liquidity constraints facing financial institutions vis-à-vis manufacturing firms during the 1970s. There is, however, no empirical support for this last point, although an attempt is made in Chapter 6 of this thesis to explore the possible effects of liquidity constraints on the timing of ATM adoption.

It is possible to go a step further in the above analysis and consider more closely three key aspects of ATM technology most frequently considered in the literature [Humphrey (1994) and Vesala (1994)]: factor bias, economies of scale and scope and ATMs as a competitive strategy. The first two aspects refer to ATM technology as a process innovation for financial institutions. The last, although not excluding the possibility of the ATM as a process innovation, allows for the potentiality of a product innovation interpretation of the ATM. Although, arguably, evolutionary economics offers a broader view of ATM technology it does not readily lend itself to such an analysis. It is therefore appropriate to examine these first two attributes within the conventional framework of a single capital-embodied process innovation as introduced in Chapter 2.

4.4.1 Factor Bias and ATM Technology

A number of commentators [most notably, Jones (1981), Podolski (1986), Scarborough and Lannon (1987), Hannan and McDowell (1984a, 1984b, 1987) and Llewellyn (1997)] have argued that the main motivation in the decision of financial institutions to adopt ATMs has been its inherent 'labour-saving' qualities and subsequent reduction in the average costs of producing deposit services. This proposition immediately raises two interrelated questions concerning the nature of technical change. Firstly, what precisely is meant by a 'labour-saving' technique? Secondly, given a definition of such a technique, what is the evidence that ATMs are inherently labour-saving?

The issue of factor bias can be addressed within a framework of analysis that interprets the ATM as a capital-embodied process innovation in the vein of Solow (1960) and Burmeister and Dobell (1970). Although ATM technology certainly contains disembodied elements (such as the creation of new organisational structures), the main technical change it invokes is arguably embodied in design changes built into new ATM machines. If this argument is accepted, then the output of deposits provided by ATMs will be given by a *vintage production function*. For the case of two factors of production - labour and capital - this type of production function has the following specification [Burmeister and Dobell (1970) and Stoneman (1983)]:²⁵

$$Q(v, t) = F[K_v(t), L_v(t), v] \quad (4.1)$$

where $Q(v, t)$ is the output of deposits from ATM machines of vintage v , $L_v(t)$ is the allocation of labour (in man-hours) to work on machines of vintage v at time t , and $K_v(t)$ is the capital stock of vintage v at time t . The production function in (4.1)

²⁵ A major objection to the specification of the production function in (4.1) for a financial institution is that it ignores the dual role that deposits play in the production process. As noted by Lawrence and Shay (1986) and Drake (1992), deposits for banks and building societies provide liquidity and transactions services but also provide an input (together with capital and labour) in the production of earning assets. The issue here, however, is not to provide a precisely defined production function (which lies outside the ambit of this thesis), but rather, as a means of providing an insight into ATM factor bias.

satisfies the usual properties of a general production function [see, for example, Gravelle and Rees (1992)].²⁶

The form of the production function represented in (4.1) pinpoints the assumption that it is the date of manufacture v of the capital good (the ATM), rather than the current time t in disembodied models, that determines factor productivity. Unfortunately, there is no neat, or completely general, definition of factor-bias that can be summarised from (4.1) and there is considerable debate amongst economists surrounding the exact meaning of bias and its measurement [Stoneman (1983)]. Factor-bias can only be defined relative to some measure of neutrality and in this context, Thirtle and Ruttan (1987) have argued strongly in favour of Salter's (1966) definition of factor neutrality at the firm and sectoral level. Using the vintage production function in (4.1), this is defined (when viewed as a continuous process) in terms of the proportional change in the capital-labour ratio at constant factor process and is measured with respect to different vintages. This gives the following expression:

$$\frac{\partial [K_v(t)/L_v(t)]}{\partial v} \left[\frac{1}{K_v(t)/L_v(t)} \right] \left| \begin{array}{l} > 0 \text{ labour - saving} \\ = 0 \text{ factor - neutral} \\ < 0 \text{ capital - saving} \\ \text{factor prices } (W/R) \text{ constant} \end{array} \right. \quad (4.2)$$

where W is the nominal wage rate and R the rental price of capital.

From (4.2), technology is factor-neutral when the capital-labour ratio does not change with factor prices remaining constant [Salter (1966)]. The adoption of a technique that is 'labour-saving' (such as ATMs) in the Salter-framework will therefore lead, at constant factor prices, to a rise in the capital-labour ratio and a fall in total unit costs of production (assuming technologies employed are those which minimise unit costs). This result does not imply that the technique does not save only labour, but simply that the proportionate saving in labour is greater than the proportionate saving in capital. Consequently, if the ATM is indeed labour-saving then it is expected that financial

²⁶ Importantly, the definitions of factor-bias below depend on the production function being linear homogenous [(Thirtle and Ruttan (1987))].

institutions would have experienced an increase in their capital-labour ratios from the time of adoption of ATMs.²⁷

There is, however, considerable debate in the literature over the sources of factor-bias and whether bias is purely technical, induced (by, for example, bottlenecks or labour disputes)²⁸ or can result from changes in relevant factor prices [Thirtle and Ruttan (1987)]. In the seminal work of Salter (1966), the hypothesis of induced-innovation is rejected and, instead, three measures of movements in 'best practice techniques' (i.e. technical change) are identified. First, the general effect of technical advance. Second, the bias effect arising out of technical advance which tend to save more of one factor than another, and third, the substitution effect reflecting changes in relative factor prices, including those arising out of technical progress in the manufacture of capital goods. Revell (1986) and Podolksi (1986) have identified the last effect as one of the main causes of wider ATM adoption in the UK financial sector. Both argue that there has been a downward trend in the price of ATMs due to technical progress in the ATM-producing and components industries and an upward trend in real wages in the financial sector since 1972, leading to institutions adopting 'labour-saving' technology. Their arguments are weakened, however, by the anecdotal nature of the evidence presented in support of their hypothesis.

Blaug (1963) and Stoneman (1976, 1983) have however, vigorously disputed the possibility that changes in relative factor prices are a source of technical change. First, in Salter's (1966) model the production frontier (or isoquant) defining the technically efficient level of output embraces all possible designs conceivable by existing scientific knowledge. According to Stoneman (1976, 1983), this not only confuses the invention and innovation stages of technical change as distinguished by Schumpeter (1934), but additionally implies that factor substitution inevitably embraces technical change. In

²⁷ This definition of factor-bias is essentially the reverse of the more frequently used economy-wide Hicksian definition of a factor-neutral technique which, with given factor proportions, raises the marginal product of labour in the same proportion as the marginal product of capital [Thirtle and Ruttan (1987)]. Thus, a Hicksian labour-saving technique with given factor proportions will decrease the marginal product of labour relative to that of capital. The two definitions do not differ in their division of labour-saving and capital-saving techniques and, as noted by Binswanger (1978), the definition of factor-bias given by Hicks (1948) is equal to that of Salter (1966) multiplied by the elasticity of substitution.

²⁸ The induced-innovation models of von Weizsacker (1966) and Ahmad (1966) invoke the possibility the existence of an 'innovation possibility frontier'. In these models changes in relative factor prices 'induce' savings in those factors which have a relatively high share in the firm's total costs.

contrast, Stoneman (1976) argues strongly that the production frontier should strictly be defined to include only commercially available techniques immediately for use.²⁹ Second, Stoneman (1976, 1983) disputes the result that a cheapening of capital goods in capital-producing industries ('exogenous substitution') will result in factor substitution in the capital-using industries. Rather, if capital goods are measured in terms of the consumption good as an input into both sectors in the vein of Rymes (1971), then there will be Harrod-type technical change in both sectors. Technical change will subsequently be greater in the capital-goods sector relative to the consumption-goods sector. Thus, Salter (1966) confuses inputs and outputs.³⁰ Third, Stoneman (1976, 1983) does not deny the possibility of 'endogenous substitution' (caused by, for example, an economy-wide autonomous rise in real wages) at the theoretical level but argues that this implies that wages have to be institutionally determined and is therefore in conflict with marginal productivity theory. Freeman and Soete (1987) have, however, argued that engineers and technologists act on the *expectation* that there is a slow, but systematic, increase in real wages (relative to the price of capital).

More contemporary views on factor bias and the associated aspect of employment changes brought about by the adoption of new technology have distinguished between the separate effects of production innovation (changes in demand for the final product) and process innovation (changes in the production function). Katsoulacos (1986), for example, has studied the short-run and long run effects of product innovation (both horizontal and vertical differentiation) in imperfect markets by introducing a price-cost margin different from unity. The theoretical results obtained by Katsoulacos indicate that product innovation is more likely to have positive effects on industry-wide levels of employment as compared to process innovations.

As discussed by Bosworth (1983), Stoneman (1987), Shy (1996) and Van Reenan (1997) the employment effects of wider process innovation adoption are difficult to predict. The majority of theoretical models investigating the employment effects of

²⁹ Rosenberg (1976a) has gone a step further and argued that firms do not face smooth continuous isoquants and, moreover, if factor substitution creates new knowledge via the necessity of R&D expenditures then, strictly, this has to be defined as technical change.

³⁰ In defence of Salter (1966), it is clear that he is aware of this issue. Salter argues that Harrod's (1948) definition of technical change, which assumes an economy-wide constant capital-output ratio, is more appropriate when interest lies in the net effects of technical change.

new technology adoption assume that markets are perfectly competitive operating under a constant returns, constant elasticity of substitution production function and invoke the Hicksian concept of neutral technical change. When industry level output is fixed in this modelling framework the effect on technical change will depend on the degree of substitutability between capital and labour. When the level of output and capital are allowed to vary, then employment effects depend on the elasticity of demand for the final product and how 'radical' the new technology is in terms of reducing average costs of production.

At the firm level of disaggregation, however, early adopters (or 'first-movers') may enjoy a high market analogous to the order and stock effect mechanism outlined in Chapter 2. Consequently, although ATM technology may have a long-run negative effect on employment levels in the financial sector this effect may be offset by the fact that early-adopters enjoy a (temporary) increase in market share. The overall employment effects of wider ATM adoption is then ambiguous and can arguably only be resolved by empirical investigation. The situation is more complicated as the overall employment outcome depends on the specific definition of labour involved in transaction services [de Wit (1990)]. To date, no empirical study has been carried out to investigate the overall employment effects of wider ATM adoption in the UK financial sector. This offers a definite opportunity for future research.

Bosworth *et al* (1996) has noted that the effects of the labour market may impinge on the diffusion process. Bosworth *et al* envisage a new technology that reduces labour per unit of output analogous to that studied by David (1969, 1975) and Davies (1979). Bosworth *et al* argue that diffusion of such a technology may reduce employment levels for a particular skilled labour and, therefore, lead to a fall in real wages. As real wages fall then adoption becomes *less* profitable for some firms rather than more profitable as predict by the majority of rank effect model outlined in Chapter 2 which, in general, assume that real wages are either unaffected by technology adoption or that real wages increase over time.³¹ According to Bosworth *et al* the effects on the industry level of employment depend on the prevasiveness of the technology and the flexibility of real wages. This perspective does, however, ignore the possibility that expectations of

³¹ The mechanism is, however, entirely identical as predicted by rank effect models.

higher market shares may dominate the effect of falling real wages [see, for example, Shliefer (1986)]. For the case of ATM technology there is anecdotal evidence to suggest that the overall effects of ATM diffusion may be positive due to the increase in labour demand in complementary sectors [Vesala (1994), The Tower Group (1997) and Retail Bank Research (1997)]. These sectors include, for example, post-sales servicing of ATMs, the electronics sector and marketing. The effect on employment of wider ATM adoption is, however, highly sensitive to the definition of labour and level of aggregation chosen for analysis.

The empirical evidence examining the effects of ATMs on factor-bias and employment that does exist is sparse and most studies have investigated the effects of IT rather than ATMs singularly. De Wit (1990) has investigated the effects of IT on factor-inputs in the Dutch Automated Clearing House (BGC). He argues that technological change in the financial sector has been 'localised' in nature, affecting one technique of production and leaving other techniques unchanged. Consequently, he argues that the production function facing financial institutions resembles a Leontief type function [Gravelle and Rees (1992)] in which factor substitution (resulting from changes relative factor prices) leaves factor levels and proportions unaffected. A model of technical change is then developed similar to that of Salter's (1966), in which the relative change in unit requirements of factors (capital and labour) are expressed in terms of the rate of technical change and a weighted sum of its biases with other factors of production. Output is measured by the volume of payment transfers. Empirical results obtained by de Wit for the Dutch banking sector indicate that the adoption of IT has increased labour inputs of non-routine labour and information technologists and decreased inputs of production labour. In addition, results suggest that IT has reduced the capital-input of buildings. Diederer *et al* (1990, 1991) calculate input coefficients for a panel data set of 100 Dutch bank branches between 1975 and 1987 by aggregating costs of inputs of bank branches and deflating by a price index and dividing by production volume. Results indicate that the adoption of IT has resulted in labour-savings in particular areas of bank services such as data input work and processing of transfers and checking of balances. Lawrence and Shay (1987) estimate a translog cost function for a panel data set of 632 American banks between 1979 and 1982. Their result show that the elasticity of substitution between ATMs and those labour-inputs involved in counter work has

increased over the sample period. Such a result undermines the assumption that new technology in the financial sector has resulted in a Leontief type production function and supports the arguments put forward by Salter (1966) that new technology increases the elasticity of substitution between capital and labour.

The empirical studies of de Wit (1990), Diedreren *et al* (1990, 1991) and Lawrence and Shay (1987) illustrate that inferences pertaining to the factor-bias effects of new technology are sensitive to the particular definition of 'labour' and 'capital' involved in the production process of financial institutions. To elucidate this, Figure 4.3 below shows the movement of the ratio of aggregate branch staff employed to total number of branches for the sample of financial institutions defined in Appendix A4.1 for the period 1972 to 1992. The graph shows that this ratio has increased over the sample period, although starts to fall slightly from 1990. Initial observation of this graph suggests that ATMs have not reduced branch staff in the UK. Such a conclusion has, however, to be qualified by noting that the main consequence of ATM technology may have been on a reduction of those labour-inputs involved in counter and data input functions. Support for this cannot, arguably, be gained from aggregate labour measures and it may be the case that labour saved by adopting ATMs could have simply be re-deployed in other commercial activities of the branch such as personal customer services and ATM daily maintenance [see, for example, evidence provided in Scarborough and Lannon (1987) and Haynes *et al* (1990)].

Moreover, the results of de Wit (1990) suggest that the adoption of IT in the financial sector has resulted in capital-saving (buildings) as well as labour-saving. As noted by Salter (1966) the possibility of a singular factor-bias technique is indeed a special case:

... advances which allow absolute savings of only labour or capital are likely to be relatively rare, for they involve the special case of a zero elasticity of substitution. The more general case, at least when such proportional advances are involved, would imply some absolute savings of both factors, though by no means equal savings. [Salter (1966), p. 34].

Vesala (1994) shows that the ATM/branch ratio (used as a measure of the proximity of institutions' services and customer convenience) for deposit taking institutions across ten European countries (including the UK) from 1983 to 1992 has fallen in nine of the ten countries sampled. Vesala argues that this supports the hypothesis that financial

institutions are substituting ATMs for branches in the provision of deposit services and, thus, that ATMs are capital-saving. In nine of the ten countries (Belgium being the exception) institutions had reduced their branch network over the sample period. Similar evidence (after adjusting for changes in bank legislation) is presented in Barthel (1992, 1993) and Humphrey (1994) for the US banking sector from 1973 to 1992. In addition, Humphrey (1994) also shows that during this sample period that there is a positive and significant relationship between the deposit/branch ratio and the ATM/branch ratio. This result suggests that ATM adoption by banks has allowed the number of branches to decline while supporting the same level of deposits. Humphrey argues, however, that this result does not imply that ATMs are being used as a substitute for branches and that a more appropriate measure is the population/branch ratio. This ratio (after adjusting for changes in bank legislation) has actually risen in the US since 1972 therefore supporting the capital-saving hypothesis.

The ratio of ATMs to the number of branches for the sample of financial institutions used in this thesis is illustrated in Figure 4.4. The ratio is drawn for the period 1972 to 1996. This indicates a sharp rise in the ratio since 1972 and illustrates the fact that there are now more ATMs than branches in the UK.³² During this period the branch network for the sample of financial institutions in the UK has fallen from 19007 in 1972 to 16660 in 1996.³³

³² Interestingly, the Spanish monetary authorities consider an ATM to be half a branch for regulatory purposes [Vesala (1994) and Matutes and Padilla (1994)].

³³ It should be noted that from 1972 to 1988 the number of branches operated by building societies in the sample increased from 1774 in 1972 to 5775 in 1988 [BSA (1996)] through a combination of branch expansion and merger and acquisition activity. If a population figure of BSA members were used the rise in the ATM/branch ratio would be even more dramatic than that contained in Figure 4.4.

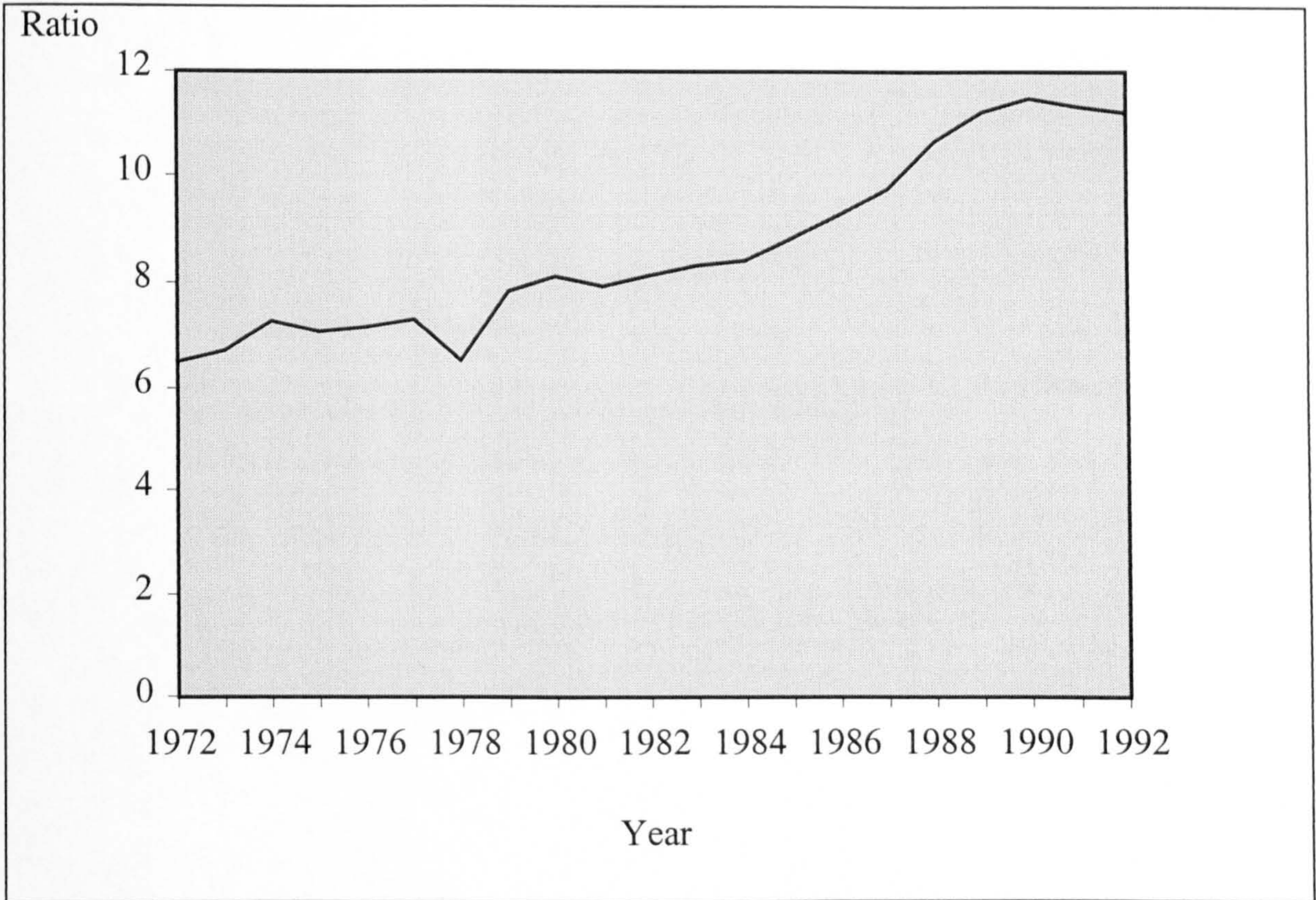


Figure 4.3: Ratio of Branch Staff to Branches in the UK

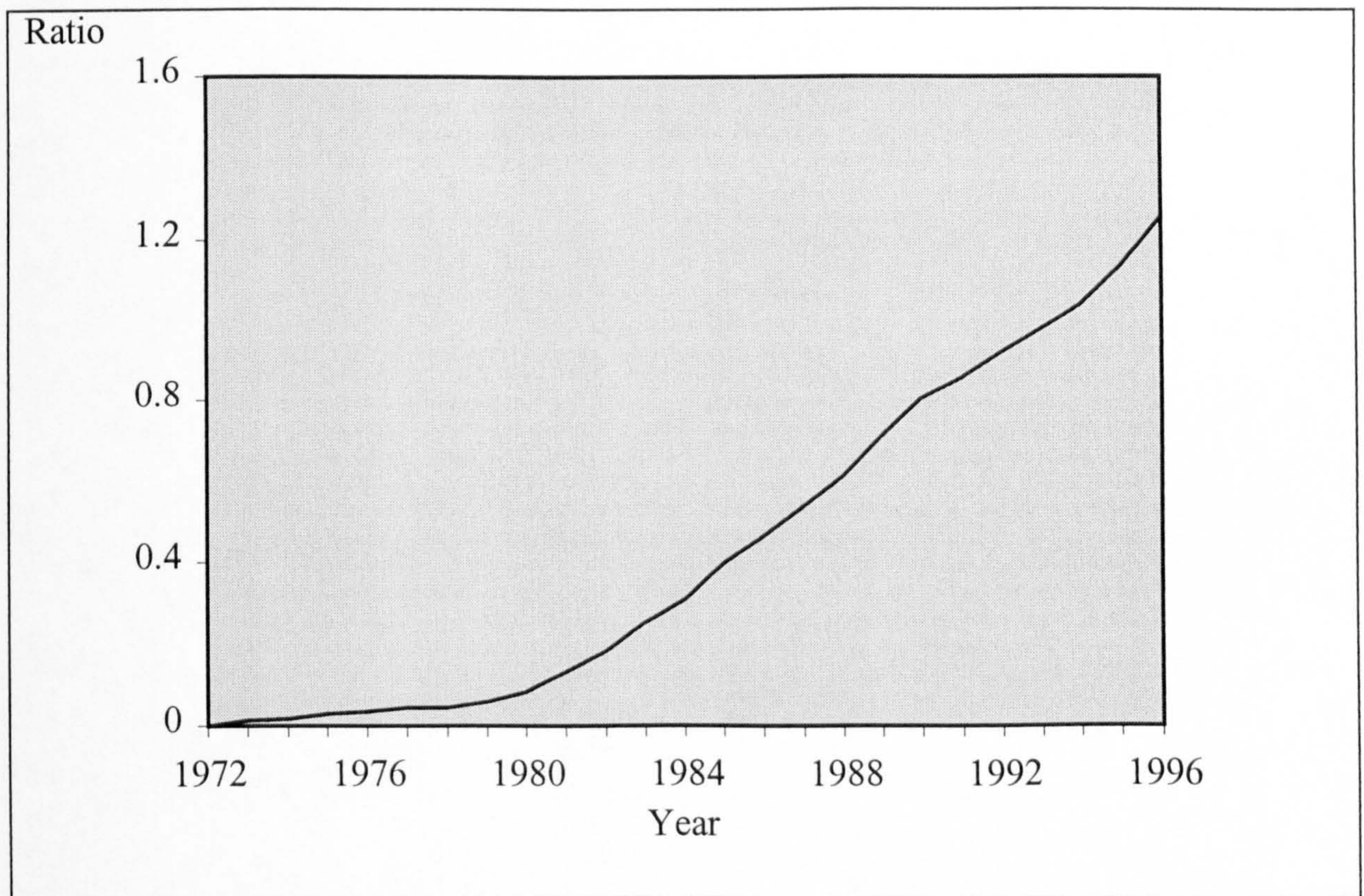


Figure 4.4: Ratio of ATMs to Branches in the UK

4.4.2 Scale and Scope Economies and ATM Technology

The potential of scale economies in IT applications in the financial sector (including ATMs) have been identified by Podolski (1986), Kirkman (1987), and most recently by, Llewellyn (1997) as being a major source of change in the financial sector, affecting the nature and type of competition between financial institutions. The source of these scale economies in the case of ATMs arise from the high fixed costs (relative to operating costs) of setting up both a proprietary and shared network. This results in the marginal cost of providing deposit services to fall over a large range of deposit volumes [Vesala (1994)].³⁴ The main debate in the literature has centred on whether these scale economies are converted into lower average costs of producing deposit services vis-à-vis providing deposit services through a branch network.

Empirical studies investigating the nature of ATM scale economies have, unfortunately, been exclusively for the US banking sector. An early study by Walker (1978), using a simple log-linear equation relating ATM total costs to ATM transactions for a cross section of US banks, found scale economies in 1975 to be 0.50. In addition, Walker (1978) and Berger (1985) find that a monthly ATM transaction volume of approximately 5,000 implied that the cost per ATM transaction was below that of a branch. Such a finding was close to the average transaction volume in US banking of 6,000 in 1992 [Humphrey (1994)] implying that there may be cost savings to be obtained from adopting ATMs.

Early studies have ignored the multi-product nature of financial institutions cost functions and with this concern in mind Humphrey (1994) has investigated the potential scope economies of ATM adoption. Estimating a composite cost function [Carroll and Ruppert (1984, 1988)] for a cross section of 161 banks in the US banking sector in 1991 and 1992, Humphrey (1994) measures cost savings from ATM adoption using the following expression:

³⁴Economies of scale for a single can be defined using the elasticity of cost with respect to output, which is measured using $E_{cy} = LMC/LAC$ where LMC and LAC are the long-run marginal and average costs of the firm respectively. The size of E_{cy} depends ultimately on the underlying technology and, more succinctly, on the returns to scale inherent in the technology defined as: $E = (\partial y / \partial s) \cdot s / y$, where y is the level of output and s is a scale parameter that defines the level of scale and multiplies each input by s . A technology displaying scale economies will additionally have increasing returns to scale, with E_{cy} and $E > 1$ [Gravelle and Rees (1992)].

$$\frac{C[q_i, B(1-\varepsilon), \varepsilon ATM, r_k] + C[q_i, \varepsilon B, ATM(1-\varepsilon), r_k] - C(q_i, B, ATM, r_k)}{C(q_i, B, ATM, r_k)} \quad (4.3)$$

where $C(\cdot)$ is the composite cost function, q_i is the deposit and loan output for bank i , B is the number of branches operated by bank i , ATM the number of ATMs a bank owns, r_k the deposit input prices a bank faces, and ε is a parameter reflecting the median quantity of ATMs or branches a bank can operate and is allowed to vary in the range $[0, 0.5]$. Intuitively, (4.3) implies that cost savings (economies of scope) are measured as the cost of using only branches plus the cost of only ATMs minus the cost of using both (the extreme right-hand expression in the numerator) divided by the cost of using both (the denominator). ATMs will then be 'cost saving' if (4.3) is positive.

The empirical results obtained by Humphrey indicates that unit costs are 2.50% higher when ATMs and branches are jointly used to deliver deposit services compared to them being separately used when $\varepsilon = 0$, the conventional method of estimating scope economies [Gravelle and Rees (1992)]. In addition, profit scope economies between branch and ATM delivery methods are investigated using a composite profit function [net income replacing total cost in (4.3)] and the above methodology. Results indicate that for $\varepsilon = 0$, there is a profit increase of 3.60% associated with ATM adoption.

Humphrey accounts for this apparent dichotomy between the positive cost-saving effects and the negative effects of profitability of ATM adoption in three ways. First, although ATMs have been found to have substantial scale economies the potential for cost savings are offset by higher a frequency of withdrawals. This argument is consistent with the Baumol (1952) model of transactions money demand (see Section 4.5.2) which predicts that lower transactions or transfer costs (proxied by the added convenience of ATMs for cash withdrawal) will, *ceteris paribus*, increase the frequency of cash withdrawal. Second, banks have to keep inventories of cash in ATMs and this imposes an opportunity cost of interest forgone. Third, 'interchange fees' from customers using a rival banks network have offset the cost effects of higher ATM usage.

Unfortunately, there are no studies that have investigated the cost effects of ATM adoption for the UK experience. However, Table 4.3 displays ATM withdrawal

statistics in the UK from end 1986 to end 1996. These figures indicate that although the average number of withdrawals has increased by approximately 50% since 1986 the real value of the amount withdrawn has remained relatively stable. It may be the case that a similar experience encountered by the US financial institutions has also occurred in the UK financial sector. Usage of a banks ATM network by its own customers, for example, is free at the point of use in the UK, although interchange fees do exist in some cases when using ATMs of a bank belonging to a different network [see APACS (1997) for details]. Clearly, however, much more research is required to arrive at any concrete conclusions concerning the cost effects of ATM adoption for the UK situation.

Table 4.4: ATM Withdrawals in the UK

Year	Average Number of Withdrawals per ATM - end year	Average Real Value Withdrawal per ATM (£) - end year
1986	48021.26	41.91
1987	50632.91	41.21
1988	53927.07	43.11
1989	56098.62	43.96
1990	58348.71	42.75
1991	59825.76	45.37
1992	62674.24	43.61
1993	64890.28	43.47
1994	66893.82	42.71
1995	70271.82	41.46
1996	72284.25	41.44

Source: APACS (1997). Note: real value of withdrawals has been calculated from nominal value of withdrawals deflated by the consumer price index (1990=100).

4.4.3 ATMs, Competitive Strategy and Regulatory Change

An alternative interpretation of ATMs, but not a mutually exclusive one, as a process innovation has been put forward by Scarborough and Lannon (1988), Vesala (1994), Haynes *et al* (1990) Akçaoglu (1996) and Llewellyn (1997) who argue that ATMs are a key component of an institution's competitive strategy and, as such, can additionally be interpreted as a product innovation. The basis for this interpretation lies that in the theory of the demand for characteristics initially conceived by Lancaster (1966, 1971) as a method of analysing consumer demand. Revell (1986), BIS (1986), Desai and Low

(1987) and Akçaoglu (1996) in their analysis of product and process innovation have most recently used this approach for the financial sector. In this approach to ATM technology, the deposit services provided by a particular institution do not, *per se*, give utility to the consumer, but rather, it is the inherent characteristics embedded in that service that yields utility. Demand for deposit services will then become a derived demand and consumers will demand services due to their characteristics, since this is the ultimate source of utility [Akçaoglu (1996)]. This implies that the consumer's utility function incorporates the essential characteristics of the deposit services that a financial institution provides. Thus, expansion and diversification of these characteristics becomes an important component of a firm's overall competitive strategy.³⁵ This approach is closely related to Schumpeter's (1934) observation that non-price competition and innovation is a more common tool for gaining competitive advantage than price.

From this framework of analysis, the adoption of ATMs by financial institutions can be seen as increasing the number of service-characteristics that a particular deposit taking institution can provide (for example, 24 hour cash withdrawal and an extensive network of ATMs). Customers benefit since, generally, ATM transactions are free at the point of use and increased convenience and reduced transactions costs (including the opportunity of time) means that customers can gain through reducing cash balances and moving their funds to interest earning time or savings deposits [Pawley (1993)]. As noted by Scarborough and Lannon (1988) and Pawley (1993), those institutions in the early 1980s that delayed their adoption of ATMs as an explicit 'wait and see' policy came to realise that ATM adoption by rival institutions was a threat to their market share of personal sector deposits. Moreover, such a competitive strategy may increase deposit funds by 'locking-in' customers to a particular deposit taking institution to whom other income generating financial services can be sold (such as insurance and shares) and which ultimately lowers the bank's funding costs [Revell (1986)]. As stated by Vesala (1994), however, the competitive use of ATMs was probably more pronounced in the early phase of ATM diffusion when network co-operation between financial institutions had not been established.

³⁵ Competitive strategy is interpreted here as being a strategy implemented by management in order to create or gain a competitive advantage in the market place in which it competes [Porter (1988)].

The nature of competition between retail banks and building societies will influence the technology strategy pursued by an institution and therefore partly determine the characteristics of technology diffusion [Freeman (1974) and Akçaoglu (1996)]. The nature of competition in the UK personal sector has arguably been strongly influenced by regulatory changes [Bank of England (1983, 1986) and Pawley (1993)]. Although a comprehensive analysis of these major regulatory changes lies outside the ambit of this thesis two main changes can be identified that have influenced the nature of ATM diffusion. First, the abolition of the last *bona fide* quantity control in the UK banking sector - the 'Corset' - in July 1980 and, secondly, the Building Societies Act (1986). The first of these changes removed quantity ceilings on the interest eligible liabilities (IBELs) that banks could raise which had previously been restricted by a system of supplementary special deposits [Spencer (1986)]. The second change allowed, *inter alia*, building societies to expand into those money transmission services previously the monopoly of banks (such as, for example, the issuing of cheque books, credit and debit cards, unsecured loans and travellers cheques). Moreover, it allowed building societies to change their status from specialised licensed deposit holders (i.e. mutual societies owned by its investors and borrowers) into commercial banks through conversion to full PLC status.

These changes had two main effects on the nature of competition between retail financial institutions. Firstly, the removal of the 'Corset' allowed banks to perform portfolio distribution between wholesale and retail deposits [Podolski (1986) and Llewellyn (1992)]. Prior to 1980 banks raised most of their funds in the wholesale market, but the removal of the 'Corset' increased competition for funds in the wholesale market and forced banks to seek alternative funding in the retail market. By 1981 banks had already entered the traditionally building society dominated mortgage market and expedited the breakdown of the building societies *de facto* interest rate setting cartel in 1983 [Spencer (1986)]. The entrance of clearing banks into this market forced building societies to compete on price (i.e. mortgage and deposit rates) whereas, previously, interest rate setting policy was aimed at securing an adequate 'spread' between mortgage lending and deposit rates to keep smaller, more marginal building societies in business [Drake (1989)]. Secondly, the Building Societies Act (1986) allowed building

societies to offer the full range of deposit services only previously offered by banks [Bank of England (1992) and Boleat (1992)]. This also had the effect of increasing competition for personal sector deposits and lead to the introduction of high interest rate cheque accounts by banks and building societies in the mid-1980s [Pawley (1993)].

Moreover, the increased competition between retail banks and building societies had the impact of making ATMs an important component of the range of services they provided and, thus, part of the overall competitive strategy for institutions. Additionally, they were seen by building societies as a means of developing a strong customer relationship and increasing the potential of cross-selling a wider range of financial services [Madden (1986) and Drake (1989)]. Although the Halifax and Alliance and Leicester were the first building societies to introduce ATMs in 1983 (see Table A4.2.2), the Building Societies Act (1986) allowed them (and others building societies) to diversify into the full range of deposit services with the ultimate effect that the share of total ATMs operated by building societies in the UK increased from less than 1% at the end of 1983 to approximately 20% at the end of 1986 (see Table 4.1).

These regulatory changes have arguably had two effects on the time path and nature of ATM diffusion in the UK financial sector. Firstly, building society adoption of ATMs may have been delayed prior to the Building Societies Act (1986) because of restrictions on the provision of a full range of deposit services by being mutual societies. As ATMs are predominately used by individuals for non-mortgage transactions, it may be the case that ATM adoption would have been unprofitable for building societies prior to 1986. Secondly, Neven (1993) and Vesala (1993, 1994) have shown that in spatial models of monopolistic competition [see, for example, Salop (1979), Tirole (1988) and Shy (1996)] financial institutions may 'oversupply'³⁶ ATMs if price competition is restricted and institutions are forced to compete exclusively on quality of service. This was certainly the scenario in the UK prior to 1986 and may, arguably, explain the forces behind the establishment of shared and eventually merging ATM networks. Drake (1989) argues that expansion of proprietary ATM networks was

³⁶ This implies that banks provide a quantity of ATMs that is Pareto sub-optimal when compared to unrestricted price competition.

becoming marginal in the mid-1980s and provides evidence in support for such a view.³⁷

4.5 Consequences of ATM Adoption for the Financial Sector

Two main consequences of the wider adoption and utilisation of ATMs by financial institutions and individual deposit holders have been identified in the literature: the potential for lowering entry barriers and the effects on the stability of the money demand function. Both of these possible effects are considered in this section.

4.5.1 Barriers to Entry

There is a vigorous and growing debate in the literature as to whether IT has reduced entry barriers into the financial sector and, in particular, in the retail end of the sector [Bank of England (1983), Revell (1986), Podolski (1986), Akçaoglu (1996) and Llewellyn (1997)]. Although this debate has tended to concentrate on the generic nature of IT a number of insights can be provided for the case of ATMs. Economists [see Tirole (1988) and references therein] have, generally, distinguished between two types of entry barrier that may exclude a potential entrant to enter a market in which supernormal profits are being made. Firstly, there are *absolute* entry barriers that may arise from some legal impediment (such as a patent) granting production rights to a particular firm or set of firms. Secondly, there are *relative* entry barriers that have three main elements: economies of scale and capital market imperfections, absolute cost advantage and product differentiation advantage. Each of these can be applied to the case of ATMs and the market for deposit services.

Firstly, the debate surrounding the existence of absolute entry barriers has focused on the pre- and post- Building Societies Act (1986) situations. As noted in Section 4.4.3, this act, *inter alia*, enabled building societies to participate fully in the provision of

³⁷ In addition, the post-1986 experience in the UK has seen a dramatic fall in the number of building society branches from 6954 at the end of 1986 to 5011 at the end of 1996 [BSA (1996)]. This lends support to the spatial model of bank competition of Neven (1993) where the removal of price regulation leads to a fall in the number of branches.

money transmission services - and therefore justified the adoption of ATMs - which were previously the domain of commercial banks. There is controversy in the literature [Drake (1989), Ash (1989) and Boleat (1992)] as to whether building societies (being mutual societies and therefore not coming under commercial banking law) were indeed inhibited in the provision of deposit services prior to 1986. As noted by Drake (1989), it was the adoption of ATMs and the issuing of ATM cards by the Halifax and Alliance Leicester Building Societies which, in many ways, lead inevitably to the 1986 Act.³⁸ Indeed, as Table A4.2.2 of Appendix A4.2 shows, 11 building societies had already adopted ATMs in 1985 in anticipation of legislative changes [Boleat (1992)].

Secondly, as observed from US empirical studies outlined in Section 4.2.2, ATM technology is characterised by scale economies which may imply that potential entrants operating on a smaller scale (relative to incumbents) cannot achieve a production scale such to achieve these economies of scale where long-run average costs are minimised. Moreover, if the capital market is characterised by imperfections [Gravelle and Rees (1992)] then even relatively large entrants may be excluded by the higher interest rates required to borrow funds for investment in a proprietary ATM network as compared to established institutions. There may, however, be three mitigating forces at work, which reduce these barriers to entry. Firstly, the establishment of co-operative ATM networks in the UK (LINK, MINT and FOUR BANKS) has allowed the relatively smaller clearing banks and building societies to participate in an ATM network and realise scale economies without the high fixed costs [Revell (1986), McKillop and Ferguson (1994) and Vesala (1994)].³⁹ Secondly, the real price of ATMs has been falling since 1972 (see evidence provided in Chapter 6) which implies that the scale economies aspects of ATM technology may becoming less important [Revell (1986) and Llewellyn (1997)]. Thirdly, before the commercialisation of ATMs in 1972 established banks and building societies had a competitive advantage (reflected in an absolute cost advantage) in the provision of deposit services because of their established network of branches which were a prerequisite of providing deposit services. This arguably was further enhanced

³⁸ This is consistent with Kane's (1981) 'regulatory dialectic' which implies interaction between the political process and economic process as the origins of the Building Societies Act (1986) were in the removal of the 'Corset' in 1981 allowing commercial banks into the mortgage market.

³⁹ As noted by Katz and Shapiro (1986), however, this may introduce a free-rider problem as smaller institutions may benefit more than relatively larger ones who offer the full range of money transmission services. Also note the anti-competitive aspects of ATM networks as outlined by Economides (1995).

by reputational effects established by customer-institutions relationship resulting from the experience good aspects of the financial services industry [Nelson (1970)]. As noted by Boleat (1987) and Llewellyn (1997), the application of IT (including ATMs) to the financial sector has implied that the competitive advantages that the branch network has traditionally brought has now diminished:

Branches were essential when there were no ATMs, because they were a method of paying in or taking out money from a bank. Now that people can get cash out of ATM, however, branches are largely redundant. [Boleat (1987), p. 33].

Indeed, as Table 4.2 indicates, branches are no longer a prerequisite for operating ATMs. This last issue is further related to the 'unbundling' of financial products [Llewellyn (1992, 1997)]. New technology allows institutions to produce a subset of characteristics (see Section 4.4.3) of what has traditionally been an 'indivisible bundle' of services. As noted by Podolski (1986) this means that the financial sector has become more contestable in terms of Baumol (1982). More succinctly, in the multiproduct setting of contestable markets an industry configuration is not suitable if cross-subsidisation takes place and this 'unbundling' effect may mean that institutions examine the current price-setting nature of their ATM services [Haynes *et al* (1990), Banking World (1992) and Llewellyn (1997)]

4.5.2 Money Demand Stability

The possible effects of ATM diffusion and their greater utilisation by the financial sector and individual deposit holders respectively on money demand has most frequently focused on the potential for ATMs to lower the 'transaction costs' associated with cash withdrawals. The lowering of transaction costs brought about by ATMs is associated with their greater convenience and availability relative to the traditional bank and building society branch. The consequences of these effects have then been analysed within the Baumol (1952) model of money demand which conceptualises money demand as an inventory of cash arising from the imperfect synchronisation of receipts and expenditures [Goldfield (1989)]. In such a framework, a lowering of transaction costs will reduce cash holdings and increase the frequency of transfer between cash and the alternative interest-bearing asset (a 'bond').

The predictions made by the Baumol (1952) model are certainly consistent with the UK experience. The velocity of narrow money (both notes and coins and M0) has certainly experienced a downward trend in the UK during the 1980s and 1990s [Breedon and Fisher (1993)]. The role of ATMs in this effect has, however, been reduced to an 'enabling' one in the literature, enabling the personal sector to hold more of their wealth in the form of high interest bank and building society accounts. In addition, ATMs have caused a move away from cash payments to employees (see Table 4.4 below) and increased the use of the debit and credit card from transactions [see evidence presented in Duca and Whitesell (1995)].⁴⁰ This enabling role that the ATM fulfils illustrates the complementarity between ATM hardware and the debit card software. This aspect of ATM technology is elaborated on in Chapter 7.

Empirical studies investigating the effects of ATMs on money demand have focused on narrow definitions of money. They have, however, been hampered by inadequate measures of the non-pecuniary transactions cost advantages associated with ATM usage and the multicollinearity problems encountered when including a subset of innovation measures. For instance, the seminal work of Johnston (1984) on money demand and innovation found that including the number of ATMs in a linear money demand equation had a positive effect on the demand for notes and coins in the UK from 1968 to 1982. He concludes, however, that this result is virtually meaningless because of the high multicollinearity between the number of ATMs and credit card variables. Viren (1992) reports an identical result for the case of Finnish narrow money demand. The multicollinearity problem has, however, been circumvented by Hall *et al* (1989), Westaway and Walton (1991) and Breedon and Fisher (1993) who all use a cumulative interest rate term to capture the 'ratchet' effects involved in cash management during periods of high inflation and economic uncertainty. These studies, although emphasising that innovation has to be taken into account for a long-run money demand equation to exist (co-integrate), merge all types of innovation within one variable and so it is impossible to identify the separate effects of ATMs.

⁴⁰ This enabling role of ATM technology reflects the compatibility between ATM hardware and debit card software.

Daniels and Murphy (1994) find that the monthly demand for currency falls with the increased probability of ATM adoption by households. Daniels and Murphy use perform a cross-sectional study of currency demand across American households. Interestingly, this result is derived from a 'double-hurdle' model [Maddala (1983)], that initially models the probability of ATM use in a logit model. The first stage results indicate that age and education have a negative and positive effect respectively on the probability of having an account with ATM access.

Table 4.5: Payment of Wages and Salaries in the UK

Means of Payment	% at end 1976	% at end 1996
Cash	58.00	15.00
Cheque	12.00	9.00
Direct to Account	26.00	71.00
Other	100.00	100.00

Source: APACS (1997). Note: percentage totals refer to the percentage of adults in full/part-time employment including self-employment.

4.6 Concluding Remarks

The results from tracing the diffusion of ATMs in the UK financial sector show that ATM adoption has been restricted to two types of institution: clearing banks and building societies. From this result a set of potential ATM adopters has been constructed as the *stock* of clearing banks and building societies at the end of 1992. It was illustrated that ATM adoption occurred earlier, and that inter-firm diffusion has been quicker, for the set of clearing banks vis-à-vis building societies. In addition, empirical support was lent (assuming a constant number of potential adopters) to the often made stylised fact that the inter-firm diffusion curve is sigmoid in shape.

The analysis of the nexus between the technical attributes of the ATM and its economic consequences for factor-bias, scale and scope economies and competitive strategy indicates that the frequent interpretation of the ATM as a simple capital-embodied process innovation may not give full consideration to its product innovation dimension. Although there is an absence of relevant UK studies, the empirical evidence from other countries indicates that the labour-saving nature of ATMs is highly sensitive to the

definition of labour-inputs and that ATMs may be capital-saving (in terms of branches). Moreover, empirical evidence supports the contention that ATMs have inherent (but reducing) scale economies, but that their cost-saving advantages may be off-set by greater frequency of use by deposit holders. Furthermore, for the UK context the importance of the nexus between regulatory change and the time path of ATM diffusion was emphasised and results indicate that the entrance of building societies into the ATM market may have made ATMs an important component in an institutions' overall competitive strategy.

To summarise, despite the paucity of evidence pertaining to the UK experience the implications of the analysis presented in this chapter are twofold. First, economists need to be aware of both the product and process aspects of technological innovation for the adopting firm. This approach is largely denied in the conventional capital-using/capital-producing framework. Second, the nexus between the market structure of the innovating industry and the regulatory framework may have an important influence on the time path of innovation diffusion.

A4.1 Appendix One: The Set of Potential Adopters

The definition employed for retail banks is identical to that used by the Bank of England (1997) which focuses on two criteria: firstly, the institution must have a 'large' branch network in the UK, and secondly, the institution must participate in the UK clearing system. At the end of 1992 there were 20 retail banks operating in the UK [Bank of England (1993)], whilst at the end of 1996 there were 31. The principal function of the retail banks is to take retail deposits (£100,000 or less) and wholesale deposits (£100,000 or more) and on-lend items in the form of overdrafts and various types of loans to both the personal and corporate sectors.⁴¹ Furthermore, they provide additional income generating financial services such as credit cards, share dealing, investment and tax advice, insurance and estate agency [Pawley *et al* (1991) and Pawley (1993)]. A distinguishing feature of retail banks, and one exploited by the Bank of England (1997) definition, is their extensive branch network.

The definition employed for building societies in this thesis is purely an 'empirical' one and utilises the members of Building Societies Association (BSA), an umbrella organisation for *all* building societies in the UK. At the end of 1996 there were a total of 86 members of the BSA [BSA (1997)]. The activities of building societies can, arguably, be categorised into pre and post The Building Societies Act (1986) situations. Prior to 1986 the main activities of building societies was the raising of funds (through the subscription of its members) for the purpose of making advances for upon security by way of mortgage or freehold leasehold estate [Pawley *et al* (1991)]. Under the Building Societies Act (1986), however, building societies have been given the power to undertake, *inter alia*, money transmission services, foreign exchange services, personal equity plans, administration of pension schemes, investment services, insurance and unit trust schemes [Drake (1989)]. These changes have had the effect of building societies offering more flexible, immediate notice interest earning accounts and a decline in the share of their traditional savings accounts [see Bank of England (1990) for other balance sheet changes].

⁴¹ This is the usual definition given in the literature. The Bank of England (1997), however, has a different definition of retail deposits and defines them as deposits which arise from a customer's acceptance of an advertised rate (including nil) for a particular product and which usually taken in the banks' branch network.

Tables A4.1.1 and A4.1.2 below list the retail banks and building societies included in the set of potential ATM at the end of 1992 respectively. The '*' indicates that the institution is an ATM adopter and included in the final sample and '**' indicates that the institution was not included in the final sample of potential adopters. Where no mark exists next to the institution, it can be assumed that the institution is included in the final sample and is a non-adopter. Finally, if the institution is a member of an inter-institution ATM network at the end of 1996 the name of it is included in parenthesis after the institution's name.

Table A4.1.1: Retail Banks - as at end of 1992	
Abbey National Group (LINK)	The Co-Operative Bank Group* (LINK)
Ulster Bank*	Banking Department at the Bank of England**
The Bank of Scotland* (LINK and FOUR BANKS)	Bank of Ireland**
The Barclays Group* (FOUR BANKS)	Clydesdale Bank Group* (LINK)
The Lloyds Group* (FOUR BANKS)	Airdrie Savings Bank** (LINK)
The Midland Group* (MINT)	Girobank*
The National Westminster Group* (MINT)	Yorkshire Bank Group* (LINK)
The Royal Bank of Scotland* (FOUR BANKS)	Allied Irish Banks** (LINK)
The Standard Chartered Group*	Northern Bank**
The TSB Group* (LINK and MINT)	Coutts Bank**

Note: TSB bank is a reciprocity partner of LINK.

Table A4.1.2: BSA Members - as at end of 1992		
Alliance and Leicester* (LINK)	Hanley Economic	Nottingham
Barnsley	Harpenden	Nottingham Imperial
Bath Investment	Heart of England	Penrith
Beverley	HFC* (LINK)	Portman* (LINK)
Birmingham Midshires* (LINK)	Hinckley and Rugby	Principality
Bradford and Bingley* (LINK)	Holmesdale	Progressive

Table A4.1.2: BSA Members - as at end of 1992 *continued*

Bristol and West* (LINK)	Ilkeston Permanent	Saffron Walden Herts & Essex
Britannia* (LINK)	Ipswich	Saint Pancras
Buckinghamshire	Kent Reliance	Scarborough
Cambridge	Lambeth	Scottish
Catholic	Leeds and Holbeck	Shepshed
Chelsea* (LINK)	Leeds Permanent*	Skipton
Cheltenham and Gloucester	Leek United	Stafford Railway
Chesham	Londonderry Provident	Staffordshire
Cheshire	Loughborough	Standard
Chorley and District	Manchester	Stroud and Swindon
City and Metropolitan	Market Harborough*	Swansea
Clay Cross Benefit	Marsden	Teachers'
Coventry* (LINK)	Melton Mowbray	The Ecology
Cumberland* (LINK)	Mercantile	The Mansfield
Darlington	Monmouthshire	Tipton and Coseley
Derbyshire* (LINK)	National and Provincial*	Tynemouth
Dudley	National Counties	Universal
Dunfermline* (LINK)	Nationwide* (LINK)	Vernon
Earl Shilton	Newbury	West Bromwich
Furness*	Newcastle	West Cumbria
Gainsborough	North of England*	Woolwich* (LINK)
Greenwich	Northern Rock* (LINK)	Yorkshire* (LINK)
Halifax* (LINK)	Norwich and Peterborough* (LINK)	

A4.2 Appendix Two: ATMs Operated by Financial Institution

Tables A4.2.1 and A4.2.2 below show the total number of ATMs operated by each retail bank and building society ATM adopter respectively and which is additionally a member of the set of potential adopters as defined in A4.1. These figures were constructed from extensive fieldwork conducted during 1993 and 1994. The relevant finance, planning and research departments at building societies and retail banks were contacted during this period and asked the specific date at which they first adopted ATMs and the subsequent number of ATMs operated for preceding years.

Table A4.2.2: Total Number of ATMs Operated by Building Societies - end 1972 to end 1992

Institution	Year										
	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Alliance and Leicester	0	16	20	12	9	12	25	45	38	72	62
Birmingham Midshires	0	0	0	0	0	0	0	21	27	27	27
Bradford and Bingley	0	0	3	34	35	51	54	55	57	59	61
Chelsea	0	0	0	6	6	6	6	6	6	6	5
Coventry	0	0	0	0	15	16	18	25	26	32	36
Cumberland	0	0	0	0	0	1	2	2	2	3	5
Derbyshire	0	0	0	0	9	12	15	17	24	24	25
Dunfermline	0	0	0	0	3	5	5	5	5	5	5
Furness	0	0	0	0	0	0	1	1	1	1	1
Halifax	0	1	100	250	357	498	813	1124	1423	1539	1581
HFC	0	0	0	0	0	0	0	15	26	16	15
Leeds Permanent	0	0	8	58	102	102	102	102	100	96	103
Market Harborough	0	0	0	2	2	2	2	2	2	2	2
Nationwide	0	0	0	6	103	201	274	357	455	607	713
North of England	0	0	0	0	0	6	6	6	6	6	6
Northern Rock	0	0	0	0	0	0	0	19	26	30	31
Norwich and Peterborough	0	0	0	3	3	4	5	6	8	9	11
Portman	0	0	0	0	0	5	5	5	5	3	3
Woolwich	0	0	10	23	49	99	120	92	211	280	348
Bristol and West	0	0	5	12	30	40	52	54	55	61	66
Brittania	0	0	0	4	22	29	36	38	49	50	45
Yorkshire	0	0	0	0	21	24	25	30	50	50	48
National and Provincial	0	0	0	0	0	21	33	36	39	44	49
Total	0	17	146	410	766	1134	1599	2063	2641	3022	3248

Table A4.2.1: Total Number of ATMs Operated by Retail Banks - end 1972 to end 1992

Institution	Year																				
	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Abbey National	0	0	0	0	0	0	0	0	0	0	0	0	0	55	222	295	413	675	858	911	946
Airdrie Savings Bank	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	6	6	6
Bank of Scotland	0	0	0	0	0	0	0	0	0	150	190	203	211	267	292	302	311	330	344	368	391
Barclays	0	0	0	0	50	85	104	131	157	268	487	683	787	887	1120	1384	1777	2193	2445	2609	2683
Clydesdale	0	0	0	0	0	0	0	12	51	104	137	194	225	252	268	280	291	298	294	301	299
Co-Operative	0	0	0	0	0	0	0	0	0	0	0	0	0	36	48	56	60	70	76	81	89
Girobank	0	0	0	0	0	0	0	0	0	0	0	0	0	125	160	207	182	185	187	187	188
Lloyds	0	230	337	568	580	679	749	819	986	1208	1400	1535	1622	1716	1728	1918	2068	2184	2272	2296	2446
Midland	0	0	0	0	0	0	0	24	140	280	518	703	858	1081	1249	1397	1581	1868	2018	1981	1945
National Westminster	0	0	0	0	46	96	109	136	229	514	765	1304	1554	2022	2240	2342	2445	2583	2677	2775	3042
Royal Bank of Scotland	0	0	0	0	0	15	38	56	133	201	260	305	502	526	550	575	613	646	703	718	751
TSB	0	0	0	0	0	0	0	0	28	39	102	445	648	1148	1224	1672	1795	1911	1966	1914	1912
Yorkshire Bank	0	0	0	0	0	0	0	6	11	41	69	92	127	139	165	187	203	225	243	297	312
Total	0	230	337	568	676	875	1000	1184	1735	2805	3928	5464	6534	8254	9266	10615	11743	13172	14089	14444	15010

Note: The Airdrie Savings Bank was not included in the set of potential adopters

CHAPTER 5

A PROFILE OF ATM DIFFUSION IN THE UK FINANCIAL SECTOR - A DURATION ANALYSIS¹

5.1. Introduction

The underlying concept of duration models, as illustrated in Chapter 3, is the hazard function which gives the instantaneous rate of 'failure' (in this thesis 'adoption') of an economic agent leaving a certain state at time t conditional on the state still being occupied at t . The crucial question for an empirical economist is how to estimate and employ this hazard function with the ultimate aim of examining the effects of covariates (i.e. firm heterogeneity). As indicated in Chapter 3 there are essentially two methods for estimating the hazard function: non-parametrically or parametrically. The former approach is purely an empirical one which makes no assumptions regarding the underlying probability distribution of adoption times, whilst the latter chooses a specific form of the hazard function from a family of functions and, hence, makes explicit assumptions regarding the distribution of adoption times. In general, previous empirical diffusion studies have been somewhat *ad hoc* in their application of these two approaches, often simply drawing from previous non-economic studies with little or no consideration of the underlying rationale for using a particular approach. As shown by Heckman and Singer (1984), Chung *et al* (1991), Crowder *et al* (1991) and Neumann (1997), such an approach can lead to model mis-specification and subsequent bias in estimated coefficients, the extent of the bias depending on the nature of the mis-specification. Thus, there is the contention that researchers should aim to examine the underlying characteristics of their data set before proceeding to more formal models that include covariates.

The aim of this chapter is to explore both non-parametric and parametric approaches to estimating the hazard, survivor and integrated hazard functions that currently exist in the literature and to apply them to a set of discrete panel data which consists of ATM

adoption histories derived from yearly records for the set of financial institutions presented in Chapter 4. These estimates are provided by assuming that all financial institutions are homogenous so that adoption histories can be conceptualised as repeated drawings from the same probability distribution. Furthermore, a distinguishing feature of the data set is that some institutions do not adopt by the end of the study in 1992 and, hence, are so-called 'right-censored'. Such observations have to be accounted for in the construction of the likelihood of adoption times.

In addition, emphasis is placed on discrimination between the various parametric models estimated and measures of their goodness-of-fit are considered. An important distinction between omnibus and directional tests is made and a number of informal and formal tests of both types are carried out.

It is reassuring to discover that the results are in accordance with *a priori* expectations based on the shape of the empirical diffusion curve. Moreover, they lead to the questioning of the often *ad hoc* use of certain distributions utilised in previous empirical studies. In general, the analysis of ATM adoption histories indicates that the empirical hazard function is non-monotonic and is best represented by non-monotonic parametric hazard functions. This lends support to the epidemic theory of diffusion.

The remainder of this chapter is set out as follows. Section 5.2 describes the data set and methodology that is employed in the empirical estimation. Section 5.3 discusses the empirical findings. Section 5.4 examines and implements some approaches used in goodness-of-fit tests and Section 5.5 concludes the chapter.

5.2. Data and Methodology

There are essentially two issues that any empirical study of innovation diffusion has to address before proceeding to analyse the characteristics of the diffusion process. The

¹ This chapter draws on Gourlay (1998b).

first centres on the definition of the innovation under investigation, while the second concerns an appropriate definition of the population of potential adopters. These two issues had to be addressed in the compiling of the data set analysed in this chapter. In addressing the first of these issues, it was decided to follow past research convention outlined in Chapter 3 and to concentrate on a generic conceptualisation of the technological innovation. An ideal data set would, of course, be one that allowed precise measurements of quality changes in the technology and for these to be taken account of in the modelling procedure. Such data requirements are seldom met and the resources (both pecuniary and time) required to pursue it can preclude this route being taken. This was found to be the case for ATMs for which accurate information on quality changes from manufacturers was impossible to gain. Consequently, the technology under investigation is defined generically as second-generation ATM technology as defined in Chapter 4. The location of the ATM is not considered as being part of the criteria.

The second issue concerns the appropriate definition of the population of potential adopters. As shown in Chapter 4, the enabling characteristic of an ATM is its ability to allow deposit holding customers to access their accounts via a personal cash card, the extent of the access depending largely on the location of the ATM. It was shown that the adoption of ATMs in the UK has been confined to retail banks and building societies. Given this, it was decided to use as a sample of potential adopters those institutions listed in Appendix Two of Chapter 4. As noted in Chapter 4, these institutions are also members of two industry-specific organisations: the British Bankers' Association (BBA) and the Building Societies Association (BSA). Both these organisations proved to be an abundant source of institution-specific information over the period of study. In addition, the coverage of institutions in both organisations is particularly comprehensive and covers all institutions in the UK that had adopted ATMs at the end of 1992 except for the *Airdrie Savings Bank* [APACS (1997)]. At the end of 1992 the BBA consisted of nine Major British Banking Groups (MBBG) and provided report and account data on four other large retail banks. In contrast, the BBA contained eighty-eight members at the end of 1992. The set of potential adopters is assumed to be the members of both these organisations, although two BSA members had to be

excluded due to incomplete records.² This results in a total sample of ninety-eight financial institutions as at the end of 1992. Of these, thirty-five institutions (35.71% of potential adopters) had adopted ATMs at the end of 1992. The adoption date for a specific institution is defined as that date at which the institution first adopted one or more ATMs. The methodology employed in the collection of these adoption dates is identical to that employed in the construction of Tables 4.1 and 4.2 of Chapter 4.

As stated by Cox and Oakes (1984) and Kiefer (1988), once an appropriate data set to investigate has been identified there are three requirements to defining a 'duration'. Firstly, a time origin must be precisely defined for each institution. In this and proceeding empirical chapters the time origin is assumed to be identical for each institution and is set at 1972, the year that second generation ATMs were first commercialised. Secondly, a time scale to measure 'durations' has to be agreed. To a large extent this is defined by the nature of the data gathered and is defined as annual calendar time starting from 1972 and ending in 1992, the end of the empirical study. Thirdly, the point at which a duration ends (the point of 'failure') must be strictly defined. This is defined as the point at which an institution first adopts one or more ATMs and is recorded by the year of adoption. Hence, it follows that a 'duration' is conceptualised as being the length of time, or 'spell', in years (measured from 1972), that an individual institution remains in a state of non-adoption until that time at which it adopts ATMs.

These considerations can be more formally represented as follows. If T is assumed to be a continuous non-negative random variable and represents the duration of an individual institution in a state of non-adoption then the data set of ATM adoption histories consists of a cross section of duration times (i.e. realisations of T), t_1, \dots, t_n , where t_i is the duration of the i th institution ($i=1, \dots, 98$) and which can be ordered as $t_1 < \dots < t_n$.³ The range of t will be $[0, 21]$. This follows from the time elapsed from start of the study in 1972 ($t = 0$) to the end of the study in 1992 ($t = 21$). It is

² These were The Ecology Building Society and the Swansea Building Society.

assumed throughout this chapter that duration times are homogenous with respect to any systematic factors and regressor variables that affect the distribution of T . Consequently, each institution's duration time will be the realisation of a random variable from the same probability distribution [Lancaster (1990) and Neumann (1997)].

Thus, from the above considerations the general methodology employed in this chapter involves estimating both non-parametric and parametric representations of three important probability distributions associated with duration data formally introduced in Chapter 3: the hazard function, the survivor function and the integrated hazard function. The continuous time representations of these three functions are respectively (as noted in Chapter 3):

$$h(t) = \frac{f(t)}{S(t)} \quad (5.1)$$

$$S(t) = \exp\left[-\int_0^t h(s) ds\right] = \exp[-\Lambda(t)] \quad (5.2)$$

$$\Lambda(t) = \int_0^t h(s) ds = -\ln S(t) \quad (5.3)$$

where t is time with $S(0) = 1$ and $\lim_{t \rightarrow \infty} S(t) = 0$. Of particular interest in the empirical work is to examine the duration dependence of (5.1). Duration dependence is said to exist if and only if the following condition holds:

$$h(t) \neq \frac{f(t)}{S(t)} \quad (5.4)$$

Furthermore, if $dh(t)/dt > 0$ at $t = t_0$ there exists *positive duration dependence* at t_0 ; if $dh(t)/dt < 0$ at $t = t_0$ there exists *negative duration dependence* at t_0 [Heckman and

³ Since no institution in the data set dis-adopt ATMs this type of duration data can be described as being *single spell* because we observe an institutions' duration of stay in a single state of non-adoption [Lancaster (1990)]. Ties in adoption times are considered in sections 5.2.1 and 5.2.2 of this chapter.

Singer (1984, 1985)]. The former implies that the probability that a duration of non-adoption will end in the next short time interval dt after t_0 increases in t . In contrast, the latter implies that the probability that a duration of non-adoption in the next short time interval dt after t_0 decreases in t .

There are, however, two characteristics of the sampling design employed in the construction of ATM adoption histories that have to be accounted for in the estimation of the functions contained in (5.1) to (5.3). The first centres on those institutions that do not adopt on or before the end of the study in 1992. In the sample set of potential adopters there are 63 non-adopters (64.29% of potential adopters) at the end of 1992. The duration times of these non-adopters are 'right-censored' [Greene (1993) and Neumann (1997)] and are a distinguishing feature of duration modelling. The second characteristic relates to the possible effects of sample design on the sampling distribution resulting from the exclusion of those institutions that have closed down ('exited') during the observation period.

These aspects of the data set may be analysed by considering the possible life histories of institutions over the observation period 1972 to 1992.⁴ This is depicted in Figure 5.1. Time is measured along the horizontal axis. The solid lines represent institutions that have not yet adopted ATMs and the dotted lines represent establishments post-adoption. The figure shows the possible life histories of six institutions, A1 to A6, classified according to their adoption behaviour and entry and exit times from 1972 to 1992. Institution A1 fails to adopt ATMs before the end of the study in 1992. A2 has the same life history as A1, with the difference that it adopts the technology at calendar time D1. Institution A3 enters the industry at time D2, but does not adopt ATMs before the end of the study. A4 enters the industry at time D3, adopts the technology at time D4 (which may equal D1) and survives beyond 1992. Institutions A5 (adopter) and A6 (non-adopter) leave the industry at dates X1 and X2 respectively and are therefore excluded from the sample observed in 1992.

⁴ The following analysis draws from Karshenas and Stoneman (1993).

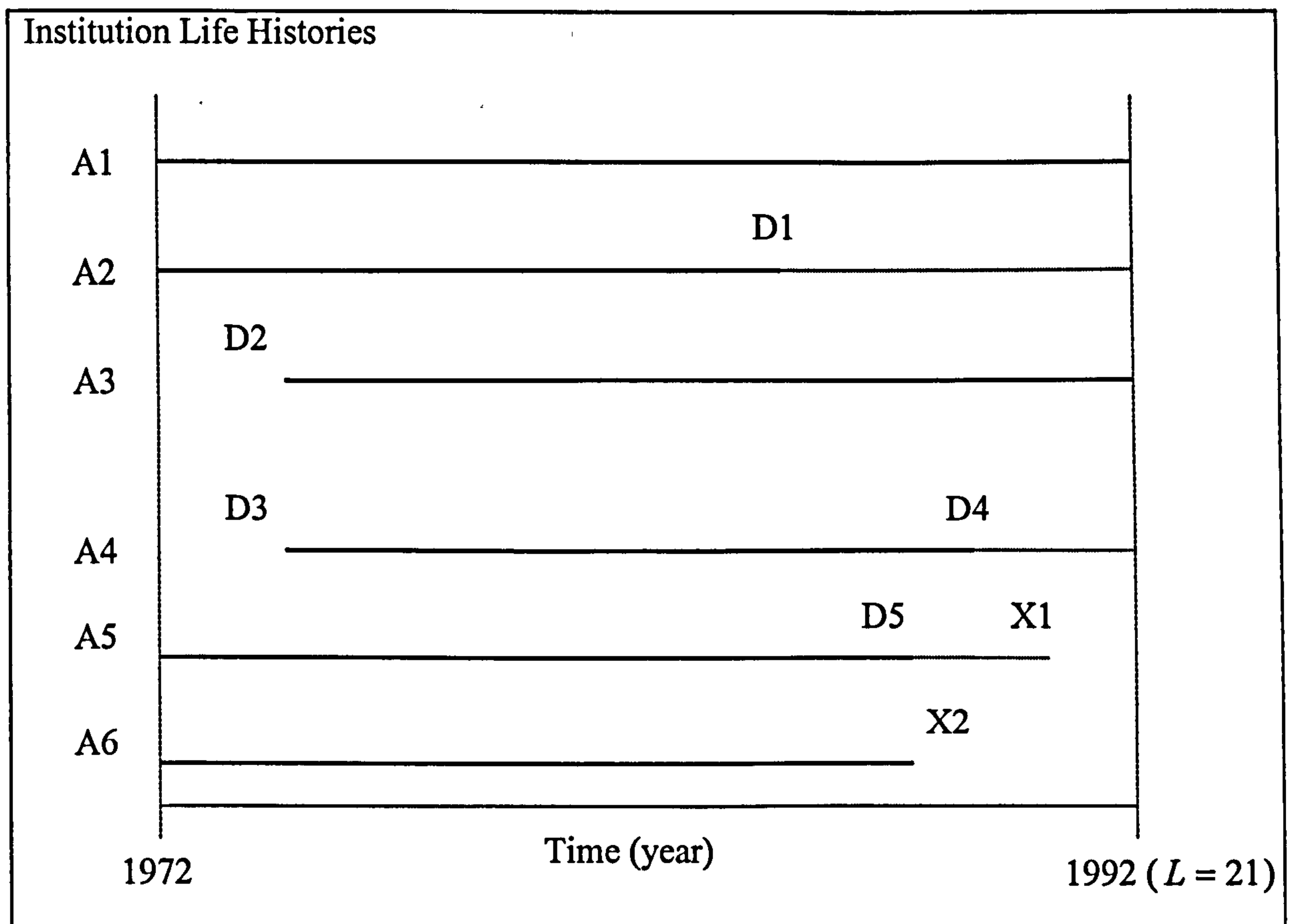


Figure 5.1: Possible Life Histories of Financial Institutions

The durations of institutions A1 and A3 are known as being *right-censored* or *singly Type I censored* [Greene (1993) and Neumann (1997)].⁵ Such durations are necessarily incomplete and are only known to exceed some predetermined length (i.e. the end of the study) set by the researcher. More succinctly, in a Type I censored sample with institutions $1, \dots, n$, we observe an institution's durations if and only if $T_i \leq L_i$, where T_i is a random variable representing the duration of an individual in the absence of censoring and L_i is the censoring time. For the set of ATM adoption histories utilised in this chapter the L_i 's are equal for all institutions and is set to $L = 21$ years.⁶ In dealing with such censoring it is convention in the modelling procedure [Karshenas and

⁵ Other possible censoring mechanisms can exist in duration modelling, such as left-censoring or Type II censoring. Type II censoring can occur, for example, if the date at which ATMs were first commercialised was unknown so that exact durations of non-adoption for each institution was not known with accuracy. See Lawless (1982) for discussion of different censoring mechanisms.

⁶ As noted by Lawless (1982) this implies that the number of observed completed durations becomes a random variable.

Stoneman (1995)] to employ an ‘indicator’ variable, δ_i , to distinguish between uncensored and censored observations. This takes the following form:

$$t_i = \min(T_i, L) \text{ and } \delta_i = \begin{cases} 1 & \text{if } T_i \leq L \\ 0 & \text{if } T_i > L \end{cases} \quad (5.5)$$

From (5.5), δ_i indicates whether the observed duration time, t_i , of the i th institution is censored or not, and equals unity if it is uncensored and zero otherwise.

As noted in Chapter 4, the sample of potential adopters has been constructed by sampling a ‘stock’ of institutions at the end of 1992. Such a retrospective study will by implication exclude those institutions that have closed down from the start of the study. These can be depicted by institutions A5 and A6 in Figure 5.1. This is relevant for BSA membership where the number of societies has fallen from 456 in 1972 to only 88 in 1992 through a combination of bankruptcy and acquisitions [BSA (1993)]. As Hoem (1985) and Karshenas and Stoneman (1993) have illustrated, such a sample design implies that those institutions that have left the industry during the period of observation have zero probability of being selected even though their durations may form part of the set of potential adopters. As a consequence, the sample set is necessarily constructed conditional on survival after 1992. A crucial question is whether the sampling design is independent of the outcome of adoption histories. For this sample design to be ignorable the probability of exit times should be independent of adoption times [Karshenas and Stoneman (1993)]. If it is not, then Hoem (1985) has indicated that individual adoption times have to be weighted according to selection probabilities. In Appendix A5.1 it is shown that the sampling design can be ignored for estimation purposes.

5.2.1 Non-Parametric Estimation

There are two non-parametric approaches to estimating the survivor, hazard and integrated hazard functions that are most frequently advocated in the literature [see, for example, Cox and Oakes (1984), Kiefer (1988) and Neumann (1997)]: the *Kaplan-Meier* [Kaplan and Meier (1958)] and *Cutler-Ederer* [Cutler and Ederer (1958)] estimates. Both of these are purely empirical approaches to estimation and have the advantage over parametric models in that they impose fewer restrictions on the data under investigation [Greene (1993)]. In particular, they are useful for displaying general features (for example, percentile points and dispersion) of the data and for preliminary analysis acting as an aid in the selection of parametric forms [Chung *et al* (1991)].

The decision to use the Kaplan-Meier or the Cutler-Ederer approach in non-parametric estimation depends crucially on the nature of the sample data. If the data are *ungrouped*, that is adoption and censoring times are known for each individual in the sample, then the Kaplan-Meier approach is appropriate. If, however, the sample data is *grouped* in nature, that is it is only known in which the intervals particular individuals adopted or are censored, then the Cutler-Ederer approach is appropriate. The data set of adoption histories utilised in this chapter is ungrouped because the year of adoption or censoring time for each individual institution is known exactly. Therefore, the Kaplan-Meier approach is the preferred non-parametric method of estimation as it provides more accurate information as to the pattern of the survivor, hazard and integrated hazard functions over the observation period.⁷

The non-parametric approach to estimation begins by assuming that there exists an uncensored sample of n *distinct* adoption times (t_1, \dots, t_n) that are observed from a homogenous population. From this assumption the sample survivor function, $\hat{S}(t)$, will be a step-function decreasing by n^{-1} immediately following each observed failure time and can be written simply as [Kiefer (1988) and Neumann (1997)]:

⁷ Details of the Cutler-Ederer approach are given in Cox and Oakes (1984).

$$\hat{S}(t) = n^{-1}[\text{number of sample points } \geq t] \quad (5.6)$$

A modification to (5.6) is, however, required when some observations are right-censored and/or ties exist in the data. Suppose the *completed* durations in the sample size n are ordered from the smallest value to the largest, $t_1 < \dots < t_k$. The number of completed durations, k , will be less than n if some observations are right-censored and/or because of the existence of ties. Ties will occur when two or more observations have identical duration times.

Following Kiefer (1988) the estimation of the survivor, hazard and integrated hazard functions under these conditions can proceed as follows. Let d_j be the number of completed spells of duration t_j for $j = 1, \dots, k$. In the absence of any ties the d_j will all be equal to one. Further, let m_j be the number of observations censored between t_j and t_{j+1} , with m_k being the number of observations with durations greater than t_j , the longest completed duration. Using these definitions the number of durations neither completed or censored before duration t_j can be given as n_j and is calculated as [Kiefer (1988)]:

$$n_j = \sum_{i \geq j}^k (m_i + d_i) \quad (5.7)$$

The expression in (5.7) is referred to as the 'risk set' and can be interpreted as defining those individuals who are eligible to adopt at time t_j [Neumann (1997)].

As shown in Chapter 4, the hazard function, $h(t_j)$, is the probability of completing a spell at time t_j conditional upon the individual reaching time t_j . A natural estimator for the hazard function is therefore:

$$\hat{h}(t_j) = d_j/n_j \quad (5.8)$$

The hazard estimator in (5.8) can then be interpreted as the number of adopters at duration t_j divided by the 'risk set' at duration t_j . This will be characterised by a step-function. In addition, the corresponding estimator for the survivor function is also a step-function and is given by:

$$\hat{S}(t_j) = \prod_{i=1}^j (n_i - d_i)/n_i = \prod_{i=1}^j (1 - \hat{h}_i) \quad (5.9)$$

The estimator in (5.9) is referred to as the *Kaplan-Meier* or the *product-limit* estimator [Kiefer (1988)]. The estimator in (5.9) is obtained by setting the estimated conditional probability of completing a spell at t_j equal to the relative frequency of completion at t_j .⁸ The estimate $\hat{S}(t)$ is then built up as a product, and each term in the product can be conceptualised as an estimate of the conditional probability of surviving past time t_j , given survival just prior to t_j . The resulting estimate will consequently be a step-function that equals 1 at $t = 0$ and drops by a factor $(n_i - d_i)/n_i$ immediately after each lifetime t_j .⁹

Finally, the integrated hazard can be estimated using the definition of the integrated hazard in equation (5.3) and is given as follows:¹⁰

$$\hat{\Lambda}(t_j) = \sum_{i \leq j} \hat{h}(t_i) \quad (5.10)$$

The integrated hazard function can never decrease and as indicated by Kiefer (1988) it will be linear in time if the hazard function has no duration dependence, a convex function in time if the hazard has positive duration dependence and a concave function in time if the hazard function has negative duration dependence.

⁸ See Kalbfleisch and Prentice (1980) for a formal derivation of (5.9).

⁹ Under conditions set out by Johansen (1978) and Lawless (1982), the estimator in (5.9) also has a maximum likelihood interpretation under a wide variety of sampling schemes.

An important aspect of estimating (5.9) is that $\hat{S}(t)$ will never reduce to zero if $m_k > 0$ as exact durations are unknown. This is the case for the data set of adoption histories where 64 institutions do not adopt on or before 1992. In this case it is convention to take $\hat{S}(t)$, $\hat{h}(t)$ and $\hat{\Lambda}(t)$ as undefined for $t > t_k$ [Kalbfleisch and Prentice (1980)].

As indicated by Lawless (1982), the estimate of $S(t)$ is subject to sampling variation and it is therefore desirable to have some idea of its precision. Under certain assumptions concerning the censoring mechanism used in the calculation for $\hat{S}(t)$ an asymptotic variance function for the Kaplan-Meier estimator in (5.9) can be derived. This is known as Greenwood's formula [Greenwood (1926)] and is given by:

$$\text{Vâr}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{i < j} \frac{d_i}{n_i(n_i - d_i)} \quad (5.11)$$

The corresponding variance function for the estimated integrated hazard function is given in Lawless (1982) as:

$$\text{Vâr}[\hat{\Lambda}(t)] = \frac{\text{Vâr}[\hat{S}(t)]}{[\hat{S}(t)]^2} \quad (5.12)$$

where $\text{Vâr}[\hat{S}(t)]$ is the estimate from (5.11).

5.2.2 Parametric Estimation

As illustrated in Chapter 3, throughout the empirical diffusion literature that has applied duration models a variety of parametric models of duration have been used to represent the hazard and survivor functions. The defining characteristic of these parametric models, which distinguishes them from non-parametric models, is that they assume a

¹⁰ Alternatively, using the relationship contained in (5.3), the integrated hazard can be estimated as minus the logarithm of the Kaplan-Meier estimator. As indicated by Kiefer (1988), these two approaches give

specific probability distribution for adoption times. Given the structure of the parametric model, this may additionally impose a particular form of duration dependence on the data under investigation. Thus, the assumed distribution of adoption times will have clear implications for the duration dependence of the hazard function. Moreover, it will provide a valuable tool for attempting to answer such a relevant question as: 'given that a financial institution has not adopted ATMs for, say, fourteen years after their commercialisation is the conditional probability of adoption in the next short time interval increasing or decreasing?' Parametric models, by imposing a specific structure on the hazard function, will have a crucial role in answering such a question and therefore provides an extremely useful characterisation of the duration data.

The literature on duration analysis provides a plethora of parametric models to select from. The most commonly used distribution in diffusion studies is the Weibull. This distribution is, however, often chosen because it has been extensively employed in non-economic studies and does not necessarily lend itself to be a good representation of diffusion data [Cox and Oakes (1984)]. A crucial question in the selection of a particular distribution is clearly its relationship to the empirical hazard function and to economic theory. Economic theory does not explicitly consider the dependence of the hazard function, although the epidemic model does imply that the hazard function will be non-monotonic and the rank effects model identifies those economic factors underlying the shape of the hazard function. These aspects are elaborated in Appendix A5.3. It is also proved in Appendix A5.3 that if the inter-firm diffusion curve is sigmoid, as is the case for ATM diffusion in the UK as summarised in Figure 4.2 of Chapter 4, then the resulting hazard will be non-monotonic. Based on this evidence, then *a priori* it is expected that the hazard function will be non-monotonic.

It was decided to choose four parametric models for estimation: the Exponential, the Weibull, the Lognormal and the Log-logistic. These models are diverse enough to allow for the existence of a wide variety of different forms of duration dependence and have the advantage (for the number of degrees of freedom) of only having at most two

similar results when the h 's are small.

parameters. Appendix A5.2 provides a summary of the specifications and properties of these models.

Once a family of duration distributions has been chosen then the data distribution is known up to a vector of parameters θ , with $\theta = \theta_1, \dots, \theta_m$. In the case of the four distributions considered in this chapter, $\theta = (\lambda, p)$. The density of a duration of length t can then be written as $f(t, \theta)$ [Greene (1993)]. If the sample data of duration lengths under investigation contains n individuals with observed *completed* spells of t_i for the i th individual ($i = 1, \dots, n$), then the likelihood can be specified as [Kiefer (1988)]:

$$L^*(\theta) = \prod_{i=1}^n f(t_i, \theta) \quad (5.13)$$

where $f(t_i, \theta)$ is the probability density for the i th individual, with f known and θ unknown and to be estimated.

The expression in (5.13) is the standard specification of the likelihood function and can be interpreted as the joint probability distribution of the sample as a function of parameters θ , given the t_i 's are independent [Cuthbertson *et al* (1992)]. The aim of maximum likelihood estimation is then to select values for the parameters θ , $\hat{\theta}$, for which (5.13) obtains a maximum value.

As indicated in Section 5.2 the data set of ATM adoption histories contains right-censored observations. When a duration spell is right-censored at time t_j the only information available to the researcher is that the duration was at least t_j . Consequently, the contribution to the likelihood for a censored observation is the value of the survivor function $S(t_j, \theta)$, the probability that the duration is longer than t_k .

Using the indicator variable in equation (5.5), the log of the likelihood, $L(\theta) = \ln L^*(\theta)$, is:¹¹

$$L(\theta) = \sum_{i=1}^n \delta_i \ln f(t_i, \theta) + \sum_{i=1}^n (1 - \delta_i) \ln S(t_i, \theta) \quad (5.14)$$

Using the fact from (5.1) that the density is the product of the hazard and the survivor function and the fact from (5.2) that the log of the survivor function is minus the integrated hazard, the log-likelihood can be written in terms of the hazard function as:

$$L(\theta) = \sum_{i=1}^n \delta_i \ln h(t_i, \theta) - \sum_{i=1}^n \Lambda(t_i, \theta) \quad (5.15)$$

In practice it is usual to estimate the parameters by maximum likelihood. Under a set of assumptions (met by all the parametric distributions estimated in this chapter) concerning the shape of the likelihood function [see, for example, Amemiya (1985)] the maximum likelihood estimator $\hat{\theta}$ will be consistent for θ . This approach can proceed using the *Newton-Raphson* method [Cuthbertson *et al* (1992)], which utilises the matrix of second derivatives (the Hessian), $\partial^2 L(\theta) / \partial \theta \partial \theta'$, from (5.15). This is an iterative procedure that makes a series of local quadratic approximations of the Hessian solving this problem and then re-computing the approximation. The variance of estimated coefficients can be estimated using the expected second derivatives of the log-likelihood, the information matrix [Greene (1993)]. The results presented in Section 5.3 were derived using this approach.

¹¹ This specification of the likelihood function assumes that the censoring mechanism is independent, which requires the pairs (t_i, δ_i) , $i = 1, \dots, n$, are independent. As shown by Kalbfleisch and Prentice (1980) this is the case for the Type I censoring mechanism employed throughout this thesis.

5.3 Results

Using the approaches set out in Sections 5.2.1 and 5.2.2, non-parametric and parametric estimates of the survivor, hazard and integrated hazard functions were obtained for the data set of ATM adoption histories. All the results were obtained using the econometric package *LIMDEP 6.0* [Greene (1994)]. It should be noted that for the parametric models estimated in Section 5.3.2 the significance of the estimated coefficients is based on a test of the null hypothesis $H_0: p \text{ or } \lambda = 0$ against the alternative $H_1: p \text{ or } \lambda \neq 0$. The test statistic, t , is given by $t = \hat{p} \text{ or } \hat{\lambda} / SE(\hat{p} \text{ or } \hat{\lambda})$, where 'SE' is the standard error of the estimated coefficient. All tests are carried out for a significance level of 0.01. Using a normal approximation this gives critical points for a two-tailed test of ± 2.5760 .

5.3.1 Non-parametric Estimates

Non-parametric estimates of the survivor, hazard and integrated hazard functions are presented in Table 5.1 and Figures 5.2, 5.3 and 5.4 show how these functions evolve overtime. As there are no new adopting institutions after 1989 (i.e. $m_k > 0$), the functions are only defined up to a duration of 17 years ($= t_k$). As expected, all three functions are step-functions implying a constant value throughout the relative time period of t_j to t_{j+1} .

For $t_j = 0$ the hazard and integrated hazard functions were both set equal to zero and the survivor function equal to unity. This follows from their formal definition in (5.1) to (5.3). Following convention [Kiefer (1988) and Neumann (1997)] it was decided that for those years (except for $t_j = 0$) where $d_j = 0$ the hazard was assumed to be constant throughout that duration and equal to the previous periods value. Standard errors were calculated using (5.11) and (5.12).

By visual inspection of Figure 5.2 and the results contained in Table 5.1, the hazard function attains a relatively small and constant value for those durations between 1 and 10 years. The function does, however, increase from 11 to 13 years, obtaining a maximum value of 0.0964 at a duration of 13 years (1985) and then declines. Thus, the conditional probability of adopting ATMs as measured by $\hat{h}(t_j)$ attains its largest values between 1983 and 1985. The shape of the plot in Figure 5.2 does appear to suggest that the hazard is non-monotonic. Given the 'noise' present in the data it is, however, not possible to rule out multi-modality of the hazard function. This 'roughness' present in the plot results from the stochastic independence of previous observations [Neumann (1997)].

The shape of the non-parametric hazard function reflects the inter-firm diffusion curve for ATMs overtime as summarised in Figure 4.2 of Chapter 4. The diffusion of ATMs is most rapid (in terms of the number of new adopters) between 1984 and 1987. During this period the number of new adopters, d_j , is increasing relative to the risk set, n_j . Hence, given the definition of the non-parametric hazard function in (5.8), the estimated hazard function will increase during the period 1984 to 1987.

The shape of the hazard function over time parallels that of the survivor function contained in Figure 5.3. Referring to Figure 5.3, it can be observed that the survivor function declines relatively slowly between durations one and eleven years. The decline is, however, more rapid between durations twelve and seventeen years. This result is to be expected given the relationship between the survivor function and the hazard function in (5.9), which implies that a higher value of the estimated hazard function will lower the estimated value of the survivor function. Intuitively, if the hazard function is increasing then this means that from (5.8) the number of adopting institutions at time t_j , d_j , is increasing relative to the risk-set at time t_j , n_j . Consequently, from (5.9) an increasing hazard function implies a lower value for the number of survivors at t_j , $(n_j - d_j)$, and hence a more rapid decline in the survivor function. The underlying

reasons for the particular evolution of the survivor function are identical to that for the hazard function.

Finally, the plot of the integrated hazard function is shown in Figure 5.4. Referring to this figure and the results contained in Table 5.1, it can be observed that the integrated hazard is relatively linear between durations of one and ten years then relatively convex to the origin between eleven and seventeen years. This is to be expected given that from (5.10) the integrated hazard function can be given an ‘accumulated hazard function’ interpretation. During periods when the hazard is displaying no duration dependence the integrated hazard will be linearly increasing over time and when the hazard displays positive duration dependence the integrated hazard will be convex to the origin [Kiefer (1988)]. Thus, the integrated hazard reflects the non-monotonic nature of the hazard function in Figure 5.2.

Table 5.1: Non-parametric Hazard, Integrated Hazard and Survivor Function Estimates - 1972 to 1992							
Ordered Duration Number j	Duration in Years t_j	Censored Observations m_j	New Adopters d_j	Risk Set n_j	Hazard $\hat{h}(t_j)$	Survivor $\hat{S}(t_j)$	Integrated Hazard $\hat{\Lambda}(t_j)$
0	0	0	0	98	0	1.0000 (-)	0 (-)
1	1	0	1	98	0.0102	0.9898 (0.0102)	0.0102 (0.0103)
2	4	0	2	97	0.0206	0.9694 (0.0174)	0.0308 (0.0179)
3	5	0	1	95	0.0105	0.9592 (0.0200)	0.0413 (0.0209)
4	7	0	3	94	0.0319	0.9286 (0.0260)	0.0732 (0.0280)
5	8	0	1	91	0.0110	0.9184 (0.0277)	0.0842 (0.0302)
6	9	0	1	90	0.0111	0.9082 (0.0292)	0.0953 (0.0325)
7	11	0	2	89	0.0225	0.8878 (0.0364)	0.1178 (0.0359)
8	12	0	4	87	0.0460	0.8469 (0.0364)	0.1638 (0.0429)
9	13	0	8	83	0.0964	0.7653 (0.0428)	0.2602 (0.0559)
10	14	0	4	75	0.0533	0.7245 (0.0451)	0.3135 (0.0622)
11	15	0	4	71	0.0563	0.6837 (0.0470)	0.3698 (0.06874)
12	16	0	1	67	0.0149	0.6735 (0.0484)	0.3847 (0.0704)
13	17	63	3	66	0.0455	0.6429 (0.0484)	0.4302 (0.0752)

Note: Figures in parenthesis are estimated standard errors; t_j , m_j , d_j and n_j are defined in section 5.2.

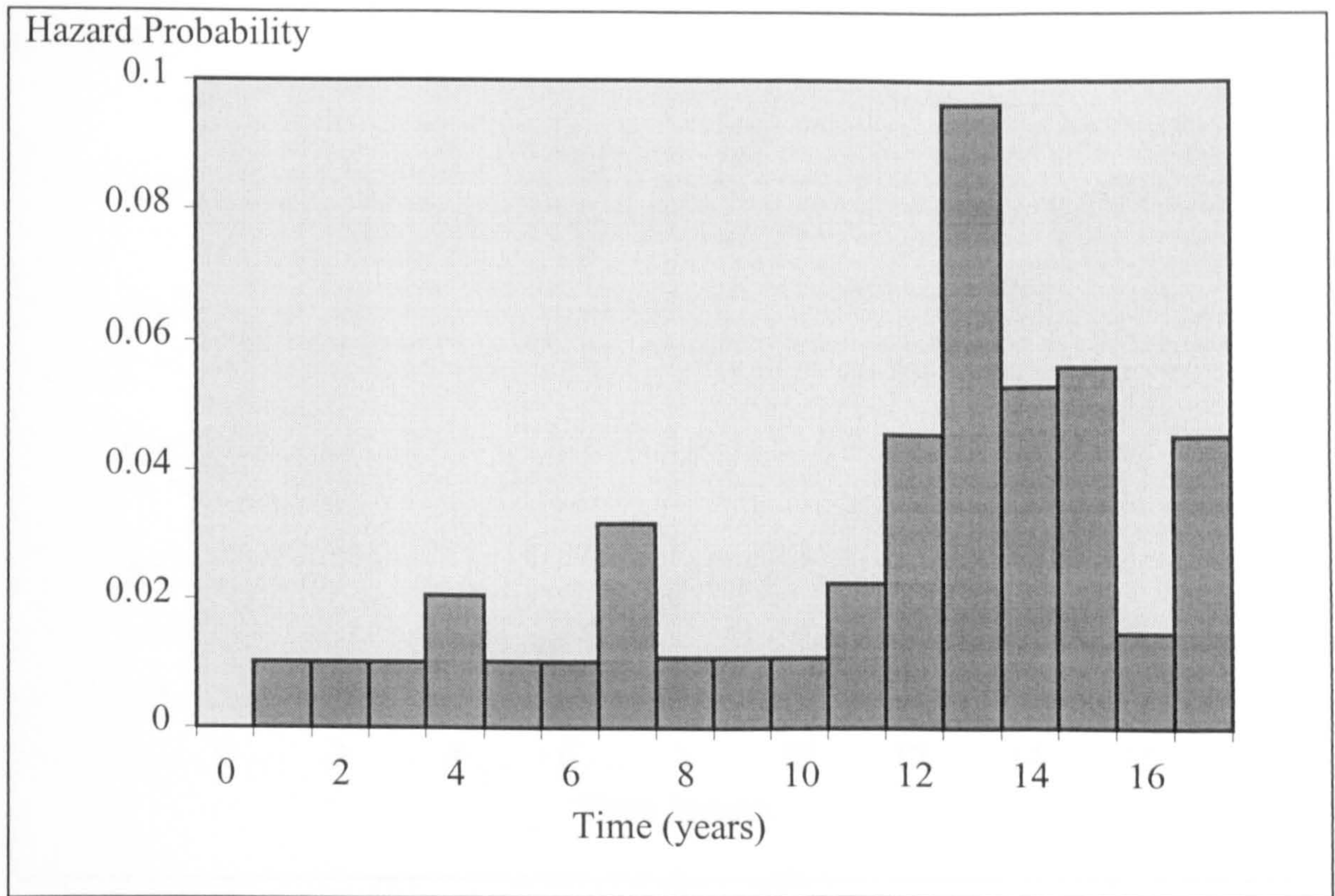


Figure 5.2: Estimated Non-parametric Hazard Function - 1972 to 1992

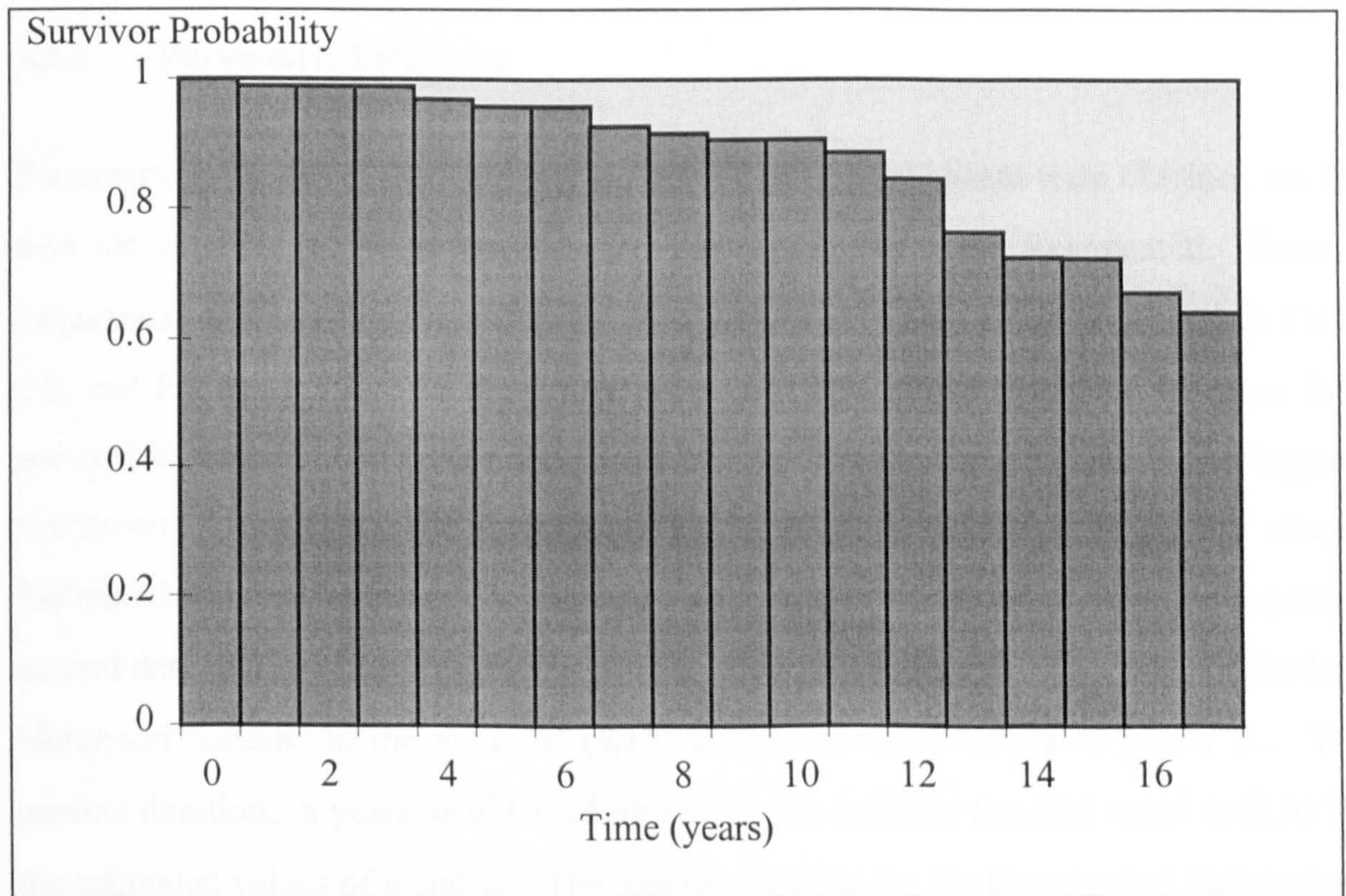


Figure 5.3: Estimated Non-parametric Survivor Function - 1972 to 1992

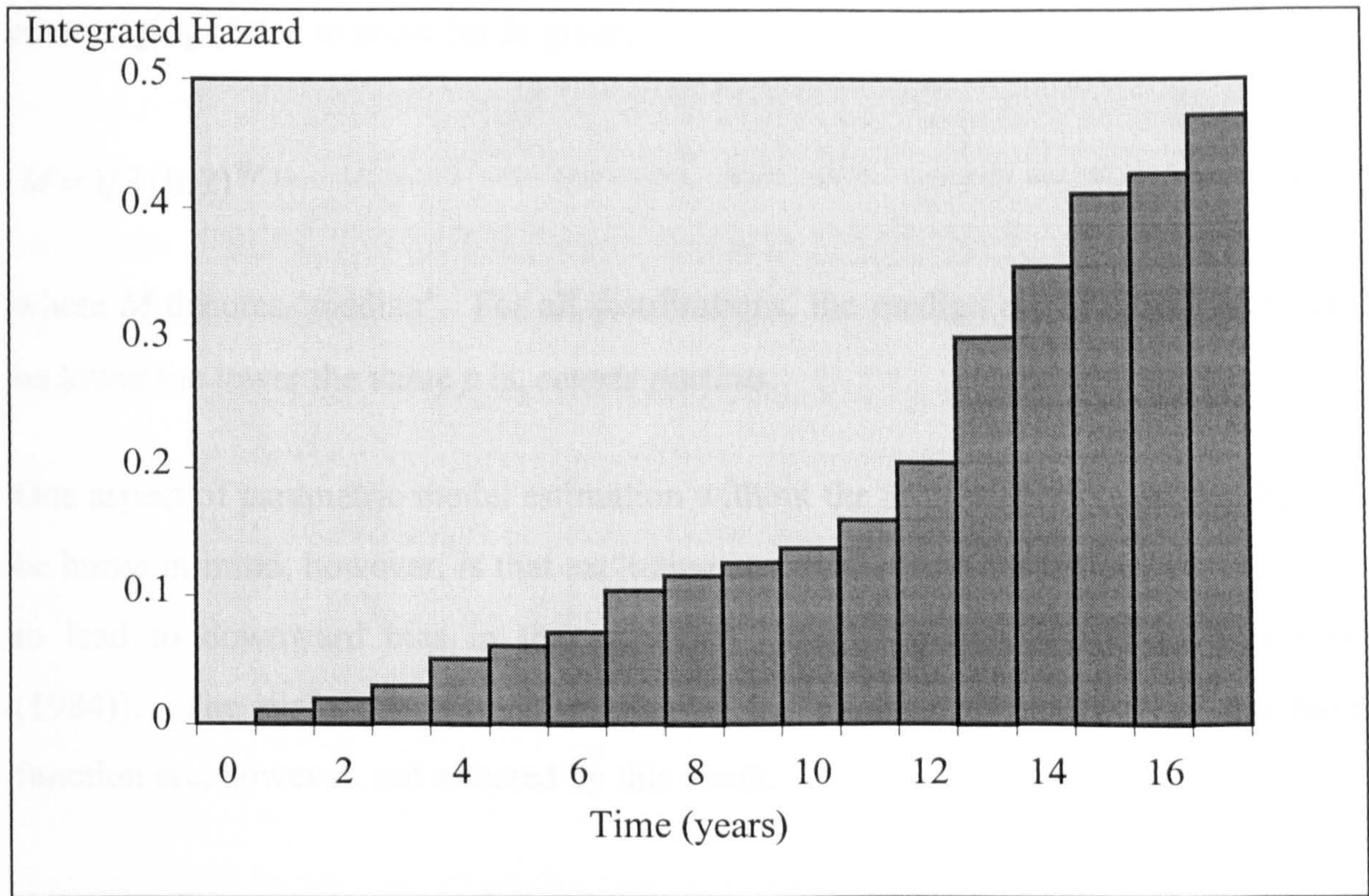


Figure 5.4: Estimated Non-parametric Integrated Hazard Function - 1972 to 1992

5.3.2 Parametric Estimates

Parametric estimates of the survivor, hazard and integrated hazard were obtained for the data set of ATM adoption histories for four distributions: the Exponential, Weibull, Lognormal and Log-logistic. The estimated values of p and λ are presented in Table 5.2, and Figures 5.5 to 5.10 show how these functions evolve overtime for these four parametric models. Unlike the non-parametric models, the functions are drawn beyond twenty-one years (i.e. beyond the year 1992) to show their overall general shape. Estimated standard errors of the estimated coefficients are obtained from the expected second derivatives of the log-likelihood (the information matrix). The value of the log-likelihood pertains to the value of (5.14) at the estimated values of p and λ . The median duration, in years, is obtained by setting the survivor function equal to 0.50 for the estimated values of p and λ . The median duration for the Exponential distribution, for example, can be found by employing the following equation [Greene (1993)]:

$$S(M) = 0.5 = e^{-(\lambda M)^p} \quad (5.16)$$

re-arranging (5.16) to solve for M gives:

$$M = 1/\lambda (\ln 2)^{1/p} \quad (5.17)$$

where M denotes 'median'. For all distributions, the median duration will necessarily be lower the lower the value p is, *ceteris paribus*.

One aspect of parametric model estimation without the inclusion of covariates that must be borne in mind, however, is that excluding covariates from the estimation is expected to lead to downward bias in the estimated hazard function [Heckman and Singer (1984)]. The main conclusions relating to the duration dependency of the hazard function are, however, not affected by this result.

Model	λ	p	Median Duration (years)	Log-likelihood
Exponential	0.0202** (0.0036)	1.0000 (-)	34.3008** (6.0644)	-89.1470
Weibull	0.0295** (0.0040)	1.6545** (0.2986)	27.1311** (3.7117)	-84.9890
Lognormal	0.0328** (0.0049)	0.9808** (0.1204)	30.5300** (4.5881)	-84.8450
Log-logistic	0.0357** (0.0044)	1.8795** (0.2944)	27.9815** (3.4599)	-84.3240
Number of Observations	98	98	98	98

Note: λ and p are defined in Appendix A5.2 and figures in parenthesis are the estimated standard errors; '**' signifies coefficients are significant at the 0.01 level. For the Exponential distribution the restriction $p=1$ is imposed.

The results presented in Table 5.2 and the subsequent sketches in Figures 5.5 to 5.10 tend to reinforce the results obtained by the non-parametric approach that the hazard function displays non-monotonicity during the observation period.

The Exponential distribution, as expected, exhibits no duration dependence in Figure 5.5 and predicts a constant hazard of 0.0202 ($= \lambda$) throughout the observational period. This implies that 2.02% of the surviving population should fail in each year after commercialisation. The resulting integrated hazard in Figure 5.9 is consequently linear against time reflecting the constancy of the hazard. By visual inspection of the predicted hazard from the Exponential distribution and the empirical one in Figure 5.2 it can be seen that the Exponential distribution gives a very poor representation of the underlying hazard function. It displays no non-monotonic behaviour and fails to predict the rise in the hazard during the 1984 to 1987 period when ATM diffusion was accelerating. Qualitatively, it is apparent that the Exponential distribution can be dismissed as a possible representation of the empirical hazard function.

The hazard, survivor and integrated hazard functions pertaining to the Weibull model are sketched in Figures 5.5, 5.7 and 5.9 respectively. The estimated model attains an estimated value of p equal to 1.6545. This implies that the resulting hazard function will be monotonic increasing. Subsequent sketches of the hazard and the integrated hazard confirm this result, the latter being convex in nature. The curve arguably gives a better representation of the empirical hazard than the Exponential but, like the Exponential, fails to predict the non-monotonic nature of the hazard.

The Lognormal distribution, by definition, is characterised by a non-monotonic hazard and this is independent of the value of p (although $p > 0$ is imposed). The resulting hazard, survivor and integrated hazard functions are depicted in Figures 5.6, 5.8 and 5.10 respectively. By visual inspection, this distribution seems to be a superior representation of the empirical hazard than either the Exponential or the Weibull. The hazard is non-monotonic and this is reflected in the initial convexity of the integrated hazard and its subsequent concavity. Qualitatively, the Lognormal hazard has the appropriate features required to represent the non-parametric hazard in that it rises then falls. The Lognormal distribution does, however, tend to over predict the hazard for durations between one and eleven years and under predicts the hazard for durations between twelve and seventeen years. Moreover, the hazard attains a maximum value of

approximately 0.030 at a duration of six years which is lower and occurs earlier than predicted by the empirical hazard.

The results obtained from the estimation of the Log-logistic model are significant because this distribution is the most flexible of all the parametric models in terms of potential duration dependence and allow for either negative duration dependence or non-monotonicity. The hazard, survivor and integrated hazard functions are sketched in Figures 5.6, 5.8 and 5.10 respectively. The estimated value of p is 1.8795, which implies that the hazard is non-monotonic. This, however, is not clear from the sketch of the hazard in Figure 5.6 or from the integrated hazard in 5.10, both of which apparently illustrate a monotonic increasing hazard function. From the summary details of the distribution contained in Table A5.2.2 of Appendix A5.2, the Log-logistic distribution reaches a maximum value at a time $t = (p - 1)^{1/p} / \lambda$. For the estimated values of p and λ given in Table 5.2 this predicts a maximum hazard is attained at $t = 26.1615$ years. This is just outside the range for t employed in the sketching of the functions and, thus, explains this apparent inconsistency. Moreover, this maximum is attained considerably after that of the non-parametric hazard. In addition, the Log-logistic distribution does suffer similar problems as the Lognormal distribution in that it tends to under predict the value of the hazard between one and four years and then between eleven and seventeen years. These problems are, however, less severe when compared to the Exponential and Weibull distributions.

One interesting aspect of the results shown in Table 5.2 is the large value obtained for the median length of duration. All distributions predict a median duration length in excess of twenty-five years which is greater than the censoring time of twenty-one years. The cause of this is the existence of a relatively high proportion of non-adopters (64.29% of potential adopters). From the specification of the likelihood function in (5.14) the contribution of these observations to the likelihood is $S(t_j, \theta)$, the probability that the duration is longer than the censoring time t_k . As indicated by Kiefer (1988), not taking account of right-censored observations will necessarily lead to upward bias in the value of p and hence a lower value for the median duration. Given that a majority of

institutions are right-censored it is, therefore, not surprising that the predicted median is in excess of the censoring time.¹²

The results obtained from the Log-logistic distribution are significant because they lend support to the hypothesis that the hazard for ATM diffusion is characterised by non-monotonicity, increasing in early years of ATM diffusion and the decreasing in later years. It is shown in Appendix A5.3 that this is indicative of the existence of epidemic effects in the diffusion process. Comparison with other empirical diffusion studies are, however, complicated because none of them to date explicitly consider the underlying duration dependence of the hazard function without the inclusion of covariates as this chapter has performed. As observed in Chapter 3, the vast majority of empirical studies allow the parametric hazard function to take the role of a 'baseline' hazard and it is the dependency of this that is generally checked. Comparison with other studies is therefore left until Chapter 6, which explicitly includes covariates in the modelling. It is, however, worth noting at this stage that the results from this chapter are not substantially changed by the inclusion of covariates and are in broad agreement with other empirical diffusion studies that have employed identical distributions to this chapter.

Although the non-parametric and parametric estimates of the hazard function indicate that the underlying hazard is non-monotonic, there is a need to assess the overall fit of the estimated models. From Table 5.2, for example, it can be seen that the Log-logistic and the Lognormal distributions give the best fit based on their lower values of the log-likelihood.¹³ As noted by Kiefer (1988), however, simply because Model A has a lower log-likelihood than, say, Model B does not imply that Model A is correctly specified. More rigorous model testing is, however, complicated because duration models do not have a direct counterpart to the set of regressor residuals with which to assess a specific specification [Greene (1993)]. With this objective in mind, the aim of Section 5.4 is to present results from a number of goodness-of-fit tests.

¹² All four models were estimated without non-adopting institutions and the predicted median duration times obtained were all below fifteen years confirming this analysis.

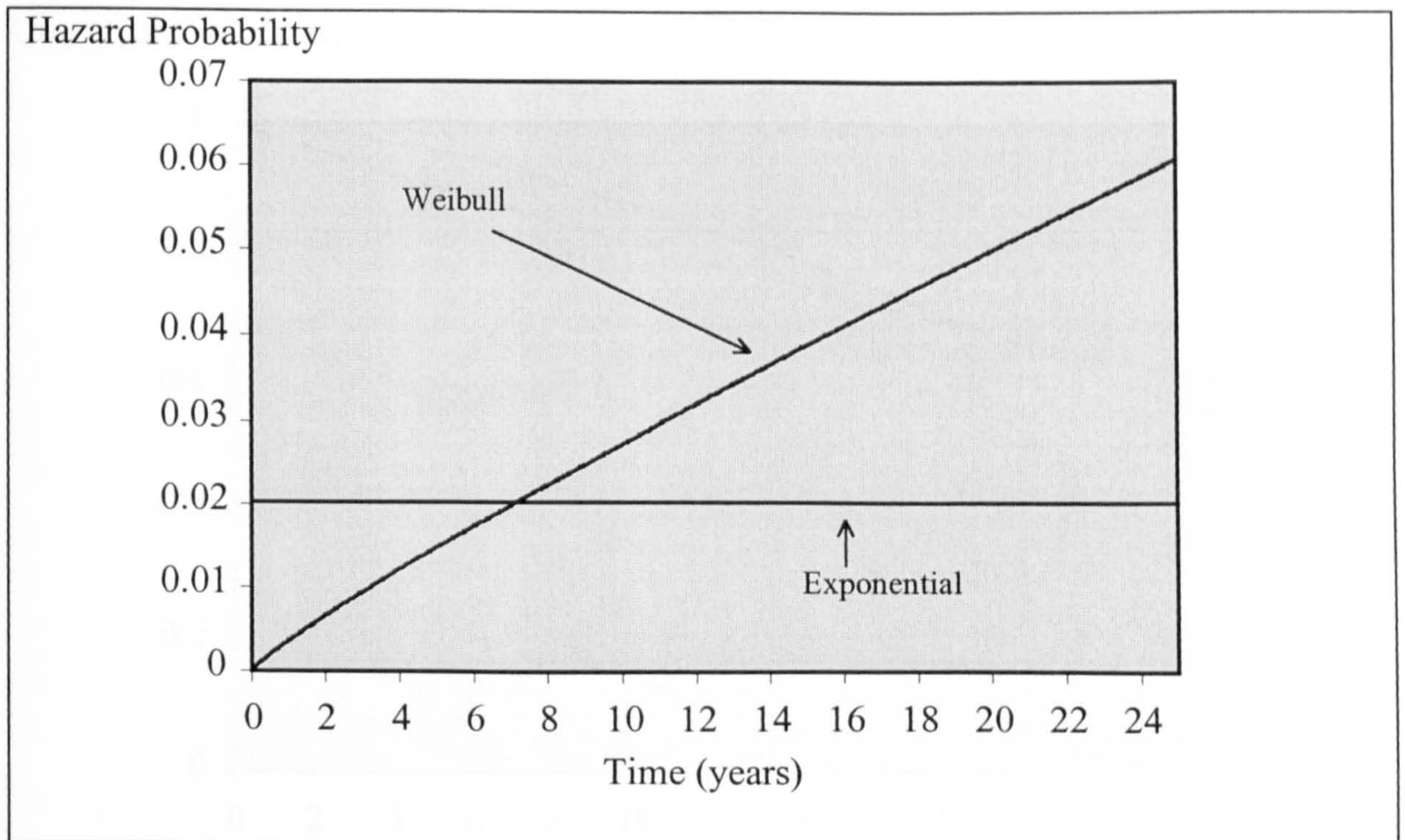


Figure 5.5: Estimated Exponential and Weibull Hazard Functions - 1972 to 1992

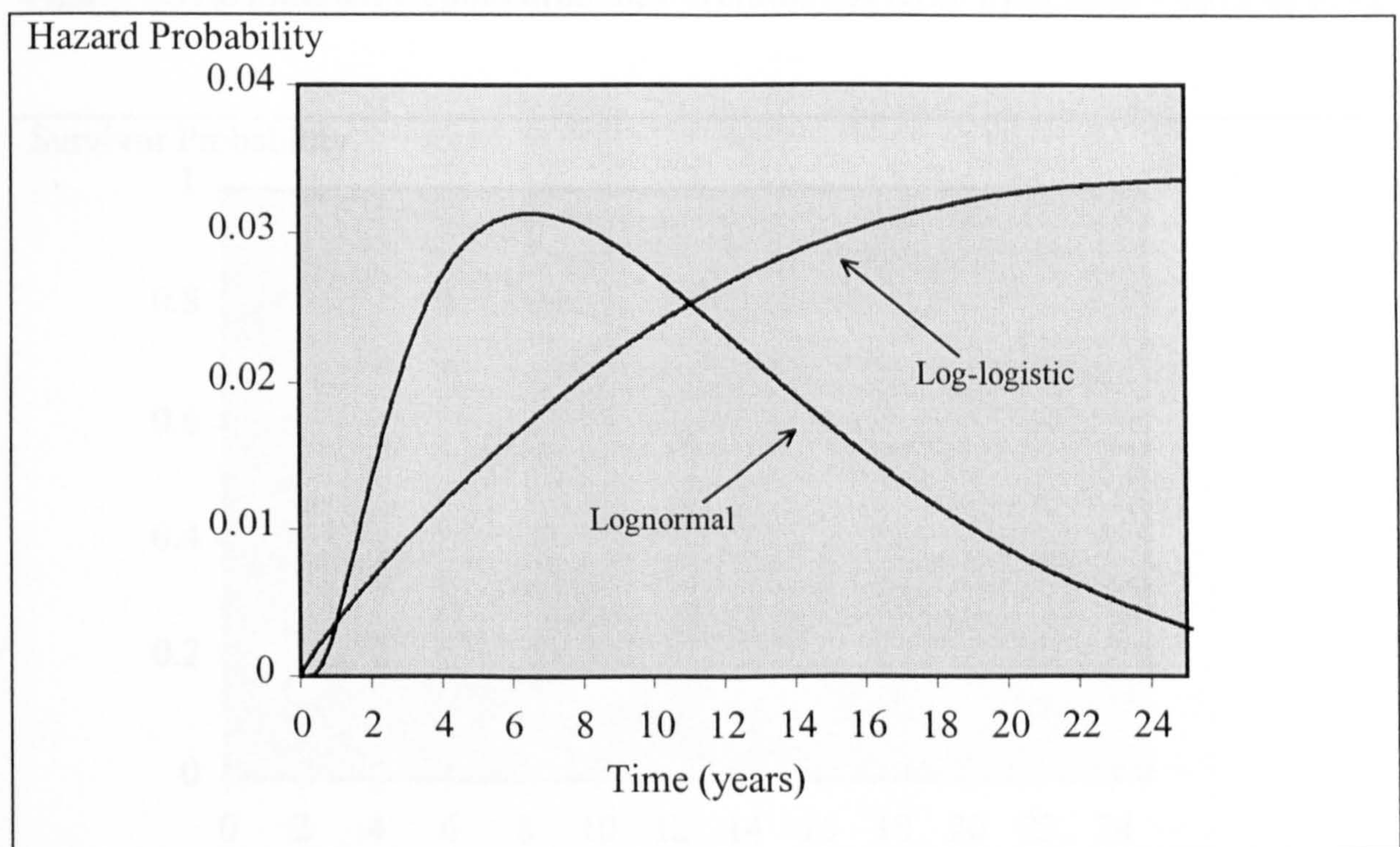


Figure 5.6: Estimated Lognormal and Log-logistic Hazard Functions - 1972 to 1992

¹³ This is the only model criteria used in Sinha and Chandrashekar (1992), for example, and is based on the fact that the value of the likelihood function results from maximising the probability of observing the sample of adoption times.

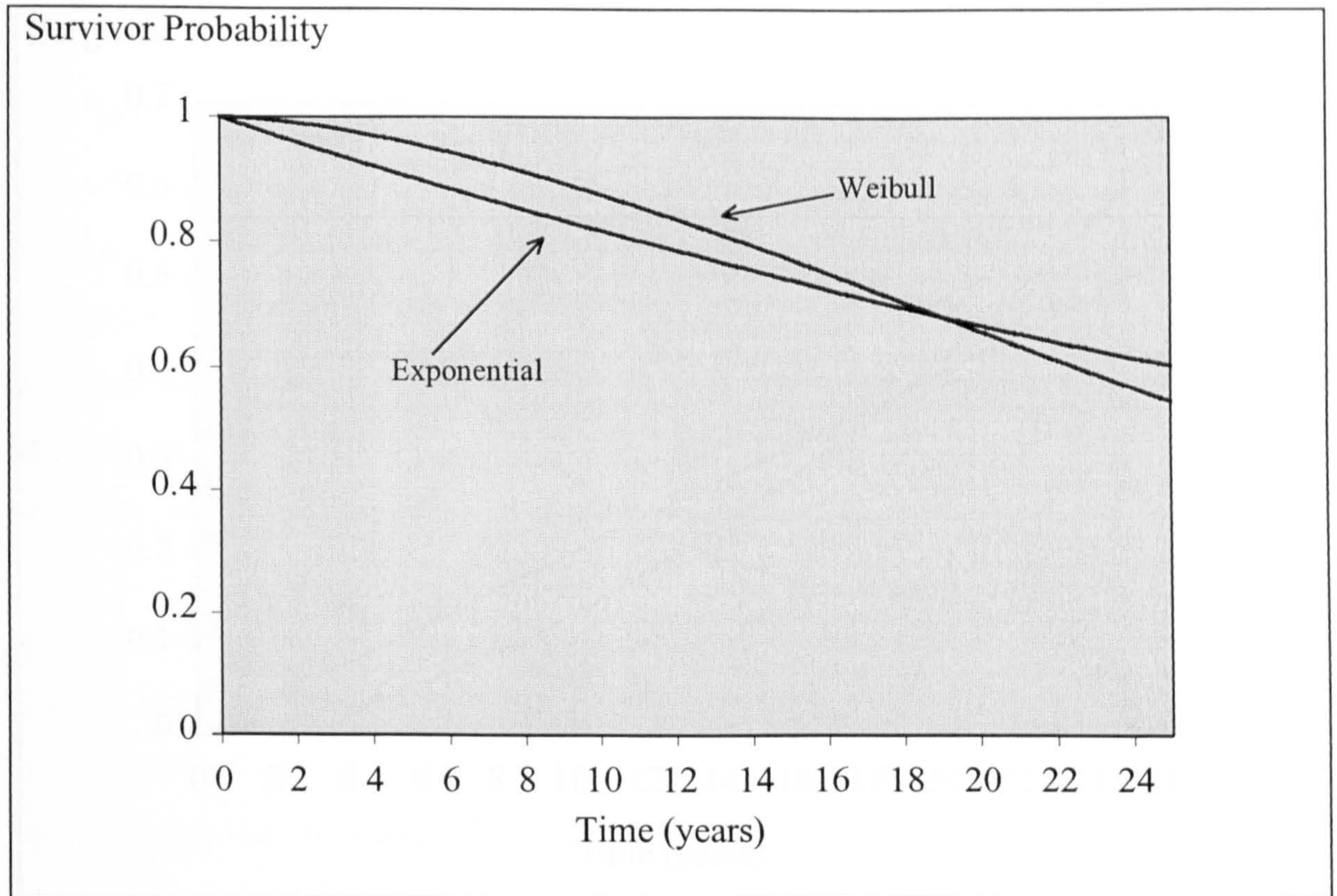


Figure 5.7: Estimated Exponential and Weibull Survivor Functions - 1972 to 1992

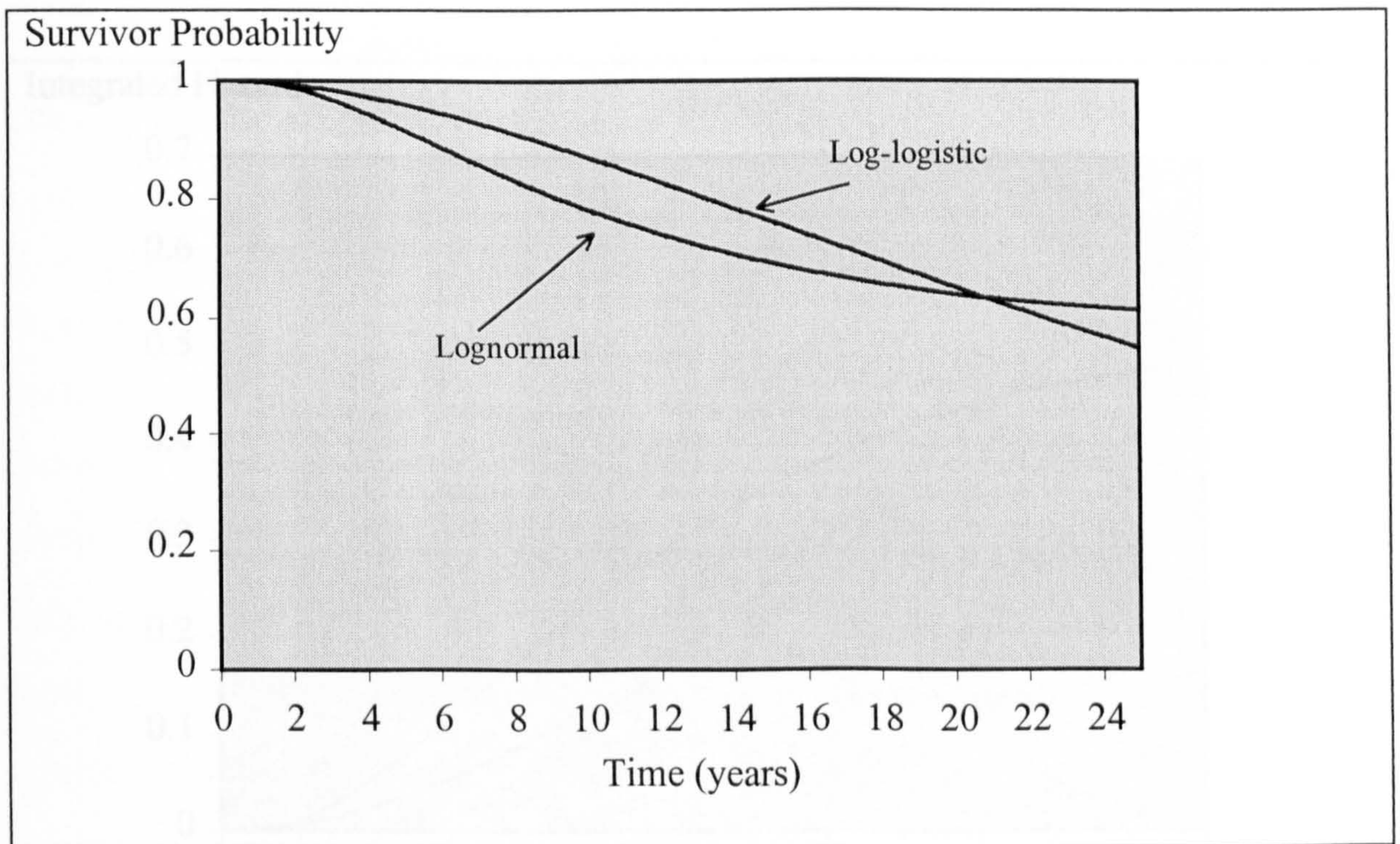


Figure 5.8: Estimated Lognormal and Log-logistic Survivor Functions - 1972 to 1992

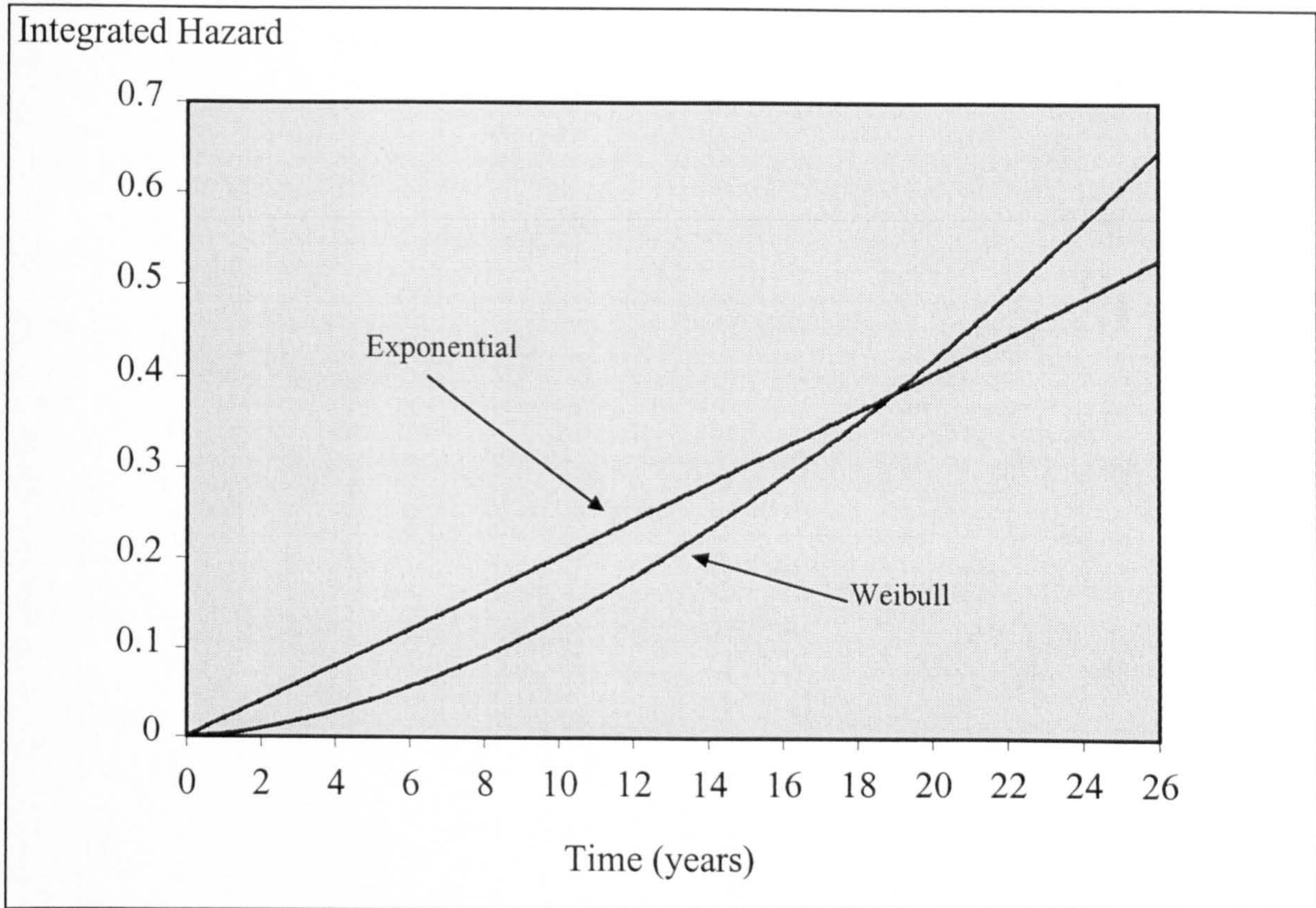


Figure 5.9: Estimated Exponential and Weibull Integrated Hazard Functions - 1972 to 1992

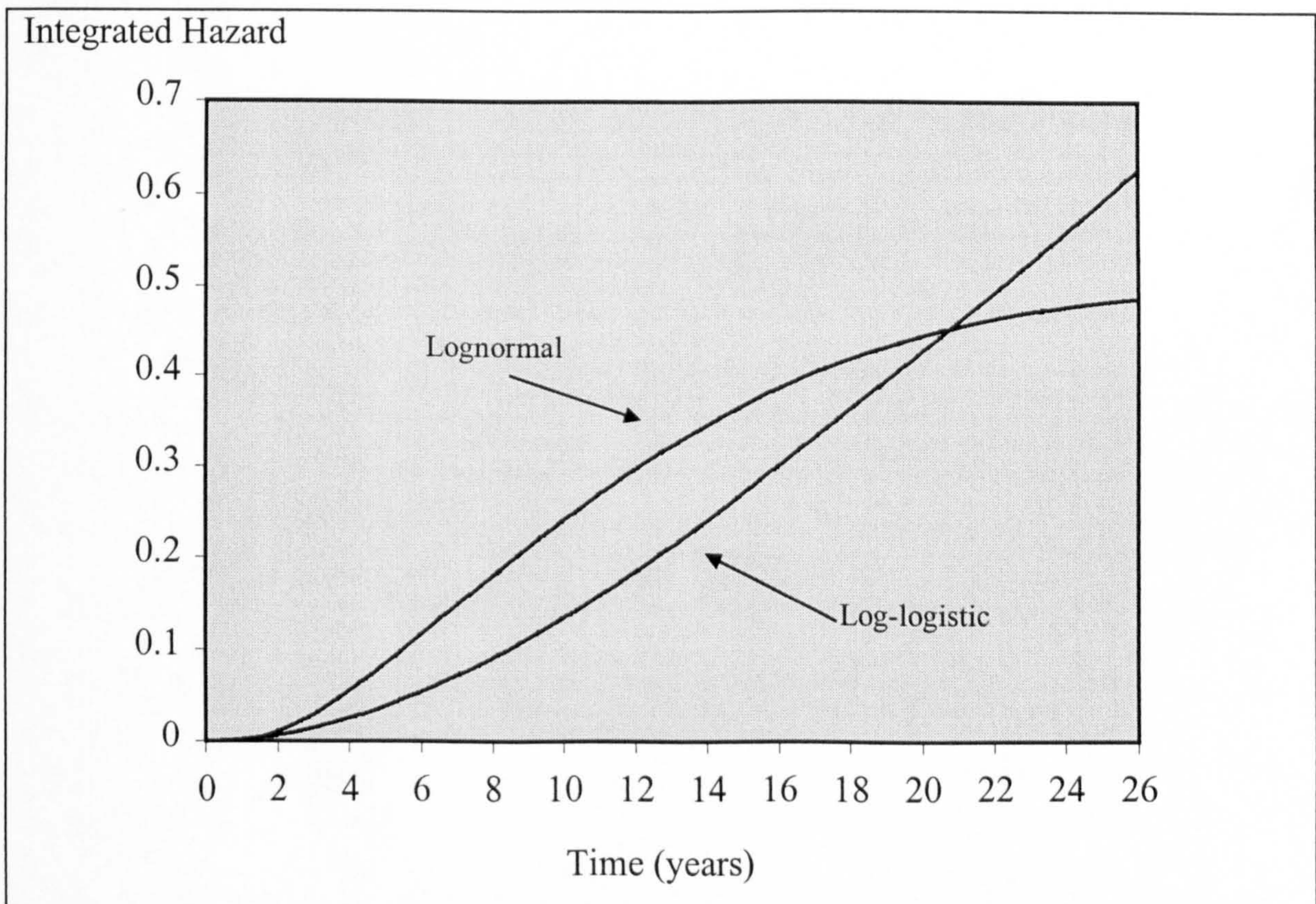


Figure 5.10: Estimated Lognormal and Log-logistic Integrated Hazard Functions - 1972 to 1992

5.4 Goodness-of-Fit Tests for Parametric Models¹⁴

It is important to check the overall statistical adequacy of the parametric models estimated in Section 5.3.2. As noted by Lawless (1982), Crowder *et al* (1991) and Neumann (1997), incorrect specification of the true population model may lead to inconsistent and misleading estimates of parameters variables and model misspecification when covariates are included. Comparison with the empirical hazard function by visual inspection is, however, not a rigorous approach to model checking and more formal statistical methods are required.

In this section a number of tests are employed on the data set of adoption histories with the aim of choosing between various parametric models and in assessing their overall fit. The central element of these tests is that of testing hypotheses about the population cumulative density function (CDF) of adoption times of the form:

$$H_0: F(t) = F_0(t) \tag{5.18}$$

where t is time, $F(t)$ is the population CDF and $F_0(t)$ is a specified family of CDFs (such as the Exponential for example).

It is usual in the literature [Conover (1971), Lawless (1982) and D'Agostino and Stephens (1986)] to refer to tests such as (5.18) as goodness-of-fit tests. It is also possible to distinguish between two classes of tests: *omnibus* and *directional* tests. The former of these are designed to be effective against wide classes of alternatives to a given $F_0(t)$. In these tests the alternative hypothesis, H_1 , is *composite* in that it gives little or no information on the distribution of the data and simply tests whether H_0 is false. In contrast, directional tests are tests that are effective at detecting certain specific types of departure from $F_0(t)$ [D'Agostino and Stephens (1986)].

¹⁴ Graphical analysis, as outlined in Kimber (1985), D'Agostino and Stephens (1986) and Crowder *et al* (1991), is a common approach in goodness-of-fit testing but these suffer similar problems to simple visual inspection of the estimated parametric function. In addition, with less than 25 distinct adoption times such methods were found to produce misleading results for extreme values of t and could not adequately deal with censored observations. Consequently, these approaches are not presented in this chapter.

Three tests are considered: a non-parametric omnibus test of duration dependency based on the Gini statistic proposed by Gail and Gastwirth (1978) and Heckman and Singer (1984, 1985); a directional test based on the likelihood ratio test; and an omnibus test based on the empirical distribution function (EDF) of adoption times.

5.4.1 A Non-Parametric Test of Duration Dependence

A non-parametric test of duration dependence has been suggested by Gail and Gastwirth (1978) and shown by Heckman and Singer (1984, 1985) to have good power against alternative distributions with monotone hazard functions. Assuming that a random sample of adoption times of $t_1 \leq \dots \leq t_r$ are observed in a sample of size of n with $n - r$ observations being right-censored, then Gail and Gastwirth (1978) have proposed the following test statistic:

$$G_{r,n} = \left(\sum_{i=1}^{r-1} iW_{i+1} \right) / (r-1) \sum_{i=1}^r W_i \quad (5.19)$$

where $t_0 = 0$ and the W_i 's are the scaled spacings defined as:

$$W_i = (n - i - 1)(t_i - t_{i-1}) \quad (5.20)$$

with $i = 1, \dots, n$. The test is based around the fact that if the adoption histories are derived from an Exponential distribution then the W_i/λ 's ($i = 1, \dots, r$) have a standard Exponential distribution [Lawless (1982)]. Thus, the test takes the following form:

$$\begin{aligned} H_0: W_i/\lambda \ (i = 1, \dots, n) \text{ has a standard Exponential distribution} \\ H_1: W_i/\lambda \ (i = 1, \dots, n) \text{ does not have a standard Exponential distribution} \end{aligned} \quad (5.21)$$

The range of $G_{r,n}$ is $[0,1]$. Values of $G_{r,n}$ close to 0 or 1 provide evidence against exponentially. Applying (5.19) to the set of adoption histories it was found that

$G_{r,n} = 0.2139$. As $n > 20$, the following normal approximation suggested by Gail and Gastwirth (1978) was applied:

$$[12(n-1)]^{1/2} (G_n - 0.5) \sim N(0,1) \quad (5.22)$$

Calculating (5.22) obtains a value of -9.7610 for $G_{r,n} = 0.2139$. For a two-tailed test at the significance level of 0.05 this gives critical points of ± 2.5670 . Since $|-9.7610| > 2.5670$, H_0 cannot be accepted and it can be concluded that evidence suggests that the sample of adoption histories does not come from an Exponential distribution. This result provides further evidence that the hazard function for ATM diffusion cannot be considered a constant.

5.4.2 Nested Tests Based on the Likelihood Ratio Test

It is possible to test the validity of imposed linear and non-linear parameter restrictions in maximum likelihood estimation by using the likelihood ratio test. This can also be interpreted as a test of overall fit because one parametric model can be nested within another through a particular restriction. Suppose the unrestricted estimated vector of parameters is $g(\hat{\theta})$ and that of the unrestricted parameters is $g(\theta')$, then the likelihood ratio test can be used to test the general restriction:

$$\begin{aligned} H_0: g(\theta) &= 0 \\ H_1: g(\theta) &\neq 0 \end{aligned} \quad (5.23)$$

Cuthbertson *et al* (1992) have outlined the conditions under which testing (5.23) can proceed and it can be noted that these are met for all the parametric models estimated in this chapter. Moreover, the simplest forms for $g(\theta)$ are, for example, $p = 0$ or $p = 1$. The likelihood ratio test is then defined as [Cuthbertson *et al* (1992)]:

$$\text{LRT} = 2 \left\{ \ln[L(\hat{\theta})] - \ln[\theta'] \right\} \sim \chi^2(m) \quad (5.24)$$

where 'LRT' is 'likelihood ratio test', $L(.)$ is the log-likelihood defined in (5.14) and m the number of restrictions.

From Appendix 5.3 it can be observed that the Weibull distribution reduces to the Exponential when the restriction $p = 1$ is imposed. Thus, the following test was carried out:

$$\begin{aligned} H_0: p &= 1 \\ H_1: p &\neq 1 \end{aligned} \tag{5.25}$$

Under H_0 the data conforms to the Exponential distribution. Using the values of the log-likelihoods in Table 5.3, with the Exponential model representing the restricted model, the following test statistic was computed: $LRT=2[-84.9890+89.1470] = 8.3160$. At a significance level of 0.05 the critical value of $\chi^2(1)$ is 6.6300. Since $LRT > 6.6300$, H_0 cannot be accepted and it is concluded that the Exponential model is rejected as a possible representation of the adoption data.

5.4.3 Tests based on the Empirical Distribution Function (EDF)¹⁵

Tests based on the empirical distribution function (EDF) are used predominately for testing the fit of a particular parametric distribution to a sample distribution and are, therefore, ideal for assessing the overall fit of the parametric models estimated in Section 5.3. In addition, these tests have the advantage of being distribution-free in that the distributions of the statistics under H_0 do not depend on the distribution being tested for [Lawless (1982)].

The tests assume that T is a continuous random variable as defined in Section 5.2 and that there exists a sample of durations t_1, \dots, t_n from the distribution for T . The tests are then:

¹⁵ Leung (1997) has recently used this approach.

$$\begin{aligned}
 H_0: t_1, \dots, t_n \text{ comes from } F(t, \theta) \\
 H_1: t_1, \dots, t_n \text{ does not come from } F(t, \theta)
 \end{aligned}
 \tag{5.26}$$

where θ is a vector of known¹⁶ parameters, and $\theta = \lambda, p$ for the parametric distributions estimated in Section 5.3.

Tests then proceed by measuring the discrepancy between the EDF and a given parametric CDF. The EDF is $F_n(t)$ and is characterised by a step function. It records the proportion of observations less than or equal to t and is defined by:¹⁷

$$\hat{F}_n(t) = \frac{\text{number of observations } \leq t}{n}; \quad -\infty < t < \infty
 \tag{5.27}$$

with:

$$\hat{F}_n(t) = 0, \quad t < T_1
 \tag{5.28}$$

$$\hat{F}_n(t) = i/n, \quad T_i \leq t \leq T_{i+1}, \quad i = 1, \dots, n-1
 \tag{5.29}$$

$$\hat{F}_n(t) = 1, \quad T_n \leq t
 \tag{5.30}$$

The EDF statistic then measures the difference between $F_n(t)$ and $F(t, \theta)$. A number of tests have been formulated for testing (5.26) and are based around different measures of this discrepancy [see D'Agostino and Stephens (1986) for a review]. For ease of computation the Kolmogorov statistic [Miller and Miller (1998)] was selected. This test statistic measures the discrepancy as the largest vertical difference when $F_n(t)$ is greater than $F(t)$ and, again, the largest vertical difference when $F_n(t)$ is smaller than $F(t)$. Formally, these are respectively:

$$D^* = \sup_t [F_n(t) - F(t, \theta)]
 \tag{5.31}$$

and

¹⁶ See D'Agostino and Stephens (1986) for approaches when θ is not fully known.

$$D^- = \sup_t [F(t, \theta) - F_n(t)] \quad (5.32)$$

where 'sup' is 'supremum'.

The Kolmogrov statistic is then given by:

$$D = \sup_t |F_n(t) - F(t)| = \max(D^+, D^-) \quad (5.33)$$

Computation proceeds by using the Probability Integral Transformation (PIT), which is defined as $Z = F(t, \theta)$. This transforms the adoption times to realisations from a parametric distribution with a known parameter vector θ . Thus, adoption times t_1, \dots, t_n will be transformed into times $Z_i = F(t_i, \theta)$, $i = 1, \dots, n$ for a particular distribution. The discrepancy between these values and the EDF in (5.27) is central to the test. For a sample of adoption times that are right-censored, t_1, \dots, t_r , with $r < n$ the Kolmogrov statistic in (5.30) is calculated as D_i^* [D'Agostino and Stephens (1986)]:

$$D_i^* = \sqrt{n} D_{i,n} + 0.19/\sqrt{n} \quad (5.34)$$

where:

$$D_{i,n} = \max_{1 \leq i \leq r} [i/n - Z_i, Z_i - (i-1)/n, t - r/n] \quad (5.35)$$

and $t = F(L, \theta)$, where L is the censoring time.

The results from calculating (5.34) for the four parametric distributions estimated in Section 3 are summarised in Table 5.3 below. The test statistics are computed at the estimated values of λ and p contained in Table 5.3.

¹⁷ It has been shown by D'Agostino and Stephens (1986) that (5.27) is a consistent estimator for $F(t)$.

Table 5.3: Summary of Kolmogrov EDF Statistics for Estimated Parametric Models

Distribution	D_i^*	Critical Value	Decision
Exponential	2.2783	1.3581	Reject H_0
Weibull	2.1515	1.3581	Reject H_0
Lognormal	1.3071	1.3581	Do Not Reject H_0
Log-logistic	1.3467	1.3581	Do Not Reject H_0

Note: Critical values are calculated at the 0.01 level of significance and are taken from Koziol and Bayer (1975).

The results show that H_0 is not accepted for the Exponential and Weibull distributions but is not rejected for the Lognormal and Log-logistic. This result provides evidence to support the hypothesis that ATM adoption histories are characterised by a non-monotonic hazard function. Those distributions having a constant - the Exponential - or a monotone - the Weibull - hazard are not supported by the evidence as being representative of the adoption data.

5.5 Concluding Remarks

There were three aims to this chapter. The first was to provide non-parametric and parametric estimates of the survivor, hazard and integrated hazard functions for the set of ATM adoption histories. The second was to explore the duration dependency of the hazard function and the third was to assess the adequacy of estimated parametric distributions using a variety of goodness-of-fit tests, which is often lacking in the current literature.

The empirical results from non-parametric estimation presented in Section 5.3.1 indicate that the hazard function is characterised by non-monotonicity, increasing during the initial stages of the diffusion process and then decreasing in latter periods. This is supported and reinforced by the parametric results presented in Section 5.3.2 and the goodness-of-fit tests carried out in section 5.4. The goodness-of-fit tests reject the Exponential and Weibull distributions, but do not reject the Lognormal or the Log-logistic. Overall, the results are indicative of the existence of epidemic effects in the

diffusion of ATMs. It was also found that the sampling design, which necessarily excludes exiting institutions during the observation period, can be ignored for estimation purposes.

Finally, it is significant to note that all the parametric results predict a median duration length in excess of the censoring time. An explanation for this lies in the relatively large number of non-adopters in the data set (64.29%) which lowers the estimated value of p in the parametric models.

To summarise, the implications of the analysis suggest that empirical economists need to be more aware of the characteristics of their data set and employ more rigorous model testing before implementing parametric estimation of duration models. The frequently used Weibull distribution, for example, is not supported by goodness-of-fit tests and overall the empirical results indicate that non-monotonic distributions are more appropriate. This empirical evidence may be interpreted as indicating the existence of epidemic effects in the diffusion of ATMs.

A5.1 Appendix One: The Effects of Sample Attrition

A5.1.1 Introduction

As discussed extensively by Hoem (1985), the systematic exclusion of institutions such as A5 and A6 in Figure 5.1 from the sample of potential adopters may introduce selection bias in the sampling distribution of the model. This occurs because such institutions have zero probability of being selected even though their corresponding durations may belong to the set of potential adopters. Lancaster (1990) has described the sampling plan that results in such bias as *selection by virtue of survival*.

The consequences of this bias for the likelihood function have been discussed in Karshenas and Stoneman (1993). They show that the likelihood function must be made conditional on survival of the institutions beyond 1992, the last year of the empirical study. Following Karshenas and Stoneman, the probability of adoption at time t conditional on the exit time x being greater than x^* [$x^* = \min\{21, (1992\text{-entry date})\}$] can be written as:

$$f(t|x > x^*) = \frac{f(t) \int_{x^*}^{\infty} g(x|t) dx}{\int_{x^*}^{\infty} \int_0^{\infty} g(x|t) f(t) dt dx} \quad (\text{A5.1.1})$$

where $g(x|t)$ is the conditional density of exit time given the adoption time. As stated by Hoem (1985), the sampling plan is ignorable if the probability density of exit time is independent of that of adoption time. If the two densities are independent then the right-hand side of (A5.1.1) becomes equal to $f(t)$ and the sample likelihood equals the population likelihood. If, however, this condition is not met then the sampling plan is not ignorable and Hoem (1985) has argued that to counteract the bias requires weighing each observation by their respective probability of being selected.

A5.1.2 Empirical Analysis

As indicated by Karshenas and Stoneman (1993), a retrospective study does not generate a rich enough source of information required to test the independence of the probability density of exit time and that of adoption time. An alternative procedure involves a follow-up survey of the original sample and then stratifying the them according to adoption dates. From this the calculation of the conditional frequencies can be carried out. This is shown in Table A5.1.1 for the follow-up period 1992 to 1996. From this table there appears to be no systematic variation of exit frequencies across adoption times, with only one institution exiting in the proceeding period of 1992 to 1996.

Adoption time (in years, 1972=0)	0-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21
Number of adopters	1	0	3	2	3	2	11	9	4	0	0
Number of exits	0	0	0	0	0	0	0	0	1	0	0
Exit frequency	0	0	0	0	0	0	0	0	0.25	0	0

To test this hypothesis more formally Box *et al* (1978) have indicated that adoption times, t , and exit times, x , will be linearly independent if $\text{Cov}(t,x) = 0$, where 'Cov' is the covariance. The covariance is not independent of scale and so the correlation coefficient, ρ , is calculated between adoption and exit times. This is given by [Box *et al* (1978)]:

$$\rho = \frac{\text{Cov}(t,x)}{\sigma_t \sigma_x} \quad (\text{A5.1.2})$$

where σ_t and σ_x are the standard deviations for adoption and exit times respectively.

The sample correlation coefficient may then be calculated as:

$$r(t_i, x_i) = \frac{\sum_{i=1}^n (t_i - \bar{t})(x_i - \bar{x}) / (n-1)}{s_t s_x} \quad (\text{A5.1.3})$$

where s_t is the sample standard deviation of adoption times and s_x the sample standard deviation of exit times. The following two-tailed hypothesis is tested:

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &\neq 0 \end{aligned} \tag{A5.1.4}$$

Under H_0 adoption times and exit times are linearly independent. Fleming and Nellis give the test statistic as follows:

$$t = \frac{r}{\sqrt{[(1-r^2)/(n-2)]}} \tag{A5.1.5}$$

which has a t distribution with $(n-2)$ degrees of freedom. Using the data contained in Table A5.1.1 it is found that $r = 0.1419$ with $t(33) = 0.8232$. With a level of significance of $\alpha = 0.05$ then for a two-tailed test; $t_{0.025}(33)$ have points ± 2.042 . Because $t(33) < t_{0.025}(33)$ H_0 cannot be rejected and it is concluded that adoption times and exit times are linearly independent. Thus, the sample design is ignorable for estimation purposes.

A5.1.3 Concluding Remarks

In this Appendix the consequences for maximum likelihood estimation of sample attrition resulting from the sample design was examined. The results indicate that adoption histories and exit times are linearly independent and, thus, the sample design is assumed to be ignorable for estimation.

A5.2 Appendix Two: Parametric Duration Models

This Appendix examines the characteristics and properties of four parametric distributions: The Exponential, the Weibull, the Lognormal and the Log-logistic. All distributions are continuous and assume that the population is homogenous to any systematic factors. As throughout the thesis T is assumed to be a random variable representing an individual institutions duration time that has a range of $[0, \infty)$ and t represents a typical point in this range.

A brief description of each distribution is given below and the characteristics and properties of each one is summarised in Table A5.2.1 For further details of these distributions see Kalbfleisch and Prentice (1980) and Lawless (1982). In all models the parameter p determines the nature of the duration dependence.

The Exponential Distribution

This is a two parameter distribution and is obtained by taking the hazard function to be constant, $\lambda(t) = \lambda > 0$, over the range of T . Consequently, the distribution exhibits no duration dependence. The conditional probability of adoption is thus independent of how long the institution has been in a state of non-adoption. With only one parameter the distribution is not flexible. The mean of the distribution is given as $E(T) = 1/\lambda$ and $\text{Var}(T) = 1/\lambda$ and so the mean and variance cannot be adjusted separately.

The Weibull Distribution

This is a two parameter distribution, with λ being a scale parameter and p being a shape parameter. It is a simple generalisation of the Exponential distribution with $p \neq 1$. It exhibits monotone duration dependence depending on the value of p (see Table A5.2.1). It has a simple form vis-à-vis the Lognormal and Log-logistic distributions and this is one reason why it has been so commonly used in empirical diffusion studies. In addition, the distribution is related to the Extreme Value distribution. If T has a Weibull distribution $\ln T$ will have an Extreme Value Distribution [Lawless (1982)].

The Lognormal Distribution

This is a two parameter distribution. It arises when T is lognormally distributed with a mean $-\ln \lambda$ and standard deviation $1/p$. The resulting hazard is non-monotonic for all values of $p > 0$.

The Log-logistic Distribution

This is a two parameter distribution and results when $\ln T$ has a logistic distribution with mean $-\ln \lambda$ and standard deviation $\pi^2/(3p^2)$. The hazard is identical to that of the Weibull except for the denominator $1 + (\lambda t)^p$. It can exhibit both negative duration dependence and non-monotonicity (see Table A5.2.1).

Table A5.2.1: Summary of Parametric Duration Models

Model	Number of parameters	Hazard Function	Survival Function	Model Restrictions	Nature of duration dependence
Exponential	1	λ	$\exp(-\lambda t)$	$p = 1$	No duration dependence
Weibull	2	$\lambda p (\lambda t)^{p-1}$	$\exp[-(\lambda t)^p]$	$p \neq 0$	Monotone increasing for $\forall t$ if $p > 1$; monotone decreasing for $\forall t$ if $p < 1$; none for $\forall t$ if $p = 1$ (reduces to exponential model)
Lognormal	2	$\frac{(p/t)\phi[(p \ln(\lambda t))]}{\Phi[-p \ln(\lambda t)]}$	$\Phi[-p \ln(\lambda t)]$	$p > 0$	Non-monotonic: increases monotonically from $t = 0$, obtains a maximum value determined by the value of p and then approaches 0 as t approaches infinity. The higher the value of p , <i>ceteris paribus</i> , the higher the maximum value of the hazard.
Log-logistic	2	$\frac{\lambda p (\lambda t)^{p-1}}{1 + (\lambda t)^p}$	$\frac{1}{1 + (\lambda t)^p}$	$p > 0$	Monotone decreasing from infinity for $\forall t$ if $p < 1$; monotone decreasing from λ for $\forall t$ if $p = 1$; increases from 0 to a maximum at $t = (p - 1)^{1/p} / \lambda$ and decreases toward 0 thereafter if $p > 1$.

Note: Φ and ϕ are the standard normal CDF and PDF respectively.

A5.3 Appendix Three: Economic Theory and Duration Dependence

A5.3.1 Introduction

The formal theories of diffusion outlined in Chapter 2 do not explicitly consider the consequences for duration dependence resulting from the assumptions that they make. From the definition of the hazard function in equation (5.1), however, it can be seen that the shape of the hazard function will depend crucially on the diffusion curve - that is, how the total number of adopters changes over time.¹⁸ Intuitively, from (5.8) if the inter-firm diffusion curve is sigmoid, which is the case of ATM diffusion in the UK as summarised in Figure 4.2 of Chapter 4, then during the early stages of diffusion, where the curve is convex, the number of adopters will be increasing relative to the risk-set and the hazard function will be increasing. In later stages of the diffusion process, when the diffusion curve becomes concave in shape, the number of adopters is decreasing relative to the risk set and hazard will therefore be decreasing. A number of insights may therefore be gained from examining the epidemic and rank effects models which do provide predictions for the shape of the diffusion curve.

A5.3.2 The Epidemic Model

Under the assumptions and reasoning presented in Section 2.3, the epidemic model predicts that the average number of adopters in a small time interval dt will be:

$$dS_t = \beta(S_t/N)(N - S_t)dt \quad (\text{A5.3.6})$$

where t is time, α is the constant of integration, β the rate of diffusion, S_t the number of adopters at t and N the number of potential adopters (assumed constant). Integrating

¹⁸ As noted by Stoneman (1983), the diffusion curve may be interpreted as the CDF derived from a particular distribution.

(A5.3.6) gives the diffusion of the new technology over time, which is a logistic curve defined as:

$$S_t = N/[1 + \exp(-\alpha - \beta t)] \quad (\text{A5.3.7})$$

Karshenas and Stoneman (1993) have shown that substitution of (A5.3.6) into (A5.3.7) yields an 'epidemic hazard function' defined as:

$$h(t; \Theta) = (dS_t/dt)/(N - S_t) = [\beta \exp(\alpha + \beta t)]/[1 + \exp(\alpha + \beta t)] \quad (\text{A5.3.8})$$

where $\Theta = (\alpha, \beta)$. Interpretation of (A5.3.8) is the same as for all hazard functions: it gives the conditional probability for a firm that has not adopted the technology at time t that it will adopt in the next small time interval $(t, t + dt)$. Employing the term $(dS_t/dt)/(N - S_t)$ as a measure of the hazard function is similar to that of the non-parametric estimate in (5.8). It gives the average number of adopters as a ratio of the risk set at time t .

From (A5.3.8), duration dependence may be examined by the value of:

$$dh(t; \Theta)/dt = \beta^2 \exp[(\alpha + \beta t)]/[1 + \exp(\alpha + \beta t)]^2 \stackrel{?}{\leq} 0 \quad (\text{A5.3.9})$$

With parameter restrictions $\alpha > 0$ and $\beta < 0$, (A5.3.9) will exhibit non-monotonic duration dependence with the hazard at first increasing and then decreasing [Karshenas and Stoneman (1993)]. As an indication of parameter signs for the set of ATM adoption histories, (A5.3.7) was estimated by Non-Linear Least Squares [Greene (1993)] over the observation period. The results are summarised in Table A5.3.2 for the period 1972 to 1992.

The results in Table A5.3.2 indicate that the parameter restrictions $\alpha > 0$ and $\beta < 0$ are supported by the data. Thus, this lends further support to the contention that the hazard

function is non-monotonic and that a possible explanation for this is the epidemic model of diffusion.

Coefficient	Estimate
α	3.8255** (0.2594)
β	-0.1709** (0.0151)
R^2	0.9299
Number of observations	21

Note: Figures in parenthesis are estimated standard errors; ‘**’ indicates estimated coefficient is significant at the 0.01 level.

A5.3.2 Rank Effects

As shown in Chapter 2, the rank effects model as outlined by Davies (1979) and David (1991), emphasise the central role that firm-specific characteristics play in the determining the gross returns from adopting new technology and, hence, determining acquisition times. The shape of the diffusion curve and, thus, the hazard function will depend on two aspects of these models. Firstly, how benefits are related to firm-specific characteristics. Secondly, how characteristics are distributed across the population. *A priori* it is not possible to predict a non-monotonic hazard function, although Stoneman (1986) has shown under what conditions a logistic diffusion curve will exist and therefore a non-monotonic curve will occur.

CHAPTER 6

THE DETERMINANTS OF ATM DIFFUSION: EVIDENCE FROM THE UK FINANCIAL SECTOR 1972 TO 1992¹

6.1 Introduction

The previous chapter estimated non-parametric and parametric forms of the survivor, hazard and integrated hazard functions under the assumption that the distribution of adoption times is homogenous to any systematic differences. This methodology was employed in order to compare and contrast estimates obtained from these two approaches and to observe the underlying nature of duration dependence in the diffusion of ATMs. In reality, however, economic data are seldom observations that can be regarded as repeated drawings from the same probability distribution [Lancaster (1990) and Neumann (1997)]. Rather, allowance must be made for measured, and possibly unmeasured, systematic differences between financial institutions. The main aim of introducing systematic differences into the modelling framework is not, however, simply an econometric procedure *per se* but, rather, to examine what economic factors have been important in the diffusion of ATMs to date.

Given this background there are three aims to this chapter. First, to consider the theoretical aspects of ATM adoption and to develop a theoretical model of adoption. Second, to consider how econometricians have incorporated regressors or covariates² into empirical duration models. Third, to empirically examine the economic determinants of inter-firm ATM diffusion in the UK financial sector within the duration modelling framework outlined in Chapter 5.

The rest of the chapter is set out as follows. Section 6.2 considers the theoretical and methodological aspects of modelling ATM adoption. Section 6.3 develops an inter-firm theoretical model of adoption and outlines its empirical representation. Section 6.4

¹ This chapter draws on work contained in Gourlay (1998c).

² These terms are used synonymously throughout this and preceding chapters.

discusses the nature of the data set. Section 6.5 discusses the estimation procedures and Section 6.6 presents the empirical results.

6.2 Theoretical and Methodological Considerations

A distinction was made in Chapter 2 between the methodology of the neo-classical approach to inter-firm diffusion and that of the evolutionary approach. The former emphasises that the (microeconomic) modelling of the adoption decision at the firm level is a pre-requisite to understanding the process of diffusion at the inter-firm level of aggregation. In contrast, evolutionary theories have stressed how ordered patterns of diffusion may emerge from apparently irrational behaviour by firms. These differences in approaches are reflected in the typology of models that have emerged in the literature. Briefly, the neo-classical approach has, in general, assumed a modelling framework in which firms (assumed either to be identical or heterogeneous) have perfect information and who adopt a single capital embodied innovation only when it is profitable to do so. Differences in optimal adoption times between firms then leads to a time-intensive diffusion process. Moreover, this formal choice-theoretic framework leads to structural models of diffusion and to testable hypotheses. The evolutionary approach, however, has highlighted situations where firms are characterised by bounded rationality and where technological variety co-exists. As Nelson and Winter (1982) and Silverberg (1991) point out, analytical solutions are often impossible to obtain in situations such as these. Consequently, the evolutionary approach has been confined to computer simulations to portray the diffusion process and, thus, empirical testing of real-life data sets is extremely problematic for this approach [Sarkar (1998)].

Given these methodological aspects and the aims of the thesis as outlined in Chapter 1, it was decided that the modelling of ATM diffusion should follow the neo-classical route. It was deemed necessary, however, that the modelling framework had to meet two criteria. Firstly, the approach had to lend itself to empirical implementation and, more succinctly, had to be capable of being implemented explicitly within a duration framework. Secondly, the approach had to be capable of incorporating the most recent

advances in the theoretical literature and the unique characteristics of ATM technology and diffusion as outlined in Chapter 4.

The first criteria was met by modelling the adoption of ATMs within a choice-theoretic framework in which profit maximising institutions have to choose an optimal time to adopt t^* (i.e. the profit maximising value of t), with t^* being measured from the time at which ATMs were first commercialised. The optimal date of adoption as thus defined also serves as the duration of non-adoption and can therefore be modelled empirically using the methodology employed in Chapter 5. Moreover, the theoretical model that is developed allows the direct derivation of an empirical model in which the dependent variable is the hazard function or the conditional probability of adoption. This point is elaborated on in more detail in Section 6.3.

The second criteria has two interrelated elements to it which are specified in order to address the current weaknesses in the empirical literature and the unique characteristics of ATM diffusion in the UK. Firstly, the empirical literature has arguably not been rigorous in its testing of the existence of price expectations (and expectations in general) and strategic elements vis-à-vis rank effects, despite recent theoretical contributions (see Chapter 2) which have emphasised their potential importance in influencing the diffusion path. Moreover, the work of Dixit and Pindyck (1994) has emphasised that conventional theories of investment decision making which focus exclusively on the net present value (NPV) criteria ignore the opportunity cost of waiting and may lead to incorrect investment decisions. As noted by Farzin *et al* (1998) there is a direct analogy between investment decisions and decisions pertaining to technology adoption. These arguments may imply that previous empirical studies of ATM diffusion [see, for example, Hannan and McDowell (1984b, 1987)] that assume adoption takes place only at that time at which the NPV is positive may be mis-specified. Thus, it was deemed important that these new developments in the literature should be incorporated into the modelling of ATM diffusion.

Secondly, as discussed in Chapter 4 the diffusion of ATMs in the UK has a number of unique elements, which are both specific to ATM technology and the UK experience itself. Firstly, the evolution of ATM design has gone through two distinct generations

and their quality has improved over time (i.e. ATM technology is vintage-specific). As noted in Chapter 4, first generation machines were commercialised in 1967 and their adoption by institutions may have led to learning-by-using effects, which have influenced their optimal date to adopt second generation machines. Such effects have been found significant for other technologies [see, for example, Weiss (1994) and Colombo and Mosconi (1995)]. The potential existence of these effects should, therefore, be investigated for the case of ATMs. Secondly, a number of commentators [see, for example, Scarborough and Lannon (1988), Vesala (1994) and Llewellyn (1997)] have stressed the potential of ATMs as a key component of institutions' competitive strategy. This raises the potential that ATM adoption involves strategic or pre-emptive behaviour. As was pointed out in Chapter 4, this aspect of ATM technology was particularly pertinent in the UK after the Building Societies Act (1986) which increased competition for retail deposits. Thirdly, as stated by Matutes and Padilla (1994), Shepard and Saloner (1995) and Economides (1996), ATM technology may have inherent positive network externalities, which affect their adoption and subsequent diffusion. This aspect of the technology has been reflected in the establishment of reciprocal networks in the UK during the late 1980s.

As will be shown in Section 6.3, the first two of these aspects are addressed in the modelling of expectations and the stock and order effects respectively. It was found, however, that the network aspects of ATM diffusion was more appropriately addressed within a separate theoretical and empirical framework and in this respect network effects are addressed exclusively in Chapter 7.

To incorporate these many facets discussed above, a theoretical model of ATM adoption was developed which is able to encompass the main elements of the theoretical literature and which lends itself to be tested empirically using duration models. Thus, empirical testing of the model is simultaneously both a test of those economic factors deemed to be important for the diffusion of ATMs and, additionally, a test of the recent theoretical literature. The discussion now turns to the theoretical model.

6.3 A Theoretical Model of ATM Adoption

The starting point in the development of the theoretical model is to assume that ATM technology is embodied in a specific capital good, which is produced by a capital-producing industry and is then purchased by a capital-using industry (i.e. the financial sector). This conforms to the 'basic-type' concept of technical change put forward by Stoneman (1987) and is arguably an accurate representation of ATM diffusion in the UK given the narrative account contained in Chapter 4. In addition, ATM technology is conceptualised as being a generic technology. This assumption follows the conventional approach of the diffusion literature [Karshenas and Stoneman (1995)]. The definition of ATM technology therefore conforms to that employed in Chapter 4 and pertains strictly to second generation machines commercialised in 1972. Specific vintages of ATMs are therefore subsumed within the general concept of ATM technology. In so far that quality changes in the technology do occur these are incorporated within the exogenously determined quality-adjusted (or 'hedonic') price of the technology. This point is elaborated on in more detail in Section 6.4. Finally, it is assumed that there is no risk or uncertainty pertaining to the characteristics of ATM technology.

Throughout the development of the theoretical model and its empirical implementation, the nature of the market supplying ATM technology is not explicitly modelled in order to focus exclusively on those economic features of the financial sector that have been significant in the diffusion of ATMs.³ The model is therefore purely demand orientated. There is no loss in detail, however, in assuming that the capital-producing industry is perfectly competitive in terms of Stoneman and Ireland (1983) and Ireland and Stoneman (1986) which experiences (non-appropriable) industry-wide learning-by-doing effects analogous to Arrow (1962b) and Jovanovic and Lach (1989).

Given the conditions that the theoretical model had to satisfy set-out in Section 6.2, it was decided that the basis of the model should follow those specifications contained in Bresnahan and David (1986), Karshenas and Stoneman (1993, 1995) and Colombo and

Mosconi (1995). The approach employed in these papers has the advantage of being capable of subsuming recent advances in the theoretical diffusion literature into one encompassing decision-theoretic model. The formulation of this 'encompassing model' then enables the empirical testing of which, if any, of the rank, stock, order, epidemic and/or strategic effects play a significant role in the diffusion of ATMs without imposing a high degree of *a priori* parameter restrictions [Colombo and Mosconi (1995)].

The basis of the theoretical model is that financial institutions are assumed to be profit maximisers⁴ and behave in accordance with perfect foresight expectations. It is assumed that an ATM can be adopted by financial institution i in industry j by purchasing one unit of the technology at price P_t at time t . Furthermore, $g_{i\tau}$ is defined as the (gross) profit obtained by an institution in period τ from the use of the ATM. The variable $g_{i\tau}$ can be further defined using the concept of quasi-rents provided by Stigler (1987). Defining π as the quasi-rents per period on the new technology and π^* as the quasi-rents on the old technology, then if the new technology is replacing an old one then $g_{ij} = \pi - \pi^*$.

It is additionally assumed that the per period profits are determined by the rank, stock and order effects as outlined in Chapter 2. Specifically, the arguments of $g_{i\tau}$ are assumed to be a vector of characteristics of the firm C_i (reflecting the rank effects), the number of firms already using the new technology at time t , K_{jt} , reflecting the stock effects and the number of previous adopters at the *date of adoption*, S_{it} , reflecting the order effects. Although, clearly $S_{jt} = K_{jt}$ by definition, both terms are initially entered as separate arguments into the determinants of g_{ij} to keep the stock and order effects

³ Although this point may be objected to given the conjectures of demand and supply models outlined in Chapter 2, there is a paucity of relevant supply-side data available in order to extend this present modelling framework.

⁴ For the case of building societies this assumption may be objected to on the grounds that their objectives before the Building Societies Act (1986) was to maximise growth [McKillop and Ferguson (1993)]. As noted by Drake (1989), however, the only source of capital for societies pre-1986 was their profit (or 'surplus'). If growth is set as an objective of the building society then an operating surplus is required in order to generate the reserves necessary to act as capital backing for an expanded balance sheet. Given these arguments the profit maximising assumption is arguably not so unrealistic.

conceptually separate. In the empirical results presented in Section 6.5 it is assumed that all institutions belong to the same industry so that by definition $j = 1$.⁵

As discussed in Chapter 2, the distinguishing feature of the stock and order effect models is that the return to adoption in each period of use of the new technology is decreasing in the number of other adopters at that date. This reflects the so-called 'rent-grabbing' effect or first-mover advantages inherent in strategic adoption behaviour [Fudenberg and Tirole (1985), Quirmbach (1986) and Tirole (1988)] and/or there being limited supplies of other factor inputs [Ireland and Stoneman (1985)]. In contrast, rank effect models assume that the returns to adoption in each period are independent of other adopters. In this class of models, returns differ between firms because firms are heterogeneous in some important characteristic that determines adoption returns. These different effects can be incorporated in a specific functional form representing the returns to adopting ATMs. For the i th institution adopting an ATM in time t , its per period (or annual) benefits in time τ from adoption at time t , $\tau \geq t$, is specified as follows:

$$g_{i\tau} = g[C_i, S_t, K_\tau] \quad \tau \geq t, \quad g_2 \leq 0, \quad g_3 \leq 0 \quad (6.1)$$

where g_2 and g_3 represent the partial derivative of $g_{i\tau}$ with respect to the second and third terms of the right-hand side of (6.1) respectively. The sign of g_1 depends on the specification of the institution-specific characteristics contained in the vector C_i (institution size, for example).

Colombo and Mosconi (1995) have further distinguished between those elements of C_i that are time-invariant and those that are time-varying. Consequently, they re-formulate the general expression in (6.1) as:

$$g_{i\tau} = g[C_i, D_{it}, S_t, K_\tau] \quad \tau \geq t, \quad g_2 \leq 0, \quad g_3 \leq 0 \quad (6.2)$$

where C_i captures time-invariant factors and D_{it} captures time-varying ones. They further distinguish between three other elements within D_{it} . Firstly, cumulative

⁵ Consequently, the subscript ' j ' used in previous terms is now ignored in the following discussion for the

learning-by-using effects which reflect the stock of knowledge, capabilities and technical and managerial skills that institution i has developed through the use of previous vintages of the technology. This applies, for example, to those institutions adopting first generation machines preceding the commercialisation of second generation machines in 1972. These learning effects are assumed to have a positive impact on g_{it} . Secondly, a term capturing the adoption of complementary technology and, thirdly, any remaining institution-specific characteristics. The former is assumed to have a positive impact on g_{it} , while the impact of the latter is determined empirically.

From this simple framework the present value of the increase in gross profits arising from adopting at time t , G_{it} , with discount factor r , in continuous time is given by [Lambert (1990) and Ostaszewski (1993)]:

$$G_{it} = \int_t^{\infty} g[C_t, S_t, K_t] \exp[-r(\tau - t)] d\tau \quad (6.3)$$

The institution is then assumed to choose an optimal time to adopt (taken from the commercialisation of the ATM for which $t=0$), t^* , which is determined by two necessary and sufficient conditions. The first is that adoption yields positive profits. The second condition is known as the *arbitrage condition* [Karshenas and Stoneman (1993, 1995)]. Defining Z_{it} as the net present value of adopting in time t , then for adoption to be profitable it is necessary that:

$$Z_{it} = -P_t + G_{it} \geq 0 \quad (6.4)$$

where P_t is the cost of acquiring one unit of the ATM technology at time t . This conforms to the conventional NPV rule of investment decisions [see, for example, Primrose (1991)]. For it not to be more profitable for the institution to wait before adopting, it is additionally necessary that:

$$y_{it} \equiv d[Z_{it} \cdot \exp(-rt)]/dt \leq 0 \quad (6.5)$$

purpose of clarity.

where y_{it} is discounted to ensure a common time basis for evaluation [Karshenas and Stoneman (1993)]. Although the expression contained in (6.5) is in the spirit of Dixit and Pindyck (1994) there is no suggestion that it follows their distinctive options value approach to investment decision making. Rather, it aims to address their contention that the basic NPV rule ignores the possibility that firms can wait for additional information. The main contribution of including the arbitrage condition is, then, that it allows institutions to delay adoption at time t in anticipation of higher profitability from adoption even though it is profitable to adopt at time t . Such a scenario is ignored in previous models of ATM adoption [see, for example, Hannan and McDowell (1984, 1987)].

These two conditions are pivotal for the subsequent process of technological diffusion in this model and determine two distinct aspects of the process. The profitability condition in (6.2), assuming institutions are profit maximisers, determines the set of potential adopters. In contrast, the arbitrage condition contained in (6.5) determines the optimal adoption times, t^* , for each adopter [Karshenas and Stoneman (1995)]. The optimal adoption date for institution i , t_i^* , is given by the following condition:

$$y_{it^*} \leq 0 \tag{6.6}$$

That is, it is not profitable for institution i to wait a short time interval. The inequality sign in (6.6) allows for the possibility of corner solutions (this could occur, for example, when it is optimal to adopt the technology in the first year of its commercialisation). Karshenas and Stoneman (1993) show that if Z_{it} is bounded from above and that if each member of the population is a potential adopter, then there will exist an optimum time to adopt, $t_i^* < \infty$, where net benefits are maximised. This additionally applies that for potential adopters the arbitrage condition in (6.5) dominates the profitability condition in (6.4).

Again, assuming that institutions are profit maximisers and conjecture their own actions do not affect the actions of other institutions, then Karshenas and Stoneman (1993,

1995) have shown that substituting (6.3) into (6.4) and differentiating $e^{-rt}Z_{it}$ with respect to t yields the following expression for y_{it} :

$$y_{it} = rP_t - p_t - g[C_i, S_t, K_t] + \int_t^{\infty} g_2[C_i, S_t, K_t] s_t \exp[-r(\tau - t)] d\tau \quad (6.7)$$

where lower case letters represent derivatives with respect to time and s_t and $p(t)$ represent *expected* changes in the number of users and the price of technology in the small time interval $(t, t + dt)$ respectively. Equation (6.7) states that the benefits from waiting for a time interval before adopting equals the interest saved, rP_t , plus any expected reduction in the cost of adoption, $-p_t$, minus the benefits forgone from not having the new technology for the time interval, plus the net present value of the changes in benefits resulting from a move down the order of adoption (the integral term) for all $\tau \geq t$ respectively.

The model outlined above makes the following predictions concerning the diffusion of new technology. Firstly, y_{it} will be a positive function of rP_t , S_t and K_t . This implies that the optimal time to adopt for institution i (after the commercialisation of ATMs) will increase with the interest saved from waiting, rP_t , the number of previous adopters and the number of adopters in time t . The last two effects are consistent with the predictions made by the stock and order effects models outlined in Chapter 2. Secondly, y_{it} will be a negative function of the expected change in the cost of acquisition p_t and, given that $g_2 \leq 0$, negatively related to the expected reduction in the number of users, s_t . This follows from the arbitrage condition in (6.5) and from the stock and order effects respectively. This result is consistent with Rosenberg (1976a, 1982) and Balcer and Lipmann (1984) who argue that expected reductions in the cost of acquisition price of technology will delay new technology adoption. Thus, faster expected increases in the number of users will lead to earlier adoption. The more adopters there are at time t the fewer adopters there will be in time t , *ceteris paribus*, (via the order effect). The effect of C_i on the optimal time to adoption depends on the elements contained in this vector.

In the derivation of (6.7) it is assumed that all conjectures as to the impact of the institution's adoption decision on the adoption decisions of other institutions are included implicitly rather than explicitly, being incorporated in the expectations term s_t . Given that institutions are assumed to behave in accordance with perfect foresight this aspect of the model may be objected to because it does not fully capture the more detailed features of the rank and stock effects models. This is arguably the case for the strategic model presented in Fudenberg and Tirole (1985) for which the type of equilibrium is sensitive to the nature of the technology and the quantity (output) setting behaviour of firms. Due to the highly stylised nature of game-theoretic models and the difficulty of identifying important decision variables, however, the extent to which strategic behaviour can be implemented empirically is, arguably, highly questionable. It is reasonable, therefore, to approximate strategic effects by the main claim of the game-theoretic approach: that is, the benefits to the marginal adopter are decreasing in the number of adopters.

The distinguishing feature of the theoretical model is that if institutions act in a myopic manner - which is the approach taken in previous models of technological adoption such as Hannan and McDowell (1984b, 1987) and Rose and Joskow (1990) - then $p_t = s_t = 0$ and $S_t = K_t$ for all $\tau > t$. This implies that the arbitrage condition in (6.6) yields an optimal adoption time, t_i^* , for institution i that would coincide with that implied by $Z_{it} = 0$ in (6.4). Thus, under myopic expectations the arbitrage and profitability conditions are identical [Karshenas and Stoneman (1995)]. It can thus be observed that myopic models necessarily ignore the arbitrage condition in technological adoption.

In the original formulation of the model by Karshenas and Stoneman (1993), the expression in (6.7) is simplified by assuming that the marginal benefit changes resulting from moving down the order of adoption at time t are independent of the level of future stock of adopters K_t for $\tau > t$. This is obtained by assuming that if the benefit function is of the form $g[C_i, S_t, K_t] = g^1[C_i, S_t] + g^2[C_i, K_t]$ then (6.7) can subsequently be rewritten in a more simplistic form as:

$$y_{it} = rP_t - p_t + g_2[C_i, S_t, K_t] s_t/r - g[C_i, S_t, K_t] \quad (6.8)$$

6.3.1 An Empirical Representation of the Theoretical Model

The expressions contained in (6.6) and (6.7) yield the optimal date to adopt, t_i^* , for institution i under the assumption that institutions form their expectations in a perfect foresight manner. As stated in Section 6.1, however, the main aim of this chapter is to empirically test the theoretical model within the explicit framework of duration models. A pre-requisite to achieve this is to generate a specification of the hazard function or the conditional probability of adoption. As shown by Mosconi and Colombo (1995) this can be achieved by introducing a stochastic element into the model by assuming that not all institutional-specific variables are known with certainty. This implies that the arbitrage condition in (6.6) that defines the optimal date of adoption is re-specified as the following:

$$y_{it} + \varepsilon \leq 0 \quad (6.9)$$

where ε is a stochastic error term whose distribution remains invariant across firms and overtime. If ε is further assumed to be distributed independent of y_i with a symmetric distribution function (ε) and that the hazard function is defined as the probability of adoption in the small time interval $(t, t + dt)$ for an institution that has not adopted ATMs by time t it follows that the hazard function is specified as [Colombo and Mosconi (1995)]:

$$h_i(t) = \Pr[y_{it} + \varepsilon \leq 0] = V[-y_{it}] \quad (6.10)$$

where $h_i(t)$ is the hazard function for institution i and $V(\cdot)$ is a decreasing function in y .

By substituting (6.8) into (6.10) and removing the artificial distinction between the stock and order effects by further assuming that $S_i = K_i$, then (6.10) may be re-written as:⁶

⁶ Note that the discount factor, r , in (6.11) has been specified as time-varying, whilst in the derivation of the theoretical model in (6.7) it was assumed to be time-invariant. As noted by Karshenas and Stoneman (1995) the basic results hold for the theoretical model if r is assumed to be time-varying, except that the

$$h_i(t) = J[r_t P_t, K_t, C_t, p_t, k_t/r_t] \quad (6.11)$$

where, from the discussion of the theoretical literature in Chapter 2 and the results obtained from the derivation of the theoretical model in Section 6.3, the following *a priori* parameter restrictions apply: $J_1 < 0$, $J_2 < 0$, $J_3 \geq 0$, $J_4 > 0$ and $J_5 > 0$.

To make the model in (6.11) fully operational at the empirical level requires two additional steps. Firstly, the elements of the rank effects contained in the vector C_t need to be specified and, secondly, the functional form of $J(\cdot)$ needs to be specified. The second of these is addressed first whilst the specification of the rank effects is left until Section 6.4.1.

The empirical implementation of the theoretical model cannot proceed until an exact functional form of $J(\cdot)$ in equation (6.11) is specified. As noted by Gourlay (1998c), this aspect of the modelling principally concerns the nature of the relationship between the duration of non-adoption, t , and the variables or, more precisely, the covariates contained in the bracketed term on the right-hand side of expression (6.11). Thus, the specification of $J(\cdot)$ cannot be separated from the wider issue of introducing systematic differences between institutions into the hazard function. There are two main approaches in the econometrics literature that has formalised the relationship between t and a vector of covariates: the non-parametric approach of Cox (1972, 1975) and the accelerated lifetime model of Kalbfleisch and Prentice (1980) and Neumann (1997). The differences between these two approaches are reflected in the specification of the relationship between duration t and the covariates.

In both the proportional hazards model and the accelerated lifetime model it is assumed that there exists a $K \times 1$ vector, X , which contains K covariates and that there is a cross-section of duration times for n institutions, t_1, \dots, t_n . The vector of covariates can contain both time-invariant or time-varying covariates. In the case of the latter, a further distinction is made in the literature between *exogenous* and *endogenous* covariates. For a covariate that is exogenous the information that an institution has not

first two terms in (6.8) are replaced by a user-cost of capital formula. In the empirical results presented in Section 6.5, r is assumed to be time-varying.

adopted in the small time interval $(t, t + dt)$ does not aid in the prediction of the path of the covariate process from $(t, t + dt)$ given its history to t . A covariate that is not exogenous is, by definition, endogenous [Lancaster (1990)]. A more formal definition of these two types of covariates is presented in Appendix One. The following specifications of these two models are made under the assumption that X contains only exogenous variables.

The non-parametric proportional hazards model of Cox (1972) assumes that the interaction between t and the covariates to be a multiplicative one. The continuous time model specifies that:

$$h_i(t|X) = h_0(t)\exp(X_i\beta) \quad (6.12)$$

where β is a $\times K$ vector of covariate parameters and $h_0(t)$ is labelled the 'baseline' hazard function [Neumann (1997)] and conforms to the condition $X = 0$. The covariates are embodied in the link function $\exp(X\beta)$.⁷

As Cox (1972, 1975) indicates, the advantage of the proportional hazards model is that the parameter vector β can be estimated *without* specifying the form or family for the baseline hazard $h_0(t)$. This is achieved by using a partial likelihood method of estimation [Amemyia (1986)]. Its disadvantage is that it cannot, to date, incorporate time-varying covariates.

In the accelerated lifetime model the effect of covariates is to re-scale time directly or, equivalently, the role of X is to accelerate (or decelerate) the time to failure. In terms of the baseline hazard introduced in the preceding section this model is specified in continuous time as follows [Kalbfleisch and Prentice (1980) and Neumann (1997)]:

$$h_i(t|X) = h_0[\phi(X_i\beta)]\phi(X_i\beta) \quad (6.13)$$

⁷ As noted by Lawless (1982), nothing requires the link function to be specified as this. A pre-requisite is, however, that the function must be restricted to those that guarantee $\exp(X\beta) > 0 \forall X$.

with the conventional specification of $\phi(\cdot)$ taken to be $\exp(\cdot)$ as in the proportional hazards model. Estimation of the accelerated lifetime model then proceeds by choosing a functional form for h_0 (from the selection of parametric models estimated in Chapter 4 for example) and maximising the resulting log-likelihood function. Right-censored observations are accommodated into the log-likelihood function using the indicator variable introduced in Chapter 5.

The advantages of using the accelerated lifetime model is that the effect of covariates on the hazard function is, arguably, more intuitive than the proportional hazards model and its ability to incorporate time-varying covariates.

The more mathematical aspects of the proportional hazards and accelerated lifetime models and the procedures employed in their estimation are presented in Appendix One. In the empirical results presented in Section 6.5 both approaches are employed to test the robustness of the results across different specifications of the $J(\cdot)$ function.

There is considerable debate in the literature as to what the baseline hazard, $h_0(\cdot)$ in (6.12) and (6.13) captures and this debate leads conveniently on to a discussion of how epidemic effects can be incorporated into the model. Recall from Chapter 2 that the epidemic model of diffusion predicts that the duration dependence of the diffusion path for the new technology will be non-monotonic: at first increasing and then decreasing. This non-monotonicity will be reflected in the values of the parameter p in the baseline hazard for the lognormal and log-logistic distributions (see Table A5.2.1 of Chapter 5). Thus, depending on the exact parametric form specified for the baseline hazard, subsequent tests of the duration dependence of $h_0(\cdot)$ will, thus, reveal if the diffusion process contains or does not contain epidemic effects. Karshenas and Stoneman (1993, 1995) have argued strongly in support of this viewpoint. They argue that epidemic effects, which they label as 'endogenous learning', cannot separately be introduced in the vector of covariates X in equations (6.12) and (6.13).

In contrast, Colombo and Mosconi (1995) have argued that $h_0(\cdot)$ may pick-up movements in the price and quality of technology over time (in a myopic model) not adequately accounted for in the model. In addition, they argue that if individual spells

are measured simply as calendar time⁸ then finding support for the existence of epidemic effects is more likely. Instead, they contend that measures of individual firm spells, the t_i 's, should take into account the date of entry of firms into the industry and that for those firms that enter *after* the date of commercialisation the two measures will not coincide.

An alternative view is that taken by Heckman and Singer (1984, 1985) and Neumann (1997). They argue that the baseline hazard captures unobserved heterogeneity (i.e., not all factors affecting optimal adoption times are accounted for in vector X) and that this will produce downward bias in duration dependence. Indeed, Karshenas and Stoneman (1993) use this fact to argue that epidemic effects are less likely to be found in duration models vis-à-vis early contributions to inter-firm diffusion modelling. As noted by Heckman and Singer (1984), however, identification of the effects of time-varying covariates on the conditional probability of adoption may be difficult to separate from duration dependence if there is not 'sufficient' cross-individual variation in the covariates in X. This problem arises from the classic multicollinearity problem [see Greene (1993) for example].

In the empirical results presented in Section 6.5, the baseline hazard is interpreted as reflecting epidemic effects for two reasons. Firstly, because the distinction between calendar and duration time made by Colombo and Mosconi (1995) does not apply to the set of potential ATM adopters employed in this study (i.e. no new potential adopters enter after 1972). Secondly, because diagnostic tests reject the hypothesis of heterogeneity in specified models.

⁸ They define calendar time for an individual firm as the maximum time of either the time from commercialisation of the technology or the firm's entrance into the industry. For those firms that have entered after the date of commercialisation these two measures will not be equal.

ATM technology is that of a second generation one. Again, this is also consistent with the definition utilised in the empirical contribution in Chapter 5. Descriptive statistics for all the covariates employed in this chapter can be found in Appendix Two. In addition, Appendix Three outlines the sources used in compiling the data.

All institution-specific and market-specific covariates are measured from the time of commercialisation of second generation ATM technology in 1972 until that time at which the institution adopted or until the end of the empirical study in 1992 for non-adopting or right-censored institutions. A list of all the covariates employed in this chapter together with a summary of their meanings are presented in Table 6.1. The descriptive statistics for these covariates and the sources used in their collection are provided in Appendices Two and Three respectively.

Since it is one of the main aims of this chapter to study the potential effect of price expectations on the diffusion of ATMs, it is important to discuss the construction and measurement of the price variable P_t . The series provided by the Office for National Statistics (ONS) was a quality adjusted one or 'hedonic price' [Deaton and Muellbauer (1980)] to reflect the evolving nature of the technology. The methodology employed in its construction followed a similar one as outlined in Lancaster's (1966) approach to consumer theory. This methodology involves viewing ATM technology as a 'bundle' of characteristics (such as services provided and speed of withdrawal and so on) and valuing these in order to obtain an estimate of the ATMs 'quality'. By dividing an individual machine's price by its quality an estimate of the quality-adjusted price is obtained. Then by averaging over all machines installed in a given year a quality-adjusted time-series price is obtained [see Stoneman (1976, 1983) for details].⁹

The construction of the profitability measure *PROFIT* and the growth in deposits covariate *GROWTH* raised two problematic issues. Firstly, the relevant decision-making unit has to be defined and, secondly, a consistent series for these covariates has to be generated over the entire sample period for all institutions.

entered after the date of commercialisation these two measures will not be equal.

⁹ Unfortunately, the exact characteristics used in the construction of this price index were not forthcoming from the ONS so that any criticism surrounding the characteristics chosen in its construction is not

The first issue centres on the relevant dimension for which the covariates should be measured. In the case of the sample of clearing banks, for example, their activities are extremely diverse, encompassing such activities as personal banking, business banking, cross-border services and international banking [see BBA (1996) for more details]. In contrast, the activities of building societies in the UK are focused on the retail and mortgage markets [BSA (1997)]. These aspects of institution-specific activity are reflected in the construction of individual report and accounts data which, in general, distinguish between overall group performance and that of its constituent activities. Moreover, this aspect raises the issue of what decision-making unit is relevant for the adoption of ATMs and, thus, for which constituent activity of the institution should be used in the construction of the *PROFIT* and *GROWTH* covariates. It may be argued, for example, that these covariates should be measured exclusively for the retail activities of the institution as these activities are more closely related to the services provided by ATMs vis-à-vis other activities. In the collection of the data, however, the unavailability of disaggregated data for the entire sample period forced the *PROFIT* covariate to be measured at the group level of aggregation. Disaggregation was, however, possible for the *GROWTH* variable. This was measured as the growth in customer deposits for clearing banks and the growth in deposit and share deposits for building societies [see Pawley *et al* (1991) for more details]. Consequently, the *GROWTH* covariate more accurately captures the retail-side of the institution's activities than the *PROFIT* covariate.

The second issue centres on changes in accounting regimes and standards during the sample period and the consequences for generating a consistent series for *PROFIT* and *GROWTH*. For clearing banks, a consistent series for both covariates was obtained from the British Bankers' Association (BBA). Similarly, a consistent series for building societies was provided by the Building Societies Association (BSA). For both building societies and clearing banks, the *PROFIT* series is defined as post-tax profit on ordinary activities and is measured on a historical cost basis. A post-tax measure of profitability was chosen in order to more accurately capture more accurately the potential effects of liquidity constraints.

possible. Given the paucity of price data from trade sources, however, it was decided that the advantages of employing this covariate outweighed the disadvantages of not using it.

For both covariates, the data pertains to end-of-year figures. This is also the case for the covariates *SIZE*, *STAFF* and *BRANCH*. This reflects the source of the data as annual report and accounts.

6.4.1 Institution-Specific Characteristics and Rank Effects

The covariates included in the rank effect have been subsumed in the vector C_i and, up to now, the exact elements of this vector have been left unspecified except for a brief discussion of the construction of the *PROFIT* and *GROWTH* covariates in the previous section. Although there are clearly many institution-specific variables that will jointly determine the optimal time to adopt those outlined below are identified as being potentially important from the discussion of ATM technology in Chapter 4. Those covariates that are time-varying have subscript t and those that are time-invariant do not.

*Size of the Institution (SIZE_{*t*})*

As stated in Chapter 2, firm size has frequently been the defining characteristic, which determines the profitability of adoption in theoretical rank models. For the specific case of ATM technology the importance of institution size can be justified on two grounds. Firstly, US evidence [Walker (1978) and Humphrey (1994)] suggests that there are considerable positive scale and scope economies in ATM technology which can only be obtained by relatively large deposit taking institutions. This suggests that adoption of ATMs is, *ceteris paribus*, likely to be more profitable for relatively larger institutions. Secondly, institution size can be taken as an indicator of the differences in relative risk from adoption faced by different-sized institutions [see, for example, the early contributions of Mansfield (1961, 1968)]. These positive effects on the conditional probability of adoption may, however, be lessened, or even reversed, by the existence of more flexible managerial attitudes in smaller financial institutions who are able to adapt their organisations to technical change [see, for example, the survey evidence in Ferguson and McKillop (1993)].

Growth in Deposits (GY_t)

As noted by Elyasiani and Mehdiian (1990), Field (1990) and Coldwell and Davis (1992), the deposits of a financial institution play a dual role in their production process. Firstly, they provide liquidity and transactions services, and secondly, they provide an input (together with capital and labour) in the production of other assets. During periods of rapid (retail) deposit growth, therefore, financial institutions may wish to expand capacity by adopting ATM technology. Indeed, as shown by Karshenas and Stoneman (1993) the adoption rule employed in adopting new technology purely for increasing capacity is less stringent than that employed for replacement reasons alone (if expectations are myopic). Moreover, financial constraints may be lessened during periods of relatively high output growth [see Kamien and Schwartz (1982) and Goodacre and Tonks (1995)]. *A priori*, the growth in deposits will have a positive effect on the conditional probability of adoption.

Staff to Branch Ratio ($STAFF_t/BRANCH_t$)

As stated in Chapter 4, a number of commentators have argued that the main motivation underlying the adoption of ATMs has been the technology's inherent 'labour-saving' qualities and subsequent reduction in the average cost of producing deposit services. To measure the potential for labour-savings is extremely problematic and, ideally, a measure of the capital-labour ratio defined over those capital and labour inputs employed directly in the provision of deposit services would be a potential candidate for inclusion [de Wit (1990)]. In the absence of such a measure, the ratio of the number of branch staff employed to the number of branches operated is employed as a measure of the potential for labour-saving. *A priori*, it is expected that those institutions with higher values of this ratio will have a positive effect on the conditional probability of adoption.

Profit to Size Ratio ($PROFITS_t/SIZE_t$)

As noted by Oakey *et al* (1988) and Seaton (1996), internally generated profits is the predominate source of investment finance in R&D for small, high technology industrial firms. This raises the question of whether similar financial constraints exist in the financial sector for the adoption of ATMs. Following the empirical work of Hannan and McDowell (1987), liquidity constraints are measured as the ratio of after-tax profits

to the value of assets (the latter being a measure of relative institution size). This is expected to have a positive effect on the conditional probability of adoption.

Institutional Factors

The empirical evidence from previous diffusion studies outlined in Chapter 3 has found that the institutional arrangements of firms as well as the effect of direct government regulations can affect the diffusion path. Two institutional factors for the case of ATM diffusion in the UK financial sector are deemed to be important. These are:

*DTAKE_{*t*}*, - this is a dummy variable and takes the value of unity if the institution has taken over another institution during the sample period and zero otherwise. The effect of this variable is likely to be positive. Horizontal takeovers may raise market power for the institution initiating the takeover and may, thus, reflect more intensive innovative behaviour by firms [Tirole (1988) and Shy (1996)].

*DSUB_{*t*}* - this is also a dummy variable and takes a value of unity if the institution is part of a larger corporate unit and zero if it is an independent unit. As noted by Karshenas and Stoneman (1993), the effect of this type of variable is likely to be ambiguous. On the one hand, independent institutions may be better positioned with regard to speed of implementation once the decision to adopt has been made. On the other, institutions that are part of a larger institution may be better informed and bare less financial risk in adopting new technology.

Learning-By-Doing Effects (DPREVIOUS)

As noted by Church and Gandal (1993) and Colombo and Mosconi (1995), the earlier adoption of previous vintage technologies may increase the marginal benefits from the adoption of later ones. This may reflect significant learning-by-doing effects. This covariate takes a value of unity if the institution has adopted first generation machines and zero otherwise. *A priori*, this covariate is expected to have a positive effect on the conditional probability of adoption.

6.5 Econometric Specification and Estimation

An identical framework employed in Chapter 5 is used in this chapter to represent the duration times for the set of potential adopters. It is assumed that T is a continuous non-negative random variable and represents the duration of an individual institution in a state of non-adoption. The data set of ATM adoption histories then consists of a cross section of duration times (i.e. realisations of T), t_1, \dots, t_n , where t_i is the duration of the i th institution ($i=1, \dots, 98$) and which can be ordered as $t_1 < \dots < t_n$. The range of t is $[0, 21]$. This follows from the time elapsed from start of the study in 1972 ($t = 0$) to the end of the study in 1992 ($t = 21$).

In contrast to Chapter 5, however, it is further assumed that the hazard function is determined by a $K \times 1$ vector of covariates which contains both time-invariant and time-varying covariates. This reflects the inherent time-invariant nature of some of the covariates, such as *DPREVIOUS*, and the time-varying nature of others, such as *SIZE*, for example. The value of these covariates at time t is denoted by vector $x(t)$ and its entire path from entry into the study ($t = 0$) to the time of adoption is denoted by $X(t)$. As noted in Appendix One, the hazard function in the presence of time-varying covariates will be conditioned on $X(t)$.

To analyse the effect of covariates on the conditional probability of adoption the accelerated lifetime model contained in expression (6.13) was initially estimated. The model was estimated in its continuous time version rather than its discrete version for two reasons. Firstly, as argued by Heckman and Singer (1984) there is no underlying reason why discrete time periods will be perfectly synchronised across individual institutions. Secondly, continuous time models have the advantage of being invariant to the time unit employed and this implies that a common set of parameters can be used to generate probabilities of events occurring in intervals of different length [Heckman and Singer (1984)]. Thus, direct comparison with the results of other diffusion studies employing duration models is then possible with this approach.

As stated by Neumann (1997), however, an extension of the basic accelerated lifetime model is required in the presence of time-varying covariates. To address this aspect of the modelling, the approach of Petersen (1986a, 1986b) is followed which allows the incorporation of time-varying covariates into the accelerated lifetime model. Briefly, this approach involves splitting the data interval of duration 0 to t_i into k exhaustive, non-overlapping intervals $t_0 < t_1 \dots < t_{k-1} < t_k$, where $t_0 = 0$ and $t_k = t_i$. The covariates are then assumed to stay constant within each of the k intervals, but may change from one interval to the next. The resulting log-likelihood is given by the following expression [Kiefer (1988)]:

$$\ln L(\beta) = \delta_i \ln f[t_i, X(t_i), \beta] - \sum_{j=1}^k \int_{t_{j-1}}^{t_j} h[s, X(t_{j-1}), \beta] ds \quad (6.14)$$

where β is a $1 \times K$ vector of covariate parameters to be estimated, δ_i is the indicator variable introduced in Chapter 5 which takes the value of unity if the duration observation is uncensored and zero otherwise and the second-term on the right-hand-side is the survivor function (for survival beyond duration t_k). The more mathematical and technical aspects of this approach together with the estimation procedure are outlined in Appendix One.

The above approach assumes that all covariates are exogenous (see Appendix One). By definition this is the case for time-invariant covariates [Lancaster (1990)]. For the set of the time-varying covariates, however, this assumption is likely to be violated by the inclusion of the price variables $r_t P_t$ and p_t and the stock of adopters at time t , K_t . The possible endogeneity of the price variables derives from the determination of $r_t P_t$ by the interaction between demand (i.e. the financial institutions) and supply (i.e. the ATM manufactures). This possible endogeneity is, however, mitigated by the fact that a large proportion of ATM technology was imported during the early years of diffusion in the UK [Kirkman (1987) and Austin (1992)]. The endogeneity of the stock covariate is more serious. As illustrated by Lee (1981), Murphy and Topel (1985) and Karshenas and Stoneman (1993), however, this issue can be tackled by employing a two-stage estimation procedure. In the first-stage a first-period time series autoregressive model is

fitted for K_t . In the second stage, estimates of parameters from the first stage are employed to obtain estimates of K_t and its first one-period ahead expectation, k_t . The estimation results from the first-order autoregressive model for K_t are given in Appendix Four.

Two specifications of the baseline hazard, $h_0(t)$, are employed in the empirical estimation of (6.13). These are the lognormal and log-logistic parametric models (see Table A5.2.1, Chapter 5 for exact specifications). This is based on the empirical results obtained in Chapter 5 that indicated these two parametric models gave the most adequate representation of the underlying distribution of adoption times.

The list of all covariates contained in the vector $X(t)$ is contained in Table 6.1 and their expected signs based on the *a priori* results of the theoretical model outlined in Section 6.1 and the arguments contained in Section 6.3.1 are summarised in Table 6.2.

Table 6.1: Definitions of Explanatory Variables

Variable		Description
p	=	Parameter of parametric models that determines duration dependence
<i>CONSTANT</i>	=	Intercept term
$SIZE_t$	=	Size of the institution, measured by the value of the total assets at time t deflated by the Producer Price Index (PPI)
GY_t	=	Growth in institutions deposits measured at time t
r_t	=	Discount rate, measured by the yield on Treasury Bills expressed as annual interest rates
k_t	=	Expected change in the cumulative number of adopters in the interval $(t, t+1)$ measured by $(K_{t+1} - K_t)$
K_t	=	Cumulative number of adopters at time t
p_t	=	Expected change of the price of ATMs measured by $(P_{t+1} - P_t)$
P_t	=	Real price of ATMs (quality adjusted) at time t , deflated by PPI
$STAFF_t$	=	Total number of part-time and full-time branch staff at time t
$BRANCH_t$	=	Number of branches operated by each institution at time t
$PROFITS_t$	=	Profitability of an institution, measured as the after-tax profits at time t
$DPREVIOUS$	=	A dummy variable taking the value of unity for previous cash dispenser adoption (1967 to 1971) and zero otherwise
$DTAKE_t$	=	A dummy variable taking the value of unity if the institution has taken over another institution in the period 1972 to 1992 and zero otherwise
$DSUB$	=	A dummy variable taking the value of unity if the institution is a subsidiary and zero otherwise

Table 6.2: Expected Sign of Explanatory Variables in the Model

Coefficient	Variable	Expected Sign	Economic Rationale
p	$TIME$	++ and >1.00 in the log- logistic model	Epidemic effects
α_1	$CONSTANT$?	N/A
β_1	$SIZE_t$	++	Risk and positive scale effects
β_2	K_t	-	Stock effects
β_3	GY_t	+	Liquidity constraints and expanding capacity
β_4	k_t/r_t	+	Order effects
β_5	$r_t P_t$	+	Interest saved from delaying adoption
β_6	P_t	+	Arbitrage condition
β_7	$STAFF_t/BRANCH_t$	+	Labour-saving arguments
β_8	$PROFITS_t/SIZE_t$	+	Liquidity constraints
β_9	$DPREVIOUS$	++	Cumulative learning-by-doing effects
β_{10}	$DTAKE_t$	+	Reflects innovative stance
β_{11}	$DSUB$?	Flexibility aspects (negative effects) versus risk and informational advantages (positive effects)

Note: '+' = positive; '++' = highly positive; '0' = no effect; '-' = negative; '--' = highly negative; '?' = no *a priori* expectation; 'N/A' = not applicable

6.6 Estimation Results

This section presents the empirical results obtained from the estimation of (6.13) for the set of potential ATM adopters.¹⁰ In all estimated models the sign of the estimated coefficient indicates the direction of the effect of the covariate on the conditional probability of adoption. Alternatively, a positive coefficient can be interpreted as implying the covariate 'accelerates' the time to adoption, whilst a negative coefficient implies the covariate 'decelerates' the time to failure. The exception to this is the parameter p , which represents a parameter of the baseline hazard and, given the arguments in Section 6.3.1, captures epidemic effects. Unfortunately, the absolute value

of the coefficients cannot be interpreted as the marginal effect of the covariate. The number of observations are identical for each model estimated and are calculated as the sum of the total number of years pertaining to completed durations (507 in total) and those pertaining to censored (or 'uncompleted') durations (1449 in total).

The statistical significance of all estimated covariates, except for p , are tested using the standard two-sided t -test [see Johnston (1987)] with critical values of 1.960 and 2.576 for the 0.05 and 0.01 levels of significance respectively. The significance of the estimated value of p is tested using a one-sided test given the restrictions on its values that each parametric baseline hazard assumes. Consequently, the 0.05 and 0.01 levels of significance for the case of p are 1.645 and 2.326 respectively [Johnston (1987)]. The critical values employed for the likelihood ratio (LR) test are dependent on the number of linear restrictions imposed and these are given after $\chi^2_{.95}(m)$ in each table, where ' m ' is the number of restrictions.

Three basic models were estimated: a fully specified model, a restricted stock model and a myopic expectations model. The restrictions are imposed on the fully specified model and the nature of these restrictions are summarised in Table 6.3 below. The fully specified model, as its name suggests, does not impose any restrictions on the value of the parameters. It includes all the rank effects outlined in Section 6.3.1 and the stock, order and expectations effects contained in expression (6.11). The results obtained from estimating this model are presented in Table 6.4. The restricted stock model imposes the restriction $K_t = 0$ on equation (6.11) and the results from this model are presented in Table 6.5. Finally, the myopic model imposes the restriction $k_t/r_t = p_t = 0$ on the fully specified model and the results from this model are presented in Table 6.6. This final model pertains to the conventional models estimated by Hannan and McDowell (1984, 1987) and Rose and Joskow (1990) which explicitly ignore the arbitrage condition in (6.5). A test of these two linear restrictions using the LR test introduced in Chapter 5 will, therefore, be a test of the significance of strategic effects and expectations in the diffusion of ATMs respectively.

¹⁰ All results in this section were obtained from the econometric package *LIMDEP 6.0* [see Greene (1994)].

Model Name	Restriction(s) Imposed	Table of Results
Fully specified	None	6.4
Restricted stock	$K_t = 0$	6.5
Myopic	$k_t/r_t = p_t = 0$	6.6

The existence of epidemic effects are discussed first. As stated in Chapter 5, the existence of epidemic effects will be reflected by duration dependence that is non-monotonic (although the converse is not necessarily true): at first increasing and then decreasing. Non-monotonicity of this type additionally implies parameter restrictions for the estimate of p in the lognormal and log-logistic baseline hazards. These will be $p > 0$ for the lognormal baseline and by $p > 1$ in the log-logistic. From Table 6.4 the t -test for these restrictions is accepted at the 0.01 level of significance for both the log-logistic and lognormal models. Moreover, epidemic effects are also found to be significant in both the restricted stock model in Table 6.5 and the myopic model in Tale 6.6. In addition, all the models are re-estimated employing an exponential baseline hazard, which imposes the restriction $p = 1$ (i.e. there are no epidemic effects). The LR test rejects this restriction at the 0.05 level for all models. Thus, the exponential model is rejected by the data. These results confirm those obtained in Chapter 5 for the case of a homogenous population: that is, epidemic effects have been significant in the diffusion of ATMs. This indicates that non-market channels have been important in diffusing the technical and economic attributes of ATM technology. These could include, for example, the transfer of technical staff between institutions (as a means of transferring tacit information) and the transfer of relevant attributes through trade journals.

Before examining the role of rank and learning-by-using effects, attention will initially be focused on the stock and order effects. It can be seen from Table 6.4 that the coefficient on K_t (representing stock effects) though having the correct sign is found to be statistically insignificant. As noted by Karshenas and Stoneman (1993), this may not necessarily indicate the total absence of such effects and may alternatively be captured by the time-varying baseline hazard. If this is indeed the case then imposing the condition $K_t = 0$ should lead to an increase in the estimated coefficient for p . As

illustrated in Table 6.5 the reverse of this effect occurs for both the lognormal and log-logistic baseline hazards. Moreover, the LR test rejects this restriction at the 0.05 level for both the lognormal and log-logistic models. The failure to find a significant impact for K_i may indicate that ATM technology has had only a insignificant impact on institutions' costs and has, therefore, not been a 'drastic innovation' analogous to Arrow (1962a). Moreover, this finding is consistent with the finding in Chapter 5 that there has been no sample attrition from the adoption of ATMs.

Order effects are captured by the coefficient on k_i/r_i . As can be seen from Tables 6.4 and 6.5, the coefficient on this variable is statistically significant and has the correct sign *a priori*. This result indicates that the expected change in the number of adopters has been significant in the diffusion of ATMs. This may reflect the view of some institutions that the adoption of ATMs by rival institutions in the near future would lead them to loose market share for retail deposits.

Turning attention to the role of rank effects, it is immediately apparent from Table 6.5 that these perform an extremely significant role in the diffusion of ATMs. The likelihood ratio-test for the joint significance of all the rank effect covariates indicate that their existence cannot be rejected at the 0.01 level in all models and for all specifications of the baseline hazard. In particular, the size of the institution, captured by $SIZE_i$, has a positive and highly significant effect on the conditional probability of adoption. This is consistent with previous studies examining the diffusion of ATMs [see, for example, Hannan and McDowell (1984a, 1984b and 1987) and Sinha and Chandrashekar (1992)] and those examining diffusion in the industrial sector.

The $STAFF_i/BRANCH_i$ ratio although having the correct sign *a priori* is not found to be statistically significant in both the lognormal and log-logistic models except for the myopic lognormal model in Table 6.6. This result may reflect the fact that this variable is an inadequate proxy for the opportunity for labour saving associated with ATM technology. As noted in Chapter 4, the labour-saving potential for ATMs may have been overstated and those branch staff previously employed exclusively in the provision of demand services may have been re-deployed to other services after the adoption of ATMs. Thus, this variable may be capturing the output mix of a financial institution

rather than the potential for labour-saving. Alternatively, the total number of branch staff employed may be a poor proxy for the labour input into providing deposit services. As found by de Wit (1990) the effects of new technology on employment in the financial sector is highly sensitive to the definition of labour inputs. A more precise measure encompassing those labour inputs providing exclusively deposit services may produce a different result. Unfortunately such disaggregated data is unavailable.

The positive and significant coefficients on $PROFITS_t/SIZE_t$ and GY_t in all models are consistent with *a priori* expectations and suggests that liquidity constraints play an important role in the diffusion of ATMs. This result is consistent with US empirical evidence [see Hannan and McDowell (1984a, 1984b, 1987) for example].

A notable aspect of the results is the role played by the learning-by-using variable, $DPREVIOUS$. This is found to have a positive and highly significant coefficient at the 0.01 level in all models and specifications of the baseline hazard. This suggests that institutions with experience in using previous vintage ATM technology (which embodies similar characteristics to second generation technology) has a positive impact on the conditional probability of adoption. This supports the empirical evidence in Colombo and Mosconi (1995).

In addition, the coefficient on $DTAKE_t$ is positive and highly significant at the 0.05 level in all models except for the lognormal restricted stock model. This implies that institutions that actively seek to increase market share by takeovers have a higher conditional probability of adoption. This arguably reflects the aim of some institutions to enhance their market power through a dual strategy of adopting ATMs and takeovers in order to capture a higher market share of retail demand deposits. The coefficient on $DSUB$, although negative, is statistically insignificant. Thus, being a subsidiary institution does not have any significant impact on the conditional probability of adoption. This implies that the flexibility aspects of being a subsidiary cancel out the risk and informational advantages.

Another notable aspect of the results contained in Table 6.4 is the role of the technology-price variable, $r_t P_t$, and that of the price expectations variable p_t . From

Table 6.4, the coefficient on $r_t P_t$ is of the correct sign as suggested by the theoretical model but is not found to be statistically significant. This result is consistent across all models and specifications of the baseline hazard. In contrast, the coefficient on the price expectations variable, p_t , is found to be both of the correct sign and statistically significant at the 0.05 level across all models. The significance of the price-expectations variable lends support to the theoretical model outlined in Section 6.3 and additionally suggests that myopic-type models employed by Hannan and McDowell (1984, 1987) and Rose and Joskow (1990) may be seriously mis-specified.

To investigate this issue in greater detail the empirical model is specified imposing the condition $k_t/r_t = p_t = 0$. The results obtained from imposing this restriction are shown in Table 6.6 and can be interpreted as a myopic expectations model. The results in Table 6.6 show that whilst the coefficient estimates of other variables remain relatively stable and of the same sign, the myopic model is rejected using the likelihood-ratio test to test the parameter restriction implied by the myopic model. This result is consistent with that of Karshenas and Stoneman (1993) and similarly suggests that myopic models, by ignoring the possible role of price expectations, may be seriously mis-specified.

As can be observed from Tables 6.4, 6.5 and 6.6 the parameter vector β remains relatively stable in terms of values and signs when moving from a lognormal specification of the baseline hazard to that of a log-logistic one. This is re-assuring and reinforces the robustness of the results obtained.

Table 6.4: Maximum Likelihood Estimates of the Fully Specified Lognormal and Log-logistic Models

Coefficient	Variable	Lognormal Model		Log-logistic Model	
p	$TIME$	4.609	(4.609)**	4.426	(3.632)**
α_1	$CONSTANT$	3.628	(11.249)**	4.100	(5.911)**
β_1	$SIZE_t$	0.033	(3.678)*	0.032	(3.565)**
β_2	K_t	-0.012	(1.188)	-0.009	(0.800)
β_3	GY_t	1.314	(2.671)*	1.819	(1.972)*
β_4	k_t/r_t	0.968	(2.562)*	1.371	(2.567)*
β_5	$r_t P_t$	0.025	(1.175)	0.023	(1.221)
β_6	p_t	-0.011	(2.052)*	-0.008	(2.278)*
β_7	$STAFF_t/BRANCH_t$	0.012	(1.727)	0.020	(1.882)
β_8	$PROFITS_t/SIZE_t$	8.942	(2.030)*	7.321	(1.955)*
β_9	$DPREVIOUS$	1.112	(4.149)**	1.309	(3.426)**
β_{10}	$DTAKE_t$	0.380	(2.342)*	0.584	(1.976)*
β_{11}	$DSUB$	-0.588	(0.860)	-0.485	(0.842)
Median duration (years)		29.120	(5.970)*	32.910	(4.320)**
Log-likelihood		-166.28		-166.22	
Number of observations		1956		1956	
Number of individual institutions		98		98	
Likelihood ratio test for the existence of epidemic effects: ($p = 1$)		14.28 ($\chi^2_{.95}(1) = 3.84$)		13.68 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for the existence of rank effects: ($\beta_1 = \beta_3 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11}$)		48.72 ($\chi^2_{.95}(7) = 14.10$)		48.12 ($\chi^2_{.95}(7) = 14.10$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 0 or 1 for the lognormal and log-logistic models respectively).

Table 6.5: Maximum Likelihood Estimates of Restricted Stock Lognormal and Log-logistic Models

Coefficient	Variable	Lognormal Model		Log-logistic Model	
p	$TIME$	3.421	(4.013)**	4.312	(3.907)**
α_1	$CONSTANT$	4.270	(9.251)**	4.476	(8.709)**
β_1	$SIZE_t$	0.039	(2.233)*	0.039	(2.313)*
β_2	K_t	—		—	
β_3	GY_t	1.161	(2.369)*	1.485	(2.526)*
β_4	k_t/r_t	1.282	(2.338)*	1.417	(2.141)*
β_5	$r_t P_t$	0.034	(1.147)	0.035	(0.959)
β_6	p_t	-0.004	(2.143)*	-0.007	(2.226)*
β_7	$STAFF_t / BRANCH_t$	0.014	(1.920)	0.019	(1.965)*
β_8	$PROFITS_t / SIZE_t$	12.865	(2.176)*	13.572	(2.086)*
β_9	$DPREVIOUS$	0.600	(1.940)*	0.711	(1.908)*
β_{10}	$DTAKE_t$	0.383	(1.705)	0.543	(1.959)*
β_{11}	$DSUB$	-0.506	(1.101)	-0.041	(1.425)
Median duration (years)		28.820	(4.250)**	32.440	(4.170)**
Log-likelihood		-167.62		-167.16	
Number of observations		1956		1956	
Number of individual institutions		98		98	
Likelihood ratio test for the existence of epidemic effects: ($p = 1$)		16.16 ($\chi^2_{.95}(1) = 3.84$)		15.08 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for the existence of stock effects: ($\beta_2 = 0$)		2.68 ($\chi^2_{.95}(1) = 3.84$)		1.88 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for the existence of rank effects: ($\beta_1 = \beta_3 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11}$)		60.14 ($\chi^2_{.95}(7) = 14.10$)		54.46 ($\chi^2_{.95}(7) = 14.10$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 0 or 1 for the lognormal and log-logistic models respectively).

Table 6.6: Maximum Likelihood Estimates of Restricted Myopic Lognormal and Log-logistic Models

Coefficient	Variable	Lognormal Model		Log-logistic Model	
p	$TIME$	4.648	(5.141)**	4.856	(4.427)**
α_1	$CONSTANT$	3.187	(12.420)**	3.271	(10.610)**
β_1	$SIZE_t$	0.004	(2.313)*	0.004	(2.313)*
β_2	K_t	-0.013	(1.642)	-0.011	(1.736)
β_3	GY_t	1.035	(2.487)*	1.241	(2.071)*
β_4	k_t/r_t	—		—	
β_5	$r_t P_t$	0.012	(0.775)	0.006	(0.829)
β_6	p_t	—		—	
β_7	$STAFF_t/BRANCH_t$	0.011	(1.962)*	0.016	(1.697)
β_8	$PROFITS_t/SIZE_t$	13.986	(2.392)*	13.410	(2.027)*
β_9	$DPREVIOUS$	0.914	(4.704)**	1.035	(4.041)**
β_{10}	$DTAKE_t$	0.328	(2.492)*	0.441	(2.241)*
β_{11}	$DSUB$	-0.209	(0.866)	-0.420	(0.884)
Median duration (years)		28.820	(10.960)**	24.360	(9.860)**
Log-likelihood		-173.56		-175.58	
Number of observations		1956		1956	
Number of individual institutions		98		98	
Likelihood ratio test for the existence of epidemic effects: ($p = 1$)		21.64 ($\chi^2_{.95}(1) = 3.84$)		23.60 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for the existence of expectation effects: ($\beta_4 = \beta_6 = 0$)		14.56 ($\chi^2_{.95}(1) = 3.84$)		18.72 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for the existence of rank effects: ($\beta_1 = \beta_3 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11}$)		47.40 ($\chi^2_{.95}(7) = 14.10$)		49.0 ($\chi^2_{.95}(7) = 14.10$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 0 or 1 for the lognormal and log-logistic models respectively).

6.6.1 Mis-Specification Tests

The aim of this section is to provide tests for three potential problems that can arise in the estimated models presented in Section 6.5. These are: unobserved heterogeneity, incorrect specification of the interaction between duration times and the covariates and multicollinearity between covariates. In general, the presence of these three features in the data set will lead to biased estimates and, thus, incorrect policy proposals being formulated. Consequently, specification testing is an important aspect of the modelling procedure and has, to date, been neglected by empirical economists.

6.6.1.1 Unobserved Heterogeneity

The empirical results presented in Section 6.4 were obtained from a modelling framework that implicitly assumes that the conditional PDF of duration times, $f(t|X)$, is true for all institutions. This assumption has two further elements to it. Firstly, there is an assumption that the functional form of $f(\cdot)$ is correctly specified. Secondly, there is an assumption that all institutions are homogenous with respect to the vector of covariates X . When employing such a framework for real life data sets the empirical economist is unlikely to be able to capture all institution-specific covariates that are important in determining the optimal date of adoption. Such covariates could include, for example, managerial attitudes or other organisational characteristics that do not lend themselves to measurement. Moreover, it is unlikely that all institutions will have the same functional form for $f(\cdot)$. Both these aspects encompass the problem of *unobserved heterogeneity* [Lancaster (1990) and Neumann (1997)].

As shown in Lancaster (1990) and Neumann (1997), unobserved heterogeneity has two main consequences for empirical duration models. Firstly, it creates downward bias in estimates of duration dependence. Secondly, it results in inconsistent maximum likelihood estimators and standard errors. Moreover, techniques to deal with unobserved heterogeneity can create additional problems of over-parameterisation

[Heckman and Singer (1984), and Neumann (1997)]. Thus, careful specification testing is an extremely important aspect of the empirical modelling.

Despite this, the literature has only just begun to consider these issues and consequently there is a definite paucity of formal specification tests available for duration models [Neumann (1997)]. The most commonly employed specification test, which is now established in the literature, is that proposed by Chesher (1984) and Lancaster (1985, 1990). The test assumes that heterogeneity enters multiplicatively into the hazard function as an institution-specific term. The basis of this test pertains to the properties of the estimated integrated hazard function $\Lambda(t_i, X_i, \hat{\theta})$, where X contains either time-invariant or time-varying covariates and $\hat{\theta}$ is an estimated vector of parameters belonging to the model. This term has been labelled the ‘generalised error’ by Lancaster (1990) by analogy with the OLS residual. In the absence of heterogeneity, the values of the generalised residuals will resemble a unit exponential distribution with a mean of zero and variance of unity. These properties can be summarised as follows:

$$\varepsilon_i = \int_0^{t_i} h(t_i, X_i, \hat{\theta}) \equiv \Lambda(t_i, X_i, \hat{\theta}) \quad (6.15)$$

with:

$$E(\varepsilon_i) = 0 \text{ and } \text{var}(\varepsilon_i) = 1 \quad (6.16)$$

where ε_i is the generalised residual.

The properties summarised in (6.15) and (6.16) have been the basis for informal residual plot tests proposed by D’Agostino and Stephens (1986) and Kiefer (1988). These tests exploit the fact that plotting (6.15) against time will produce a 45 degree line if the model is correctly specified. As noted by Neumann (1997), however, these tests are inaccurate because there is an absence of critical test statistic values to judge whether the residuals are ‘significantly’ large. The more formal numerical tests of Chesher (1984) and Lancaster (1985, 1990) are therefore followed. These numerical tests are

'score statistics' [see Greene (1993)] and are based on the generalised residual variance $s^2 - 1$ which is defined as:

$$s^2 - 1 = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^2 - 2 \quad (6.17)$$

where $\hat{\varepsilon}$ is the estimated residual from maximum likelihood estimation. Given the properties of ε in (6.16), it follows that in the presence of unobserved heterogeneity s^2 will be significantly greater than one. The term in (6.17) is transformed into a score statistic by dividing it by its asymptotic variance [Kiefer (1988)]. The nature of this variance term and the resulting score statistic are given in Appendix Five for the purpose of clarity.

The hypothesis to be tested then takes the following form:

$$\begin{aligned} H_0: & \text{unobserved heterogeneity not present in the data (i.e. } s^2 = 1) \\ H_1: & \text{unobserved heterogeneity present in the data (i.e. } s^2 > 1) \end{aligned} \quad (6.18)$$

The rejection region in favour of the hypothesis of heterogeneity is then the right tail area under the normal distribution $N(0, 1)$. The results obtained from each model estimated in Section 6.5 are presented in Table 6.7.

Table 6.7: Summary of Generalised Error Specification Tests for Estimated Accelerated Lifetime Models

Model	Baseline Hazard	Test Statistic	Decision
Fully specified	lognormal	0.879	Accept H_0
Fully specified	log-logistic	0.901	Accept H_0
Restricted stock	lognormal	0.801	Accept H_0
Restricted stock	log-logistic	0.905	Accept H_0
Myopic expectations	lognormal	1.011	Accept H_0
Myopic expectations	log-logistic	1.107	Accept H_0

Note: Critical values for all tests are one-sided and are at a 0.05 level of significance with critical value of 1.645 [Johnston (1987)].

The results summarised in Table 6.7 demonstrate that the hypothesis of no unobserved heterogeneity in the data cannot be rejected at the 0.05 level for all estimated models and specifications of the baseline hazard. Consequently, unobserved heterogeneity does not appear to be a problem for estimated models.

6.6.6.2 Robustness of the Estimates

The estimation of the accelerated lifetime model in Section 6.5 explicitly assumes that the interaction between duration times and the vector of covariates is to re-scale time. This specification provided an intuitive appeal for the subsequent interpretation of the signs of the estimated coefficients. As in the case of unobserved heterogeneity, however, the consequences of assuming an incorrect form for this interaction are inconsistent estimators [Lancaster (1990) and Horowitz and Neumann (1992)]. Ideally, it would be possible to check whether the underlying generating process is consistent with the assumptions made by the empirical model. Unfortunately, formal methods for testing this aspect of the model's specification have not been thoroughly worked out for the accelerated lifetime model and even less so in the presence of censoring and time-varying covariates [Neumann (1997)].

In the absence of formal tests, a method frequently employed by empirical economists [see, for example, Karshenas and Stoneman (1993), Colombo and Mosconi (1995) and Saloner and Shepard (1995)] is to examine the robustness of the results for different functional forms of the $J(.)$ function in expression (6.11). As noted in Section 6.3.1, a natural alternative to the accelerated lifetime model is the proportional hazards model of Cox (1972, 1975). In contrast to the accelerated lifetime model the proportional hazards model defines the interaction between duration times and covariates to be multiplicative in nature. This can be seen from the specification of the model in (6.12). Again, formal tests that compares the performance of the proportional hazards model with the accelerated hazards model have not yet appeared, although this is a current theme in the econometrics literature [Horowitz (1992) and Neumann (1997)].

The continuous time version of the proportional hazards model in (6.12) was estimated for the set of potential adopters. As this model has not yet been formulated in the literature for the presence of time-varying covariates the model was estimated employing time-invariant covariates measured at the time of adoption for adopting institutions and the time of censoring (1992) for non-adopting institutions. This methodology follows convention [see, for example, Karshenas and Stoneman (1995)]. The model was estimated by partial likelihood assuming the baseline hazard to be arbitrary in order to impose the minimum amount of restrictions (see Appendix One). The three separate specifications of the theoretical model as summarised in Table 6.3 are estimated and the results are shown in Tables 6.8, 6.9 and 6.10 respectively. All 't' subscripts are removed from the covariate names to indicate their time-invariance.

Four aspects of the results obtained from the proportional hazards model should be noted. Firstly, because the model is homogenous of degree zero in X [see Greene (1993)] those covariates which remain constant across institutions do not enter the partial likelihood and subsequently drop-out of the estimation procedure (see Appendix One). Consequently, all models are estimated without an intercept term. Secondly, given the differences in functional forms between the accelerated lifetime and proportional hazard models, the absolute values of the coefficients in Tables 6.8, 6.9 and 6.10 are not directly comparable with those summarised in Tables 6.4, 6.5 and 6.6. Thirdly, the model could not be estimated with the inclusion of *DTAKE* and *DSUB*. This occurred because there was 'insufficient' variation in the X vector resulting in the Hessian becoming singular [Heckman and Singer (1984)]. Fourthly, as the baseline hazard is arbitrarily specified it is not possible to test for the existence of epidemic effects in this approach.

In general, the results obtained from the proportional hazards model support those obtained from the accelerated lifetime model. Observation of Tables 6.8 and 6.9 show a significant role for rank effects in the diffusion process. The LR test for their joint significance rejects the null hypothesis that all coefficients are zero at the 0.05 level in all three specified models. Moreover, all rank effect covariates have the same sign as found in the accelerated lifetime model and conform to the *a priori* expectations as set-

out in Section 6.4.1. No statistically significant role, however, is found for the *STAFF/BRANCH* ratio. This result is also consistent with that obtained for the accelerated lifetime model and, again, may indicate that this variable is a poor proxy for the labour-saving potential of ATMs.

No significant role is found for stock effects. From Table 6.9 the LR test cannot reject the restriction $K = 0$ at the 0.05 level. In addition, a significant role for the price of technology is found in all three models and has the correct sign *a priori*. This result does differ from that obtained from the accelerated lifetime model which failed to register a significant effect for this covariate. This outcome may occur because of the methodology used in the measurement of covariates for non-adopting institutions. The real quality-adjusted price of ATMs has fallen monotonically since their commercialisation and so, by definition, those institutions not adopting ATMs at the end of 1992 will necessarily have a lower value for this covariate relative to those institutions that do not adopt.

From Tables 6.8 and 6.9 it can be observed that the coefficients on k and p are both found to be statistically significant and have the correct sign based on the results of the theoretical model outlined in Section 6.3. Moreover, in Table 6.8 the LR test rejects the restriction $k/r = p = 0$ at the 0.05 level of significance and so the myopic model specification is not supported by the data. This lends further support to the contention that myopic models of technology diffusion may be mis-specified.

Overall then, the results from the proportional hazards model by and large support those obtained from the accelerated lifetime model. With the exception of the covariate rP , the estimated vector of coefficients β in terms of sign and statistical significance closely follow those of the accelerated lifetime model.

Table 6.8: Maximum Likelihood Estimates of the Fully Specified Proportional Hazards Model			
Coefficient	Variable	Estimates	
α_1	<i>CONSTANT</i>	3.001	(4.679)**
β_1	<i>SIZE</i>	0.046	(6.360)**
β_2	<i>K</i>	-4.581	(0.607)
β_3	<i>GY</i>	2.597	(2.081)*
β_4	<i>k/r</i>	2.728	(2.414)*
β_5	<i>rP</i>	0.068	(3.733)**
β_6	<i>p</i>	-0.122	(8.744)**
β_7	<i>STAFF/BRANCH</i>	0.015	(1.280)
β_8	<i>PROFITS/SIZE</i>	32.952	(2.785)**
β_9	<i>DPREVIOUS</i>	1.875	(4.567)**
Log-likelihood		89.32	
Number of observations		98	
Number of individual institutions		98	
Likelihood ratio test for the existence of rank effects: ($\beta_1 = \beta_3 = \beta_7 = \beta_8 = \beta_9$)		23.07 ($\chi^2_{.95}(5) = 11.10$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '**' means significant at the 0.05 level; '***' means significant at the 0.01 level.

Table 6.9: Maximum Likelihood Estimates of the Restricted Stock and Myopic Proportional Hazards Model

Coefficient	Variable	Restricted Stock		Myopic	
α_1	<i>CONSTANT</i>	2.875	(3.876)**	2.980	(3.541)**
β_1	<i>SIZE</i>	0.036	(5.670)**	0.028	(4.888)**
β_2	<i>K</i>	—		-4.302	(0.316)
β_3	<i>GY</i>	2.483	(2.080)*	2.045	(2.016)*
β_4	<i>k/r</i>	2.728	(2.550)*	—	
β_5	<i>rP</i>	0.068	(6.281)**		
β_6	<i>p</i>	-0.101	(5.678)**	—	
β_7	<i>STAFF/BRANCH</i>	0.015	(0.775)	0.012	(0.854)
β_8	<i>PROFITS/SIZE</i>	27.550	(2.903)**	17.402	(2.314)*
β_9	<i>DPREVIOUS</i>	2.031	(5.272)**	2.201	(5.721)**
Log-likelihood		90.63		94.25	
Number of observations		98		98	
Number of individual institutions		98		98	
Likelihood ratio test for the existence of stock effects: ($\beta_2 = 0$)		2.83 ($\chi_{.95}^2(1) = 3.84$)		—	
Likelihood ratio test for the existence of expectation effects: ($\beta_4 = \beta_6 = 0$)		—		10.04 ($\chi_{.95}^2(2) = 3.84$)	
Likelihood ratio test for the existence of rank effects: ($\beta_1 = \beta_3 = \beta_7 = \beta_8 = \beta_9$)		19.71 ($\chi_{.95}^2(5) = 11.10$)		18.79 ($\chi_{.95}^2(5) = 11.10$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level.

6.6.6.3 Collinearity between the Covariates

In the estimation of the empirical duration models in Section 6.5 it is implicitly assumed that the vector of covariates, X , has full rank¹¹ In the case of classical OLS regression, if this assumption is violated and there do exist exact linear relationships amongst some

¹¹ The rank of a matrix is the largest number of linearly independent column vectors it contains. A matrix which has 'full rank' is one in which the column rank is equal to the number of columns it contains [Johnston (1987)].

or all the exogenous variables then the resulting variance of the estimated coefficients will have infinite variance and the parameters of the models will be unidentified [Greene (1993)]. Consequently, there will exist an infinite number of possible parameter vectors, which are consistent with the same expected value of the independent variable [Johnston (1987) and Greene (1993)]. As noted by Kiefer (1988) and Neumann (1997) duration models are certainly not immune to this potential problem and the consequences for estimation are identical to that in the OLS case.

Such 'perfect multicollinearity' is, however, rare and the result of poor modelling rather than an inherent data 'quality' problem. A more commonly encountered problem arises when there is 'near multicollinearity' between covariates which is characterised by high correlation coefficients between covariates. This near collinearity presents three well known problems for estimation. Firstly, small changes in the data may produce dramatic changes in the signs and/or values of covariate estimates. Moreover, as noted by Johnston (1987), the higher the correlation between the covariates the less precise the covariate coefficients can be estimated. Secondly, estimated coefficients will have very high standard errors and low significance levels in spite of being jointly significant via the LR test for example. Thirdly, coefficients will have the incorrect sign based on a priori expectations and/or have an implausible magnitude.

In the empirical results presented in Section 6.5 there are potentially two sources of multicollinearity: firstly, high correlation between market-specific covariates such as $r_t P_t$, p_t and K_t and, secondly, between institution-specific covariates. A notable problem may arise between $r_t P_t$ and K_t , the former decreasing monotonically over the time and the latter increasing monotonically over time. As noted by Hsiao (1985, 1986), however, the employment of panel data often mitigates against multicollinearity problems because it introduces cross-sectional differences and, thus, provides more information relative to purely time-series data.

Given all this, there is, however, a paucity of formal tests available for multicollinearity and those that do exist are frequently difficult to interpret and do not lend themselves to the derivation of critical values required for statistical inference [Maddala (1987)]. A

more commonly taken route is to consider the symptoms of multicollinearity as outlined above and observe if estimated models suffer from any or all of these symptoms. This methodology was employed to investigate the multicollinearity issue in more detail.

Taking the results presented in Section 6.5 together, it can be observed that all estimated coefficients have the correct sign based on the underlying theoretical model presented in Section 6.3 and the arguments relating to the rank effects as laid forth in Section 6.4.1. In addition, the magnitudes of the estimated coefficients are not unrealistic given the descriptive statistics for the covariates in Appendix Two. Although the covariates K_t and $r_t P_t$ are found to be statistically insignificant in the fully specified model, restricting their coefficients to zero in Tables 6.5 and 6.6 did not lead to any changes in the signs of other covariates or their statistical significance. Moreover, the LR test in Table 6.6 cannot reject the restriction $K_t = 0$ and this supports the result of the t -test for its significance in the fully specified model. Thus, there does not appear to be a problem of multicollinearity for the expectations and stock covariates despite their time-series nature.

As noted by Karshenas and Stoneman (1995), multicollinearity is most likely to arise in empirical models of diffusion from the inclusion of firm size in the model as this may be highly correlated with other rank effect covariates such as proxies for financial constraints and output growth measures for example. Indeed, as indicated in Chapter 3 multicollinearity problems deriving from the inclusion of firm size were encountered in the early contributions of Mansfield (1961, 1968) and Romeo (1975). In order to investigate this issue further, the fully specified model in Table 6.4 was re-estimated for the period 1972 to 1992 imposing the restriction $\beta_1 = 0$. The results from the estimation are given in Table 6.10 below.

Table 6.10 Maximum Likelihood Estimates of Restricted Size Lognormal and Log-logistic Models					
Coefficient	Variable	Lognormal Model		Log-logistic Model	
p	$TIME$	4.111	(3.896)**	4.129	(3.278)**
α_1	$CONSTANT$	3.590	(7.596)**	4.054	(4.977)**
β_1	$SIZE_t$	—		—	
β_2	K_t	-0.014	(1.690)	-0.009	(0.765)
β_3	GY_t	1.462	(2.436)*	1.958	(2.779)**
β_4	k_t/r_t	1.099	(2.385)*	1.489	(2.830)**
β_5	$r_t P_t$	0.017	(1.053)	0.003	(1.124)
β_6	p_t	-0.010	(2.216)*	-0.006	(2.471)*
β_7	$STAFF_t/BRANCH_t$	0.013	(1.325)	0.024	(1.727)
β_8	$PROFITS_t/SIZE_t$	9.585	(2.180)*	10.561	(2.062)*
β_9	$DPREVIOUS$	1.276	(4.735)**	1.430	(4.087)**
β_{10}	$DTAKE_t$	0.423	(2.296)*	0.663	(1.793)
β_{11}	$DSUB$	-0.456	(0.810)	-0.442	(0.831)
Median duration (years)		29.77	(6.708)**	32.45	(5.670)**
Log-likelihood		-176.99		-176.58	
Number of observations		1956		1956	
Number of individual institutions		98		98	
Likelihood ratio test for the existence of epidemic effects: ($p = 1$)		9.98 ($\chi^2_{.95}(1) = 3.84$)		10.36 ($\chi^2_{.95}(1) = 3.84$)	
Likelihood ratio test for existence of institution size effects: ($\beta_1 = 0$)		21.42 ($\chi^2_{.95}(1) = 3.84$)		21.26 ($\chi^2_{.95}(1) = 3.84$)	

As can be observed from comparing the results in Table 6.4 with those in Table 6.10, the vector of estimated coefficients, $\hat{\beta}$, remains relatively stable in terms of the magnitudes of values estimated and their statistical significance. Although there is a slight increase in the significance of GY_t and k_t/r_t in the log-normal model and a decrease in the estimated coefficient on p_t in both the lognormal and log-logistic models, the substance of the empirical results presented in Section 6.5 remain the same.

In addition, the restriction $\beta_1 = 0$ is rejected by the LR test in both the lognormal and log-logistic models at the 0.05 level of significance. This lends further empirical support to the importance of institution size in the rank effects. An identical exercise to that carried out for $SIZE_t$, was also carried out for $PROFITS_t/SIZE_t$, and, again, there was no change in the substance of the results presented in Section 6.5 with the $\hat{\beta}$ remaining relatively stable. These covariate-deletion tests, although arguably *ad hoc*, do support the contention that collinearity between the covariates is not a major problem and is certainly not sufficient to alter the substance of the results contained in Section 6.5. Moreover, they lend support to the contention put forward by Hsiao (1985, 1986) that panel data models can mitigate against the potential problem of multicollinearity by allowing cross-sectional variation.¹²

6.6.2 Extensions to the Model

Three extensions to the basic model were considered. Firstly, different measures for the covariates. Secondly, the potential role and significance for different expectations regimes formed on the price of technology and, thirdly, the potential significance of non-linear size effects.

Due to a lack of available and consistent data the only covariates that could be re-defined were $STAFF_t/BRANCH_t$ and $PROFITS_t/SIZE_t$. The former covariate was re-defined by excluding part-time staff in its definition. The sign and statistical significance remained positive and significant. The $PROFITS_t/SIZE_t$ covariate was re-defined by measuring the numerator as net income as opposed to post-tax profits. This was employed in order to address the arguments in Kamien and Schwartz (1982) that

¹² An alternative methodology to that of deleting potentially troublesome covariates often put forward in the literature is testing the stability of the $\hat{\beta}$ vector across sub-periods of the total sample period [Maddala (1987)]. Such a methodology is, arguably, more appropriate for time-series data than panel data. In the case of the ATM diffusion data, estimating the models in sub-periods will inevitably lead to significant differences in estimated parameter values as estimation will exclude not only time-periods but also cross-sectional units. If, for example, estimation was performed exclusively for the latter years of the diffusion process this approach would, by definition, exclude those institutions adopting the technology shortly after its commercialisation. In this scenario, there are likely to be significant departures in the values of estimated coefficients from their full sample values.

the net income measure more accurately captures the liquidity constraints facing the firm. In the fully specified model this re-defined covariate remained positive and significant although its estimated coefficient declined slightly.

Two different expectation regimes formed on the price of technology were considered in order to allow a more flexible approach to expectations formation: adaptive and Goodwin expectations [see Maddala (1987)]. As noted by Karshenas and Stoneman (1993), however, if the price of technology declines monotonically overtime then the arbitrage condition contained in expression (6.5) is non-binding and the theoretical model is transformed into a myopic one. As the real quality-adjusted price of ATMs has fallen monotonically since 1972 this aspect of the model is already captured by estimating the model imposing the constraint $p_t = 0$. Consequently, this route was not pursued.

As shown in Chapter 3, the empirical contributions of Globerman (1975), Benvignati (1982) and Hannan and McDowell (1987) have all indicated that including an squared term for firm size results in a reverse effect on the hazard function to that of firm size alone. This suggests that although the conditional probability of adoption is increasing in firm size the rate of increase actually decreases in size. To investigate this issue further, $(SIZE_t)^2$ was included in the regression models. In the fully specified lognormal model this squared term obtained a value of -0.000014 with t -statistic of 0.727, whilst in the fully specified log-logistic model it obtained a value of -0.000011 with t -statistic of 0.801. With a two-sided 0.05 significance level of 1.96 both these estimated coefficients are statistically insignificant. Thus, no empirical support is found in support of non-linear effects in firm size.

6.7 Concluding Remarks

The empirical results indicate that rank effects play an extremely significant role in the diffusion of ATMs, thus supporting probit-type theoretical models and the majority of non-industrial and industrial sector studies of diffusion. Institution size, growth in deposits and profitability were all found to have a positive and statistically significant effect on the conditional probability of adoption. Moreover, the results suggest that early adoption of previous vintage technologies resulting in learning-by-doing effects play a significant role in fostering faster diffusion. Thus, the former 'technology history' of the firm affects current adoption decisions. No significant role is found, however, for the labour-saving potential of ATMs and a dummy covariate for whether the institution had been involved in takeovers.

No empirical support is found for the existence of stock effects in the diffusion process, although order effects entered the empirical model with the correct sign and was found to be statistically significant.

Furthermore, the empirical results lend support to the existence of epidemic effects in the diffusion of ATMs and it was illustrated that this was not due to the potential collinearity between stock effects and the time-varying nature of the baseline hazard. This supports the results obtained in Chapter 5.

It was also found that expectations formed on the number of adopters and the price of technology have a significant role to play in the diffusion process, although the real quality-adjusted price of technology fails to register a significant effect. This lends further support to the work of Karshenas and Stoneman (1993) that models specifying myopic expectations may be seriously mis-specified.

Reassuringly, the results were found to be robust across different specifications of the baseline hazard. Finally, formal and informal tests indicated that estimated models did not suffer from unobserved heterogeneity and multicollinearity, further supporting the robustness of the results obtained.

A6.1 Appendix One: The Mathematics of Duration Models

A.6.1.1 Introduction

In Chapter 5 several parametric distributions often employed in the literature to represent the hazard, survival and integrated hazard functions were estimated for the sample of ATM adoption histories. It was explicitly assumed in the estimation that these distributions were specified for a homogenous population. It is more usual, however, that there exists explanatory variables or covariates (i.e. institutional- and market-specific characteristics) upon which the duration of non-adoption may depend. This appendix outlines two classes of models that have been employed in the empirical literature: the non-parametric proportional model of Cox (1972) and the parametric accelerated lifetime model of Kalbfleisch and Prentice (1980) and Nuemann (1997).

Throughout this appendix it is assumed there exists a $K \times 1$ vector of covariates $X = (X_1, \dots, X_K)$. The principal modelling problem is to determine the relationship between t and X . Introducing covariates into the modelling framework is, however, complicated by the existence of two distinct types of covariates: time-invariant and time-varying. The former does not change over time. In the presence of these covariates assembled in vector X , the continuous time hazard function at time t is defined as being conditional on the value of X . Thus, the hazard function in continuous time can be written as:

$$h(t|X) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t, X)}{dt} \quad (\text{A6.1.1})$$

Thus, from (A6.1.1) $h(t|X)dt$ gives the fraction of the survivors at time t who leave in the short time interval from t to $t + dt$ in a large population of people who are homogenous with respect to X .

In the case of time-varying covariates, however, a further distinction has to be made between *exogenous* and *endogenous* covariates analogous to the distinction made in Engle *et al* (1983) and Heckman and Singer (1984).¹³ Lancaster (1990) has shown that a covariate process $X(t)$ is exogenous for T if and only if the following condition holds:

$$\Pr[X(t, t + dt) | T \geq t + dt, X(t)] = \Pr[X(t, t + dt) | X(t)] \quad \forall t \geq 0 \text{ and } dt \geq 0 \quad (\text{A6.1.2})$$

where $X(t)$ is the path of the covariates from 0 to t . The definition contained in the expression (A6.1.2) implies that information that an institution has not adopted to $t + dt$ does not aid in the prediction of the path of the covariate process from t to $t + dt$ given its history to t . Any covariate that is not exogenous is, by definition, endogenous [Lancaster (1990)].

In the case of exogenous variables the conditioning in (A6.1.2) at time t will be the entire path of X . Thus, the specification of the hazard function in (A6.1.1) can be re-specified in the presence of time-varying exogenous covariates as follows:

$$h(t | X(t)) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t, X + dt)}{dt} \quad (\text{A6.1.3})$$

This definition is, however, not possible for endogenous covariates where proper treatment may require modelling the joint probability of T and $X(t)$ [Neumann (1997)]. The treatment of time-varying covariates in duration models is in its infancy and the current debate centres on questions about the precise timing of exits on the covariate path [see Neumann (1997)]. This issue discussed in more detail in relation to the accelerated lifetime model. Unless otherwise specified in the following discussion it is assumed that X is a vector of time-invariant covariates and that $h(t | X)$ represents the hazard function at time t for a firm with covariates X .

¹³ Kalbfleisch and Prentice (1980) and Neumann (1997) use the alternative terms 'external' and 'internal'

A6.1.2 The Proportional Hazards Models

The non-parametric proportional hazards model of Cox (1972) assumes that the interaction between t and X to be a multiplicative one. The continuous time model specifies that:

$$h(t|X) = h_0 \exp(X\beta) \quad (\text{A6.1.4})$$

where β is a $\times K$ vector of covariate parameters and $h_0(t)$ is labelled the ‘base-line’ hazard function [Neumann (1997)] and conforms to the condition $X = 0$. The covariates are embodied in the link function $\exp(X\beta)$.¹⁴

The advantage of the proportional hazards model is that, as Cox (1972, 1975) shows, the parameter vector β can be estimated *without* specifying the form or family for the base-line hazard $h_0(t)$. This is achieved by using a partial likelihood method of estimation [Amemyia (1986)]. Specifically, the procedure orders the durations of non-adoption in ascending order such that $t_1 < t_2 \dots < t_n$. The conditional probability that the first duration ends at t , given that any of the n durations could have ended, is then specified (ignoring censoring) as follows:

$$\frac{h(t, X_1, \beta)}{\sum_{i=1}^n h(t_i, X_i, \beta)} \quad (\text{A6.1.5})$$

which, given the expression in (A6.1.4), reduces to the following [Kiefer (1988)]:

$$\frac{\exp(X_1\beta)}{\sum_{i=1}^n \exp(X_i\beta)} \quad (\text{A6.1.6})$$

definitions respectively.

¹⁴ As noted by Lawless (1982), nothing requires the link function to be specified as this. A pre-requisite is, however, that the function must be restricted to those that guarantee $\exp(X\beta) > 0 \forall X$.

The quantity in (A6.1.6) is then the contribution of the shortest duration observed to the partial likelihood. Similarly, the contribution of the j th shortest duration is simply the ratio of the hazard for the firm completing a spell of non-adoption at time t_j to the sum of the hazards for firms whose spells were in progress just prior to t_j . In each case, the contribution to the likelihood is the ratio of the hazard for the firm whose spell was completed at duration t divided by the sum of the hazards for firms whose durations were still in progress just prior to t . The partial likelihood function is then specified as the product of these individual contributions. The approach is also able to accommodate right-censored observations. A firm whose duration is censored between times t_j and t_{j+1} appears in the summation in the denominator of (A6.1.5) for observations 1 through to j , but not in any others. Censored observations do not enter the numerator of a contribution to the likelihood at all [Kiefer (1988) and Neumann (1997)].

Kiefer (1988) notes that the proportional hazard specification has a linear model interpretation. More succinctly, the model satisfies:

$$-\ln(\Lambda_0(t)) = X\beta + \nu \quad (\text{A6.1.7})$$

where $\Lambda_0(t)$ is the base-line integrated hazard function, ν is a random variable distributed as a unit extreme value with CDF of $F(\nu) = 1 - \exp(-\exp(\nu))$ with $-\infty < \nu < \infty$ [see D'Agostino and Stephens (1986)].

Discrete time versions of the model are outlined in Kalbfleisch and Prentice (1980) and Han and Hausman (1990). To date, the proportional hazards model has only been specified for time-invariant covariates.

A6.1.3 The Accelerated Lifetime Model

In this model the effect of covariates is to re-scale time directly or, equivalently, the role of X is to accelerate (or decelerate) the time to failure. In terms of the base-line hazard introduced in the preceding section this model is specified in continuous time as follows [Kalbfleisch and Prentice (1980)]:

$$h(t|X) = h_0(\phi(X\beta))\phi(X\beta) \quad (\text{A6.1.8})$$

with the conventional specification of $\phi(\cdot)$ taken to be $\exp(\cdot)$. Estimation of the accelerated lifetime model then proceeds by choosing a functional form for h_0 (based on *a priori* estimation of the Kaplan-Meier estimator as a measure of the structural hazard function for example) and maximising the log-likelihood function. Right-censored observations are accommodated into the log-likelihood function using the indicator variable introduced in Chapter 5. The resulting log-likelihood has the following form:

$$L(\beta) = \sum_{i=1}^n \delta_i \ln f[t_i, X_i, \beta] - \sum_{i=1}^n (1 - \delta_i) \ln S[t_i, X_i, \beta] \quad (\text{A6.1.9})$$

where δ_i takes a value of unity if the adoption time for an individual institution is uncensored and unity if censored. As shown by Nuemann (1997), the log-likelihood is concave and the negative inverse of the second derivative matrix can be used as an approximate covariance matrix for the estimator of β .

As noted by Kiefer (1988) the accelerated lifetime model also has a linear model interpretation as specified in (A6.1.7) except that the left-hand side is replaced by $-\ln t$ and $F(v)$ is a continuous but unspecified CDF. Thus, the accelerated lifetime and proportional hazards models make different assumptions concerning the features of duration data. The proportional hazards allows some general transform of duration time to be linearly related to $X\beta$, but completely specifies the distribution of the error term. In contrast, the accelerated lifetime model completely restricts the transform of duration time, but allows an arbitrary structure for the error term. In this respect, Ridder (1990)

has proposed a 'Generalised Accelerated Failure Time Model' that nests both model specifications. As noted by Horowitz (1992) and Neumann (1997), however, that Ridder's approach cannot, to date, handle censored observations and therefore has limited empirical use.

Petersen (1986a, 1986b) has considered the case in which X contains time-varying exogenous covariates. Petersen assumes that an individual duration can occur only at a set of discrete time points. This approach assumes that the duration in a state of non-adoption for an individual institution, t_k , can be divided into k non-overlapping but adjacent segments of time which need not be the same length. Letting $t_0 = 0$ and $t_0 < t_1 < \dots < t_j < \dots < t_k$ then considering the duration interval t_{j-1} to t_j where the covariates in X stay constant at $X(t_{j-1})$, the survivor function becomes:

$$\Pr[T \geq t_j | T \geq t_{j-1}, X(t_j)] = \exp\left\{- \int_{t_{j-1}}^{t_j} h[u, X(t_{j-1})] du\right\} \quad (\text{A6.1.10})$$

The expression in (A6.1.10) defines the probability of surviving beyond duration t_j given survival at duration t_{j-1} and given the path taken by X to duration t_j [Petersen (1986a)].

The novel aspect of the Petersen approach is that integration over the time path of the covariates in X is not required because they stay constant in the duration interval t_{j-1} to t_j . Thus, the total duration in a state can always be divided into sub-periods of time within which all the step-function covariates stay constant.

Petersen (1986a, 1986b) has further shown that the survivor function for surviving beyond duration t_k is:

$$S[t_k | X(t_k)] = \exp\left\{- \sum_{j=1}^k \int_{t_{j-1}}^{t_j} h[u, X(t_{j-1})] du\right\} \quad (\text{A6.1.11})$$

where $t_0 = 0$. This is derived by factoring the product of the conditional survivor functions, given in (A6.1.10), for non-overlapping but adjacent segments of time from duration 0 to t_k . The number and length of such segments will vary between observations. Estimation utilises a non-linear least squares algorithm and proceeds by maximum likelihood [see Petersen (1986a) and Greene (1993) for technical details]. Censored observations are dealt with in the conventional manner.

A6.2 Appendix Two: Descriptive Statistics

Table A6.2.1 below gives the descriptive statistics of all the variables used in this chapter. Table A6.2.2 gives two measure of dispersion: skewness and kurtosis. The skewness of a random variable X_i ($i = 1, \dots, n$) with mean and variance \bar{X} and σ_x^2 respectively is defined as [Newbold (1991)]:

$$S(X_i) = n^{-1} \left[\sum_{i=1}^n (X_i - \bar{X})^3 / \sigma_x^3 \right] \quad (\text{A6.2.12})$$

The kurtosis of X_i is defined as [Newbold (1991)]:

$$K(X_i) = n^{-1} \left[\sum_{i=1}^n (X_i - \bar{X})^4 / \sigma_x^4 \right] \quad (\text{A6.2.13})$$

The normal distribution has skewness equal to zero, as do all other symmetric distributions. Positive skewness will result if the distribution is skewed to the right because the average cubed deviations will be positive. Skewness will be negative for distributions skewed to the left. In contrast, the normal distribution has a kurtosis value of 3, but fat-tailed distributions with supplementary probability mass in the tail area will have higher or even infinite kurtosis.

For time-varying covariates the statistics refer to the values prevailing at the time of adoption for adopting institutions and the date of censoring for non-adopting institutions.

Table A6.2.1: Descriptive Statistics of the Explanatory Variables							
Variable	Units	All Institutions		Adopting Institutions		Non-Adopting Institutions	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
<i>CONSTANT</i>	N/A	1.00	0.00	1.00	0.00	1.00	0.00
<i>SIZE_t</i>	£million	2137.80	454.60	5041.70	6228.40	570.67	2056.90
<i>GY_t</i>	% growth	0.10	0.14	0.14	0.12	0.08	0.15
<i>K_t</i>	units	27.46	9.62	17.98	10.89	32.58	2.08
<i>k_t/r_t</i>	units	0.13	0.22	0.37	0.23	0.00	0.00
<i>r_t</i>	% rate	8.18	2.92	11.40	2.87	6.48	0.00
<i>r_tP_t</i>	(1980=100)	48.43	23.93	73.70	25.20	34.79	3.07
<i>p_t</i>	(<i>P_{t+1} - P_t</i>)	-1.72	1.59	-1.42	2.65	-1.88	0.30
<i>STAFF_t</i>	units	2224.15	7236.09	6185.58	11284.75	86.23	209.89
<i>PROFIT_t</i>	£million	23.43	52.83	58.33	75.11	4.59	17.01
<i>BRANCH_t</i>	units	433.73	2256.35	1206.68	3723.01	16.59	31.78
<i>STAFF_t/BRANCH_t</i>	units	6.24	5.53	8.88	7.48	4.82	3.43
<i>PROFIT_t/SIZE_t</i>	£million	0.98	0.48	1.19	0.45	0.87	0.46
		$\times 0^{-2}$	$\times 0^{-2}$	$\times 0^{-2}$	$\times 0^{-2}$	$\times 0^{-2}$	$\times 0^{-2}$
<i>DEPOSIT_t</i>	£million	1601.86	3578.71	3838.20	5083.31	394.94	1361.67
<i>DPREVIOUS</i>	dummy	0.05	0.22	0.15	0.36	0.00	0.00
<i>DTAKE_t</i>	dummy	0.06	0.24	0.15	0.36	0.02	0.13
<i>DSUB</i>	dummy	0.03	0.17	0.06	0.24	0.02	0.13
Number of observations		98					
Number of adoptions		35					
Number of censored observations		63					

Note: 'S.D.' means 'standard deviation'; 'N/A' means not applicable

Variable	Units	All Institutions		Adopting Institutions		Non-Adopting Institutions	
		Skew.	Kurt.	Skwe.	Kurt.	Skew.	Kurt.
<i>CONSTANT</i>	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<i>SIZE_t</i>	£million	2.87	10.38	1.46	3.72	6.93	52.01
<i>GY_t</i>	% growth	-4.56	39.42	0.69	7.57	-5.94	43.16
<i>K_t</i>	units	-1.62	4.23	-0.18	1.60	-5.30	29.06
<i>k_t/r_t</i>	units	1.53	3.88	0.10	1.62	0.00	0.00
<i>r_t</i>	% rate	1.46	3.90	0.10	2.22	7.69	60.05
<i>r_tP_t</i>	(1980=100)	1.73	5.13	0.57	2.60	7.69	60.05
<i>p_t</i>	($P_{t+1} - P_t$)	-1.43	4.12	-0.55	3.13	-7.69	59.05
<i>STAFF_t</i>	units	4.33	21.24	2.24	6.58	5.63	38.78
<i>PROFIT_t</i>	£million	3.27	14.37	1.76	5.46	7.08	53.71
<i>BRANCH_t</i>	units	8.73	81.69	5.02	27.63	4.23	25.07
<i>STAFF_t/BRANCH_t</i>	units	3.84	23.03	2.62	11.63	3.88	21.88
<i>PROFIT_t/SIZE_t</i>	£million	0.45	5.84	1.88	7.39	-0.16	4.18
<i>DEPOSIT_t</i>	£million	2.99	11.41	1.50	4.00	6.87	57.42
<i>DPREVIOUS</i>	dummy	4.04	17.27	1.96	4.83	7.69	60.05
<i>DTAKE_t</i>	dummy	3.62	14.09	1.96	4.83	7.69	60.05
<i>DSUB</i>	dummy	9.65	94.03	5.48	31.09	0.00	60.05
Number of observations		98					
Number of adoptions		35					
Number of censored observations		63					

Note: 'Skew.' means 'Skewness'; 'Kurt' means 'Kurtosis'; 'N/A' means not applicable

A6.3 Appendix Three: Data Sources

The data sources employed in this chapter were as follows:

- The yield on Treasury Bills was obtained from Table 7.1H of various copies of *Financial Statistics*, Office for National Statistics (ONS), Stationary Office, London.
- Producer Price Index (PPI) was obtained from Table 2.1 of various copies of *Economic Trends*, ONS, Stationary Office, London.
- Quality adjusted price of ATMs was supplied by the ONS, London.
- Data on retail banks and building societies were obtained from previous editions of the *Annual Abstract of Banking Statistics*, British Bankers' Association, London and the *Building Societies Yearbook*, Building Societies Association, London and from direct correspondence with the institutions themselves.
- Data on the time of adoption of ATMs for both retail banks and building societies were obtained from direct correspondence with the relevant institutions.

A6.4 Appendix Four: Estimating the Cumulative Number of Adopters

The estimated cumulative number of adopters at time t , K_t , is derived from a first-order autoregressive model of the form:

$$K_t = a_0 + a_1 K_{t-1} + u_t \quad (\text{A6.4.1})$$

where u_t is an error term assumed to be $u_t \sim IN(0, \sigma^2)$. The OLS estimation results for the time-series model in (A6.4.1) for the set of potential ATM adopters is given in Table A6.4.1 below.

Coefficient	Variable	Estimates	
<i>CONSTANT</i>	a_0	1.5103	(2.1308)
K_{t-1}	a_1	1.0247	(26.4098)
R^2	0.97718		
Durbin's h -statistic	1.0298		

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; Durbin's h -statistic is a test for first order autocorrelation in the error term when the X matrix of explanatory variables is non-stochastic which occurs when lagged variables are included in the regression. The significance level of Durbin's h -statistic 1.6450 [Johnston (1987)].

A6.5 Appendix Five: A Mis-Specification Test

Lancaster (1985, 1990) has shown that the asymptotic variance of (6.17) can be obtained from the information matrix I pertaining to the estimated parametric model.

For the case of a two-parameter model, $\theta = (\theta_1, \theta_2)$, the information matrix in partitioned form is defined as [Johnston (1984) and Cuthbertson *et al* (1992)]:

$$I = \begin{pmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} \end{pmatrix} \equiv \begin{pmatrix} I_{\eta\eta} & I_{\eta\gamma} \\ I_{\gamma\eta} & I_{\gamma\gamma} \end{pmatrix} \quad (\text{A6.5.14})$$

where η and γ are the true values of θ_1 and θ_2 respectively. The variance of the score statistic is then given by the following expression [Lancaster (1990)]:

$$= I_{\eta\eta} - I_{\eta\gamma} I_{\gamma\gamma}^{-1} I_{\gamma\eta} \quad (\text{A6.5.15})$$

The resulting score statistic is then defined as follows with η and γ replaced by their maximum likelihood estimates [Lancaster (1985)]:¹⁵

$$\frac{\sqrt{n}(s^2 - 1)}{V} \sim N(0, 1) \quad (\text{A6.5.16})$$

The information matrix I was obtained for each estimated model by estimating the negative Hessian at the estimated values of θ [see Cuthbertson *et al* (1992)].

¹⁵ The score statistic in expression (A6.5.3) has been also been interpreted by Chesher (1984) and Neumann (1997) as an Information Matrix test statistic [White (1982)] for unobserved heterogeneity when the variance of heterogeneity is small.

CHAPTER 7

ATM DIFFUSION AND NETWORK EXTERNALITIES: AN EMPIRICAL INVESTIGATION

7.1 Introduction

A current theme in the industrial organisation literature is the economic analysis of network technologies. Mature and developed economies abound with examples of these technologies from consumer durables, such as videocassette recorders and facsimile machines, to the provision of information services, such as telecommunications. The defining characteristic of these technologies is that they all exhibit, to varying degrees, positive consumption and production externalities, frequently labelled by the generic terms 'network externalities' or 'network-effects' [Economides (1996)]. The theoretical literature has suggested significant implications for, *inter alia*, pricing strategies and market structure, product standardisation and the diffusion process in those markets and for those technologies that exhibit these externalities [see, for example, Laffont *et al* (1998)].

Despite these innovations in the literature, economists have not, to date, been rigorous in their empirical testing of the potential effects of network externalities for the adoption and diffusion of new technology. Those studies that do exist are mainly based on case study evidence [see, for example, David (1985), Foray and Grubler (1990) and Postrel (1990)]. Although these case studies provide an interesting and indispensable insight into observed adoption behaviour in the presence of network externalities, they arguably do not provide a robust test of theoretical models.

There are three aims to this chapter. First, to outline the sources and economic implications of network externalities for the diffusion process. Second, to outline the distinctive features of network externalities that pertain to ATM technology. Third, to empirically test for the existence of network externalities in the diffusion of ATMs in the UK financial sector.

The chapter is structured as follows. Section 7.2 provides an economic analysis of network externalities and their consequences for the diffusion process. Section 7.3 considers the sources of these externalities for the case of ATMs. Section 7.4 provides a theoretical model of ATM adoption in the presence of network externalities and outlines its empirical representation. Section 7.5 discusses the data set. Section 7.6 presents the empirical results and, finally, concluding remarks are gathered in Section 7.6.

7.2 Network Externalities

The seminal work of Katz and Shapiro (1985) define network technologies¹ as those for which the flow of benefits (or utility) that an adopter derives from adoption (or consumption) of the technology *increases* with the number of other economic agents adopting (or consuming) that technology. For network technologies, the presence of new adopters therefore affects directly and positively the utility function of every other adopter.

The presence of network externalities has significant implications for the demand structure pertaining to network technologies. Indeed, the definition of network technologies provided by Katz and Shapiro (1985) at first appears to be counterintuitive to economists because it raises the possibility that market demand for the technology will slope upwards with respect to its price. In formal theoretical models [see, for example, Economides (1993, 1996b) Economides and Himmelberg (1995)], however, it is the *expected* number of users (or adopters) that determines the value of adopting the network technology. Consequently, the demand curve for the technology remains downward sloping but shifts outwards with increases in the number of units expected to be sold (adopted). The resulting market demand curve is then defined as an ‘expectations fulfilled equilibrium’, along which expectations of the network size are fulfilled [Katz and Shapiro (1994)].

In general, the demand-side implications of network externalities are to produce a minimum, but significantly sized network that can be sustained in equilibrium. Consequently, the diffusion of network technologies over time may be characterised by discontinuous ‘jumps’ in the number of adopters. Moreover, perfect competition can be shown to be inefficient in the presence of network externalities [Economides (1996)]. Extensive discussion of the demand-side implications of network externalities lie outside the ambit of this chapter, but the nature of their demand structure is analysed more fully in Appendix One.

The networks literature has identified two main sources of positive network externalities [Katz and Shapiro (1985, 1986, 1994)]. First, externalities may be generated through a direct physical effect derived from the total number of adopters. This can most clearly be illustrated in the case of a telecommunications system. The utility that an adopter of a telephone derives, for example, depends on the number of other economic agents that join the telephone system. As noted by Rohlfs (1974) and Oren and Smith (1981), if a monopoly telephone network exists for which there are n adopters connected at time t , then there are a maximum number of $n(n-1)$ potential calls (or goods). If an additional agent adopts a telephone and joins the network then this provides direct (positive) externalities to all other users by adding $2n$ potential calls (or goods). These effects are then defined within the domain of consumption externalities [Kreps (1990)].

Second, there are indirect consumption and production externalities resulting from the so-called ‘software-hardware’ paradigm [Farrell and Saloner (1985), Katz and Shapiro (1994) and Farrell (1998)]. These effects pertain to those technologies that require complementary components from vertically related products. An example of such technology is computer technology that requires software to perform certain tasks. As shown by Church and Gandal (1992), in the presence of economies of scale in the production of software, the amount and variety of software that will be supplied will be an increasing function of the number of hardware units that have been sold.² Katz and Shapiro (1994) and Economides (1995) have labelled this software-hardware paradigm

¹ As implied in Section 7.1, the definition of ‘technology’ employed in the networks literature embraces both product and process innovations and product standards. This definition is retained in this chapter unless otherwise stated.

as a 'virtual network' because externalities are indirectly derived from the number of total adopters. In addition, indirect consumption externalities may derive from the quality and availability of post-adoption services being positively related to the number of adopters [Farrell and Saloner (1985)].

The ultimate source for these direct and indirect network externalities is the complementarity between the 'links' and 'nodes' of the network [Economides and Salop (1992) and Economides and White (1994)]. It is inherent in the structure of a network that many components are required to consume or adopt the final goods or services. This can most clearly be illustrated by continuing the telecommunications example. Figure 7.1 below shows the main features of a simplified 'star' or 'two-way' telephone network [Economides (1996)]. A call from agent, or node, 'A' to 'B', for example, is composed of two distinct links: access to the switch of 'A', 'AS' and access to the switch of 'B', 'BS'. Economides and White (1994) have argued that these links are actually complementary goods even though they ultimately have identical industrial classifications.

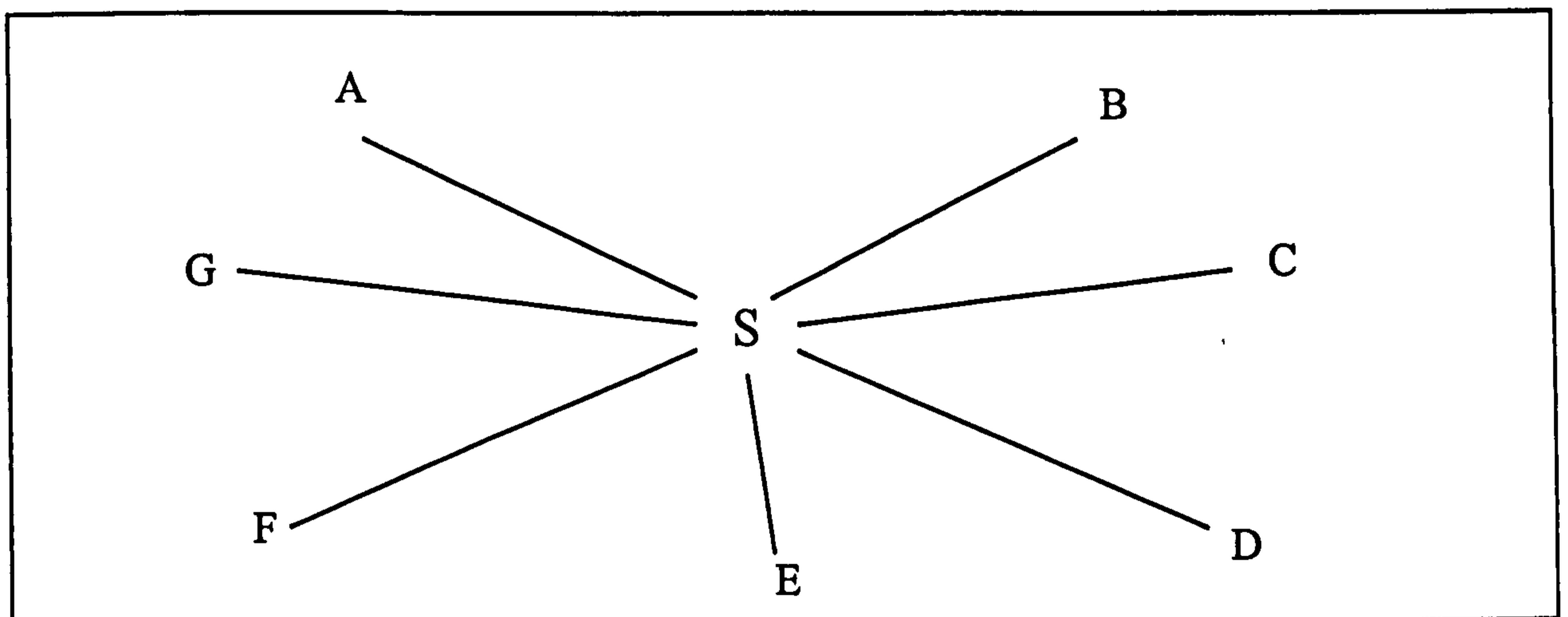


Figure 7.1 A Simple Star Network

When one of the links 'AB' or 'BA' are technically unfeasible, or is not provided because it is not profitable, then the network is labelled as being 'one-way'. Consequently, these 'one-way' networks are less likely to exhibit direct externality

² Despite the existence of these positive feedback effects it is typically assumed that the cycle of effects will not explode [see, for example, Agliardi (1995) and Economides (1996)].

effects [Katz and Shapiro (1994)]. In general, 'one-way' networks contain two types of components and only combining a component (not always in fixed proportions) of each type forms composite goods. Although separate components are complements, final composite goods will be substitutes for each other [Economides and Salop (1992)]. Katz and Shapiro (1994) have referred to the collection of complementary products that are combined (usually through an interface) to work together as a 'forming system'. As will be illustrated in Section 7.3, ATM networks adhere to this 'one-way' classification.

An important aspect of the network externalities paradigm is that compatibility between the links of the network is a pre-requisite for their complementarity [Farrell and Saloner (1986)]. For some network goods and services it may be the case that the links are inherently compatible. In general, however, this will not be the case and compatibility will only be achieved through adherence to specific technical compatibility standards or the development of adapters [Katz and Shapiro (1985)].³ The study of the economic incentives for providers of network related goods to make their products partially or fully compatible with the components produced by rival firms is an extremely important theme in the networks literature [see, for example, Farrell and Saloner (1985, 1986), Gandal (1995) and Matutes and Padilla (1995)]. Although these issues lie largely outside the ambit of this chapter reference will be made to the compatibility issue in the context of ATM networks in Section 7.3.

Discussion now turns to the implications of network externalities for the diffusion of new technology.

7.2.1 Network Externalities and the Diffusion of New Technology

The implications of network externalities for the diffusion of new technology have not, to date, been fully rationalised in the theoretical literature vis-à-vis issues surrounding the economic incentives for compatibility and standardisation and pricing strategies in markets that exhibit these externalities. Consequently, the implications for the diffusion process tend to be implicit rather than explicit. The majority of models that have been

³ See Economides (1995) for a case study into compatibility in the credit card market.

formulated conceptualise the diffusion process as involving the competition between two or more network technologies. Moreover, the adoption decision is most frequently modelled as the choice between an old vintage and a new vintage or a choice between two recently commercialised network technologies.

There are two theoretical approaches in the literature that examines the diffusion process in the presence of network externalities: the 'micro-approach' and the evolutionary approach [Economides (1996)]. The micro approach attempts to identify the underlying micro-structural reasons for network externalities (i.e. the sources of complementarity) and then attempts to model their consequences for the diffusion process. The underlying rationale of the micro-approach is neo-classical, characterised by rationality and maximising behaviour. In contrast, the evolutionary approach follows a similar methodology employed in the study of non-network technological diffusion (as outlined in Chapter 2) and focuses on behaviour under bounded rationality. The implications for the diffusion process tend to be more fully worked-out for the micro approach.

Before the discussion of these approaches it is important to address the proposition put forward by Metcalfe and Lissoni (1994) that the presence of positive network externalities substitutes the decreasing returns to adoption assumption, inherent in the stock and order effects models (see Chapter 2), with an increasing returns assumption. This proposition is arguably a gross simplification as it mis-interprets the concept of 'technology' employed in the networks literature. As noted in Section 7.2, the networks literature uses the term 'technology' to embrace *both* product and process innovations and product standards. In the presence of positive consumption externalities, the returns to agents consuming the final product network technology will be weakly increasing in the number of other adopters. It does not necessarily follow, however, that the returns to the process technology employed to produce these products will be increasing in the number of other (firm) adopters. There may, however, be indirect consequences for the diffusion of non-network process technologies in the presence of positive externalities on the product-side. Indeed, this aspect underlies the theoretical model of ATM adoption outlined in Section 7.4. Consequently, a careful distinction between those

externalities arising on the demand-side and those arising on the supply-side must be made.⁴

7.2.1.1 The Micro-Approach

Perfectly Competitive Models

An early paper analysing the diffusion process in the presence of network externalities is that of Cabral (1990). His model examines the diffusion of a non-durable product innovation with positive (consumption) externalities and assumes that both consumers and firms make production and consumption decisions based on expected sizes of the installed-base and on changes in its size overtime. He shows that under conditions of perfect competition and when network externalities are particularly strong, the diffusion path may be discontinuous not only at the start of the network (i.e. at the critical-mass level of adoption) but also at other points along the diffusion path. Economides (1996) has criticised Cabral's early contribution on the grounds that the presence of discontinuities in the diffusion path overtime implies an unrealistic infinite size of adoption at those points at which a discontinuity occurs.⁵

Economides and Himmelberg (1995) have extended the static network model of Katz and Shapiro (1985) into a dynamic setting with perfect competition on the supply-side. Positive network externalities are assumed to occur on the product-side with adoption of a durable good whose marginal costs of production fall exogenously overtime. Potential adopters are assumed to form (rational) expectations over the present value and future time paths of the expected network size, $n^e(t)$, and network good price, $p^e(t)$ and take these into account in calculating their optimal (utility maximising) date of adoption. The discontinuity problem encountered in Cabral's (1990) model is

⁴ Stoneman and Diederer (1994) have argued that positive externalities may exist for some technologies that require skilled labour, software, standardised inputs or a service network to operate efficiently. This conceptualisation of 'technology' is broader than the single-innovation concept conventional diffusion models employ and consistent with the evolutionary approach (see Chapter 2).

⁵ This argument is consistent with the Marshallian dictum *natura non facit saltum* (i.e. 'nature does not take a leap'). As noted by Sarkar (1998), however, it was Schumpeter (1934) who first highlighted the possibility of discontinuities in the process of technological change.

circumvented by assuming that supply of the network good is finite elastic.⁶ The subsequent dynamic diffusion path is then defined as a 'fulfilled expectations equilibrium path' along which the paths of $p(t)$ and $n(t)$ are such that expectations are fulfilled and supply is equal to demand at every point.⁷ This is shown to be smooth and continuous.

The implications for the diffusion process are embedded in Economides and Himmelberg's specification of the willingness-to-pay function, which encompass a general non-network diffusion model. The willingness-to-pay function is assumed to have the following functional form:

$$h(n^e) = k + \delta f(n^e) \quad (7.1)$$

where k gives the benefits of the technology in the absence of other adopters, δ is an indicator variable taking the value of unity if network-effects exist and zero otherwise and $f(n^e)$ measures the network-effect. It is further assumed that $f(0) = 0$, so that a network of size zero has no effect on the willingness-to-pay. Additionally, $f'(\cdot) > 0$ and $f''(\cdot) \leq 0$, so that network externalities are positive, but the marginal network externality is not increasing in network size.

From equation (7.1), if $\delta = 0$ there are no network externalities and diffusion proceeds purely via continuous falls in the price of the durable good analogous to the rank effects models outlined in Chapter 2. If, however, $\delta = 1$, so that externality effects are present, then Economides and Himmelberg show that the subsequent diffusion path is steeper and obtains a maximum level of adoption much faster compared to the diffusion path with no network externalities.⁸ This occurs because decreases in marginal cost over time not only makes adoption profitable for some firms but also has positive feedback effects via the network-effect which induces higher adoption. Their model does, however, suffer from two problems. Identification of network externalities and their

⁶ Economides and Himmelberg do not address the potential contradiction between a finite elastic supply curve and marginal cost pricing.

⁷ In this respect their model has similarities to David and Olsen's (1986) demand and supply rank model outlined in Chapter 2.

effects on the diffusion path depend heavily on *a priori* functional forms and the model has no closed form expression, which makes empirical implementation difficult.

As noted by Economides (1996), modelling the diffusion process in the presence of network externalities is more complex when departures from the assumption of perfect competition are made. Accordingly, analysis in this mould tends to be in the form of two-period game-theoretic models. Moreover, these models have been developed with more of a process innovation interpretation in mind rather than a product innovation.

Game Theoretic Approaches

Farrell and Saloner (1985) model the adoption a new process technology⁹ that exhibits positive network externalities as a sequential two-period game between two firms with incomplete information. Prices are assumed to exogenous and constant. All firms are assumed to initially employ the old technology prior to the start of the game and hold a varying 'preference parameter', k , formed on the new technology. The value of k is defined to lie between the closed interval $[0, 1]$ and is assumed to be uniformly distributed. Firms with a k value close to zero prefer the old technology, whilst those with k close to unity have a strong preference for the new technology. In addition, firms are assumed not to know with certainty the other firms' preference parameter.

Denoting the present value from using the old technology by $u_k(j)$, where $j = 1, 2$ is the network size and the present value (net of adoption costs) of switching to the new technology by $v_k(j)$, then positive network externalities are assumed by Farrell and Saloner to exist for both technologies if the following inequalities hold:

$$u_k(2) > u_k(1) \quad \text{and} \quad v_k(2) > v_k(1) \quad (7.2)$$

⁸ In terms of the epidemic model presented model in Chapter 2, the effect of network externalities is to increase the value of β in equation (2.4).

⁹ Farrell and Saloner's model can also be applied to a problem of compatibility between product standards.

where $u_k(\cdot)$ and $v_k(\cdot)$ are continuous and strictly increasing in k . An additional assumption is that those firms who hold a very strong preference for a particular technology prefer adopting that technology independently of the adoption choice that the other firm makes. This assumption implies the following inequalities:

$$v_1(1) > u_1(2) \quad \text{and} \quad u_0(1) > v_0(2) \quad (7.3)$$

Firms can switch to the new technology either in period 1 or 2, and switching is assumed irreversible. The 'willingness-to-adopt' is captured by the term $[v_k(2) - u_k(1)]$ and pay-offs are assumed to accrue at the end of period 2. Farrell and Saloner show that potential adopters fall into three categories¹⁰ according to the strategy they employ: never switch, whatever the behaviour of the other firm in the first-period, switch in period 2 if the other switches in period 1 (i.e. 'jump on the bandwagon') and switch in period 1 (i.e. 'start the bandwagon'). These strategies are summarised in Figure 7.2 below.

		Firm 2	
		New Technology 'N'	Old Technology 'O'
Firm 1	New Technology 'N'	$u_k(2)$	$u_k(1)$
	Old Technology 'O'	$v_k(1)$	$v_k(2)$

Figure 7.2 A Sequential Adoption Game

There are two critical values of k in this model, which determine the type of Nash equilibrium of the game and thus the nature of the diffusion path. The first is that value of k for which the willingness-to-adopt becomes positive. This condition is denoted by k^* . The second critical value for k is that k which defines indifference between the strategy of switching in period 2 if the other switches in period 1. This condition is denoted by \tilde{k} . As shown by Farrell and Saloner, these two critical value of k partition

¹⁰ There is potentially a fourth, that is: to switch in period 2 if the other firm has not switched. As noted by Beath *et al* (1995), however, this strategy is dominated by the third.

the parameter space in this model into three sets: firms who will never switch lie in the interval $[0, k^*)$, those who will switch unilaterally lie in the interval $[\tilde{k}, 1]$ and those firms who 'jump on the bandwagon' lie between $[k^*, \tilde{k})$.

The implications of this parameter space for the diffusion process are significant. If both firms have preference parameters marginally below \tilde{k} then both firms want to adopt the new technology since $v_k(2) > u_k(2)$, but in equilibrium they do not. Both firms are unwilling to start the network alone even though both are willing to 'jump on the bandwagon' if the network begins. Consequently, diffusion of the new technology will not begin. Farrell and Saloner refer to this equilibrium as 'excess-inertia' and argue that it can be overcome through communication among the firms or through co-ordination in committees.¹¹ Similarly, 'excess-momentum' is an equilibrium in which both firms adopt the new technology even though the sum of the firms' benefits is negative. This occurs when one of the firms favours the switch and, although the other opposes it strongly, the latter prefers switching to remaining alone with the old technology. Furthermore, there is also equilibrium if both firms are marginally above k^* for which $v(1) < 0$. These firms start the bandwagon rolling, but if it turns out that the other firm had a preference parameter below \tilde{k} (so that their lead is not followed), they regret their adoption decision *ex post*. All these three possible equilibria are inefficient as social benefits and private benefits diverge in every case.

The model of Farrell and Saloner (1985) is 'timeless' in the sense that the flow of benefits to adopters is determined only by *who* has adopted a new technology and not *when* the technology is adopted. This methodology ignores the possibility of first-mover advantages that are inherent in non-network stock and order diffusion models as outlined in Chapter 2. This aspect is addressed by Farrell and Saloner (1986) who retain the basic framework of Farrell and Saloner's (1985) model, but allow for potential adopters to arrive continuously over time. It is assumed that before time T^* only the old technology exists, but at T^* a new, superior, technology becomes available. Strategies available for potential adopters who arrive after T^* consist of a

¹¹ See Farrell and Saloner (1988) for mechanisms to achieve co-ordination.

choice between the old and new technologies. Network externalities pertaining to the new technology are assumed to be direct and linear in the number of adopters. These externalities are captured by the following equation¹²:

$$u(x) = a + b(x) \quad (7.4)$$

where x is the number of adopters at time t . The parameter a captures the network-independent effects that occur in a one-way network for example.

Nash equilibrium is then defined as that strategy which is optimal for each new firm given the strategies of existing firms. Farrell and Saloner show that a new technology is more likely to be successfully diffused the lower the size of the installed-base, the more quickly benefits of the new technology are realised and the relative superiority of the new technology. In particular, Farrell and Saloner indicate that even when information is complete, allowing for an initial installed-base can result in an equilibrium that represents excess-inertia or excess-momentum. These inefficiencies arise from two externalities which are absent from the model in Farrell and Saloner (1985). First, when an incoming firm switches to the new technology its rival loses some of the network benefits while they still use the old technology. These effects are ignored in the purely private adoption decision. Second, in the case where firms unanimously favour a switch, each adopter may still prefer the other to switch first. Consequently, switching may be delayed and the installed-base of the old technology can act as an 'entry barrier' to the diffusion of the new technology. Farrell and Saloner (1986) label this externality the 'penguin effect'.

Katz and Shapiro (1986) examine the adoption of two competing (and incompatible) network technologies in a two-period game in which the benefits accruing to homogenous adopters depends on the final network size. Adoption decisions are assumed to be based on a comparison between the benefits of joining an established network and the relative price difference between the old technology and the new, emerging, technology which becomes cheaper in the second period. Katz and Shapiro

¹² As adopters are assumed to arrive continuously, then the number of adopters, x , in equation (7.4) can be replaced by time, t .

show that the subsequent diffusion path depends crucially on whether there is free entry in the supply of technologies (an 'unsponsored' technology) or whether they are legally protected by patent law (an 'sponsored' technology). If both technologies are unsponsored and sell at their marginal cost, Katz and Shapiro show that there is market bias towards adopting that technology which is cheaper in period 1. From a social welfare perspective this may prove to be the 'wrong' choice. This possibility arises because firms adopting in period 1 ignore the negative effect of the non-matching on period 2 adopters. In the case of the new technology being sponsored then market adoption is biased towards that technology which has a net advantage (network benefits minus the relative price difference) in the second period. This is likely to be the sponsored technology because a monopoly supplier of this technology is able to sell it below marginal cost in period 1, build a large network, and then appropriate the returns to this penetration pricing by making second-period sales at a price above its marginal cost.¹³ Consequently, there may be 'over-adoption' of the 'wrong' technology from a welfare perspective.

Choi (1994) extends the model of Katz and Shapiro (1986) by assuming that potential adopters arrive sequentially, adoption decisions are irreversible and technology (assumed to be incompatible) evolves stochastically overtime. Choi's model is developed for the case of consumer durables and network externalities are assumed to occur on the product-side. Player 1 is assumed to maximise expected pay-offs and has to decide whether to delay adoption and wait and see how technologies evolve in period 2. Choi shows that the asymmetry in arrival times of potential adopters causes conflicts between different generations of adopters. There is a 'forward' externality in which early adoption of a technology deprives later adopters of an opportunity to co-ordinate efficiently based on better information. There is also a 'backward' externality, in which early adopters can be stranded inefficiently by later adopters who do not take earlier adopters' preferences into account. The forward externality induces period 1 players to adopt 'too early' compared with the social optimum¹⁴, whereas the backward externality deters them from adopting one of the available technologies through fear of being stranded. Choi shows that the forward externality dominates the backward externality

¹³ Katz and Shapiro assume that the monopoly supplier cannot credibly sell the technology at below marginal cost in the second period.

¹⁴ Social welfare is defined as the sum of the utilities of the two potential users.

and adoption of network technologies will be too early relative to the social optimum. Moreover, player 2 may be forced to adopt an inferior technology to ensure compatibility. Player 1 ignores this negative externality.

Chio and Thum (1998) show that the bias towards early adoption of network technologies is increased when the new technology is provided by a single producer with market power because any positive value created via exercising the option-to-wait can be appropriated by *ex post* by the monopolist. Choi and Thum argue that a monopolist can counter this by using licensing as a credible commitment mechanism not to expropriate future consumer surplus.

7.2.1.2 The Evolutionary Approach

The distinguishing difference between the evolutionary approach and to the diffusion of network technologies and the micro-approach (i.e. the neo-classical approach) is that the former approach has a much broader view of technology as a set of interrelated hardware pieces, software packages and human skills [Stoneman and Diedrern (1994)]. Network externalities remain positive in the evolutionary approach, but the definition of technology is not simply restricted to human artefacts (machines and material), but can also encompass organisational elements and localised technical progress [Lissoni and Metcalfe (1994) and David (1993)].

Despite the differences in the definition of technology between the evolutionary and micro-approaches, the evolutionary approach reaches similar conclusions concerning the diffusion of network technologies to the micro-approach. That is, the diffusion process may be path-dependent in the sense that the diffusion of a new (network) technology will depend on the production and adoption decisions pertaining to earlier vintages and older vintages may be 'locked-in' restricting the diffusion of new, superior, vintages [David (1993)]. The evolutionary approach, however, obtains these outcomes from a different modelling methodology than the micro-approach.

In general, evolutionary models are based on sophisticated applications of probability theory, the most popular being Arthur's *et al* (1987) 'density-dependent' Polya Urn scheme [Sarkar (1998)]. Polya Urn schemes are applicable to infinite populations; a feature that makes them suitable for the study of very broadly defined technologies [see, for example, David and Bunn (1987) and Cowan (1990, 1991)]. In Arthur (1988, 1989), for example, heterogeneous agents derive from the adoption of a specific network technology 'unconditional' and 'conditional' benefits, the first being independent of the number of other adopters, the second being an increasing function of it (this captures the positive network externalities).¹⁵ Uncertainty in Arthur's model derives from the assumption of adoption occurring in a sequential manner. At every time instant, t , only one agent is assumed to adopt analogous to agents being randomly selected from an urn. Uncertainty then corresponds to the preferences (or 'types') of agent that are extracted from the urn. If there are two technologies competing, A and B , then the repeated extraction of A -orientated agents at the beginning of the diffusion process cause the conditional benefits of adoption technology A to rise, so that an increasing number of later adopters will choose A irrespective of their unconditional benefits. Indeed, only one technology will eventually end up to dominate the market. Thus, the diffusion process may be driven into the 'gravitational orbit' of one of two possible outcomes: A gaining a monopoly or B gaining a monopoly. Which competing technology subsequently gets 'locked-in' therefore depends on the preferences of early adopters and the random history of adoption decisions.

In the evolutionary approach the preferences of early adopters may be determined by non-economic factors, such as minor historical accidents or cultural aspects [David (1993) and Sarkar (1998)]. Consequently, there is no *a priori* guarantee that the 'best' technologies (i.e. those with the greatest long-term development potentials) will diffuse.¹⁶ Rather, the diffusion path is determined by the short-term preferences of so-called 'diffusion agents' [Rogers (1995)] who may adopt inferior technologies. The diffusion path may therefore be characterised by 'potential inefficiency' and, thus, the main task of technology policy turns out to be preventing inferior technologies from

¹⁵ These assumptions are analogous to the relationship contained in equation (7.4).

¹⁶ Lane and Vescovni (1996) have shown that technology lock-in may occur in evolutionary models when there is also informational feedback from existing adopters, the basic mechanism being that an agent

becoming dominant at the beginning of the diffusion process [David (1987) and Sarkar (1998)].

As discussed by Witt (1997), the evolutionary approach may exaggerate the potentiality of technology lock-in. Witt argues that the results obtained by Arthur (1988) are heavily dependent on the, arguably, restrictive initial condition that competing technologies *A* and *B* enter simultaneously into a 'virgin-market' (i.e. a market that had not previously existed) and the assumption that there is an infinite number of adopters. Witt shows that if these assumptions are relaxed then lock-in is not the only possible equilibrium. Moreover, empirical testing of evolutionary models is extremely difficult. Detailed historical case studies on the diffusion of the QWERTY keyboard by David (1985) and the VHS video cassette recorder by Arthur (1989) are often put forward in support of the evolutionary approach [see, for example, Lissoni and Metcalfe (1994) and Sarkar (1998)]. Actual occurrence of technology lock-ins are, however, extremely rare and so evolutionary approaches tend to be counter-factual. As shown in Section 7.2.1, lock-ins may be explained by thoroughly neo-classical approaches without the restrictive assumptions identified in the evolutionary approach by Witt (1997).

7.3 ATM Technology and Network Externalities

The micro-approach to analysing network externalities relies on identification of the underlying micro-structural sources of complementarity inherent in network technologies. Accordingly, the micro-approach predicts that the sources of network externalities are likely to be technology-specific. This prediction is in contrast to the evolutionary approach, which embraces such a broad definition of technology as to make its applicability to an empirical case study of ATMs extremely limited. Consequently, it was decided to employ the micro-approach to analyse the existence, or otherwise, of network externalities in the diffusion of ATMs.

makes his choice of technology on the basis of private information obtained from sampling some previous adopters.

A pre-requisite to formal modelling is identification of the possible sources of network externalities. As stated in Section 7.2, the ultimate source of network externalities is the complementarity between the separate components (or 'links' and 'nodes') of the network the specific technology forms a part of. The literature has identified two (not mutually exclusive) potential sources of complementarity for the case of ATM technology. First, there is complementarity between the ATM hardware owned and operated by financial institutions and the debit card¹⁷ software held by the personal sector that enables agents to access their retail deposit accounts [Economides and Salop (1992) and Economides (1995)]. These two *compatible* components are then combined together to produce the range of transaction services outlined in Chapter 4. Second, there is complementarity between the retail deposits held by the personal sector and their holding of debit cards. These two components are combined via the ATM 'interface' which then allows retail deposits and debit cards to work together to produce transaction services [Katz and Shapiro (1994)]. The combination of these two components working together via an ATM can be described as the 'forming system' of the network [Katz and Shapiro (1994)]. Figure 7.3 below illustrates schematically how the various components involved in the provision of transaction services combine.

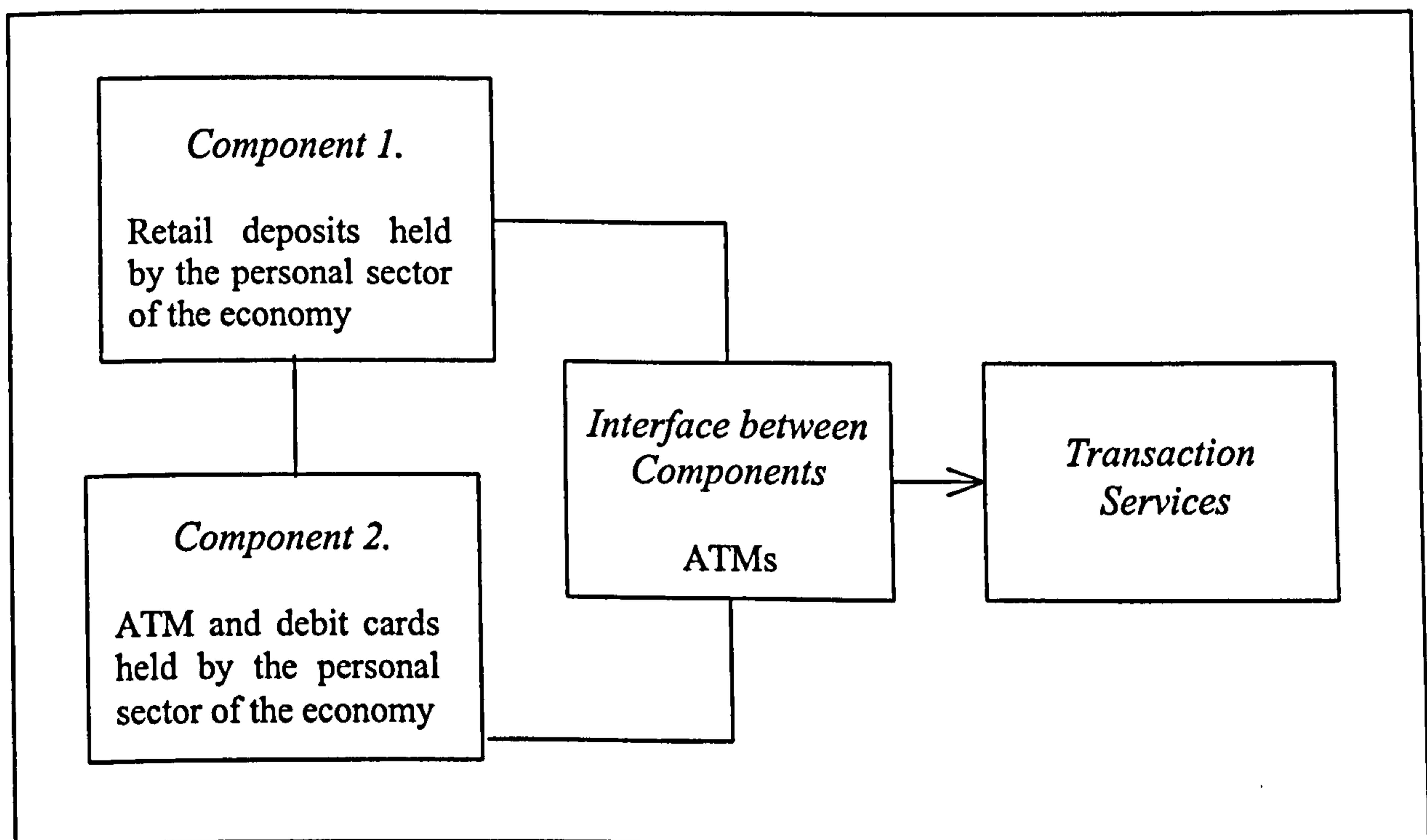


Figure 7.3 The ATM 'Forming System'

¹⁷ The term 'debit card' is used throughout this chapter to denote any card that can be used to access retail deposit accounts.

Identification of the sources of complementarity present in ATM technology is, however, only one-step in finding which link(s) or node(s) in the provision of transaction services by ATMs may represent the potential network externality and whether this externality originates on the demand-side or the supply-side. This requires closer inspection of how ATM networks are currently operated in the UK financial system.

Economides (1995, 1996a) has argued that ATM networks are inherently 'one-way' networks in the sense that not all potential connections between the various links and nodes of the network are demanded or produced. The rationale underlying this argument can be illustrated by examination of Figure 7.4 below, which represents a simplified one-way version of ATM networks currently operating in the UK [Howells and Hine (1993) and Retail Bank Research (1997)].

From Figure 7.4, a transaction (a cash withdrawal, for example) by an individual deposit holder from ATM A_i ($i = 1, \dots, 5$) from bank B_j ($j = 1, \dots, 4$) can be denoted by the sequence of connections $A_i S_A S_B B_j$ with direction going from A_i to B_j . Points S_A and S_B are network switches that route transaction messages between the ATM and the deposit holder's bank. Transactions such as $A_1 S_A A_2$ and $B_1 S_B B_2$, for example, although potentially technically feasible, are only viable if deposits are held either within ATMs themselves or intra-institutions fund transfers are completed via ATM networks. No financial system adheres to this set-up and therefore these connections are not observed in real-world ATM networks. Given the structure of connections in Figure 7.4, ATM networks can therefore be interpreted as being one-way networks.

The components of an ATM network summarised in Figure 7.4 illustrate a potential source of network externalities present in ATM technology. The deposit holder will benefit from a greater number of geographically dispersed ATMs (i.e. a greater number of A_i 's) because (as discussed in Chapter 4) they provide quicker and more convenient

access to their deposits relative to human-teller operations within branches.¹⁸ In addition, a larger and more geographically dispersed ATM network is likely to reduce expected transport costs when cash is demanded unexpectedly because deposit holders are not constrained to withdrawing cash from their own holding branch [Humphrey (1994)]. Moreover, Matutes and Padilla (1994) have argued that this positive 'transport' externality is likely to be larger the higher the proportion that those transactions that can be performed by ATMs are of total transactions because, in general, ATMs allow access for longer time periods vis-à-vis human-teller services.

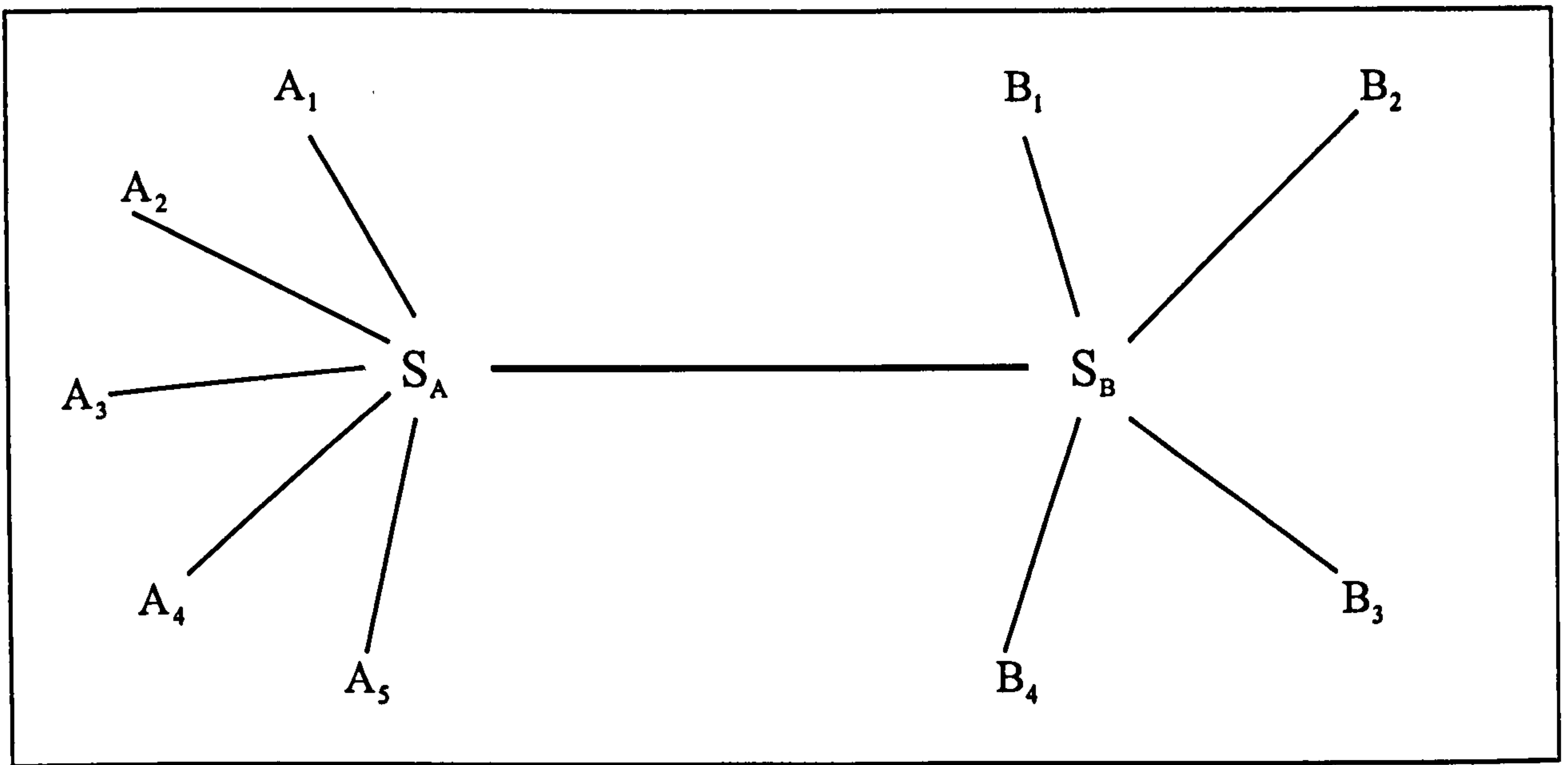


Figure 7.4 A Simplified ATM 'One-Way' Network

These positive externalities associated with a larger ATM network originate on the demand-side and are direct (i.e. physical) in nature because externalities increase purely in the total number ATM units the deposit holder can access. Although this direct externality occurs in other networks (in telecommunications, for example), the unique characteristic of ATM technology is that the direct externality increases in the number of physical hardware pieces rather than in the number of other users. Moreover, Saloner and Shepard (1995) have noted that the utility of an ATM user is likely to decrease in the number of other users as the increased usage of ATMs may increase waiting time

¹⁸ Survey evidence performed by Retail Bank Research (1997) supports this contention for the UK banking sector and for the US banking sector by The Tower Group (1997).

and thus diminish ATMs advantages over human-teller services.¹⁹ This aspect of ATM technology is in contrast to the theoretical contributions of Katz and Shapiro (1985, 1986) and Farrell and Saloner (1986) who model network externalities as being a positive function of the number of other users.

There may also exist network positive externalities in ATM technology that originate on the supply-side (i.e. those benefits appropriated by a financial institution through greater adoption by other institutions), but these are likely to be indirect and difficult to measure. Such externalities may include, for example, information learning-by-using spillovers from early adopters, improved after-sale mainframe servicing and improved software as the hardware market grows [Banking World (1992), Katz and Shapiro (1994) and The Tower Group (1997)]. Moreover, the empirical evidence presented in Chapter 6 indicates that order effects have been significant in the diffusion of ATMs in the UK suggesting that financial institutions adopted ATMs with the expectation that returns-to-adoption decreased in the number of other adopters. This result lends support to the contention that positive externalities on the supply-side are insignificant for the adoption decision.

Discussion now turns to the development of a theoretical model of ATM adoption that incorporates the demand-side externalities discussed in this section.

7.4 A Model of ATM Adoption with Network Effects

The aim of constructing the theoretical model is to examine the potential effects that network externalities have on the optimal timing of ATM adoption by financial institutions. The time dimension is a particularly important aspect of the model because it is necessary to implement the model empirically within a duration framework. The construction of the model follows the micro-approach to analysing network externalities [Katz and Shapiro (1985, 1986, 1994) and Farrell and Saloner (1985, 1986)] and focuses on the underlying micro-structural sources of externalities inherent in ATM

¹⁹ This argument does ignore, however, the potential of positive feedback effects between the greater use of ATMs, their profitability and their subsequent wider diffusion.

technology as outlined in Section 7.3. The basic assumptions of the model conform to those employed in the construction of the optimal-timing model presented in Chapter 6. They are as follows. First, ATM technology is assumed to be embodied in a specific capital good that is produced by a capital-producing industry that is then purchased by a capital-using industry (i.e. the financial sector). In addition, the definition of ATM technology embraces a second-generation definition (as outlined in Chapter 4) and it is further assumed that there is no risk or uncertainty pertaining to the economic or technical attributes of the technology. Consistent with Chapter 6 the structure of the market supplying ATM technology is not explicitly modelled in order to focus exclusively on those network externalities that originate in the market for transaction services.

Throughout the development of the theoretical model and its empirical implementation, it is assumed that the number of branches an individual institution operates measures the network externality at time t . This aspect of the empirical modelling is elaborated on in Section 7.4.1.2.²⁰ The role of expectations is, however, not incorporated into the theoretical model in order to focus exclusively on the role of network externalities on adoption timing.

The starting point in the development of the theoretical model is to assume that network externalities occur only on the demand-side for end-users (i.e. for individual card holders) and are increasing monotonically in the number of locations from which an end-user can access their retail account. It is further assumed that the benefits occurring to an end-user are independent of the number of other end-users. Following Katz and Shapiro (1985) and Farrell and Saloner (1986), an end-user's per-period benefits are assumed to be generated according to the following equation²¹:

$$u(N) = a + b(N) \tag{7.5}$$

²⁰ Note that the number of ATMs cannot be employed as a measure of externalities because the theoretical model is formulated with the principle aim of exploring how network externalities affect the adoption decision. Thus, the location of ATMs (proxied by the number of branches an institution has) is employed as the measure of network externalities in the empirical implementation of the model.

where $u(N)$ is a user's per-period benefit when the network consists of N locations from which the end-user can access and a represents the 'network independent' benefits. These network independent benefits are analogous to those outlined in Farrell and Saloner (1986) and pertain to the value that a user attaches to being able to only access an ATM at the location (i.e. the branch) the agent 'usually' uses for deposit related transactions. The presence of network externalities implies that $u'(N) > 0$ in equation (7.5) for the reasons outlined in Section 7.3. Although the sign of $u''(N)$ does not affect the major implications of the theoretical model, it is assumed for simplicity that network benefits increase linearly in N so that $u''(N) = 0$. In addition, it is assumed that $b(0) = 0$ so that a zero sized network displays no network externalities.

Equation (7.5) then defines the benefits of an ATM network for an individual end-user. The number of end-users for the i th institution is assumed to be constant and exogenously determined and denoted by n_i . The aggregate per-period value of the ATM network for institution i can then be denoted by the term $n_i[a + b(N_i)]$. Moreover, the narrative account of ATM diffusion in the UK as outlined in Chapter 4 indicated that the range of services (and quality) that ATMs provide have increased significantly since their commercialisation in 1972. To incorporate these technical changes into the model it is assumed that the aggregate network benefits from ATM technology are subject to a growth factor, θ , which increases the benefit flow to end-users and for which $\theta \geq 1$. Technical change is assumed to affect all ATMs equally. The flow of benefits that an individual end-user obtains from the ATM network during time period t and who holds deposits at institution i is therefore denoted by $n_i[a + b(N_i)]\theta^t$.

As noted by Economides (1996) the absence of supply-side externalities does not imply that product-side externalities will have no effect on the optimal timing of adoption. Indeed, if the revenue stream from adopting the new technology is dependent upon (or proportional to) the subsequent value attached to that technology by final users then

²¹ The definition of network externalities subsumed in equation (7.5) differs from the conventional measurement of network externalities as the willingness-to-pay [Economides (1996)]. The functional form of (7.5) does, however, conform to the conventional representation of network externalities.

demand-side externalities are likely to be significant for the adoption decision [Karshenas and Stoneman (1995)]. This is likely to be the case for ATM technology for two reasons. First, there is case study evidence to suggest that ATM technology has been adopted by financial institutions as a strategy of capturing higher shares of the retail deposit market by attracting more deposit holders [see, for example, Scarborough and Lannon (1988), Vesala (1994) and Llewellyn (1997)]. This strategy will only be successful if deposit holders value larger ATM networks from which to access their accounts, *ceteris paribus*. Second, the returns to investment in ATM technology are additionally dependent on the degree of utilisation by existing and new deposit holders [Kirkman (1987) and Haynes *et al* (1991)]. Returns to ATM technology for financial institutions are therefore likely to increase as more deposit holders switch from using human-tellers to withdraw cash to ATMs given the assumed cost advantages of ATMs. Consequently, deposit holders will be more likely to switch to using ATMs the larger number of ATMs they can access as this enables unforeseen demand for cash to be met [Matutes and Padilla (1994)].²²

The effect of demand-side externalities on the adoption decision is modelled by assuming that the per-period increase in revenues to the institution is proportional to the per-period benefits accruing to depositors. More succinctly, the institution's revenues are assumed to equal the per-period benefits to end-users multiplied by a constant factor λ for which $\lambda \leq 1$. The present value [Primrose (1991)] of institution i 's revenues evaluated at time T from adopting an ATM network at time T in discrete time are then given by the following term:

$$g_{iT} = \sum_{t=0}^{\infty} \lambda n_i [a + b(N_i)] r^t \theta^{T+t} \quad (7.6)$$

where g_{iT} denotes the present value of institution i 's revenue stream and r is the discount factor, which takes a value greater than zero. Note that in equation (7.6) the present value is increasing in both n_i and N_i . Equation (7.6) also incorporates the simplifying assumptions that institution i adopts ATMs at all N locations at time T and

²² Humphrey (1994) finds, however, that for US financial institutions that the profitability of adopting ATMs may actually fall beyond a certain number of transactions.

that other ATM networks are incompatible with each other. Therefore in the model N represents only the institution's own ATM locations. The consistency of this assumption with the current UK ATM network operations and its implications for empirical estimation are discussed in Section 7.4.1.2.

As noted in Chapter 6, a necessary, but not sufficient, condition for optimal adoption timing is the profitability condition. This condition states that a profit maximising institution will adopt at the first time, t , at which the net present value (NPV) becomes positive [Karshenas and Stoneman (1994, 1995)]. Therefore, to convert equation (7.6) to an NPV and to derive a term for the optimal time to adopt requires consideration of the cost dimension of ATM adoption. In Chapter 6 this cost dimension is captured by the time-varying 'hedonic' (i.e. quality-adjusted) price of adopting one unit of ATM technology. This is not, however, applicable to an exploration of adoption that explicitly incorporates network externalities. This is because changes in the number of end-users, n , and the number of ATM locations, N , may have significant implications for the cost structure of ATM adoption. An increase in n , for example, will reduce the average fixed costs of adopting an ATM network as costs are 'spread' over a greater number of end-users and therefore transaction volumes. This cost-side effect is ignored in the optimal-timing model presented in Chapter 6 because the number of end-users is not incorporated into the institution's adoption decision (although the number of other adopters is). This cost-side dimension cannot, therefore, be ignored in a model that incorporates demand-side network externalities.

In order to capture the potential effects of the number of end users (n) and the number of locations (N) on the cost of ATM adoption a distinction is made between the (short-run) variable and fixed costs associated with ATM adoption. The variable cost component of ATM adoption is assumed to consist of costs incurred with each transaction (i.e. the output measure conventionally employed for ATM technology) and embraces, for example, the cost of issuing paper receipts after an transaction has been completed [Haynes *et al* (1991) and Humphrey (1994)]. Assuming that each depositor makes the same number of transactions, then variable costs will be proportional to the number of depositors. Following Saloner and Shepard (1995) it is assumed that variable costs are incorporated into the parameter λ so that the expression

$\lambda n[a + b(N)]g'$ represents the variable profit function [Krepps (1994)] for an adopting institution in period t .

The fixed cost element of ATM adoption encompasses, for example, the alteration costs to branches required for the accommodation of ATMs, the costs of establishing ATM packet-switches, the cost of purchasing or leasing ATMs themselves and after-sales servicing costs. As noted by Saloner and Shepard (1995), total fixed costs can further be divided into two separate components. The first depends positively on the number of ATM locations, N , and embraces, for example, the costs of installation and serving. The second element is so-called 'system-costs' and is independent of the value of N . System-costs include, for example, the marketing costs associated with promoting an institution's network.

The present value of the cost of adopting an ATM network of size N_i for institution i at time T can then be denoted by the following linear cost function:

$$C_i(N_i, T) = S(T) + N_i c(T) \quad (7.7)$$

where $S(T)$ and $c(T)$ denote system-costs and cost-per-location respectively.

The inter-firm diffusion process in this model then proceeds by a similar mechanism to the theoretical model of optimal adoption timing model outlined in Chapter 6. That is, the fixed cost of adoption, $C_i(N_i, T)$, is assumed to fall monotonically (at a decreasing rate) over time due to a process of positive industry-wide learning-by-doing effects in the supplying industry analogous to Arrow (1962b).²³ As the fixed cost of adoption declines, those marginal institutions that previously found ATM adoption to be unprofitable find adoption profitable and adopt.

The NPV of institution i 's gross profit from adoption of an ATM network at time T in discrete time is then given by the following expression:

²³ These supply-side effects are not explicitly modelled in this chapter.

$$G_{iT} = \sum_{i=0}^{\infty} \lambda n_i [a + b(N_i)] r^i g^{T+i} - C_i(N_i, T) \quad (7.8)$$

where G_{iT} is the NPV for institution i .

As noted in Chapter 6 the optimal adoption time (i.e. the profit-maximising adoption time), t^* , for an institution after commercialisation of ATM technology depends on two necessary and sufficient conditions being met. The first is that adoption is profitable at t^* , which implies that $G_{iT} \geq 0$ obtains in equation (7.8). The second condition is the so-called arbitrage condition, which requires that the net benefits from adoption are not increasing in time at t^* . Following Saloner and Shepard (1995), the arbitrage condition may be expressed in discrete time as follows. Institution i will with n_i depositors and N_i locations earns higher profits from adopting at time T than from waiting until time $T+1$ if:

$$\frac{\lambda \theta^T n_i [a + b(N_i)]}{1 - r\theta} - C(N_i, T) > r \left\{ \frac{\lambda \theta^{T+1} n_i [a + b(N_i)]}{1 - r\theta} - C(N_i, T+1) \right\} \quad (7.9)$$

re-arranging (7.9) obtains the following:

$$\lambda n_i [a + b(N_i)] \theta^T > C(N_i, T) - rC(N_i, T+1) \quad (7.10)$$

The inequality condition in equation (7.10) may then be interpreted as implying that an institution will adopt at that time t when pre-period variable profits exceed the cost savings of waiting an additional time period.

Saloner and Shepard (1995) are not explicit in their assumptions concerning the mathematical characteristics of the NPV function contained in equation (7.8). Using the formal proofs in Karshenas and Stoneman (1994, 1995), however, the conditions for an optimal adoption time to exist (i.e. there exists a t_i such that $t_i^* < \infty$) may be stated for this model. These are that the per-period present value in equation (7.6) is bounded

from above at \bar{G} and that there is a lower bound for the cost of technology in equation (7.7) given by \underline{C} . These two conditions then imply that there is an upper bound on the NPV in equation (7.8). Assuming that all institutions are potential adopter's (i.e. $G_{it} \geq 0$ for $\forall i$) it then follows that G_{it} will achieve its maximum at some $t < \infty$.²⁴ It then follows from these assumptions that the smallest T that satisfies equation (7.9) is the optimal time to adopt.

As noted by Saloner and Shepard (1995) there are two special cases of the general adoption rule contained in equation (7.10). The first occurs when $C(N_i, T)$ is constant over time (i.e. $\partial C(N_i, T)/\partial T = 0$) so that technical change, θ , is the only non-constant factor that determines the optimal time of adoption. In this case equation (7.10) reduces to the following:

$$\frac{\lambda n_i [a + b(N_i)] \theta^T}{1 - r} > C(N_i) \quad (7.11)$$

Equation (7.11) can be interpreted as specifying that institution i will adopt at that t at which the NPV of variable profits exceeds the (constant) cost of adoption. Since the cost of adoption does not decline over time there is no cost advantage to waiting.

The second special case of equation (7.10) occurs when there is an absent of technical change (i.e. $\theta = 1$) and the only non-constant factor determining the optimal adoption time is the declining cost of the technology. In this case equation (7.10) becomes the following:

$$\lambda n_i [a + b(N_i)] > C(N_i, T) - rC(N_i, T + 1) \quad (7.12)$$

The right-hand side of (7.12) represents the cost saving from delaying adoption by one period. The institution will then adopt at that time t at which the cost savings are less than the (constant) per-period variable profit.

²⁴ This also implies that the arbitrage condition dominates the profitability condition.

An important aspect of the model is that it predicts the existence of production-side economies of scale in the adoption of ATM technology. In the general case of optimal adoption timing given in equation (7.10), the number of end-users, n , enters on the left-hand side only. That is, an increase in n increases the benefit flow accompanying the adoption of ATMs. Dividing equation (7.10) by n then shows that the institution's per-period per-depositor variable profits increase only due to technical change, but that total costs per-depositor decline in n . This characteristic illustrates that the existence of production-side scale economies leads to adoption being earlier the larger the number of depositors, *ceteris paribus*.²⁵

As discussed in Saloner and Shepard (1995), the cost-side effects present in this theoretical model have important implications for the empirical implementation of the effect of network externalities on the optimal time to adopt.²⁶ A simple empirical test of the significance of demand-side externalities for the adoption of ATMs would, for example, involve varying the number of ATM locations, N , while holding the number of depositors, n , constant in equation (7.10). This approach can, however, lead to an ambiguous result because of the cost-side effects that n has. From equation (7.10) an increase in N increases the left-hand side of (7.10) because of the existence of positive externalities, but this increase in N additionally decreases the right-hand side via the effect on the cost-per-location, $Nc(T)$, in equation (7.7). The overall effect on the timing of adoption then depends on the relative weight of these two effects. The aggregate effect cannot be determined *a priori* and, thus, the effect on optimal adoption timing is ambiguous. An empirical test that holds n constant while varying N will therefore tend to underestimate the effect of network externalities on adoption timing.

An alternative approach is to hold the number of depositors per location constant while increasing the number of locations. This approach can be interpreted as holding the ratio $n/N \equiv \nu$ constant. This approach, however, tends to overestimate the effects of network externalities on the timing of adoption. This can be illustrated by dividing

²⁵ The contention that ATM technology has economies of scale is supported by the case study evidence presented in Chapter 4 and by the empirical finding in Chapter 6 that institution size (measured by total assets) has a positive and statistically significant effect on the conditional probability of adoption.

²⁶ The subscript ' i ' is ignored in the following discussion.

equation (7.10) by n and separating total fixed costs into its constituent components, system-costs and cost-per-location. This gives the following expression:

$$\lambda[a + b(N)]\theta^T > \frac{[S(T) - S(T+1)]}{n} + \frac{[c(T) - rc(T+1)]}{v} \quad (7.13)$$

Observation of equation (7.13) shows that increasing N by, for example, ΔN , while holding v constant requires adding an extra $v(N + \Delta N) - n$ depositors to the network. The first term on the left-hand side of equation (7.13) involving $S(\cdot)$ is decreasing in n because system-costs are 'spread' over a larger number of depositors. Consequently, if empirical results show that institutions with more locations (holding v constant) adopt earlier this result could therefore be the outcome of strong increasing returns to system-costs rather than an indication of the existence of positive network externalities.

The implications of these cost-side effects for the empirical exploration of the adoption of ATMs can be illustrated by Figure 7.5 below. The figure is constructed as follows. The left-hand side of equation (7.13) is denoted by the series of curves $B(T, N)$ and the right-hand side is denoted by the set of curves $A(T, N, n)$, where N is the number of branches and n the number of depositors an institution has respectively. The set of $A(\cdot)$ curves shows how the benefits from adopting ATM network changes over time for a given network size N . Given the underlying assumptions of the theoretical model, the $A(\cdot)$ curve is a (monotonically) increasing convex function of time T . The curve $A(\cdot)$ shifts upward for an increase in the number of branches, N , because demand-side externalities are positive. In contrast, the curve $B(\cdot)$ represents the change in the pre-period, per-depositor cost of ATM adoption at time T for a given network size. Given the assumptions of the theoretical model the $B(\cdot)$ curve will shift downwards for an increase in the number of depositors, n , as system-costs are 'spread' over a larger number of depositors. The curve will, however, increase in the number of branches, N (given $v \equiv n/N$), as the cost-per-location of having additional branches increases. The profit-maximising ATM adopter will then adopt at that time t at which equation (7.13) pertains to equality. This implies that an optimal adoption time in Figure 7.5 is depicted

by that t where the curves $A(\cdot)$ and $B(\cdot)$ intersect for a given number of locations and depositors.

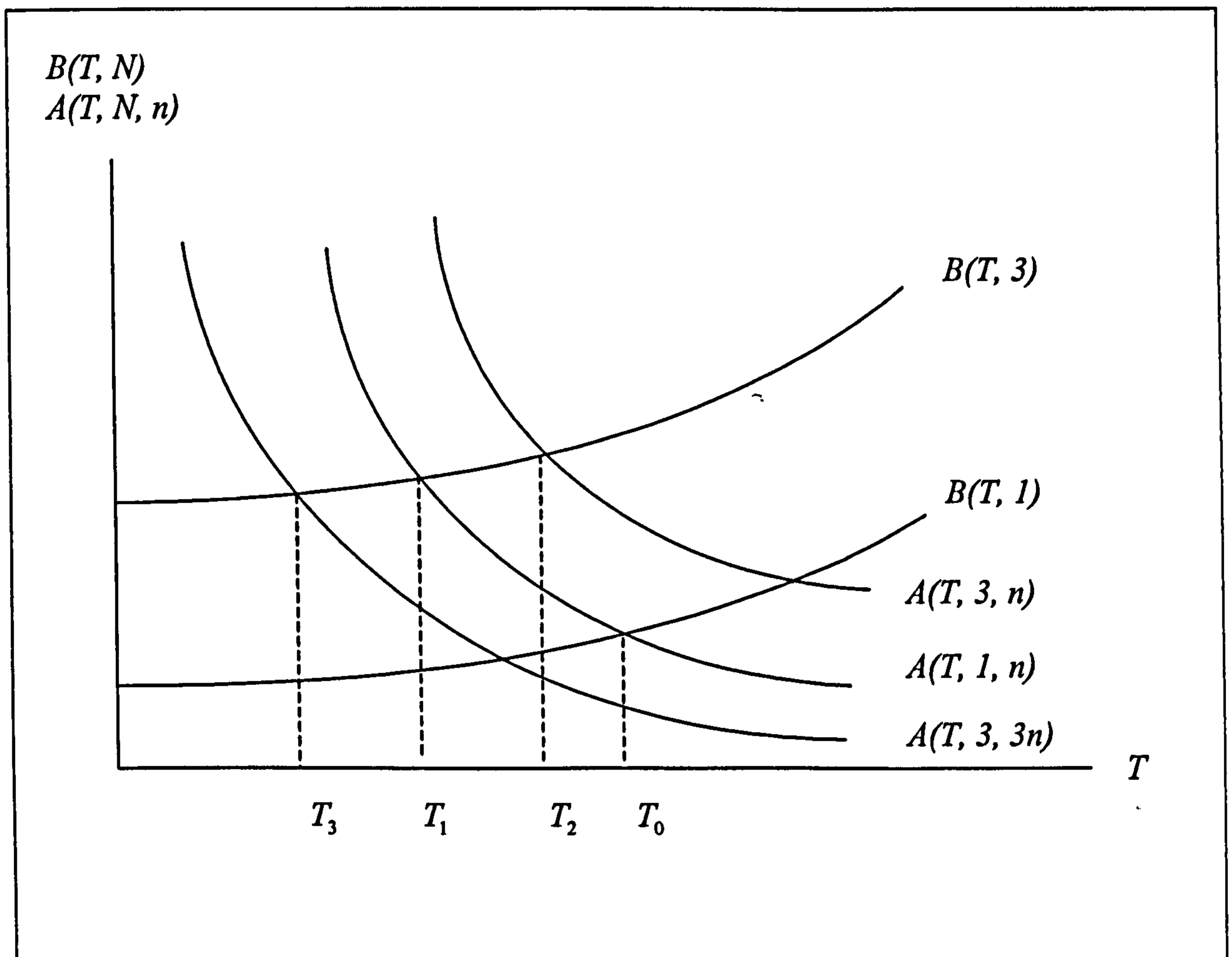


Figure 7.5 Separating Network Externalities and Cost-Side Effects

A precise measure of the effects of network externalities on the optimal timing of adoption is obtained by exploring the effects of increasing network size while holding constant the change in the per-depositor cost of adoption. This can be illustrated by reference to Figure 7.5. An institution with one branch and n depositors will adopt at the profit maximising time T_0 where the curves $A(T, 1, n)$ and $B(T, 1)$ intersect. An institution with an identical per-depositor cost of adoption, but with three branches rather than one, for example, will adopt at T_1 where the curves $A(T, 1, n)$ and $B(T, 3)$ intersect. The difference between T_0 and T_1 gives the marginal effects on the timing of adoption by adding two additional branches to an institution. Adding two additional branches, however, increases the right-hand of equation (7.13) via the increase in ν and

shifts the curve $A(T, 1, n)$ to $A(T, 3, n)$. Holding n constant and varying N therefore leads to the marginal effect of increasing N to be measured by the time difference $T_0 - T_2$. As $(T_0 - T_2)$ is greater than $(T_0 - T_1)$ the approach of holding n constant leads to an underestimation of the effect of network externalities on optimal adoption timing. A measure of this underestimation is given by $(T_2 - T_1)$. In contrast, if the number of depositors per location, v , is held constant then the curve $A(T, 1, n)$ shifts downward to $A(T, 3, 3n)$.²⁷ The optimal time to adopt then becomes T_3 , the point at which $A(T, 3, 3n)$ intersects $B(T, 3)$. Since the value $(T_0 - T_3)$ is less than $(T_0 - T_1)$ the effect of varying N and holding v constant is to overestimate the effect of network externalities on the optimal timing of adoption. A measure of this overestimation is given by the value of $(T_1 - T_2)$. Overall, the estimation of the effects of network externalities is closer to the upper-bound estimate (denoted by T_2) the higher are location-costs as a proportion of total fixed costs. This is because from equation (7.13) only system-costs decrease by the movement from curve $A(T, 1, n)$ to curve $A(T, 3, 3n)$.

To summarise, examining the propensity to adopt in the presence of network externalities by holding n constant understates the impact of network externalities, while holding n/N constant overstates their effect. In the empirical results presented in Section 7.6 these lower and upper bounds predicted by the theoretical model are estimated directly from the relevant partial derivatives of the estimated model.

7.4.1 Empirical Representation of the Theoretical Model

The expression contained in (7.13) yields the optimal date of adoption, t_i^* , for institution i under the assumption that institutions with identical number of depositors and ATM locations attach the same valuation to an ATM network. As shown in Chapter 6, however, there are likely to a wide range of institution-specific characteristics other than the number of depositors and branches an institution has that

²⁷ Since n appears in the denominator of the two right-hand terms in equation (7.13), multiplying n and N by the same amount implies that the curve $A(T, 3, 3n)$ lies below $A(T, 1, n)$.

determines the optimal date of adoption. To take account of heterogeneity between institutions the methodology employed by Saloner and Shepard (1995) in deriving the conditional probability of adoption in this model is followed. Heterogeneity is assumed to be measured by the following relationship:

$$\varepsilon_i = \bar{\Psi}(n, N) - \Psi_i(n, N) \quad (7.14)$$

where $\Psi_i(n, N)$ denotes the per-period profits of the i th institution with n depositors and N locations and $\bar{\Psi}(n, N)$ denotes the mean per-period profits of all potential adopters with an identical number of depositors and locations. From equation (7.14) the larger ε_i is the lower the institutions propensity to adopt is relative to the mean institution. The NPV of the i th institution's profit from adopting at time T is then:

$$\Psi_i = \frac{\lambda \theta^T n [a + b(N)]}{1 - r\theta} - C(N, T) - \frac{\varepsilon_i}{1 - r} \quad (7.15)$$

Then substituting (7.15) into (7.13) obtains the following adoption rule:

$$\varepsilon_i < n\lambda[a + b(N)]\theta^T - [C(N, T) - rC(N, T + 1)] \quad (7.16)$$

Equation (7.16) then predicts that those institutions with large net benefits relative to the mean (i.e. low values of ε_i) adopt relatively early while those institutions with relatively smaller net benefits relative to the mean (i.e. high values of ε_i) adopt late. The smallest $T_i = T(n, N, \varepsilon_i)$ that satisfies equation (7.16) is then the optimal date for adoption for the i th institution [Saloner and Shepard (1995)]. The diffusion process in this model is then similar to a rank-effects one encountered in Chapter 2. Consequently, the shape of the inter-firm diffusion curve will depend on how the cost and benefits of adoption change over time and on the distribution of ε_i . Although the exact shape of the inter-firm diffusion curve lies outside the ambit of this chapter, the mechanism at work can be seen from using equation (7.16) to define:

$$\varepsilon^*(n, N, T) \equiv n\lambda[a + b(N)]\theta^T - [C(N, T) - rC(N, T + 1)] \quad (7.17)$$

where $\varepsilon^*(n, N, T)$ is the ε of an institution with n depositors and N locations that is just indifferent between adopting and not adopting an ATM network at time T . The hazard function of this model can then be derived for this model as the conditional probability that institution i will adopt in the next short time period dT given that it has not adopted at time T [Lancaster (1990)]. The conditional probability is then given by the following:

$$h(T) = \frac{F[\varepsilon^*(n, N, T + 1)] - F[\varepsilon^*(n, N, T)]}{1 - F[\varepsilon^*(n, N, T)]} \quad (7.18)$$

where $h(T)$ is the hazard function and $F(\cdot)$ is the cumulative distribution function (CDF). The behaviour of $h(T)$ overtime will then depend on the distribution of $F(\cdot)$ and the rate at which $\varepsilon^*(\cdot)$ changes over time.

Given the assumptions of the theoretical model the reduced-form hazard function of the theoretical model for institution i may be stated as follows:

$$h_i(T) = J(n_i, N_i, R_i) \quad (7.19)$$

where R_i is a vector of other institution-specific characteristics. From the results obtained from the derivation of the theoretical model the following *a priori* parameter restrictions apply: $J_1 > 0$, $J_2 > 0$ and $J_3 \geq 0$. The sign of J_1 captures the effects of network externalities on the conditional probability of adoption. Equation (7.19) then forms the basis of the empirical work.

7.4.1.1 Econometric Implementation

The theoretical model of ATM adoption outlined in Section 7.4 conjectures that the marginal effect of network externalities on ATM adoption can only be estimated to lie between a lower and upper bound. The difficulty of obtaining a precise estimate for the effects of network externalities on adoption timing derives from the cost-side effects of increasing the number of ATM locations an institution has. The lower bound is defined by the marginal effect on adoption timing when the number of ATM locations (proxied by the number branches) is varied while holding all other institution-specific variables constant. In contrast, the upper bound is defined as the marginal effect on adoption timing from varying the number of ATM locations an institution has, but holding constant the ratio of the number of depositors to the number of locations.

Given the prominence that the theoretical model gives to these upper and lower bounds it was decided to employ a different approach to econometric modelling in this chapter than the approach used in Chapter 6. Recall that the aim of Chapter 6 was to explore the statistical significance or otherwise of a set of institution-specific and market-specific covariates on the conditional probability of adoption. The model chosen for estimation in Chapter 6 was the accelerated lifetime model [Kalbfleisch and Prentice (1980) and Neumann (1997)] because of its ability to incorporate time-varying covariates and its intuitive appeal deriving from the assumption that covariates re-scale the time to adoption. Given the more sophisticated structure of the accelerated lifetime model vis-à-vis the proportional hazards model, however, estimated covariates in the accelerated lifetime model cannot be interpreted as a marginal effect analogous to estimated coefficients in linear OLS models [Kiefer (1988)]. This aspect of the accelerated lifetime model therefore limits its application to an exploration of the effects of network externalities on adoption timing.

Given the limitations of the accelerated lifetime model two alternative approaches to econometric modelling are employed in this chapter that allow the calculation of the lower and upper bounds predicted by the theoretical model. The first approach involves an extension of the non-parametric proportional hazards model proposed by Cox (1972, 1975) and by Cox and Oakes (1984). The extension involves relaxing the conventional

assumption that the baseline hazard is left arbitrary, but instead specifies a particular parametric form for the baseline hazard. Recall from Appendix One of Chapter 6 that the proportional hazards model assumes that the interaction between the duration of non-adoption, t , and a set of time-invariant covariates, X , is a multiplicative one. The continuous time model then specifies that for the i th institution that the hazard function takes the following form:

$$h_i(t|X) = h_0(t) \exp(X_i \beta) \quad (7.20)$$

where X is a $K \times 1$ vector of time-invariant covariates, β is a $1 \times K$ vector of covariate parameters to be estimated, $h_0(t)$ is the baseline hazard function and conforms to the condition $X_i = 0$ and $h_i(t|X)$ is the hazard function for institution i conditional on the vector of covariates [Neumann (1997)]. The covariates are embodied in the link function $\exp(\cdot)$, which ensures non-negativity of the hazard function.

Two parametric forms of the baseline hazard are selected in the empirical implementation of (7.20): the Weibull distribution and the log-logistic distribution [see Greene (1993)]. Although the empirical evidence presented in Chapter 5 indicates that the Weibull distribution is a relatively poor representation of the non-parametric hazard function it is selected for empirical implementation in this chapter because estimable models with lognormal baseline hazard functions are not yet available in econometric software packages. The inclusion of empirical results that assume a Weibull baseline hazard function therefore serves as a comparative exercise and act as an informal test of the sensitivity of the empirical results and parameter stability to different functional forms. A summary of the specification and properties of the Weibull and log-logistic distributions are provided in Appendix One of Chapter 5 and as they are identical to those employed in this chapter they are not repeated here.

The vector of covariates, X , is then assumed to enter the link function $\exp(X\beta)$ in equation (7.20) linearly for all institutions such that the following relationship holds:

$$h_i(t|X) = h_0(t) \exp(\beta_{i0} + \beta_{i1}X_{i1} + \beta_{i2}X_{i2} + \dots + \beta_{iK}X_{iK}) \quad (7.21)$$

The assumption that the vector of covariates enters the link function linearly is a conventional one employed in the empirical diffusion literature [see, for example, Levin *et al* (1987), Rose and Joskow (1990) and Karshenas and Stoneman (1994)] and is the one employed in Chapter 6.

Taking (natural) logs of equation (7.21) and assuming that the baseline hazard, $h_0(t)$, is a Weibull distribution obtains the following specification:

$$\ln h_i(t|X) = \ln(\lambda p) + (p-1)\ln(\lambda t) + X_i\beta \quad (7.22)$$

where λ is a scale parameter and p is a shape parameter whose value determines duration dependence (see Appendix One of Chapter 5).

In contrast, when the baseline hazard is a log-logistic distribution the log of the hazard function is equation (7.20) obtains the following form:

$$\ln h_i(t|X) = (p-1)\ln[\lambda p(\lambda t)] - \ln[1 + (\lambda t)^p] + X_i\beta \quad (7.23)$$

where λ and p have an identical interpretation to those contained in the Weibull distribution.

From equation (7.22) and (7.23) and using the linear relationship in equation (7.21), the one-period marginal effect on $\ln h_i(t|X)$ with respect to a change in X_{iK} can be obtained by calculating the partial derivative $\partial \ln h_i(t|X) / \partial X_{iK}$. From equation (7.21) this partial derivative is equal to β_K . Thus, estimated coefficients in the extended proportional hazards model have a marginal effect interpretation. This characteristic of the proportional hazards model is used in the procedure to calculate the marginal effects of increasing the number of branches on the conditional probability of adoption in the empirical results presented in Section 7.6. If any of the terms in X_i contain higher-power terms or ratio's then convention is followed [Hannan and McDowell (1984b,

1987) and Saloner and Shepard (1995)] and marginal effects are evaluated at the sample means for that covariate for the observations included in the associated regression.

The proportional hazards model is estimated using time-invariant covariates, as models that incorporate time-varying covariates are, to date, not available in estimable form. This implies that all covariates are necessarily exogenous as defined in Appendix One of Chapter 6. The log-likelihood function for the proportional hazards model is then defined as follows:

$$\ln L(\beta) = \sum_{i=1}^n \delta_i \ln f(t|X_i, \beta) - \sum_{i=1}^n (1 - \delta_i) S(t|X_i, \beta) \quad (7.24)$$

where n is the number of institutions in the sample and δ_i is an indicator variable that takes a value of unity for non-censored observations (i.e. for those institutions that adopt on or before the end of the sample period) and zero for right-censored observations (i.e. for those institutions that fail to adopt on or before the end of the sample period), $f(\cdot)$ is the PDF and $S(\cdot)$ is the survivor function. Censoring is assumed to be independent as defined in Chapter 5. Estimation of the vector β then proceeds by maximising the log likelihood. Kiefer (1988) has shown that the log-likelihood function in (7.24) is concave and so numerical maximisation²⁸ can be used to obtain consistent estimators of the β vector [Cuthbertson *et al* (1992)].

The second approach employed in empirical estimation is the exponential regression model [Kalbfleisch and Prentice (1980) and Kiefer (1988)] as used by Hannan and McDowell (1984b, 1987) in their study of ATM diffusion in the US banking sector. This approach assumes that the baseline hazard in equation (7.20) is constant and is equal to unity so that the conventional proportional hazards models represented in equation in (7.20) reduces to the following specification:

$$h_i(t|X) = \exp(X_i \beta) \quad (7.25)$$

²⁸ The Newton-Raphson method is employed in this chapter to obtain the empirical results.

where all variables have identical definition and interpretation as those in the proportional hazards model. In contrast to the extended proportional hazards model the exponential regression model can be estimated using time-varying covariates. Indeed, given that the baseline hazard is assumed to be time-invariant and equal to unity then variations in the hazard function over time can be captured by the inclusion of time-varying covariates [Saloner and Shepard (1995) and Neumann (1997)]. Incorporating time-varying covariates into the exponential regression model in (7.25) results in the following formulation of the exponential regression model:

$$h_i[t|X(t)] = \exp[X_i(t)\beta] \quad (7.26)$$

The marginal effects of varying covariates is calculated using a different procedure in the exponential regression model than that used in the proportional hazards model because of the time-varying nature of the covariates. There are in fact two distinct marginal effects present in the exponential regression model. The first pertains to the derivative of the one-period conditional probability of adoption with respect to any given X_k . As shown by Hannan and McDowell (1984b, 1987) this is calculated as (ignoring the subscript i) the value of $\exp(X\beta)\beta_k$. The term $\exp(X\beta)$ is then measured at the mean values of the covariates. The second marginal effect relates to the change in the conditional probability of adoption at the end of the sample period with respect to a change in X_k . This conditional probability can be calculated using the relationship between the survivor function and the hazard function presented in Chapter 5. Recall from Chapter 5 that the survivor function gives the conditional of adoption at the end of period T and may be expressed as follows:

$$S(T) = \exp\left[-\int_0^T h(s) ds\right] \quad (7.27)$$

Using the relationship between the survivor function and the hazard function in (7.27) and the specification of the exponential model in (7.26), the partial derivative of $X(t)$ with respect to any X_k affecting all T periods and evaluated at mean values, \bar{X} , may then be expressed in discrete time as follows:

$$\{T \exp[-T \exp(\bar{X}\beta)] \exp(\bar{X}\beta)\} \beta_K \quad (7.28)$$

where β_K is the coefficient X_K .

In the empirical results presented in Section 7.6 both marginal effects are calculated for the exponential model.

7.4.1.2 Measurement of ATM Network Externalities

An important issue that has to be addressed in an empirical exploration of the effects of network externalities on the diffusion of ATMs is how to measure the degree of externality, N , in the theoretical model. Following the construction of the theoretical model, the relevant network size an institution has is the number of ATM locations that the institution is expected to have in equilibrium. In the empirical results presented in Section 7.6 the number of ATM locations is proxied by the number of branches an institution has.

The simplified representation of an ATM network in Figure 7.4 indicates that the number of locations, A_i , from which a deposit holders can access their accounts can be employed to capture demand-side externalities that are present in ATM networks. This schematic interpretation of an ATM network lends support to the proxy employed in the theoretical model. This externality can only be realised by deposit holders if, however, there is compatibility between the debit card held by the deposit holder and the ATM operated by the financial institution. Thus, there is the inherent assumption contained in the construction of Figure 7.4 that all deposit holders, irrespective of the bank at which they hold their deposits, can all use ATM locations A_i ($i = 1, \dots, 5$). This assumption is not, however, consistent with the current situation in the UK financial sector and this aspect has to be considered before empirical implementation of the theoretical model.

As shown in Chapter 4, the diffusion of ATMs in the UK financial sector can be divided into two distinct time periods. From 1972 to 1986 financial institutions developed and invested in their own proprietary networks. Growth in the number of new adopters and in the total number of ATMs in operation was most rapid during this period. In contrast, from 1986 (the date at which the LINK and MATRIX networks were established) institutions started to consolidate the growth of their own networks through a process of sharing agreements and joint ventures with other institutions and vertically related non-financial organisations. These shared reciprocal networks continue today [see APACS (1997) for further details]. Accordingly, the total number of new adopters fell from the period 1986 to 1994 relative to the preceding period 1972 to 1985.²⁹ The important aspect of this pattern of diffusion for the measurement of demand-side externalities is that pre-1986 deposit holders were restricted to using ATMs operated by the institution that issued the debit card they held. This limited the degree of externalities to an institutions own proprietary network. This situation is consistent with the construction of the theoretical model in Section 7.5, which assumes that the number of branches, N , pertains to the number of proprietary branches an institution has. From 1986, however, deposit holders were able to use ATMs owned by other institutions in the shared network for cash withdrawal as well the institution they held they deposits with. Moreover, Chapter 4 showed that reciprocal agreements between rival networks both at the domestic and international levels were made during the 1990s in the form of interchange and inter-bank transaction fees therefore enabling customers (at a monetary cost) access to a larger number of ATMs.

Given this pattern of ATM diffusion in the UK it was decided to estimate the extended proportional hazards model, first, for the period 1972 to 1986 and, secondly, for the period 1972 to 1992 to observe if any significant changes in parameter estimates occur. In the first period, 1972 to 1986, the number of branches (as a proxy for the number of ATM locations) are measured as the institution's proprietary ATM network at the date of adoption. For the second period, 1972 to 1992 the number of branches is measured as the institutions own proprietary network if they adopt ATMs on or before 1986 and

²⁹ As noted by Vesela (1994), Retail Bank Research (1997) and McAndrews and Rob (1997) this pattern of ATM diffusion is not unique to the UK case and has been experienced in other developed financial systems such as France, Spain and Germany.

as the total number of ATMs of the shared network if they adopt after 1986.³⁰ Implementation of the exponential regression model is more complicated because of its incorporation of time-varying covariates. This aspect of the model raises the question of what designates the proxy measure of the network externality – the institutions own proprietary branch network or the future shared branch network? Should, for example, the network externality in 1972 of an institution that adopts after 1986 be measured as the institution's proprietary branch network in 1972 or the size of the shared branch network that the institution eventually joins post-1986? The theoretical model gives a clear indication that the externality should be measured as an institution's proprietary network. This issue is therefore resolved by restricting the estimation of the exponential regression model to the period 1972 to 1986 for which the network externality is measured as an institution's proprietary branch network.

Objection may be raised at the employment of the number of branches as a proxy for the number of ATM locations. As shown in Chapter 4, however, at the end of 1991 approximately 82% of all ATMs in the UK were located at branches, either 'through the wall' type or in customer areas [APACS (1997)]. It is argued, therefore, that the number of branches is an accurate proxy for the number of ATM locations for the sample period under investigation.

7.4.1.2 Measurement of Other Covariates

In the construction of the theoretical model the variable, n , should ideally be implemented empirically as the total number of depositors an institution has. It was found in the collection of the data set, however, that for clearing banks the number of (retail) deposits per-institution is considered commercially sensitive by banks and consequently was not revealed. Consequently, the measurement for total depositors is proxied by the value of deposits as employed in Chapter 6 to measure the growth covariate GY . For clearing banks total deposits are measured as total customer deposits and for building societies as total value of deposit and share accounts. In the empirical results presented in Section 7.6 total deposits are denoted by the covariate

³⁰ Note that all institutions adopting ATMs after 1986 joined shared ATM networks.

DEPOSITS. The accuracy of this proxy covariate for total depositors, n , ultimately depends on how much variation there is in the proportion of individual retail accounts in *DEPOSITS* across institutions. Without any additional information pertaining to this proxy covariate, however, it is not possible to predict its accuracy.

The theoretical model also predicts a role for technical change, captured by the parameter θ in equation (7.10), in determining the timing of ATM adoption. As the role of order effects for adopting timing are not explored in this chapter the hedonic price of ATMs cannot be employed to proxy improvements in ATM technology. It was therefore decided to capture technical improvements by the time-varying baseline hazard, $h_0(t)$, and the growth in deposits GY . Using the time-varying baseline hazard as a means of capturing unobservable technical change has been employed previously in empirical diffusion studies by Levin *et al* (1987) and by Colombo and Mosconi (1995). The use of the growth-in-deposits covariate, GY , as a proxy for technical change is based on the argument that an increase in the benefits to ATM adoption, via the benefit function in equation (7.6), may increase if the number of depositors is expected to increase.

The covariates included in vector R_i in equation (7.19) follow those included in the exploration of rank-effects in Chapter 6. There are two covariates not included in R_i that are included in the rank-effects covariates in Chapter 6. These are the size of the institution, $SIZE$, and the dummy covariate representing the ownership status of the institution $DSUB$. The covariate $SIZE$ is not included in regressions because of high multicollinearity encountered between this covariate and *DEPOSITS*.³¹ As *DEPOSITS* is capturing scale-economies effects the exclusion of the size covariate is not considered critical in the interpretation of the results. The covariate $SIZE$ does, however, enter as the denominator of the liquidity ratio $PROFITS/SIZE$.³²

All covariates included in the results presented in Section 7.6 are summarised in Table 7.1 below together with their associated covariate name. The economic rationale

³¹ The correlation coefficient between the time-invariant covariates $SIZE$ and *DEPOSITS* is found to be 0.8043 for the period 1972 to 1986 and 0.7643 for the period 1972 to 1992.

underlying the expected signs of the covariates follow from the arguments presented in Chapter 6 and the theoretical model of adoption presented in Section 7.4 of this chapter.

Variable	Coefficient	Description	Expected Sign
p	p	Parameter of parametric models that determines duration dependence	++ and >1.0 in all models
<i>CONSTANT</i>	α_1	Intercept term	N/A
<i>DEPOSITS</i>	β_1	Value of total deposits for each institution measured at time t	+
<i>BRANCH</i>	β_2	Number of branches operated by each institution at time t	++
<i>DEPOSITS/BRANCH</i>	β_3	Ratio of the value of deposits to the number of branches measured at time t	+
<i>GY</i>	β_4	Growth in institutions deposits measured at time t	+
<i>STAFF/BRANCH</i>	β_5	Ratio of total number of part-time and full-time branch staff to total branches measured at time t	+
<i>PROFITS/SIZE</i>	β_6	Profitability of an institution, measured as the after-tax profits at time divided by total assets measured at time t	+
<i>DPREVIOUS</i>	β_7	A dummy variable taking the value of unity for previous cash dispenser adoption (1967 to 1971) and zero otherwise	++
<i>DTAKE</i>	β_8	A dummy variable taking the value of unity if the institution has taken over another institution in the period 1972 to 1986 or 1992 and zero otherwise	+

Note: '+' = positive; '++' = highly positive; '0' = no effects; '-' = 'negative'; '- -' = highly negative; '?' = no *a priori* expectation; 'N/A' = not applicable

³² The correlation coefficient between the time-invariant covariates *SIZE* and *PROFITS/SIZE* is found to be 0.1206 for the period 1972 to 1986 and 0.1401 for the period 1972 to 1992.

7.5 The Data Set

The set of potential adopters analysed in this chapter is identical to those analysed in Chapters 5 and 6. Discussion of the methodology employed in the construction can be found in Appendix One of Chapter 4 and is not repeated here. The Girobank was, however, excluded from the set of potential adopters. This institution has an average of 21,663 branches for the period 1972 to 1992 compared with an average of 433 for all potential adopters. It was found that due to a relatively small sample (98 potential adopters) its inclusion in regressions with time-invariant covariates distorted the estimation of the extent of network externalities. The results contained in Section 7.6 do not therefore include figures for the Girobank and consequently the number of potential adopters is reduced to 97 institutions.

For results pertaining to the extended proportional hazards model all institution-specific covariates are measured at the time of adoption if the institution is an adopter. If the institution is a non-adopter then the covariate is measured at the time of censoring either at 1986, for those results pertaining to the period 1972 to 1986, or at 1992, for those results pertaining to the period 1972 to 1992. For the exponential regression model all covariates are measured from the time of commercialisation of second-generation ATMs in 1972 until the end of the sample period in 1986. Thus, the definition of ATM technology is that of a second-generation definition. This definition is consistent with those employed in Chapters 5 and 6.

7.6 Estimation Results

This section presents the empirical results from estimating the extended proportional hazards model in equation (7.20) and the exponential regression model in equation (7.26) for the set of potential adopters compiled in Chapter 4. All results were obtained from using the econometric package *STATA 5.0*.

In all tables of results the sign of the estimated coefficient indicates the direction of the effect that the covariate has on the conditional probability of adoption. The network

externality is measured as the number of branches denoted by *BRANCH*. Marginal effects on the conditional probability of adoption are denoted at the foot of each table and are calculated using the procedures outlined in Section 7.4.1.1 for each specific model estimated. Marginal effects for the covariate *BRANCH* given in the tables, however, pertain only to the lower bound estimate. The upper bound estimate is given in the subsequent discussion of the results. A summary of the lower and upper bound marginal effects for all estimated models are provided in Table 7.7.

The parameter p represents a parameter of the baseline hazard (see Appendix One of Chapter 5) and its value denotes the nature of duration dependence present in the sample of adoption dates. The empirical results obtained in Chapters 5 and 6 suggest that the underlying hazard function is non-monotonic, at first increasing and then decreasing over time. In the Weibull model this characteristic of the hazard function implies the parameter restriction $p > 1$.³³ In the log-logistic model the non-monotonic character of the hazard function also implies the parameter restriction $p > 1$. The significance of the estimated value of p is therefore tested using a one-sided test given the restrictions on its values that each parametric baseline hazard assumes. Consequently, the 0.05 and 0.01 levels of significance for the case of p are 1.645 and 2.326 respectively [Johnston (1987)].

The exponential regression model imposes the restriction that $p = 1$ so that by definition there is no duration dependence in this model. This restriction additionally implies that the conditional of adoption is assumed constant over time. Consequently, no formal testing of duration dependence in the exponential regression model is possible.

The statistical significance of all other estimated covariates are tested using the standard two-sided t -test [see Johnston (1987)] with critical values of 1.960 and 2.576 for the 0.05 and 0.01 levels of significance respectively. The critical values employed for the likelihood ratio (LR) test are dependent on the number of linear restrictions imposed and these are given after $\chi^2_{.95}(m)$ in each table, where ' m ' is the number of restrictions.

³³ Note that Weibull model cannot predict a strictly non-monotonic hazard function, but as the maximum value of the hazard function is found to obtain nearer to the end of the sample period then this characteristic should reflect itself in positive duration dependence as opposed to a negative one.

In all tables the 'E' operator indicates the power by which 10 is raised and subsequently multiplied by the estimated coefficient. A coefficient multiplied by E-04, for example, implies that the coefficient is multiplied by 10^{-4} .

The presentation of the results is divided into two separate sections. Section 7.6.1 presents the results obtained for the period 1972 to 1986, and Section 7.6.2 presents the results obtained for the period 1972 to 1992.

7.6.1 Estimation Results for the Period 1972 to 1986

Two basic models are estimated for the period 1972 to 1986: a fully specified model and a restricted model. The fully specified model, as its name suggests, does not impose any restrictions on the values of the parameters and includes all the rank effect covariates outlined in Table 7.1. In contrast, the restricted model imposes the parameter restriction $\beta_2 = \beta_3 = 0$ in tables 7.2 and 7.4 below and is imposed in order to focus initially on the relationship between the number of depositors, as proxied by *DEPOSITS*, and the conditional probability to adopt. A test for this linear restriction is performed using the likelihood (LR) test introduced in Chapter 5. Furthermore, the existence of network externalities is tested by imposing the parameter restriction $\beta_2 = 0$ on the covariate *BRANCH* and re-estimating the fully specified model. This test is analogous to estimating the fully specified model without the covariate *BRANCH*. Although parameter estimates from this restriction are not provided in this chapter the value of the LR test of this restriction is given at the foot of the appropriate table.

To test for robustness to functional form, results from the extended Cox model with a Weibull and log-logistic baseline hazard and those obtained from the exponential regression model are compared.

The results pertaining to the restricted model are discussed first. Turning to the results in Table 7.2 and Table 7.4, the sign of the estimated coefficient on the covariate *DEPOSITS* is positive and statistically significant at the 0.01 level. This result implies that the log of the hazard function is an increasing function of *DEPOSITS*. This finding is consistent with empirical evidence obtained by Hannan and McDowell (1984b, 1987)

and Saloner and Shepard (1995) for the US banking sector. The marginal effect of increasing deposits by £1 million is reported at the bottom of Table 7.2 by the value of $\partial \ln(\text{hazard}) / \partial \text{BRANCH}$. The value of $\partial \ln(\text{hazard}) / \partial \text{BRANCH}$ of 0.94E-04 in both the Weibull model and the log-logistic model implying that an increase in an institution's total value of deposits by £1 million above the sample mean leads to an approximately 0.0094% increase in the hazard function. A marginally higher figure for the marginal effects of increasing *DEPOSITS* of 0.0095% is obtained for the exponential regression model.

Turning to the role of other rank effects, it is immediately apparent from Table 7.2 and 7.4 that these effects perform a significant role in the diffusion of ATMs in the UK and lend further support to the empirical results obtained in Chapter 6. Estimated coefficients tend to have higher values in the log-logistic Cox model relative to the Weibull model. Estimated coefficients are highest in the exponential model apart from those on *STAFF/BRANCH* and *PROFITS/SIZE*. This outcome reflects the time series nature of the covariates employed in the estimation of the exponential regression.

The *STAFF/BRANCH* covariate although having the correct sign based on *a priori* expectations is not found to be statistically significant in any of the estimated models. This result is consistent with that obtained in Chapter 6 and may reflect the fact that this covariate is an inadequate measure for the opportunity for labour saving associated with ATM technology.

A positive and significant coefficient is found on *PROFITS/SIZE* in all estimated models. This result is again consistent with that obtained in Chapter 6 and suggests that liquidity constraints play an important role in the diffusion of ATMs in the UK.

The learning-by-using covariate, *DPREVIOUS*, is found to have a positive and statistically significant coefficient at the 0.01 level in all models. This lends further support to the empirical result obtained in Chapter 6 that suggests that institutions with experience in using previous ATM vintages has a positive and significant impact on the conditional probability of adoption.

An insignificant role is, however, played by the growth-in-deposits covariate, *GY*, and the horizontal takeover dummy covariate *DTAKE*, although the signs of the estimated coefficients on these covariates are correct *a priori*. This result contradicts that obtained in Chapter 6 which finds a statistically significant role for both these rank effects.

In all estimated restricted Cox models the value of the shape parameter p is found to be significantly different from unity at the 0.01 level. This implies positive duration dependence in the extended Weibull model and non-monotonic duration dependence in the log-logistic extended model. These results lend further support to the empirical finding in Chapter 5 and Chapter 6 that epidemic effects have played a significant role in the diffusion of ATMs in the UK.

Table 7.3 and the last column of Table 7.4 presents the regression results obtained from relaxing the restriction $\beta_2 = \beta_3 = 0$ in the extended Cox models and the exponential regression model respectively. The covariate *BRANCH* captures the effect of network externalities on the conditional probability of adoption, while the *DEPOSITS / BRANCH* ratio takes account of location-specific costs of adoption. The LR test rejects the restriction $\beta_2 = \beta_3 = 0$ in all estimated models at the 0.05 level of significance.

The theoretical model outlined in Section 7.4 predicts that the sign on the *DEPOSITS / BRANCH* ratio should be positive as an increase in depositors per location will 'spread' location-specific costs over a greater number of end users (and therefore transaction volumes). Moreover, if location-specific costs are high relative to system fixed costs then the coefficient on *DEPOSITS / BRANCH* may capture a large share of the effect captured by *DEPOSITS* forcing *DEPOSITS* to become statistically insignificant.³⁴

From Table 7.3 and Table 7.4 the inclusion of the *DEPOSITS / BRANCH* ratio increases the estimated coefficient on *DEPOSITS* in all estimated models, but the coefficient on *DEPOSITS* remains statistically significant at the 0.01 level. Thus, the

³⁴ This potential outcome can be seen from equation (7.13). If the second term on the right-hand side of (7.13) is large relative to the first term on the right-hand side then this may be reflected in the empirical model by *DEPOSITS / BRANCH* capturing a large share of the effect of *DEPOSITS*.

effect of the number of depositors on the adoption decision comes through both the *DEPOSITS / BRANCH* ratio and the covariate *DEPOSITS*. This result suggests that location-specific costs do not dominate system costs for the adoption decision. Moreover, this result is consistent with the inclusion of relatively late adopters in the sample of potential adopters. As noted by Banking World (1987, 1988) and Kirkman (1987) institutions adopting ATM technology for the first time after 1983 adopted a high proportion of on-line machines rather than the off-line machines adopted by institutions in the early 1970s. This adoption pattern reflected the inherent advantages of on-line ATM technology vis-à-vis off-line technology as outlined in Chapter 4. Saloner and Shepard (1995) have argued that on-line technology has a smaller location-specific cost component relative to off-line technology. The argument put forward by Saloner and Shepard is a potential explanation of why the *DEPOSITS / BRANCH* ratio does not dominate the effects captured by *DEPOSITS*.³⁵ Estimating the model in subset periods could enable the testing of this hypothesis, but with a relatively small size of potential adopters this potential avenue was not explored.

There is a relatively small increase in the marginal effect on an increase in the value of deposits given by the value of $\partial \ln(\text{hazard}) / \partial \text{BRANCH}$ in the unrestricted Weibull model, but a decrease in the marginal effect for the log-logistic model. The marginal effect on the conditional probability of adoption from an increase in deposits by £1 million is found to be 0.0095% in the Weibull model and 0.0070% in the log-logistic model.

A notable aspect of the results is the positive and significant coefficient obtained on the covariate *BRANCH* in all estimated models. This result lends support to the hypothesis that the existence of positive network externalities (proxied by the number of ATM locations) increases the conditional probability of adoption. To investigate the significance of network externalities further, the fully specified model was estimated imposing the restriction $\beta_2 = 0$. The LR test of this restriction is rejected in all models at the 0.05 level of significance further lending support to the existence of positive network externalities.

³⁵ Empirical results obtained by Saloner and Shepard (1995) for the US banking sector indicate that for this sector *DEPOSITS / BRANCH* dominates *DEPOSITS*. This is partly explained by the sample period selected by Saloner and Shepard, 1971 to 1979, which covers the early period of the diffusion process (15% of potential adopters in their sample have adopted by 1979).

The subsequent estimation of the model with the restriction $\beta_2 = 0$ imposed leads to no change in the statistical significance and sign of estimated coefficients relative to the unrestricted model. In particular, the potential problem of multicollinearity between the covariate *DEPOSITS* and *BRANCH* does not appear to be a significant problem for estimation.³⁶ The correlation coefficient between *DEPOSITS* and *BRANCH* is 0.5053 for the period 1972 to 1986 and 0.4907 for the period 1972 to 1992. Consequently the results from these additional regressions are not provided in this chapter.

The marginal effect of increasing the number of branches for each estimated model is calculated using the procedures outlined in Section 7.4.1.1. These marginal effects can be considered as the 'network effect'. The lower bound estimate of the network effect (as defined by the theoretical model of adoption constructed in Section 7.4) is given in Table 7.3 and Table 7.4 as the value of $\partial \ln(\text{hazard}) / \partial \text{BRANCH}$. This lower bound estimate measures the marginal effect on the log of the hazard function from increasing the number of branches (i.e. ATM locations) that the average institution has by one unit, while keeping all other factors constant. From Table 7.3 the lower bound marginal effect for the Weibull Cox model is 0.0014, which implies that adding one additional branch to the average institution increases the one-period conditional probability of adoption by 0.14 percentage points, *ceteris paribus*. An identical interpretation can be given to the marginal effects estimated for the log-logistic Cox model and the exponential regression model.

Estimates of the lower bound network effect for the period 1972 to 1986 are summarised in Table 7.7. The range of the estimates for the marginal effect on the one-period conditional probability is 0.142% to 0.254%. The Weibull model produces the lowest estimate of 0.142%, while the extended log-logistic model gives the highest estimate of 0.254%. The average of all estimates is 0.197%.

The average lower bound network effect found for the UK is lower than the one found by Saloner and Shepard (1995) in their exploration of network externalities in the US banking sector. Saloner and Shepard estimate the lower bound range to be 5%. This

³⁶ Most notably there is an increase in the significance and a fall in the value of estimated coefficients for the *PROFITS/SIZE* ratio and *DEPOSITS*. The conclusions resulting from the unrestricted model are,

difference in results reflects the average branch size employed in the respective empirical studies. In Saloner and Shepard's study the average potential adopter has 6.372 branches, while the set of potential adopters used in this chapter has an average of 433.730 branches. This difference in the average number of ATM locations is reflected in the calculation for both the lower and upper network effect.

The lower bound estimate of the network effect does, by definition, underestimate the network effect present in ATM adoption because the per-depositor cost of adoption declines as the number of ATM locations increase. In contrast, the upper bound estimate is calculated by increasing the number of branches the average institution has by unit, while increasing the number of depositors by an amount just sufficient to keep the ratio of depositors per branch constant. As its name suggests the upper bound estimate of the network effect overestimates the extent of network externalities. The upper bound estimate of the network effect for the period 1972 to 1986 for all estimated models was calculated using the procedures outlined in Section 7.4.1.1. The results of calculating the upper bound are again summarised in Table 7.7.

The results in Table 7.7 show that the range of the upper bound estimate is from 0.191% to 0.968%. This implies that a one unit increase in the number of branches the average sized institution has, while keeping the ratio of depositors to branches constant, increases the one-period conditional probability from between 0.191% to 0.968%. The extended Weibull Cox model produces the lowest estimate of 0.191%, while the exponential regression model gives the highest estimate of 0.968%. The average upper bound network effect is found to be 0.492%. The average network effect is again found to be lower for the UK than the one calculated by Saloner and Shepard and the explanation for this is identical to one given in explaining differences in the lower bound estimate.

A one unit increase in the number of branches, for the average institution, affecting all ten periods of the sample period (i.e. 1972 to 1986) is calculated from the results obtained from the exponential model using and using equation (7.28). The ten-period increase in the conditional probability is found to be equal to 1.140%. This figure can

however, not changed substantially.

be interpreted as the marginal effects on the conditional probability of adoption for the average institution from increasing the number of branches in 1972 by one unit and which subsequently affects all ten periods up to 1986, *ceteris paribus*. There is no lower or upper bound on this estimation analogous to the one-period estimation of the network effect.

As suggested by the empirical results obtained in Chapter 6, the conditional probability may be a function of non-linear rank effects. To investigate these issues further, $(DEPOSITS)^2$ and $(BRANCH)^2$ were included in the unrestricted regression models. It was found, however, that the likelihood functions failed to converge to a global maximum indicating that there may either exist either perfect collinearity between the covariates or that the variation in the squared covariates produces 'too large' standard errors [Greene (1993)].

The inclusion of the covariates *DEPOSITS* and *DEPOSITS/BRANCH* in the unrestricted regressions appears to have no substantial affect on either the sign or statistical significance of the other rank effect covariates relative to the estimates obtained from the restricted models. There is a marginal increase in the statistical significance in the estimated coefficient for *DPREVIOUS* in the log-logistic extended Cox model and the growth-in-deposits covariate, *GY*, becomes marginally less significant in the exponential regression model. Overall, however, the main results for the other rank effects included in the model are the same those obtained for the restricted model.

Table 7.2: Maximum Likelihood Estimates of the Restricted Branch Weibull and Log-logistic Models – end 1972 to end 1986

Coefficient	Variable	Weibull Model		Log-logistic Model	
p	<i>TIME</i>	1.786	(4.314)**	2.292	(5.013)**
α_1	<i>CONSTANT</i>	-4.785	(8.886)**	-4.586	(9.021)**
β_1	<i>DEPOSITS</i>	0.94E-04	(3.678)**	0.94E-04	(3.342)**
β_2	<i>BRANCH</i>	—		—	
β_3	<i>DEPOSITS/BRANCH</i>	—		—	
β_4	<i>GY</i>	0.350	(2.275)*	0.511	(2.372)*
β_5	<i>STAFF/BRANCH</i>	0.029	(1.160)	0.033	(1.541)
β_6	<i>PROFITS/SIZE</i>	24.087	(2.572)*	30.843	(2.443)*
β_7	<i>DPREVIOUS</i>	0.259	(2.101)*	0.498	(2.145)*
β_8	<i>DTAKE</i>	0.354	(1.204)	0.323	(1.509)
$\partial \ln(\text{hazard}) / \partial \text{DEPOSITS}$		0.94E-04		0.94E-04	
$\partial \ln(\text{hazard}) / \partial \text{BRANCH}$		—		—	
Median duration (years)		35.443	(4.263)**	33.491	(5.400)**
Log-likelihood		-74.27		-72.99	
Number of observations		97		97	
Number of individual institutions		97		97	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 1 for both the Weibull and log-logistic models).

Coefficient	Variable	Weibull Model		Log-logistic Model	
p	<i>TIME</i>	2.818	(3.583)**	3.397	(3.194)**
α_1	<i>CONSTANT</i>	-4.048	(9.187)**	-3.950	(9.002)**
β_1	<i>DEPOSITS</i>	0.29E-04	(2.678)**	0.14E-04	(3.289)**
β_2	<i>BRANCH</i>	0.38E-02	(2.175)*	0.48E-02	(2.262)*
β_3	<i>DEPOSITS/BRANCH</i>	0.014	(1.930)*	0.012	(2.737)**
β_4	<i>GY</i>	0.846	(1.137)	0.435	(1.155)
β_5	<i>STAFF/BRANCH</i>	0.027	(1.091)	0.026	(1.251)
β_6	<i>PROFITS/SIZE</i>	20.865	(2.351)*	30.843	(2.443)*
β_7	<i>DPREVIOUS</i>	0.534	(2.660)**	0.517	(2.145)*
β_8	<i>DTAKE</i>	0.315	(1.280)	0.323	(1.509)
$\partial \ln(\text{hazard}) / \partial \text{DEPOSITS}$		0.95E-04		0.70E-04	
$\partial \ln(\text{hazard}) / \partial \text{BRANCH}$		0.14E-02		0.25E-02	
Median duration (years)		36.575	(4.827)**	33.108	(5.876)**
Log-likelihood		-60.88		-61.62	
Number of observations		97		97	
Number of individual institutions		97		97	
Likelihood ratio test for the existence of branch effects: ($\beta_2 = \beta_3 = 0$)		26.78 ($\chi^2_{.95}(2) = 5.99$)		22.74 ($\chi^2_{.95}(2) = 5.99$)	
Likelihood ratio test for the existence of network externalities: ($\beta_2 = 0$)		11.64 ($\chi^2_{.95}(1) = 3.84$)		11.32 ($\chi^2_{.95}(1) = 3.84$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 1 for both the Weibull and log-logistic models).

Coefficient	Variable	Restricted Model		Fully Specified Model	
α_1	<i>CONSTANT</i>	-4.906	(8.863)**	-4.941	(6.920)**
β_1	<i>DEPOSITS</i>	0.28E-04	(2.678)**	0.26E-04	(2.402)*
β_2	<i>BRANCH</i>	—		0.63E-02	(2.320)*
β_3	<i>DEPOSITS/BRANCH</i>	—		0.017	(2.801)**
β_4	<i>GY</i>	0.830	(1.914)*	1.271	(1.652)
β_5	<i>STAFF/BRANCH</i>	0.021	(1.093)	0.032	(1.337)
β_6	<i>PROFITS/SIZE</i>	23.008	(2.635)**	21.753	(2.548)*
β_7	<i>DPREVIOUS</i>	0.865	(2.673)*	1.267	(2.153)*
β_8	<i>DTAKE</i>	0.486	(1.387)	0.532	(1.607)
$\partial \ln(\text{hazard}) / \partial \text{DEPOSITS}$		0.95E-04		0.70E-04	
$\partial \ln(\text{hazard}) / \partial \text{BRANCH}$		—		0.19-02	
Median duration (years)		27.995	(4.408)**	25.412	(4.043)**
Log-likelihood		-151.87		-140.09	
Number of observations		1223		1223	
Number of individual institutions		97		97	
Likelihood ratio test for the existence of branch effects: ($\beta_2 = \beta_3 = 0$)		—		23.56 ($\chi^2_{.95}(2) = 5.99$)	
Likelihood ratio test for the existence of branch effects: ($\beta_2 = 0$)		—		12.86 ($\chi^2_{.95}(1) = 3.84$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level.

7.6.2 Estimation Results for the Period 1972 to 1992

Two basic models are also estimated for the period 1972 to 1992: a fully specified model and a restricted model. Both the fully specified model and the restricted model have an identical specification as those described in Section 7.6.1. To test for functional form, results from the estimated Cox model with a Weibull and log-logistic baseline

hazard are compared. Estimation is not extended to the exponential regression model for the reasons given in Section 7.4.1.2.

The network externality for this period is measured as the institutions own proprietary system if the institution adopts before 1986 and the total number of branches belonging to the members of the shared ATM network if the institution adopts after 1986. Both measures of the network externality are measured at the time of adoption and assumed to be time-invariant.

The results from estimating the restricted Weibull and log-logistic models are given in Table 7.3, while those for the unrestricted models are given in Table 7.6. The results pertaining to the restricted model are discussed first. The results obtained from the estimation of the restricted Weibull and log-logistic models for the sample period 1972 to 1992 are essentially the same as those obtained for the period 1972 to 1986. The sign of the estimated coefficient on *DEPOSITS* is positive and statistically significant from zero at the 0.01 level. This result implies that the log of the hazard function is an increasing function of *DEPOSITS*. This result is consistent with that obtained for the period 1972 to 1986. As reported at the bottom of Table 7.5, the estimate of the marginal effect $\partial \ln(\text{hazard}) / \partial \text{DEPOSITS}$ implies that increasing an institution's total value of deposits by £1 million above the sample mean leads to an approximately 0.008% increase in the hazard function in the Weibull model specification and 0.0094% in the log-logistic model specification. This compares to a marginal effect of 0.0094% for both the Weibull and log-logistic models for the period 1972 to 1986.

Turning to the role of other rank effects in the diffusion process, it is apparent from Table 7.5 that these effects have played a significant role in the diffusion of ATMs and lend further support to the empirical results obtained in Chapter 6. The sign and significance of the various institution-specific characteristics presented in Table 7.5 do not, in general, differ from those contained in Table 7.2 for the period 1972 to 1986. The significance of *PROFITS/SIZE* increases for the period 1972 to 1986 and this may

reflect the declining profitability of smaller building societies after 1986 [see, for example, BSA (1993)].³⁷

The growth-in-deposits covariate *GY* becomes statistically insignificant in both the Weibull and log-logistic models, although the estimated coefficient on this covariate remains the correct sign based on *a priori* expectations. There is an increase in the estimated coefficient on *DPREVIOUS* and the statistical significance of this covariate increases over the sample period 1972 to 1992 relative to that obtained for the period 1972 to 1986. There is also an increase in the estimated coefficient on *DTAKE*, but this covariate remains insignificant at the 0.05 level.

The results obtained for estimation of the fully specified models are presented in Table 7.6. It is apparent from Table 7.6 that as in the case for estimation over the sample period 1972 to 1986 the inclusion of the covariates *DEPOSISTS* and *DEPOSITS/BRANCH* has no substantial effect on either the sign or statistical significance of the other rank effects included in the model. Comparing the results presented in Table 7.5 and Table 7.6 it can be observed that there is a decrease in the statistical significance in the estimated coefficient for *PROFITS/SIZE* and *DPREVIOUS*, but both covariates remain significant at the 0.01 and 0.05 levels of significance respectively. The covariates *GY*, *STAFF/BRANCH* and *DTAKE* remain insignificant on the unrestricted model, but still have the correct sign based on *a priori* expectations.

The covariate *BRANCH* enters the fully specified models with a positive and statistically significant coefficient. This result lends further support to the hypothesis that the existence of positive network externalities (proxied by the number of ATM locations) increases the conditional probability of adoption. To investigate the significance of network externalities further, the fully specified model was estimated imposing the restriction $\beta_2 = 0$. The LR test of this restriction is rejected in all models at the 0.05 level of significance further lending support to the existence of positive

³⁷ Note that relatively smaller building societies are more likely to be censored observations than relatively larger ones.

network externalities. The LR test for the joint significance of $\beta_2 = \beta_3 = 0$ is also rejected at the 0.05.

Summaries of the upper and lower estimates of the network effects obtained for the period 1972 to 1992 are given in the lower half of Table 7.7. The lower bound estimate of the network effect, given by the value of $\partial \ln(\text{hazard}) / \partial \text{BRANCH}$ in Table 7.6, is calculated to be 0.066% in the Weibull model and 0.205% in the log-logistic model. These figures represent a reduction in the estimation of the lower bound from 0.142% and 0.191% respectively for the period 1972 to 1986. There is also a reduction in the upper bound estimate of the network effect for the sample period 1972 to 1992 from 0.254% to 0.205% in the Weibull model and 0.298% to 0.291% in the log-logistic model. The average lower bound for the sample period 1972 to 1992 is found to be 0.136%, while the average for the upper bound is found to be 0.227%. This compares with an average lower bound estimate of 0.198% and an upper bound estimate of 0.245% for the estimated Cox models for the period 1972 to 1986 (see summary in Table 7.7).

The finding that there may have been a reduction in the network effect reflects the distinctive nature of ATM networks: that is, the network externality is increasing in the number of ATMs (or, rather, the number of ATM locations) rather than the number of depositors. The employment of the total number of branches belonging to the members of the shared ATM network as a measure of the network externality if the institution adopts after 1986 is necessarily higher than if measured by the institution's own proprietary branch network. This distinction between pre-1986 and post-1986 adopters explains why the coefficient on *BRANCH* and the subsequent network effect declines when the sample period is extended from 1972 to 1992.

In addition, the development of shared networks such as MINT and FOUR BANKS in the UK are likely to reduce the importance of the network effects for individual institutions because a relatively large branch network is no longer required to

appropriate the returns to positive demand-side externalities.³⁸ Indeed, Economides (1995) has argued that all that is required for a relatively small institution to appropriate the returns to the network effect is gaining access to the shared network. The theoretical model presented by Matutes and Padilla (1994), that examines the incentives facing financial institutions in their decision to join compatible ATM networks also supports such a view. Matutes and Padilla argue that the development of shared ATM networks means that no individual institution obtains a network advantage, but does imply that institutions become better substitutes for each other because transactions costs for deposit holders wishing to switch deposit accounts are lower under compatibility agreements. Institutions may then circumvent this substitution effect by imposing interchange and withdrawal fees on deposit customers (see discussion in Chapter 4).

Further exploration of this aspect of the results does, however, require the development of a more sophisticated theoretical model because the benefits from technology adoption when ATM networks are shared implies that there is an additional positive externality on other institution's adoption decisions. This has to be accounted for in the construction of a theoretical model of ATM adoption.

³⁸ As noted in Chapter 4, during the period 1972 to 1992 the branch network for the sample of financial institutions employed as the set of potential adopters in this thesis has fallen from a total of 19007 in 1972 to 16660 in 1996

Table 7.5: Maximum Likelihood Estimates of the Restricted Branch Weibull and Log-logistic Models – end 1972 to end 1992					
Coefficient	Variable	Weibull Model		Log-logistic Model	
p	<i>TIME</i>	2.142	(6.166)**	2.585	(6.674)**
α_1	<i>CONSTANT</i>	-4.310	(9.780)**	-4.114	(11.150)**
β_1	<i>DEPOSITS</i>	0.80E-04	(2.678)**	0.96E-04	(3.135)**
β_2	<i>BRANCH</i>	—		—	
β_3	<i>DEPOSITS/BRANCH</i>	—		—	
β_4	<i>GY</i>	0.662	(1.067)	0.518	(1.155)
β_5	<i>STAFF/BRANCH</i>	0.017	(1.776)	0.014	(1.579)
β_6	<i>PROFITS/SIZE</i>	28.283	(3.369)**	34.796	(3.011)*
β_7	<i>DPREVIOUS</i>	0.636	(2.334)*	0.439	(2.475)*
β_8	<i>DTAKE</i>	0.485	(1.319)	0.541	(1.402)
$\partial \ln(\text{hazard})/\partial \text{DEPOSITS}$		0.80E-04		0.96E-04	
$\partial \ln(\text{hazard})/\partial \text{BRANCH}$		—		—	
Median duration (years)		27.394	(7.322)**	26.681	(8.483)**
Log-likelihood		-81.34		-81.65	
Number of observations		97		97	
Number of individual institutions		97		97	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; ‘*’ means significant at the 0.05 level; ‘**’ means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 1 for both the Weibull and log-logistic models).

Table 7.6: Maximum Likelihood Estimates of the Fully Specified Weibull and Log-logistic Models – end 1972 to end 1992					
Coefficient	Variable	Weibull Model		Log-logistic Model	
p	<i>TIME</i>	1.806	(4.314)**	3.900	(3.931)**
α_1	<i>CONSTANT</i>	-4.310	(9.780)**	-3.552	(9.733)**
β_1	<i>DEPOSITS</i>	0.80E-04	(3.678)**	0.35E-04	(3.733)**
β_2	<i>BRANCH</i>	0.26E-02	(2.020)*	0.37E-02	(2.4340)*
β_3	<i>DEPOSITS/BRANCH</i>	0.030	(2.038)*	0.024	(2.970)**
β_4	<i>GY</i>	0.701	(1.029)	0.885	(1.096)
β_5	<i>STAFF/BRANCH</i>	0.029	(1.021)	0.033	(1.541)
β_6	<i>PROFITS/SIZE</i>	26.012	(2.893)**	30.361	(2.812)**
β_7	<i>DPREVIOUS</i>	0.340	(2.211)*	0.269	(2.085)*
β_8	<i>DTAKE</i>	0.240	(1.104)	0.301	(0.864)
$\partial \ln(\text{hazard}) / \partial \text{DEPOSITS}$		0.21E-03		0.15E-03	
$\partial \ln(\text{hazard}) / \partial \text{BRANCH}$		0.66E-03		0.21E-02	
Median duration (years)		28.456	(7.612)**	31.410	(9.166)**
Log-likelihood		-65.94		-66.66	
Number of observations		97		97	
Number of individual institutions		97		97	
Likelihood ratio test for the existence of branch effects: ($\beta_2 = \beta_3 = 0$)		30.80 ($\chi^2_{.95}(2) = 5.99$)		29.98 ($\chi^2_{.95}(2) = 5.99$)	
Likelihood ratio test for the existence of branch effects: ($\beta_2 = 0$)		9.16 ($\chi^2_{.95}(1) = 3.84$)		13.81 ($\chi^2_{.95}(1) = 3.84$)	

Note: Figures in parenthesis refer to the standard $|t|$ statistics of coefficient estimates; '*' means significant at the 0.05 level; '**' means significant at the 0.01 level (in the case of estimated values of p these levels apply to p greater than 1 for both the Weibull and log-logistic models).

Time Period	Model Estimated	Lower Bound Estimate (%)	Upper Bound Estimate (%)
1972 to 1986	Extended Weibull	0.142	0.191
	Extended Log-logistic	0.254	0.298
Average effect for Cox models		0.198	0.245
	Exponential Regression	0.194	0.986
Average effect for all models		0.197	0.492
1972 to 1992	Extended Weibull	0.066	0.163
	Extended Log-logistic	0.205	0.291
Average effect for Cox models		0.136	0.227

7.7 Concluding Remarks

There were three aims to this chapter. First, to outline the sources and economic implications of network externalities for the diffusion process. Second, to outline the distinctive features of network externalities that pertain to ATM technology. Third, to empirically test for the existence of network externalities in the diffusion of ATMs in the UK financial sector.

The main empirical result obtained in this chapter is that UK financial institutions with relatively large branch networks have adopted ATMs earlier than those institutions with relatively fewer branches, adjusting for the number of depositors. This result indicates that network externalities have played a significant role in the diffusion of ATMs in the UK financial sector, thus supporting the network-type theoretical models of innovation diffusion.

The range of the network effect on the one-period conditional probability of adoption employing the extended Cox model for the period 1972 to 1986 ranges from an average lower bound estimate of 0.198% to an upper bound estimate of 0.245%. The corresponding range for the period 1972 to 1992 is found to be 0.136% to 0.227% respectively. The apparent decline in the network effect over time reflects the development of shared ATM networks in the UK and the distinctive nature of ATM

technology (that is, that the network externality is increasing in the number of ATMs rather than in the number of end-users).

In addition, the empirical results support the contention that rank effects have played an extremely important role in the diffusion of ATMs in the UK. Institution profitability, the value of deposits and the ratio of deposits to the number of branches were all found to have a positive and statistically significant effect on the conditional probability of adoption. Moreover, the results suggest that early adoption of previous vintage technologies play a significant role in fostering faster diffusion of later vintages. No significant role is found, however, for the labour-saving potential of ATMs, the growth in deposits an institution's propensity to participate in horizontal takeovers.

A7.1 Appendix One: The Demand for Technology in the Presence of Network Externalities

As stated in Section 7.2, the defining characteristic of network technologies is that the utility (or benefits) that an individual adopter derives from adopting technology depends positively on the number of other adopters. The existence of these positive network externalities has significant implications for the demand structure pertaining to network technologies.

Katz and Shapiro (1985), Economides and Himmelberg (1995) and Economides (1996) have attempted to capture the demand-side effects of network externalities in a static one-period framework through the concept of an ‘expectations fulfilled equilibrium.’ In the derivation of this equilibrium it is assumed that potential adopters hold expectations pertaining to the size of the network, n^e , normalised to lie between 0 and 1, $0 \leq n^e \leq 1$.³⁹ A ‘network externalities function’ is then defined that captures the influence of expectations on the willingness-to-pay for the technology provided by the network [Economides and Himmelberg (1995)]. The willingness-to-pay for the n th unit of the technology when n^e is expected to be sold is assumed to be given by the general function $p(n, n^e)$.⁴⁰ This is assumed to be a decreasing function in n but increasing function in n^e . The latter assumption captures the existence of positive network externalities. The exact nature of the expectation formation regime is left unspecified.

In a one-period framework expectations are fulfilled when $n = n^e$. This condition defining the fulfilled expectations demand as $p(n, n)$. The term $p(n, n)$ can then be conceptualised as defining the size of the network that can be supported by a fulfilled expectations equilibrium for a given price [Economides (1996)]. It is additionally assumed that $\lim_{n \rightarrow 1} p(n, n) = 0$. This implies that in order to achieve a large size, a network has to include adopters of a very low willingness-to-pay.

³⁹ In this formulation of the model both n and n^e represent market shares rather than absolute quantities.

⁴⁰ See Katz and Shapiro (1985) and Economides and Himmelberg (1995) for specific examples of functional forms. In particular, Katz and Shapiro assume an additive form, whilst Economides and Himmelberg assume a multiplicative form allowing for a distribution of consumer types.

Figure 7.1 below illustrates the construction of a typical fulfilled expectations demand curve. The vertical axis measures the price of adopting the network, p , and the marginal cost of providing the network, c . The curves $p(n, n_1^e)$ and $p(n, n_2^e)$ give the willingness-to-pay for adopting the technology, given different sizes of the ‘installed-base’ (i.e. overall network size) that potential adopters expect to emerge in equilibrium, where $n_2^e > n_1^e$. The point labelled ‘E₁’ on the first curve represents the point at which $n = n_1^e$, and analogously, point ‘E₂’ on the second curve represents the point at which $n = n_2^e$. The locus of all such points traces out the fulfilled expectations demand curve.

It is important to note that the fulfilled expectations demand curve $p(n, n)$ is quasi-concave⁴¹ with a single maximum occurring at the marginal cost of C^0 . In addition, this curve includes the entire vertical axis at zero, as indicated by the thicker line. This is because at any marginal cost $c > p$ a network of zero is a fulfilled expectations equilibrium. Figure A7.1.1 is sketched for the special case when network size, k , is zero. Thus, in general, the fulfilled expectations demand curve will consist of the vertical axis above k and the inverted-U curve that starts at k .

The network is referred to having a ‘positive critical-mass’ if and only if $p(n, n)$ is increasing in n in the neighbourhood of $n = 0$, i.e. if $\lim_{n \rightarrow 0} dp(n, n)/dn > 0$. In Figure A7.1.1 the point of critical-mass is indicated by network size n^0 . This is the size at which $p(n, n)$ obtains a maximum. Economides and Himmelberg (1995) show that the fulfilled expectations demand is increasing for small n if either one of three conditions hold. First, if the utility of every adopter in a network of size zero is zero so that the network has no intrinsic value (i.e. $k = 0$). This is necessarily true for a two-way network. Second, if there are immediate and large external benefits to network expansion for very small networks. Third, if there is a significant density of high willingness-to-pay consumers who are just indifferent on joining a network of approximately zero. If none of these three conditions are met then the demand curve will be monotone decreasing for all values of n .

⁴¹ As noted by Economides (1996) the shape of $p(n, n)$ depends on the distribution of adopter preferences and the functional form of the network externality function.

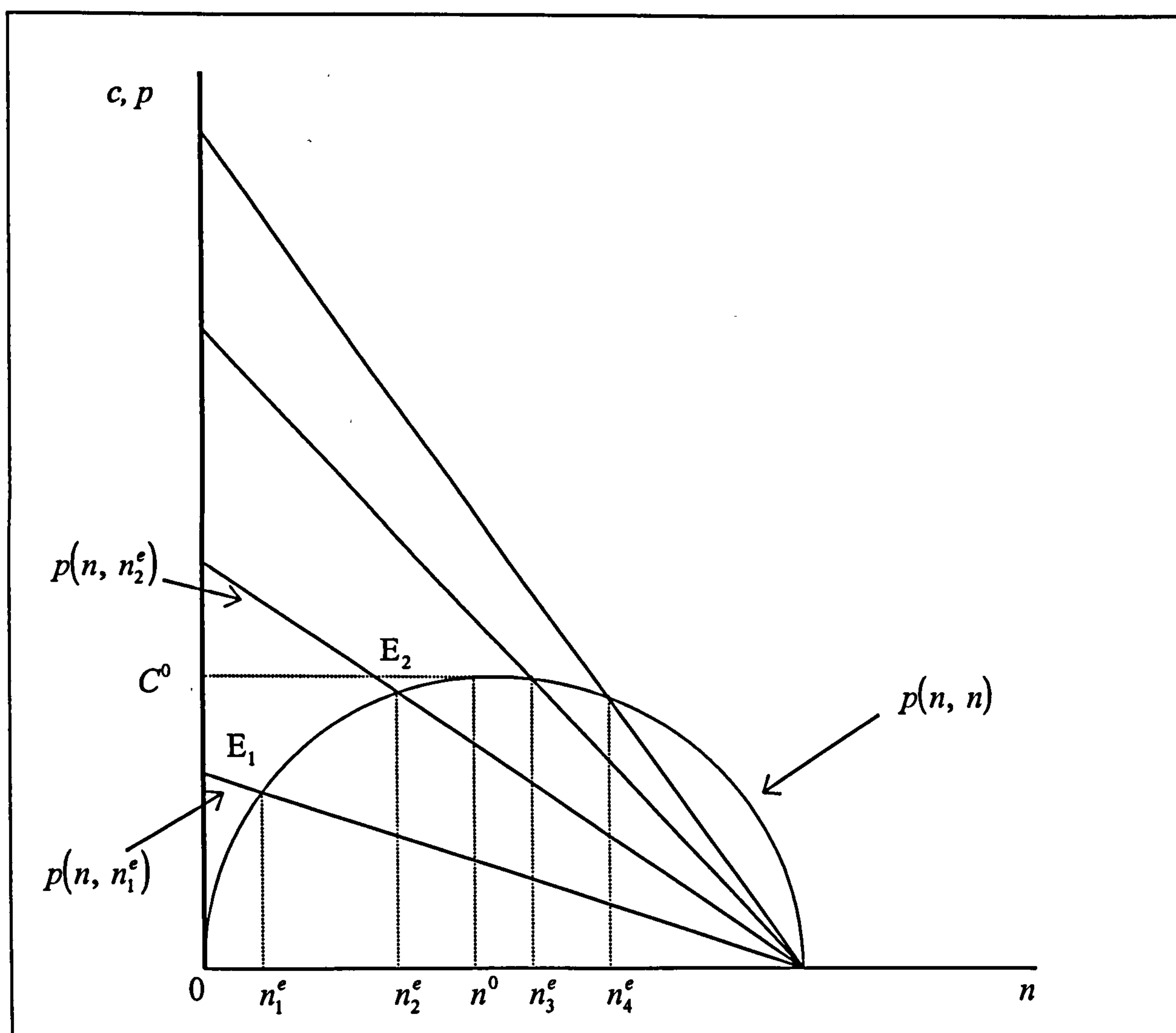


Figure A7.1.1 The Fulfilled Expectations Demand Curve

Economides and Himmelberg (1995) have interpreted the existence of a critical-mass network size as the smallest network that can be sustained in equilibrium. They argue that this implies adoption of network technology may not ‘take-off’ because the initial installed-base is of insufficient size. The underlying reason for this possible outcome is that potential adopters’ expectations of a zero network may be fulfilled in equilibrium. In addition, they show that if the supply-side is characterised by perfect competition then there are three possible equilibria. If marginal costs are above C^0 in Figure A7.1.1 then a zero sized network will result. If, however, marginal costs are below C^0 then there are two possible equilibria: an unstable, lower network size below n^0 or a Pareto dominate size above n^0 . Economides and Himmelberg interpret this result as illustrating the inefficiency of perfect competition in the presence of network externalities. This inefficiency arises because the marginal social benefit of expansion

is greater than the benefit that can be appropriated by an individual firm. Thus, perfect competition will provide a smaller network than is socially optimal and for some relatively high marginal costs will not produce the technology when it is socially optimal to do so. They show that price discrimination can overcome this inefficiency.

Katz and Shapiro (1985) relax the assumption of perfect competition by assuming that firms on the supply-side act in a standard Cournot manner. That is, firms take consumer expectations concerning the size of the network as given and assume that actual output of other firms is fixed. They show that if all firms produce incompatible goods⁴² that three equilibria are possible: symmetric oligopoly with all firms producing positive and equal levels of output, symmetric oligopoly with only a fraction of firms producing output and asymmetric oligopoly with all firms producing positive but different levels of output. They then use this framework to examine the private and social incentives that exist in moving towards compatibility.

⁴² When firms produce incompatible goods then each firm makes up its own network through its total sales of the network good.

CHAPTER EIGHT

CONCLUSION

It has become almost customary for economists to introduce their analysis of innovation diffusion by initially citing Schumpeter's trilogy of technical change and then proceeding to emphasise that the diffusion process is an integral component of wider economic welfare and growth - the current thesis is no exception. It is important to acknowledge the significance and implications of innovation diffusion in this concluding chapter in order to encourage future research into both the economic factors that underlie the diffusion process and its consequences.

As noted in Chapter 2, it was the seminal work of Schumpeter (1934, 1939) that distinguished between three distinct, time-intensive and sequential stages in the process of technical change at the economy-wide level, these being: invention, innovation and diffusion. Although research proceeding Schumpeter has illustrated that these three stages are not strictly unidirectional and that inventive activity does respond to economic incentives, the contention remains amongst economists [see, for example, Greenaway (1994)] that the diffusion of new products and processes is decisive for economic growth and welfare. In other words, it is not the invention and commercialisation of new products and/or processes *per se* that brings major benefits at the industry- or economy-wide level but, rather, their widespread use.

Despite these arguments, recent technology policy initiatives in the UK, such as the White Paper 'Realising Our Potential' [Cabinet Office (1993)] and the consultation document 'Innovating for the Future: Investing in R&D' [DTI (1998)], have largely bypassed opportunities to improve the diffusion process. Instead, policy-makers have focused their attention on improving the inventive and innovation stages of technical change. Consequently, the main objectives of current UK technology policy centre on

improving those infrastructures deemed important in facilitating knowledge transfers [OST (1996) and DTI (1998)].¹

The absence of explicit diffusion policies across most OECD countries is also paralleled by the current state of the academic literature. Economists have invested significant resources into the economic analysis of R&D and consequently the literature in this area is now wide ranging and extensive [see, for example, Griliches (1995) and Cohen (1995)]. In contrast, the literature on diffusion is relatively small and arguably fragmented. This aspect has been most recently noted by Keely and Quah (1998) who argue in the context of the economics of growth that:

... we think there has been over-emphasis here on the supply side of the economy, i.e. technology to push back the functions of the production function. Instead the consumption and diffusion of new technology are arguably just as important. [Keely and Quah (1998), p. 17].

As stated in Chapter 1, there are arguably two main weaknesses with the current literature. Firstly, the vast majority of the empirical literature has investigated the diffusion of new technology in the industrial sector with little attention being paid to diffusion in the service sector. Secondly, empirical modelling of the diffusion process has arguably lagged behind advances in the theoretical literature.

The former of these weaknesses seems paradoxical and somewhat surprising given that mature economies are becoming increasingly dominated by the production of services. Indeed, current research themes have focused on the problems of measuring productivity gains in these sectors [see, for example, Colwell and Davis (1992) and Griliches (1997)] and the associated consequences of knowledge production becoming *the* critical factor input in economic growth [Quah (1997) and Keely and Quah (1998)].²

These shifts in the economic structure of mature economies are reflected for the UK economy in Figures 8.1 and 8.2 below. These figures indicate the increasing proportion

¹ As noted by Gourlay *et al* (1997c) these have encompassed the transfer of basic research from universities to private industry, the tax treatment of R&D, improving skills (i.e. human capital) and altering the arrangements for intellectual property.

of labour employed in service industries and in the finance, banking and insurance sector respectively. Employment changes in these sectors are also paralleled by concomitant increases in their relative importance in overall economic activity. Papers by OST (1997) and HM Treasury (1998), for example, has estimated that the contribution of the financial services to GDP at the end of 1993 was almost 7%, approximately double the estimated contribution at the end of 1979. Furthermore, Anderton (1995), Goodacre and Tonks (1995) and Akçaoglu (1996) have emphasised that the process of financial intermediation plays a crucial role in directing funds for innovative activity. Thus, it would appear that more attention should be paid to the analysis of new technology diffusion in the service sector and, in particular, to diffusion in the financial sector.

The blame for the lack of economic interest in new technology diffusion in the service sector does not, however, lie entirely with economists. The paucity of economic research in this area may be partially explained by the distinct shortage of data pertaining to the adoption behaviour of individual firms and institutions. There is a shortage of not only microdata sets containing information related to individual adoption behaviour but also of more highly aggregated data sets pertaining to the economy-wide diffusion. Consequently, researchers into innovation diffusion have often relied on case studies and qualitative data.

Although qualitative data sets provide an interesting and indispensable insight into observed adoption behaviour and diffusion patterns, problems may arise since the results are consequently firm-specific and, therefore, any generalisation made from their analysis should be treated with caution. Moreover, the information gained in these studies may not take account of retrospective information and the behaviour of other firms so that testing of theoretical models is extremely problematic.

The aim of this thesis has been broadly to redress this imbalance by exploring the inter-firm diffusion of ATMs in the UK financial sector for an annual panel data set of

² Most importantly, knowledge production forms the basis of so-called endogenous growth theories of economic growth [see Aghion and Howitt (1998)].

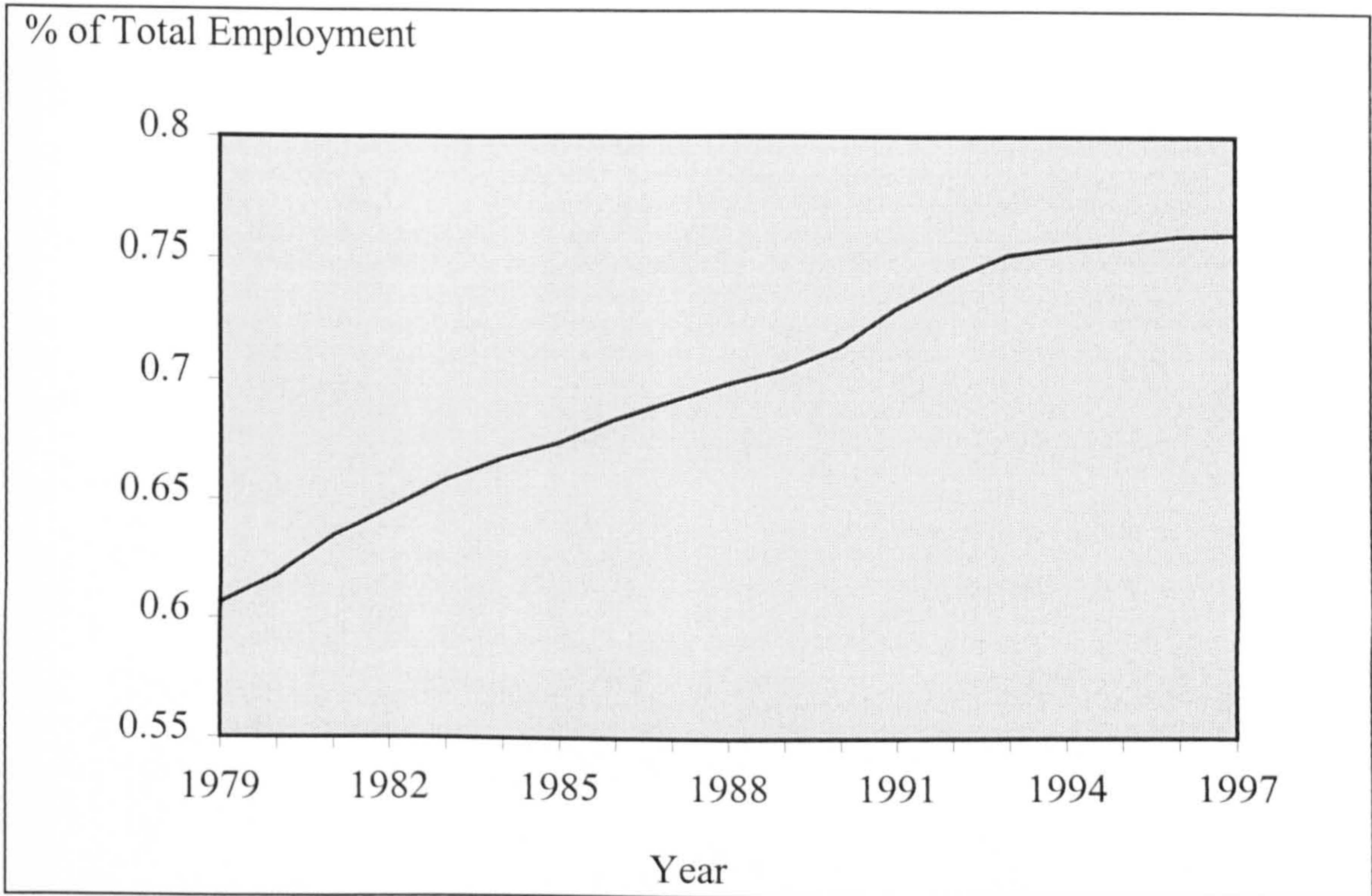


Figure 8.1: UK Total Employment in Service Industries³

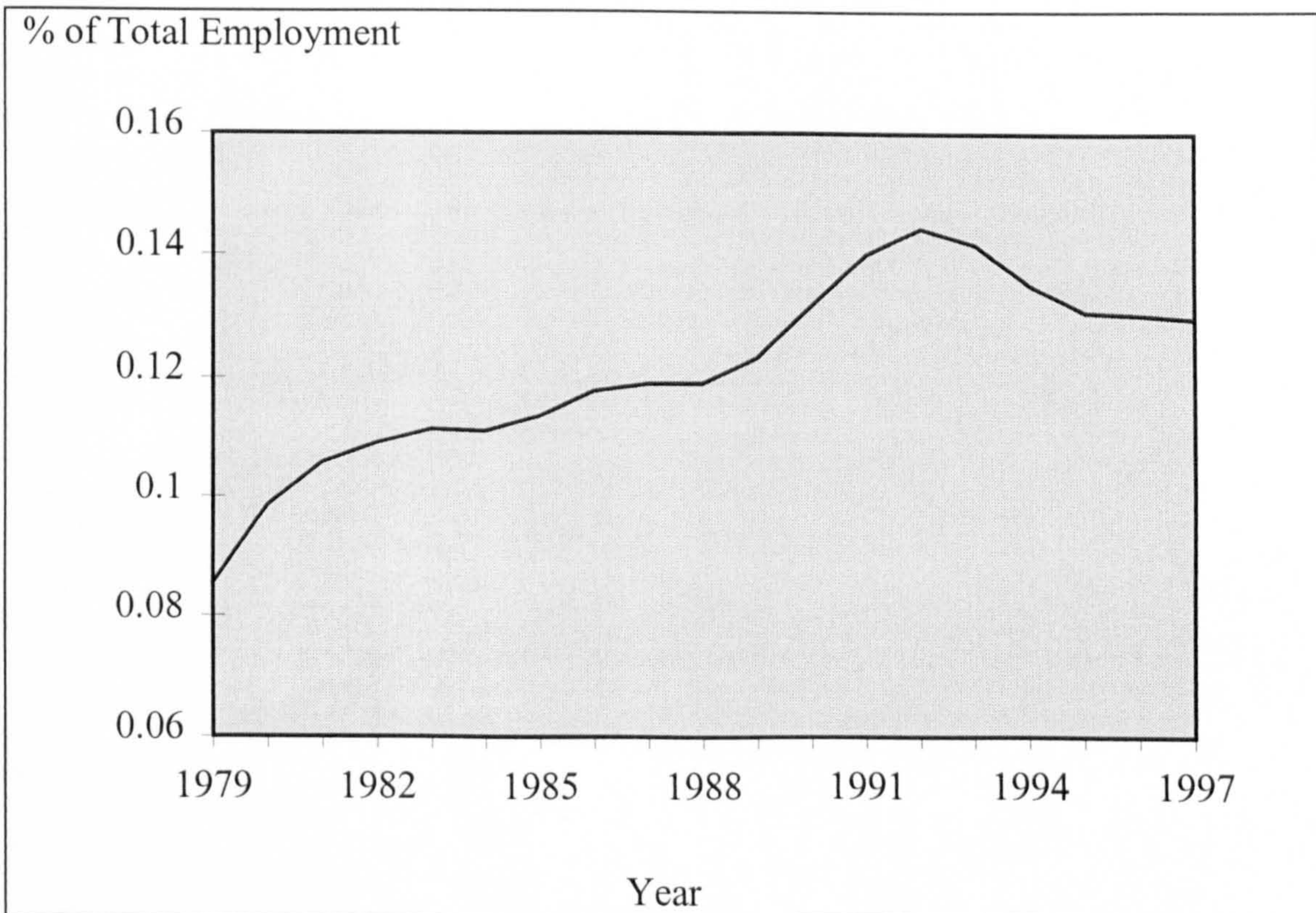


Figure 8.2: UK Total Employment in Banking, Finance and Insurance

³ The data presented in Figures 8.1 and 8.2 are taken from Table 3.1 in *Social Trends (1990-1997)* and are measured at end-year figures. From 1981 to 1985 figures pertain to 1980 SIC and from 1986 pertain to 1990 SIC definitions of relevant sectors.

potential adopters. To be specific, the series of models presented have explored the microeconomic aspects of an individual institution's incentive to adopt ATMs. The modelling approach has been explicitly set within a duration framework in order to more effectively capture the time-intensive nature of the diffusion process and to incorporate censored observations. The methodology intentionally embraced the neo-classical approach to innovation diffusion [see Sarkar (1998)] in which individual institutions are assumed to be profit maximisers and choose an optimal time to adopt determined by institution- and market-specific characteristics.

The empirical analysis presented in Chapter 5 suggests that the diffusion of ATMs has been characterised by non-monotonic duration dependence with the conditional probability initially increasing over time and then decreasing as the diffusion process slows. The analysis conducted in Appendix A5.3 illustrates that the empirical results can be interpreted as lending support to the existence of epidemic-type effects over the diffusion path. Furthermore, the results indicate that previous empirical models that assume monotonic dependence may be mis-specified and thus lends support to the contention that economists need to be more rigorous in their testing of parametric forms.

The empirical results pertaining to the influence of institution-specific and market-specific characteristics (or covariates) on the conditional probability of adoption are presented in Chapter 6. The results indicate that rank and epidemic effects and expectations formed on the price of technology and the future number of adopters have played a significant role in the diffusion of ATMs. Moreover, these effects are found to have the correct directional effect on the conditional probability of adoption as predicted by the underlying choice-theoretic model and the arguments outlined in Section 6.4.1. No empirical support was found, however, for the existence of stock effects or the price of technology on the timing of adoption. Given the analysis of sample attrition in Appendix A5.1 of Chapter 5, this latter result suggests that ATM technology is non-drastic and has had insignificant effects on the cost structures of financial institutions.

The nature of these results highlight two further potential areas of interest in which future research should arguably be directed. Firstly, although some progress has been

made in this thesis in modelling strategic behaviour, there is still along way to go in order to catch-up empirically with the advances made in game-theoretic models. Indeed, testing for the existence of strategic behaviour in the diffusion process remains by and large unexplored. This route may, however, prove to be problematic given the often highly stylised settings of these models and the inability to identify key decision variables. Secondly, there is a need to explore in greater detail both at the theoretical and empirical level the role of expectations in the diffusion process. The model of optimal adoption timing presented in Chapter 6 gives a prominent role to expectations and is therefore an advance on previous research. The model does, however, constrain the nature of expectations formation and is consequently unable to incorporate different hypotheses concerning the type of the expectations regime. It is particularly important to explore this issue because of its implications for the design of technology policy. As noted by Stoneman and Diederer (1994), there is a trade-off in welfare maximising diffusion paths between the benefits of overcoming the market failure of imperfect information on the one hand and the potential adverse effects of supplying more information in retarding diffusion by its impact upon expectations on the other. A greater understanding of expectations formation at the empirical level is therefore essential.

In addition to extending the present research to incorporate the more sophisticated aspects of strategic behaviour and expectations formation, it also appears to be the case that the role of the supply side in the diffusion process is a further aspect to be explored more fully. As noted in Chapter 6, the empirical modelling assumes that the price of technology is an exogenous covariate. This assumption is based on the fact that a large proportion of the supply of ATMs in the UK during the early years of the diffusion process consisted of imports. There has been a partial recognition in the theoretical literature of the role of the supply side in the diffusion of new technology in the 1980s, but this is still an area, which is characterised by slow growth. Karshenas and Stoneman (1995), who survey the theoretical and empirical literature, remark that:

... there is still a considerable amount of work to be undertaken on the treatment of diffusion as the result of supply demand interaction. There is a great gap in the literature on the modelling of the supply side. [Karshenas and Stoneman (1995), p. 292].

Thus, the acknowledgement of the significance of the cost and market structure (and subsequent pricing policy of capital goods producers) pertaining to the supply side may stir interest into this important area of diffusion economics. This potential route may, however, be severely constrained by the paucity of relevant supply side data. This was found to be the case for ATM diffusion in the UK.

The empirical analysis conducted in Chapter 7 indicates that there exist significant positive network externalities in the diffusion of ATM technology. It was shown that financial institutions with a larger number of branches adopt ATMs earlier than those with fewer branches, adjusting for the number of depositors. The source of these positive externalities is derived from the benefits accruing to the depositor. An ATM network is more valuable to depositors when it has many more geographically dispersed ATMs because of the greater (positive) utility they provide. This result was found to be robust under a number of different specifications of the baseline hazard.

The existence of these positive externalities raises an interesting dilemma for policy makers. As shown in Katz and Shapiro (1985, 1986) and Economides (1996), the diffusion of network technologies may suffer from market failure in that the diffusion process may not take-off because early adopters will not find it profitable to install the technology given the size (and their expectations of size) of the network at the date of the decision to install. This implies that if the wider diffusion of ATM technology is perceived as being desirable from a welfare perspective then policy makers should intervene in the diffusion process. As noted by Stoneman and David (1986) and Stoneman and Diederer (1994), the form of this intervention depends crucially on the price elasticity of supply of the new technology. If the price elasticity is relatively low then an alternative approach to subsidies is to set technical standards early on in the diffusion process. In the case of ATM technology, technical standards could encompass packet-switching technology and software compatibility.

As the inter-firm diffusion of ATMs has now appeared to have reached saturation level in the UK, these policy implications are perhaps less relevant to the UK situation. As noted by Akçaoglu (1996), however, diffusion of information technology in developing countries are still in their infancy and the results obtained in Chapters 6 and 7 may provide valuable insights into the design of policies in these countries.

A current theme in the literature [see, for example, Stoneman and Kown (1996)] is the interaction between innovation, adoption timing and firm performance. The relationship between firm performance and the diffusion process is very much under researched and those studies that do exist [Geroski *et al* (1993) and Dos Santos (1995)] tend to focus their attention on narrowly defined financial indicators of performance (such as market share and profitability) and ignore the technical efficiency aspects of adoption timing. Given the stress placed on increasing technical efficiency in the financial sector in the paper by OST (1997), this area also presents abundant research opportunities.

To summarise, the research conducted in this thesis presents a selection of original empirical results relating to the diffusion of ATMs in the UK financial sector. The results obtained indicate strong support for the existence of rank and epidemic effects and expectations formed on the price of technology and the future number of adopters in the diffusion of ATMs. In addition, evidence was found for the existence of positive externalities present in ATM diffusion. Moreover, these findings suggest that technology policy needs to re-focus its attention on the diffusion of new technology, particularly in the service sector. This thesis provides a valuable contribution to this current debate.

REFERENCES

- Adam, C. S. (1987) *Money Demand and Financial Innovation in the UK 1975-1986*, M.Phil. Dissertation, Nuffield College, Oxford.
- Adam, C. S. (1991) 'Financial innovation and the demand for £M3 in the UK 1975-86', *Oxford Bulletin of Economics and Statistics*, 53(4), pp. 401-424.
- Aghion, P. and P. Howitt (1998) *Endogenous Growth Theory*, MIT Press, Cambridge, Massachusetts.
- Agliardi, E. (1995) 'Discontinuous adoption paths with dynamic scale economies', *Economica*, 62, pp. 541-549.
- Ahmad, S. (1966) 'On the theory of induced invention', *The Economic Journal*, 76, pp. 344-357.
- Akçaoglu, E. (1996) *Financial Innovation in Turkish Banking*, M.Phil. Dissertation, Department of Economics, Loughborough University.
- Akhtar, M. A. (1983) 'Financial innovations and their implications for monetary policy: an international perspective', *BIS Economic Paper No 9*, Bank for International Settlements, Basle.
- Alchain, A. A. (1950) 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, 58, pp. 211-221.
- Alderman, N. and S. Davies (1990) 'Modelling regional patterns of innovation diffusion in the UK metalworking industries', *Regional Studies*, 24(6), pp. 513-528.
- Amemiya, T. (1981) 'Qualitative response models: a survey', *Journal of Economic Literature*, 19, pp. 1483-1536.
- Amemiya, T. (1986) *Advanced Econometrics*, Basil Blackwell, Oxford.
- Amendola, M. and J-L, Gaffard (1988) *The Innovative Choice: An Economic Analysis of the Dynamics of Technology*, Basil Blackwell, Oxford.
- Andersen, E. S. (1994) *Evolutionary Economics: Post-Schumpeterian Contributions*, Pinter, London.
- Andersen, E. S. (1997) 'The Schumpeterian trade-off in an evolutionary model of economic growth and development', paper presented at the Competition and

- Industrial Dynamics Conference, Danish Unit for Industrial Dynamics, Skagen, March.
- Anderton, B.** (1995) *Current Issues in Financial Services*, Macmillan, London.
- Anderton, B.** (1995) *Current Issues in Financial Services*, Macmillan, London.
- Antonelli, C.** (1990) 'Induced adoption and externalises in the regional diffusion of information technology', *Regional Studies*, 24(1), pp. 31-40.
- Antonelli, C., P. Petit and G. Tahar** (1992) *The Economics of Industrial Modernization*, Academic Press, London.
- Arestis, P., G. Hadjimatheous and G. Zis** (1992) 'The impact of financial innovations on the demand for money in the UK and Canada', *Applied Financial Economics*, 2, pp. 115-123.
- Arestis, P., I. B-F. Mariscal, H. Hagemann, P. Howells and H-M. Trautwein** (1993) 'Financial innovation and the demand for money in the United Kingdom and in West Germany: a comparative application of cointegration and error-correction modelling', *Working Paper No 83/93*, Universität Hohenheim.
- Arrow, K. J.** (1962a) 'Economic welfare and the allocation of resources for invention', in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER, Princeton University Press, Princeton.
- Arrow, K. J.** (1962b) 'The economic implications of learning by doing', *Review of Economic Studies*, 31, pp. 155-173.
- Arthur, W. B.** (1988) 'Self-reinforcing mechanisms in economics', in P. W. Andersen, K. J. Arrow, D. Pines (editors), *The Economy as an Evolving Complex System*, Addison-Wesley, Redwood City.
- Arthur, W. B.** (1989) 'Competing technologies, increasing returns and lock-in by historical events', *The Economic Journal*, 99, pp. 116-131.
- Arthur, W. B., Y. M. Ermoliev and Y. Kaniovski** (1987) 'Path-dependent processes and the emergence of macro-structure', *European Journal of Operational Research*, 30, pp. 294-303.
- Arulampalam, W. and M. B. Stewart** (1995) 'The determinants of individual unemployment durations in an era of high unemployment', *The Economic Journal*, 105, pp. 321-332.
- Ash, N. I.** (1989) 'Societies relying on more new technology', *Building Societies Association (BSA) Yearbook*, BSA, London.

- ✓ Association for Payment Clearing Services (APACS) (1997) *Yearbook of Payment Statistics*, APACS Statistical Unit, London.
- Audretsch, D. B. and M. P. Feldman (1996) 'R&D spillovers and the geography of innovation and production', *American Economic Review*, 86(3), pp. 630-640.
- Audretsch, D. B. and P. E. Stephan (1996) 'Company-scientist locational links: the case of biotechnology', *American Economic Review*, 86(3), pp. 641-652.
- Audretsch, D. B. and T. Mahmood (1995) 'New firm survival: new results using a hazard function', *Review of Economics and Statistics*, 77(1), pp. 97-103.
- Austin, D. (1992) 'Needles and PINs', *Banking Technology*, November, pp. 59-63.
- Ayres, R. U. and I. Ezekoye (1991) 'Competition and complementarity in diffusion: the case of octane', *Technological Forecasting and Social Change*, 39(1-2), pp. 145-158.
- Baba, Y. and S. Takai (1990) 'Information technology introduction in the big banks: the case of Japan', in C. Freeman and L. Soete (editors), *New Explorations in the Economics of Technical Change*, Pinter, London.
- Baba, Y., D. F. Hendry and R. M. Starr (1992) 'The demand for M1 in the USA, 1960-1988', *Review of Economic Studies*, 59(1), pp. 25-61.
- Bailey, N. (1957) *Mathematical Theories of Epidemics*, Griffin, London.
- Bain, A. D. (1964) *The Growth of Television Ownership in the UK since the War: a lognormal Model*, Cambridge University Press, Cambridge.
- Balcer, Y. and S. A. Lippman (1984) 'Technological expectations and adoption of improved technology', *Journal of Economic Theory*, 34(2), pp. 292-318.
- Bank of England (1983) 'The nature and implications of financial innovation', *Bank of England Quarterly Bulletin*, June, Bank of England.
- Bank of England (1986) 'The UK approach to financial regulation', *Bank of England Quarterly Bulletin*, 26(1), pp. 48-50.
- Bank of England (1997) *Statistical Abstract*, Bank of England, London.
- Banking Information Service (1982) *The Banks and Information Technology*, British Bankers' Association, London.
- Banking Technology (1986) 'The age of the automated teller', *Banking Technology*, November, pp. 14-19.
- Banking World (1984) '1984 survey of ATM installations', *Banking World*, November, pp. 48-49.

- Banking World** (1987) 'High street banks' ATM installations', *Banking World*, pp. 48-53.
- Banking World** (1988) 'The high street banks' ATMs', *Banking World*, November, pp. 52-58.
- Banking World** (1989) 'The ATM market comes of age', *Banking World*, November, pp. 35-43.
- Banking World** (1990) 'ATM 1990 - a year of consolidation', *Banking World*, November, pp. 33-41.
- Banking World** (1991) 'Another year of consolidation', *Banking World*, November, pp. 28-35.
- Banking World** (1992) 'The annual ATM report', *Banking World*, November, pp. 26-34.
- Bantel, K. A. and S. E. Jackson** (1989) 'To management and innovations in banking: does the composition of the top team make a difference?', *Strategic Management Journal*, 10, pp. 107-124.
- Barlow, R. E. and F. Proschan** (1960) 'A note on tests for monotone failure rate based on incomplete data', *Annals of Mathematical Statistics*, 40(2), pp. 595-600.
- Barthel, M.** (1992) 'ATM growth put a lid on branch costs', *American Banker*, 12, pp. 1-3.
- Barthel, M.** (1993) 'ATM fees rise for usage at customer's bank', *American Banker*, 9, pp. 1-17.
- Bass, F. M.** (1969) 'A new product growth for model consumer durables', *Management Science*, 15(5), pp. 215-227.
- Baumol, W. J.** (1952) 'The transactions demand for cash: an inventory theoretic approach', *Quarterly Journal of Economics*, 66(4), pp. 545-556.
- Baumol, W. J.** (1982) 'Contestable markets: an uprising in the theory of industry structure', *American Economic Review*, 72(1), pp. 1-15.
- Beard, T. R., S. B. Caudill and D. M. Gropper** (1997) 'The diffusion of production processes in the US banking industry: a finite mixture approach', *Journal of Banking and Finance*, 21(5), pp. 721-740.
- Beath, J., Y. Katsoulacos and D. Ulph** (1995) 'Game-theoretic approaches to modelling of technological change', in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.

- Bell, M. and K. Pavitt** (1993) 'Technological accumulation and industrial growth: contrasts between developed and developing countries', *Industrial and Corporate Change*, 2, pp. 157-210.
- Belleflamme, P.** (1998) 'Adoption of network technologies in oligopolies', *International Journal of Industrial Organization*, 16, pp. 415-444.
- Benvignati, A. M.** (1982) 'Interfirm adoption of capital-goods innovations', *Review of Economics and Statistics*, 64(2), pp. 330-335.
- Berger, A.** (1985) 'The economics of electronic funds transfers', *Working Paper No 85/11*, Board of Governors of the Federal Reserve System, Washington.
- Berry, L. L. and J. H. Donnelly** (1975) *Marketing For Bankers*, American Institute of Banking, New York.
- Binswanger, H. P.** (1953) 'Induced technical change: evolution of thought', in H. Binswanger and V. Ruttan (editors), *Induced Innovation*, John Hopkins University Press, Baltimore.
- Binswanger, H. P.** (1978) 'The microeconomics of induced technical change', in H. Binswanger and V. Ruttan (editors), *Induced Innovation*, John Hopkins Press, Baltimore.
- BIS** (1986) *Recent Innovations in International Banking*, Bank of International Settlements (BIS) Publication, Basle.
- Blaug, M.** (1963) 'A survey of the theory of process-innovations', *Economica*, 63, pp. 13-32.
- Blundell, R., R. Griffith and J. V. Reenen** (1995) 'Dynamic count data models of technological innovation', *The Economic Journal*, 105, pp. 333-344.
- Blundell, R., R. Griffith and J. Van Reenan** (1994) 'Dynamic count data model of technological innovation', *Institute of Fiscal Studies Paper No. W94/10*, London.
- Blundell-Wignall, A., F. Browne P. Manasse** (1990) 'Monetary policy in the wake of financial liberalisation', *Working Paper No 77*, Organisation of Economic Co-operation and Development (OECD), Paris.
- Boleat, M.** (1987) 'The competitor's view', *Banking World*, November, pp. 33-34.
- Boleat, M.** (1992) 'Building societies: competition and innovation in the 1990s' in K. Ray (editor), *New Issues in Financial Services*, Blackwell, Oxford.

- Bonus, H.** (1973) 'Quasi-Engle curves, diffusion, and the ownership of major consumer durables', *Journal of Political Economy*, **81**, pp. 655-677.
- Bosworth, D. L.** (1983) *The Employment Consequences of Technical Change*, Macmillan, London.
- Bosworth, D. L. and R. Wilson** (1980) 'Returns to scale in R&D: empirical evidence from the world chemical industry', *Discussion Paper No 80*, CIEBR, University of Warwick, Coventry.
- Bosworth, D. L., C. Greenhalgh, P. Hayward and M. Longland** (1996) 'Opening the black box: developments in the measurement of innovation activity for the econometric modelling of firm performance', Paper Presented at the Economics and Econometrics of Innovation Conference, Strasbourg.
- Bosworth, D. L., P. Dawkins and T. Stromback** (1996) *The Economics of the Labour Market*, Longman, Harlow.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter** (1978) *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley, New York.
- Breedon, F. J. and P. G. Fisher** (1993) M0: Causes and Consequences, *Working Paper Series No 20*, Bank of England, London.
- Bresnahan, T. and P. David** (1986) 'The diffusion of automated teller machines across US banks', Paper presented at the Innovation Diffusion Conference, Venice, March.
- Bridges, E., Coughlan, A. T. and S. Kalish** (1991) 'New technology adoption in an innovative market-place: micro- and macro-level decision making models', *International Journal of Forecasting*, **7**, pp. 257-270.
- British Bankers' Association (BBA)** (1985) *Annual Abstract of Banking Statistics*, British Bankers' Association, London.
- British Bankers' Association (BBA)** (1996) *Annual Abstract of Banking Statistics*, British Bankers' Association, London.
- Building Societies Association (BSA)** (1993) *Building Societies Yearbook*, Building Societies Association, London.
- Building Societies Association (BSA)** (1996) *Building Societies Yearbook 1995-1996*, Building Societies Association, London.

- Building Societies Association (BSA) (1997)** *Building Societies Yearbook 1996-1997*, Building Societies Association, London.
- Burmeister, E. and A. R. Dobell (1970)** *Mathematical Theories of Economic Growth*, Macmillan, London.
- Cabinet Office (1993)** *Realising Our Potential, A Strategy for Science, Engineering and Technology*, HMSO, London.
- Cabral, L. (1990)** 'On the adoption of innovations with network externalities', *Mathematical Social Sciences*, 19, pp. 229-308.
- Cameron, A. and P. Trivedi (1986)** 'Econometric models based on count data: comparisons and applications of some estimators and tests', *Journal of Applied Econometrics*, 1, pp. 29-53.
- Cameron, H. M. and J. S. Metcalfe (1988)** 'On the economics of technological substitution', *Technological Forecasting and Social Change*, 32, pp. 147-162.
- Carlsson, B. and S. Jacobsson (1994)** 'Technological systems and economic policy: the diffusion of factory automation in Sweden', *Research Policy*, 23(3), pp. 235-248.
- Carlton, D. and M. J. Klammer (1983)** 'The need for co-ordination among firms, with special reference to network industries', *University of Chicago Law Review*, 50, pp. 446-465.
- Carroll, R. and D. Ruppert (1984)** 'Power transformations when fitting theoretical models to data', *Journal of the American Statistical Association*, 79(6), pp. 321-328.
- Carroll, R. and D. Ruppert (1988)** *Transformations and Weighting in Regression*, Chapman and Hall, New York.
- Central Statistical Office (1997)** *Financial Statistics*, HMSO, London. *****
- Cesarano, F. (1990)** 'Financial innovation and demand for money: some empirical evidence', *Applied Economics*, 22(10), pp. 1437-1442.
- Chesher, A. (1983)** 'The information matrix test - simplified calculation via a score test interpretation', *Economics Letters*, 13(1), pp. 45-48.
- Chesher, A. (1984)** 'Testing for neglected heterogeneity', *Econometrica*, 52(4), pp. 865-872.
- Chiang, A. C. (1984)** *Fundamental Methods of Mathematical Economics*, Third Edition, McGraw-Hill, Singapore.

- Choi, J. P.** (1994) 'Irreversible choice of uncertain technologies with network externalities', *Rand Journal of Economics*, 25(3), pp. 382-401.
- Choi, J. P. and M. Thum** (1998) 'Market structure and the timing of technology adoption with network externalities', *European Economic Review*, 42(2), pp. 225-244.
- Chow, G.** (1967) 'Technological change and the demand for computers', *American Economic Review*, 57, pp. 1117-11130.
- Chung, C-F., P. Schmidt and P. D. Witte** (1991) 'Survival analysis: a survey', *Journal of Qualitative Criminology*, 7(1), pp. 59-98.
- Church, J. and N. Gandal** (1992) 'Network effects, software provision and standardization', *Journal of Industrial Economics*, 40(1), pp. 85-104.
- Church, J. and N. Gandal** (1993) 'Complementary network externalities and technological adoption', *International Journal of Industrial Organisation*, 11, pp. 239-260.
- Cochran, S. J. and R. H. Defina** (1995) 'Duration dependence in the US stock market cycle: a parametric approach', *Applied Financial Economics*, 5(5), pp. 309-318.
- Cohen, W.** (1995) 'Empirical studies of innovative activity', in P. Stoneman (editor), *The Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Coldwell, R. J. and E. P. Davis** (1992) 'Output, productivity and externalities - the case of banking', *Working Paper Series No 3*, Bank of England, London.
- Coldwell, R. J. and E. P. Davis** (1992) 'Output, productivity and externalities - the case of banking', *Working Paper Series No. 3*, Bank of England, London..
- Colombo, M. G. and R. Mosconi** (1995) 'Complementary and cumulative learning effects in the early diffusion of multiple technologies', *The Journal of Industrial Economics*, 43(1), pp. 13-48.
- Conover, W. J.** (1971) *Practical Nonparametric Statistics*, John Wiley, New York.
- Cooke, I. and P. Mayes** (1998) *Introduction to Innovation and Technology Transfer*, Artesch House, Boston.
- Coombs, R., P. Saviotti, and V. Walsh** (1987) *Economics and Technical Change*, Macmillan, Basingstoke.
- Cowan, R.** (1990) 'Nuclear power reactors: a study in technological lock-in', *Journal of Economic History*, 50, pp. 541-567.

- Cowan, R.** (1991) 'Tortoises and hares: choice among technologies of unknown merit', *The Economic Journal*, 101, pp. 801-814.
- Cox, D. R.** (1972) 'Regression models and life-tables', *Journal of the Royal Statistical Society*, 34, pp. 187-220.
- Cox, D. R.** (1975) 'Partial likelihood', *Biometrika*, 62, pp. 269-276.
- Cox, D. R. and D. Oakes** (1984) *Analysis of Survival Data*, Chapman and Hall, New York.
- Cramer, J. S.** (1969) *Empirical Econometrics*, North Holland, Amsterdam.
- Crampes, C.** (1991) 'Research and development and patent licensing', in J-J. Laffont and M. Moreaux (editors), *Dynamics, Incomplete Information and Industrial Economics*, Basil Blackwell, Oxford.
- Crowder, M. J., A. C. Kimber, R. L. Smith and T. J. Sweeting** (1991) *Statistical Analysis of Reliability Data*, Chapman and Hall, London.
- Cuthbertson, K., S. G. Hall, and M. P. Taylor** (1992) *Applied Econometric Techniques*, Harvester Wheatsheaf, London.
- Cutler, S. J. and F. Ederer** (1958) 'Maximum utilization of the life table in analysing survival', *Journal of Chronic Diseases*, 11, pp. 699-712.
- D'Agostino, R. B. and M. A. Stephens** (1986) *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- Daniels, K. N. and N. B. Murphy** (1994) 'The impact of technical change on the currency behavior of households: an empirical cross-section study', *Journal of Money Credit and Banking*, 26(4), pp. 867-874.
- Dasgupta, P.** (1988) 'Patents, priority and imitation or, the economics of races and waiting games', *The Economic Journal*, 98, pp. 66-80.
- David, P. A.** (1969) 'A contribution to the theory of diffusion', *Center for Research in Economic Growth Memorandum No. 71*, Stanford University.
- David, P. A.** (1975) 'The landscape and the machine: technical interrelatedness, land tenure and the mechanization of the harvest in Victorian Britain', in P. David (editor), *Technical Choice Innovation and Economic Growth*, Cambridge University Press, Cambridge.
- David, P. A.** (1985) 'Clio and the economics of QWERTY', *American Economic Review*, 75(2), Papers and Proceedings, pp. 332-337.
- David, P. A.** (1991) *Behind the Diffusion Curve*, Westview Press, Oxford.

- David, P. A.** (1993) 'Path-dependence and predictability in dynamic systems with local network externalities: a paradigm for historical economics', in D. Foray, C. Freeman (editors), *Technology and the Wealth of Nations*, Pinter, London.
- David, P. A. and J. A. Bunn** (1987) 'The economic of gateway technologies and network evolution: lessons from electricity supply history', CEPR Paper presented at the 15th Annual Telecommunications Policy Conference', Airlie, 28 September.
- David, P. A. and S. Greenstain** (1990) 'The economics of compatibility standards: an introduction to recent research', *Economics of Innovation and New technology*, 1, pp. 3-42.
- David, P. A. and T. E. Olsen** (1986) 'Equilibrium dynamics of diffusion when incremental technological innovations are foreseen', *Ricerche Economiche*, 60(4), pp. 738-770.
- Davies, S.** (1979) *The Diffusion of Process Innovations*, Cambridge University Press, Cambridge.
- de Wit, G. R.** (1990) 'The character of a technological change and employment in banking: a case-study of the Dutch automated clearing house (BGC)', in C. Freeman and L. Soete (editors), *New Explorations in the Economics of Technical Change*, Pinter, London.
- Deaton, A. and J. Muellbauer** (1980) *Economics and Consumer Behaviour*, Cambridge University Press, Cambridge.
- Department of Trade and Industry** (1998) 'Innovating for the future: investing in R&D', Consultation Document, HM Treasury, London.
- Dertouzos, M. L., R. K. Lester and R. M. Solow** (1989) *Made in America: Regaining the Productive Edge*, MIT Press.
- Desai, M. and W. Low** (1987) 'Measuring the opportunity for product innovation', in M. De Cecco (editor), *Changing Money - Financial Innovation in Developed Countries*, Basil Blackwell, Oxford.
- Diederer, P. J. M., R. P. M. Kemp, J. Muysken and G. R. de Wit** (1990) 'Diffusion of information technology in banking: the Netherlands as an illustrative case', in C. Freeman and L. Soete (editors), *New Explorations in the Economics of Technical Change*, Pinter, London.

- Diederer, P. J. M., R. P. M. Kemp, J. Muysken and G. R. de Wit (1991)** 'Diffusion of process technology in Dutch banking', *Technological Forecasting and Social Change*, 39(1-2), pp. 201-219.
- Dixit, A. K. (1990)** *Optimisation in Economic Theory*, second edition, Oxford University Press, Oxford.
- Dixit, A. K. (1992)** 'Investment and hysteresis', *Journal of Economic Perspectives*, 6, pp. 107-132.
- Dixit, A. K. (1993)** *The Art of Smooth Pasting*, Harwood, Churchill.
- Dixit, A. V. and R. S. Pindyck (1994)** *Investment Under Uncertainty*, Princeton University Press, Princeton.
- Dixon, R. (1980)** 'Hybrid corn revisited', *Econometrica*, 48(6), pp. 1451-1461.
- Dodgson, M. and J. Bessant (1996)** *Effective Innovation Policy*, International Thomson Business Press, London.
- Dodgson, M. and R. Rothwell (1994)** *The Handbook of Industrial Innovation*, Edward Elgar, Aldershot.
- Dodson, J. A. and E. Muller (1978)** 'Models of new product diffusion through advertising and word-of-mouth', *Management Science*, 24(15), pp. 1568-1578.
- Doessel, D. P. and S. M. Strong (1991)** 'A neglected problem in the analysis of the diffusion process', *Applied Economics*, 23(8), pp. 1335-1340.
- Dooley, M. P. and F. Spinelli (1989)** 'The early stages of financial innovation and money demand in France and Italy', *Manchester School*, 57(2), pp. 107-124.
- Dos Santos, B. L. and K. Peffers (1995)** 'Rewards to investors in innovative technology applications: first movers and early followers in ATMs', *Organizational Science*, 6(3), pp. 241-259.
- Dosi, G. (1983)** 'Technological paradigms and technological trajectories', in C. Freeman (editor), *Long Waves in the World Economy*, Butterworths, London.
- Dosi, G. (1988)** 'Sources, procedures, and microeconomic effects of innovation', *Journal of Economic Literature*, 26(3), pp. 1120-1171.
- Dosi, G. and L. Orsenigo (1988)** 'Co-ordination and transformation: an overview of structures, behaviours and change in evolutionary environments', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (editors), *Technical Change and Economic Theory*, Pinter, London.

- Dosi, G., C. Freeman, R. Nelson, G. Silverberg and L. Soete (1988) *Technical Change and Economic Theory*, Pinter, London.
- Dosi, G., K. Pavitt and L. Soete (1990) *The Economics of Technical Change and International Trade*, Harvester Wheatsheaf, London..
- Dotsey, M. (1988) 'The demand for currency in the United States', *Journal of Money Credit and Banking*, 20(1), pp. 22-40.
- Douchet, P. and P. B. Sloep (1992) *Mathematical Modeling in the Life Sciences*, Ellis Horwood, London.
- Drake, L. (1986) 'Economies of scale in the building society industry', *Research Paper No 26*, Banking Centre, Loughborough University.
- Drake, L. (1989) *The Building Society Industry in Transition*, Macmillan, Basingstoke.
- Drake, L. (1992) 'Economies of scale and scope in the UK building societies: an application of the translog multiproduct cost function', *Department of Economics Research Paper No 92/10*, Loughborough University.
- Duca, J. V. and W. C. Whitsell (1995) 'Credit cards and money demand: a cross-sectional study', *Journal of Money, Credit and Banking*, 27(2), pp. 604-623.
- Economides, N. (1993) 'Mixed bundling in duopoly', *Discussion Paper EC-93-29*, Stern School of Business, New York University.
- Economides, N. (1995) 'Features of credit card networks', *Federal Reserve Bank of St. Louis*, November-December, pp. 60-63.
- Economides, N. (1996) 'The economics of networks', *International Journal of Industrial Organization*, 14(6), pp. 673-699.
- Economides, N. and C. Himmelberg (1995) 'Critical mass and network size with application to the US fax market', *Discussion Paper No. EC-95-11*, Stern School of Business, New York University.
- Economides, N. and L. J. White (1993) 'One-way networks, two-way networks, compatibility, and anti-trust', *Discussion Paper EC-93-14*, Stern School of Business.
- Economides, N. and L. J. White (1994) 'Networks, two-way networks, compatibility and anti-trust', *Discussion Paper EC-93-14*, Stern School of Business.
- Economides, N. and S. C. Salop (1992) 'Competition and integration among complements and network market-structure', *Journal of Industrial Economics*, 40(1), pp. 105-123.

- Economides, N. and W. Lehr (1995)** 'The quality of complex systems and industry structure', in W. Lehr (editor), *Quality and Reliability of Telecommunications Infrastructure*, Lawrence Erlbaum, Hillsdale, New Jersey.
- Edmonds, R. G. and J. B. Meisel (1992)** 'Explaining rates of technology diffusion: a need for caution', *Quarterly Review of Economics and Finance*, 32(2), pp. 85-99.
- Edquist, C. (1997)** *Systems of Innovation: Technologies, Institutions and Organizations*, Pinter, London.
- Elliott, R. F. (1991)** *Labor Economics: A Comparative Text*, McGraw-Hill, Maidenhead.
- Ellis, R. P. and Zimmerman, M. B. (1983)** 'What happened to nuclear-power: a discrete choice model of technology adoption', *Review of Economics and Statistics*, 65(2), pp. 234-242.
- Elysiyani, E. and S. Mehdian (1990)** 'Efficiency in the commercial banking industry, a production frontier approach', *Applied Economics*, 22, p. 539-551.
- Engel, R. F., D. F. Hendry and J. F. Richard (1983)** 'Exogeneity', *Econometrica*, 51, pp. 277-304.
- Epstein, B. (1960a)** 'Testing for the validity of the assumption that the underlying distribution of life is exponential: part I', *Technometrics*, 2(1), pp. 83-101.
- Epstein, B. (1960b)** 'Tests for the validity of the assumption that the underlying distribution of life is exponential: part II', *Technometrics*, 2(1), pp. 167-183.
- Ernst, H. (1997)** 'The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry', *Small Business Economics*, 9(4), pp. 361-381.
- European Commission (1995)** Green Paper on Innovation, Brussels.
- Evenson, R. (1993)** 'Patents, R&D and invention potential: international evidence', *American Economic Review*, 83(2), pp. 463-468.
- Farell, J. and G. Saloner (1985)** 'Standardization, compatibility, and innovation', *Rand Journal of Economics*, 16(1), pp. 70-83.
- Farell, J. and G. Saloner (1986)** 'Installed base and compatibility - innovation, product preannouncements, and predation', *American Economic Review*, 76(5), pp. 940-955.

- Farrands, C.** (1998) 'Interpretations of the Diffusion and Absorption of Technology', in Talalay, M, C. Farrands and R. Tooze (editors), *Technology, Culture and Competitiveness*, Routledge, London.
- Farrell, J. and G. Saloner** (1985) 'Standardisation, compatibility and innovation', *Rand Journal of Economics*, 16(1), pp. 70-83.
- Farrell, J. and G. Saloner** (1986) 'Installed base and compatibility: innovation, product preannouncements and predation', *American Economic Review*, 76, pp. 940-955.
- Farrell, J. and G. Saloner** (1988) 'Coordination through committees and markets', *Rand Journal of Economics*, 19(2), pp. 235-252.
- Farrell, J. and G. Saloner** (1992) 'Converters, compatibility and the control of interfaces', *Journal of Industrial Economics*, 40(1), pp. 235-252.
- Farzin, Y. H., K. J. M. Huisman and P. M. Kort** (1998) 'Optimal timing of technology adoption', *Journal of Economic Dynamics and Control*, 22, pp. 779-799.
- Faulkner, W. and J. Senker** (1995) *Knowledge Frontiers*, Clarendon press, Oxford.
- Feder, G.** (1982a) 'Adoption of interrelated agricultural innovations: complementarity and impact of risk, scale and credit', *American Journal of Agricultural Economics*, 64, pp. 94-101.
- Feder, G.** (1982b) 'On information and innovation diffusion: a bayesian approach', *American Journal of Economics*, 64, pp. 141-145.
- Field, K.** (1990) 'Production efficiency of British building societies', *Applied Economics*, 22, pp. 415-425.
- Fleming, M. C. and J. G. Nellis** (1994) *Principles of Applied Statistics*, Routledge, London.
- Foray, D. and A. Grubler** (1990) 'Morphological analysis and lock-out of technologies: ferrous casting in France and the FRG', *Research Policy*, 19, pp. 535-550.
- Freeman, C.** (1974) *The Economics of Industrial Innovation*, Penguin Books, Harmondsworth.
- Freeman, C.** (1984) 'Prometheus unbound', *Futures*, 16(5), pp. 494-507.

- Freeman, C.** (1988) 'Diffusion: the spread of new technology to firms, sectors, and nations', in Hererje, A. (editor), *Innovation, Technology, and Finance*, Basil Blackwell, Oxford.
- Freeman, C. and L. Soete** (1987) *Technical Change and Full Employment*, Pinter, London.
- Freeman, C. and L. Soete** (1997) *The Economics of Industrial Innovation*, Third Edition, Pinter, London.
- Friedman, A.** (1974) 'Choosing among s-shaped curves for diffusion processes', *CIEBR Working Paper No 76*, Warwick University, Coventry.
- Fudenberg, D. and J. Tirole** (1985) 'Preemption and rent equalization in the adoption of new technology', *Review of Economic Studies*, 52(3), pp. 383-401.
- Gail, M. H. and J. L. Gastwirth** (1978) 'A scale-free goodness of fit test for the exponential distribution based on the Gini statistic', *Journal of the Royal Statistical Society Part B*, 40(3), pp. 350-357.
- Geroski, P.** (1995) 'Markets for technology: knowledge, innovation and appropriability', in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Geroski, P., S. Machin and J. Van Reenan** (1993) 'The profitability of innovation firms', *Rand Journal of Economics*, 24(2), pp. 198-211.
- Glaister, S.** (1974) 'Advertising policy and returns to scale in markets where information is passed between individuals', *Economica*, 41, pp. 139-156.
- Globerman, S.** (1975) 'Technological diffusion in the Canadian tool and die industry', *Review of Economics and Statistics*, 57, pp. 428-434.
- Goel, K. and D. P. Rich** (1997) 'The adoption of new technologies', *Applied Economics*, 29, pp. 513-518.
- Gold, B.** (1980) 'On the adoption of technological innovations in industry: superficial models and complex decision processes', *Omega*, 8(5), pp. 505-516.
- Gold, B.** (1981) 'Technological diffusion in industry: research needs and shortcomings', *The Journal of Industrial Economics*, 29(3), pp. 247-269.
- Goldfield, S. M.** (1989) 'Demand for money: empirical studies', in J. Eatwell, M. Milgate and P. Newman (editors), *Money*, The New Palgrave Series, Macmillan, Basingstoke.

- Goodacre, A. and I. Tonks (1995)** 'Finance and technological change', in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Goodhart, C. A. E. (1986)** 'Financial innovation and monetary control', *Oxford Review of Economic Policy*, 2(4), pp. 79-102.
- Gort, M. and S. Klepper (1982)** 'Time paths in the diffusion of product innovations', *The Economic Journal*, 92, pp. 630-653.
- Goureroux, C., A. Monfort and A. Trongnon (1984a)** 'Pseudo maximum likelihood methods: application to Poisson models', *Econometrica*, 52(3), pp. 701-720.
- Goureroux, C., A. Monfort and A. Trongnon (1984b)** 'Pseudo maximum likelihood methods: theory', *Econometrica*, 52(3), pp. 681-700.
- Gourlay, A. R. (1998a)** 'The market for university generated intellectual property and national systems of innovation: evidence from the UK, USA and Germany', paper presented at the *Intellectual Property in University-Industry Relations Seminar*, National Institute of Economic and Social Research, London.
- Gourlay, A. R. (1998b)** 'Duration dependence and the diffusion of new technology: the case of ATMs', *Economic Research Paper No 98/2*, Department of Economics, Loughborough University.
- Gourlay, A. R. (1998c)** 'The determinants of technology diffusion: evidence from the UK financial sector 1972 to 1992', *Economic Research Paper No 98/5*, Department of Economics, Loughborough University.
- Gourlay, A. R. (1998d)** 'Network Externalities and the Diffusion of ATMs', *Economic Research Paper Forthcoming*, Department of Economics, Loughborough University.
- Gourlay, A. R., J. Seaton and T. Weyman-Jones (1998)** 'Core competence and transaction costs in technology intensive organisations', paper presented at the High Technology and Small Firms Conference, University of Twente, Netherlands.
- Gourlay, A. R., T. S. Weyman-Jones and J. Seaton (1997a)** 'Innovation, productivity and growth - analytical ideas and empirical regularities concerning the role of universities', Paper presented at the Intellectual Property Conference, St Peter's College, Oxford.

- Gourlay, A. R., T. S. Weyman-Jones and S. Massey (1997b)** *Engineering and Physical Sciences Research Council Exploitation Audit Final Report*, Loughborough University, Loughborough.
- Gourlay, A. R., S. McCracken, J. Seaton, T. Weyman-Jones (1997c)** 'The decision to undertake research', paper presented at the Annual Western Economic Association International Conference, Seattle, July.
- Gourlay, A. R., S. McCracken, J. Seaton, T. Weyman-Jones (1997d)** 'The management of intellectual property in UK universities: a compendium of results', *Department of Economics Research Paper No. 97/28*, Loughborough University.
- Gourlay, A. R., S. McCracken, J. Seaton, T. Weyman-Jones and K. E. Hargreaves (1998e)** 'The market for university generated intellectual property: current themes and perspectives', presented at the *UK Economics of Law Study Group Meeting*, Loughborough University.
- Grübler, A. (1991)** 'Diffusion: long term patterns and discontinuities', *Technological Forecasting and Social Change*, 39(1-2), pp. 159-180.
- Gravelle, H. and R. Rees (1992)** *Microeconomics*, Second Edition, Longman, London.
- Greenaway, D. (1994)** 'The diffusion of new technologies: editorial note', *The Economic Journal*, 104, pp. 916-917.
- Greene, W. H. (1993)** *Econometric Analysis*, Second Edition, Macmillan, New York.
- Greene, W. H. (1994)** *LIMDEP 6.0: Users Manual*, Econometric Software, New York.
- Greenwood, M. (1926)** *The Natural Duration of Cancer*, Reports of Public Health and Medical Subjects No 33, HMSO, London.
- Griliches, Z. (1957)** 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, 25(4), pp. 501-522.
- Griliches, Z. (1980)** 'Hybrid corn revisited: a reply', *Econometrica*, 48(6), pp. 1463-1465.
- Griliches, Z. (1995)** 'R&D and productivity: econometric results and measurement issues', in P. Stoneman (editor), *The Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Griliches, Z. (1997)** 'The productivity slowdown: is a growing unmeasurable sector the culprit?', Comment, *Review of Economics and Statistics*, 79(3), p. 371.

- Gruber, H.** (1990) 'Product innovation and persistence of leadership: theory with evidence from the semiconductor industry', *Working Paper No 2*, The Economics of Industry Group, London School of Economics.
- Gruber, H.** (1998) 'The diffusion of innovations in protected industries: the textile industry', *Applied Economics*, 30, pp. 77-83.
- Grübler, A.** (1991) 'Diffusion: long term patterns and discontinuities', in N. Nakicenovic and A. Grübler (editors), *Diffusion of Technologies and Social Behaviour*, Springer-Verlag, Berlin.
- Guglielmo, C.** (1996) 'Here come the super-ATMs', *Fortune*, 134(7), p. 232.
- Gujarati, D.** (1988) *Basic Econometrics*, Second Edition, McGraw-Hill, London.
- Guy, K.** (1987) 'The UK tertiary service sector', in C. Freeman and L. Soete (editors), *Technical Change and Full Employment*, Basil Blackwell, Oxford.
- Håkanson, S.** (1974) 'Special presses in paper making', in L. Nasbeth and G. F. Ray (editors), *The Diffusion of New Industrial Processes: An International Study*, Cambridge University Press, Cambridge.
- Hall, B. H.** (1998) 'Innovation and market value', Presented at the NIESR Conference on Productivity and Competitiveness, London, February.
- Hall, S. G., S. G. B. Henry and J. B. Wilcox** (1989) 'The long-run determination of the UK monetary aggregates', *Working Paper No 41*, Bank of England, London.
- Han, A. and J. A. Hausman** (1990) 'Flexible parametric estimation of duration and competing risk models', *Journal of Applied Econometrics*, 5, pp. 1-28.
- Hannan, T. H. and J. M. McDowell** (1984a) 'Market concentration and the diffusion of new technology in the banking firm', *Review of Economics and Statistics*, 66(4), pp. 686-691.
- Hannan, T. H. and J. M. McDowell** (1984b) 'The determinants of technology adoption: the case of the banking firm', *Rand Journal of Economics*, 15(3), pp. 328-335.
- Hannan, T. H. and J. M. McDowell** (1987) 'Rival precedence and the dynamics of technology adoption: an empirical analysis', *Economica*, 54, pp. 155-171.
- Harrod, R. F.** (1948) *Towards a Dynamic Economics: Some Recent Developments of Economic Theory and Their Application to Policy*, Macmillan, London.

- Hausman, J. B., Hal and Z. Griliches** (1984) 'Econometric models for count data with an application to the patents R&D relationship', *Econometrica*, **52(4)**, pp. 909-938.
- Haynes, R. M., S. W. Brown and P. W. Finch** (1990) 'Bridging the gap between ATM myth and reality', *Journal of Retail Banking*, **12(4)**, pp. 37-40.
- Heckman, J. J. and B. Singer** (1984) 'Econometric duration analysis', *Journal of Econometrics*, **24**, pp. 63-132.
- Heckman, J. J. and B. Singer** (1985) 'Social science and duration analysis', in , J. J. Heckman and B. Singer (editors), *Longitudinal Analysis of Labour Market Data*, Cambridge University Press, Cambridge.
- Heckman, J. J. and G. J. Borjas** (1980) 'Does unemployment cause future unemployment? Definitions, questions, and answers from a continuous time model of heterogeneity and state dependence', *Economica*, **47**, pp. 247-283.
- Henderson, R. M. and K. B. Clarke** (1990) 'Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms', *Administrative Science Quarterly*, **35(1)**, pp. 9-30.
- Hendry, D. F. and N. R. Ericsson** (1991) 'Modeling the demand for narrow money in the United Kingdom and the United States', *European Economic Review*, **35(4)**, pp. 833-881.
- Hernes, G.** (1976) 'Diffusion and growth - the non-homogenous case', *Scandinavian Journal of Economics*, **78**, pp. 427-436.
- Hicks, J. R.** (1963) *Theory of Wages*, Second Edition, Macmillan, London.
- Hirsch, B. T. and J. T. Addison** (1986) *The Economic Analysis of Unions: New Approaches and Evidence*, Allen Unwin, London.
- Hodgson, G. M.** (1993) *Economics and Evolution*, Polity Press, London.
- Hodgson, G. M.** (1998) 'Evolutionary and competence based theories of the firm', *Journal of Evolutionary Studies*, **25(1)**, pp. 25-56.
- Hoem, J. M.** (1985) 'Weighting, misclassification, and other issues in the analysis of survey samples of life histories', in J. J. Heckman and B. Singer (editors), *Longitudinal Analysis of Labour Market Data*, Cambridge University Press, Cambridge.

- Horowitz, J. L.** (1992) 'Semiparametric estimation of a regression model with an unknown transformation of the dependent variable', *Working Paper No 92-28, Department of Economics, University of Iowa, Iowa.*
- Horowitz, J. L. and G. R. Neumann** (1992) 'A generalized moments specification test of the proportional hazards model', *Journal of the American Statistical Association*, 87(417), pp. 234-240.
- Howells, J. and J. Hine** (1993) *Innovative Banking: Competition and the Management of a New Networks Technology*, Routledge, London.
- Hsiao, C.** (1986) *Analysis of Panel Data*, Cambridge University Press, Cambridge.
- Humphrey, D. B.** (1994) 'Delivering deposit services: ATMs versus branches', *Federal Reserve Bank of Richmond Quarterly*, 80, pp. 59-81.
- Independent** (1997) 'ATMs Enter a New Age', *Independent*, 1 December.
- Ireland, N. J. and P. Stonemam** (1985) 'Order effects, perfect foresight and intertemporal price discrimination', *Recherches Economiques de Louvain*, 51(1), pp. 7-20.
- Ireland, N. J. and P. Stoneman** (1986) 'Technological diffusion, expectations and welfare', *Oxford Economic Papers*, 38(2), pp. 283-304.
- Ireland, P. N.** (1995) 'Endogenous financial innovation and the demand for money', *Journal of money Credit and Banking*, 27(1), pp. 107-123.
- Issidoridis, G.** (1986) 'Optimal diffusion: a theoretical and empirical analysis of the diffusion of innovations', D.Phil. Thesis, University of Sussex.
- J. Johnston** (1987) *Econometric Methods*, Third Edition, McGraw Hill, Singapore.
- Jaggia, S.** (1991) 'Tests of moment restrictions in parametric duration models', *Economics Letters*, 37(1), pp. 35-38.
- Jensen, R.** (1982) 'Adoption and diffusion of an innovation of uncertain profitability', *Journal of Economic Theory*, 27(1), pp. 182-193.
- Jensen, R.** (1983) 'Innovation adoption and diffusion where there are competing innovations', *Journal of Economic Theory*, 29(1), pp. 161-171.
- Jensen, R.** (1984) 'Adoption of an innovation of uncertain profitability with costly information', *Working Paper 84-8*, Department of Economics, Ohio State University.
- Jensen, R.** (1988a) 'Information capacity and innovation adoption', *International Journal of Industrial Organization*, 6(3), pp. 335-350.

- Jensen, R. (1988b) 'Information cost and innovation adoption policies', *Management Science*, 34(2), pp. 230-239.
- Jensen, R. (1992) 'Innovation adoption and welfare under uncertainty', *The Journal of Industrial Economics*, 40(2), pp. 173-180.
- Johansen, S. (1978) 'The product limit estimator as a maximum likelihood estimator', *Scandinavian Journal of Statistics*, 5, pp. 195-199.
- Johnston, R. B. (1984) 'The demand for non-interest bearing money in the United Kingdom', *Working Paper No 28*, HM Treasury, London.
- Jones, D. (1981) 'ATMs invade the high street', *The Bankers' Magazine*, November, pp. 29-32.
- Jovanovic, B. and G. M. Macdonald (1994) 'Competitive diffusion', *Journal of Political Economy*, 102(1), pp. 24-52.
- Jovanovic, B. and S. Lach (1989) 'Entry, exit, and diffusion with learning by doing', *American Economic Review*, 79(4) pp. 690-699.
- Judd, J. P. and J. L. Scadding (1982) 'The search for a stable money demand function: a survey of the post-1973 literature', *Journal of Economic Literature*, 20(3), pp. 993-1023.
- Judge, G. G. (1985) *The Theory and Practice of Econometrics*, Second Edition, Wiley, New York.
- Just, R. E. and D. Zilberman (1983) 'Stochastic structure, farm size and technology adoption in developing agriculture', *Oxford Economic Papers*, 35(2), pp. 307-328.
- Kalbfleisch, J. D. and R. L. Prentice (1980) *The Statistical Analysis of Failure Time Data*, John Wiley, New York.
- Kamien, M. L. and N. L. Schwartz (1982) *Market Structure and Innovation*, Cambridge University Press, Cambridge.
- Kane, E. J. (1981) 'Accelerating inflation, technological innovation and the decreasing effectiveness of banking regulation', *Journal of Finance*, 36(2), pp. 355-367.
- Kaplan, E. L. and P. Meier (1958) 'Nonparametric estimation from incomplete observations', *Journal of American Statistical Association*, 53, pp. 457-481.
- Karshenas, M. and P. Stoneman (1992) 'A flexible model of technological diffusion incorporating economic factors with an application to the spread of colour television ownership in the UK', *Journal of Forecasting*, 11(4), pp. 577-601.

- Karshenas, M. and P. Stoneman** (1993) 'Rank, stock, order, and epidemic effects in the diffusion of new process technologies: an empirical model', *Rand Journal of Economics*, 24(4), pp. 503-528.
- Karshenas, M. and P. Stoneman** (1995) 'Technological Diffusion', in P. Stoneman (editor) *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Katsoulacos, Y.** (1986) *The Employment Effect of Technical Change: A Theoretical Study of New Technology and the Labour Market*, Wheatsheaf, Brighton.
- Katz, M. L. and C. Shapiro** (1985) 'Network externalities, competition and compatibility', *American Economic Review*, 75, pp. 424-440.
- Katz, M. L. and C. Shapiro** (1986) 'Technology adoption in the presence of network externalities', *Journal of Political Economy*, 94, pp. 822-841.
- Katz, M. L. and C. Shapiro** (1994) 'Systems competition and network effects', *Journal of Economic Perspectives*, 8(2), pp. 93-115.
- Kay, R.** (1977) 'Proportional hazard regression models and the analysis of censored survival data', *Applied Statistics*, 26(3), pp. 227-237.
- Keely, L. C.** (1996) 'Patents: do they matter: a literature survey', mimeo, Centre for Economic Performance, London School of Economics.
- Keely, L. C. and D. Quah** (1998) 'Technology in growth', *Discussion Paper No 391*, Centre for Economic Performance, London School of Economics, London.
- Kiefer, N. M.** (1988) 'Economic Duration data and hazard functions', *Journal of Economic Literature*, 26(2), pp. 646-679.
- Kimber, A. C.** (1985) 'Tests for the exponential, Weibull and Gumbel distributions based on the stabilized probability plot', *Biometrika*, 72(3), pp. 661-663.
- Kingaby, T.** (1992) 'End of the line for the ATM?', *The Banker*, May, pp. 61-63.
- Kirkman, P.** (1987) *Electronic Funds Transfer Systems: The Revolution in Cashless Banking and Payments Methods*, Basil Blackwell, London.
- Kmenta, J.** (1986) *Elements of Econometrics*, Second Edition, Macmillan, New York.
- Koziol, J. A. and D. P. Byar** (1975) 'Percentage points of the asymptotic distributions of one and two sample k-s statistics for truncated or censored data', *Technometrics*, 17, pp. 507-510.
- Kreps, D. M.** (1990) *A Course in Microeconomic Theory*, Harvester Wheatsheaf, Hemel Hempstead.

- Kreps, D. M. and R. Wilson (1982) 'Sequential equilibria', *Econometrica*, 50, pp. 863-894.
- Labson, B. S. and P. Gooday (1994) 'Factors influencing the diffusion of electric arc furnace steelmaking technology' *Applied Economics*, 26(9), pp. 917-925.
- Laffont, J-J., P. Rey and J. Tirole (1998) 'Network competition: I. overview and nondiscriminatory pricing', *Rand Journal of Economics*, 29(1), pp. 1-37.
- Lambert, P. J. (1990) *Advanced Mathematics for Economists*, Third Edition, Basil Blackwell, Oxford.
- Lancaster, K. J. (1966) 'A new approach to consumer theory', *Journal of Political Economy*, 74, pp. 132-157.
- Lancaster, K. J. (1979) *Variety, Equity and Efficiency*, Columbia University Press, New York.
- Lancaster, T. (1985) 'Generalised residuals and heterogeneous duration models: with applications to the Weibull model', *Journal of Econometrics*, 28(1), pp. 155-169.
- Lancaster, T. (1990) *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.
- Lancaster, T. and A. Chesher (1985a) 'Residual analysis for censored duration data', *Economics Letters*, 18(1), pp. 35-38.
- Lancaster, T. and A. Chesher (1985b) 'Residuals, test and plots with a job matching illustration', *Annals de l'INSEE*, 59, pp. 47-69.
- Lane, D. and R. Vesconvi (1996) 'Decision rules and market share: aggregation in an information model', *Industrial and Corporate Change*, 5(1), pp. 127-146.
- Lawless, J. F. (1982) *Statistical Methods for Lifetime Data*, John Wiley, New York.
- Lawrence, C. and R. P. Shay (1986) 'Technology and financial intermediation in multiproduct banking firms: an econometric study of US banks, 1979-1982', in C. Lawrence and R. P. Shay (editors), *Technological Innovation, Regulation, and the Monetary Economy*, Ballinger, Cambridge, Massachusetts.
- Lee, L. F. (1981) 'Simultaneous equation models with discrete and censored variables', in C. F. Manski and D. McFadden (editors), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Massachusetts.

- Lekvall, P. and C. Wahlbin (1973) 'A study of some assumptions underlying innovation diffusion functions', *Swedish Journal of Economics*, 75, pp. 362-377.
- Leung, W-A. (1997) 'The duration of international joint ventures and foreign wholly-owned subsidiaries', *Applied Economics*, 29, pp. 1255-1269.
- Levin, S. G., S. L. Levin and J. B. Meisel (1987) 'A dynamic analysis of the adoption of a new technology: the case of optimal scanners', *Review of Economics and Statistics*, 64(1), pp. 12-17.
- Lewis, M. K. and K. T. Davis (1987) *Domestic and International Banking*, Philip Allan, London.
- Lindley, D. V. (1965) *Introduction to Probability and Statistics*, Cambridge University Press, Cambridge.
- Lindsey, I. (1994) *Credit Cards: The Authoritative Guide to Credit and Payment Cards*, Publisher, Rushmere Wynne, Leighton Buzzard.
- Link, A. N. and P. Kapur (1994) 'A note on the diffusion of flexible manufacturing systems technology', *Economics Letters*, 46(4), pp. 369-374.
- Lissoni, F. and J. S. Metcalfe (1994) 'Diffusion of Innovation Ancient and Modern: A Review of the Main Themes', in M. Dodgson and R. Rothwell (editors), *The Handbook of Industrial Innovation*, Edward Elgar, Aldershot.
- Llewellyn, D. T. (1992) 'Financial Innovation: A Basic Analysis', in H. Cavana (editor), *Financial Innovation*, Routledge, London.
- Llewellyn, D. T. (1997) 'Technology transforms the banking industry', *Financial Technology Review*, pp. 1-4, Euromoney Publications, Colchester.
- Loine, A., B. Nixon and D. Collison (1993) 'Internal and external financial constraints on investment in innovative technology', in P. Swann (editor), *New Technologies and the Firm*, Routledge, London.
- Maddala, G. S. (1983) *Limited Dependent Variables and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- Maddala, G. S. (1988) *Introduction to Econometrics*, Macmillan, New York.
- Madden, N. (1986) 'Technology brings societies a new image', *Building Societies Association (BSA) Yearbook*, BSA, London.
- Mahajan, V. and Y. Wind (1986) *Innovation Diffusion of New Product Acceptance*, Ballinger Publishing Company, Cambridge.

- Mahajan, V., E. Muller and F. M. Bass** (1990) 'New product diffusion models in marketing: a review and directions for research', *Journal of Marketing*, **54(1)**, pp. 1-26.
- Mansfield, E.** (1961) 'Technical change and the rate of imitation', *Econometrica*, **29(4)**, pp. 741-766.
- Mansfield, E.** (1963) 'The speed of response of firms to new techniques', *Quarterly Journal of Economics*, **77(2)**, pp. 290-311.
- Mansfield, E.** (1964) 'Size of firm, market structure, and innovation', *Journal of Political Economy*, **71**, pp. 556-576.
- Mansfield, E.** (1968) *The Economics of Technical Change*, Norton, New York.
- Mansfield, E.** (1969) *Industrial Research and Technological Innovation: an Econometric Analysis*, Longmans, New York.
- Mansfield, E.** (1989) 'The diffusion of industrial robots in Japan and the United States', *Research Policy*, **18(4)**, pp. 183-192.
- Mariotti, M.** (1989) 'Being identical, behaving differently: a theorem on technological diffusion', *Economics Letters*, **30**, pp. 275-278.
- Mariotti, M.** (1992) 'Unused innovations', *Economics Letters*, **38**, pp. 367-371.
- Marti, J. and A. Zeilinger** (1982) 'Micros and money, new technology in banking and shopping', Discussion Paper No. 608, Policy Studies Institute, London.
- Mata, J. and P. Portugal** (1994) 'Life duration of new firms', *Journal of Industrial Economics*, **42(3)**, pp. 227-245.
- Matutes, C. and A. J. Padilla** (1994) 'Shared ATM networks and banking competition', *European Economic Review*, **38(5)**, pp. 1113-1138.
- Matutes, C. and P. Regibeau and K. Rockett** (1996) 'Optimal patent design and the diffusion of innovations', *Rand Journal of Economics*, **27(1)**, pp. 60-83.
- McAndrews, J. J. and R. Rob.** (1996) 'Shared ownership and pricing in a network switch', *International Journal of Industrial Organization*, **14**, pp. 727-745.
- McCardle, K. F.** (1985) 'Information acquisition and the adoption of new technology', *Management Science*, **31(11)**, pp. 1372-1389.
- McKillop, D. and C. Ferguson** (1993) *Building Societies: Structure, Performance and Change*, Graham and Trotman, London.
- Meade, N.** (1984) 'The use of growth curves in forecasting market development: a review and appraisal', *Journal of Forecasting*, **3(4)**, pp. 429-451.

- Meade, N.** (1985) 'Forecasting using growth curves - an adaptive approach', *Journal of Operational Research Society*, 36(12), pp. 1103-1115.
- Meidan, A.** (1984) *Bank Marketing Management*, Macmillan, London.
- Metcalf, J. S.** (1981) 'Impulse and diffusion in the study of technical change', *Futures*, 13(5), pp. 347-359.
- Metcalf, J. S.** (1988) 'The diffusion of innovation: an interpretative survey', in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (editors), *Technological Change and Economic Theory*, Pinter, London.
- Metcalf, J. S.** (1993) 'Technology Policy and Innovative Systems from a Evolutionary Perspective', in K. Hughes (editor) *The Future of UK Competitiveness and the Role of Industrial Policy*, Policy Studies Institute, London.
- Metcalf, J. S.** (1995) 'The economic foundations of technology policy: equilibrium and evolutionary perspectives', in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Metcalf, J. S. and M. Boden** (1992) 'Evolutionary epistemology and the nature of technology strategy', in R. Coombs, P. Saviotti and V. Walsh (editors), *Technological Change and Company Strategies*, Academic Press, London.
- Miller, I. and M. Miller** (1998) *Mathematical Statistics*, Sixth Edition, Prentice Hall International, New Jersey.
- Miller, M. H. and D. Orr** (1966) 'A model of the demand for money by firms', *Quarterly Journal of Economics*, 80(2), pp. 413-435.
- Molyneux, P.** (1990) *Banking: An Introductory Text*, Macmillan, London.
- Molyneux, P. and N. Shamroukh** (1996) 'Diffusion of financial innovations: the case of junk bonds and note issuance facilities', *Journal of Money, Credit and Banking*, 28(3), pp. 502-522.
- Mowery, D.** (1995) 'The practice of technology policy', in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Murphy, K. M. and R. H. Topel** (1985) 'Estimation and inference in two-step econometric models', *Journal of Business and Economic Statistics*, 3, pp. 370-379.
- Nabseth, L. and G. F. Ray** (1974) *The Diffusion of Industrial Processes: An International Study*, Cambridge University Press, London.

- Nakicenovic, N.** (1991) 'Diffusion of pervasive systems: a case of transport infrastructures', *Technological Forecasting and Social Change*, 39(1-2), pp. 181-200.
- Nelson, P.** (1970) 'Information and consumer behaviour', *Journal of Political Economy*, 23, pp. 311-329.
- Nelson, R. R.** (1968) 'A diffusion model of international productivity differences in manufacturing industry', *American Economic Review*, 58, pp. 1219-1248.
- Nelson, R. R.** (1992) National innovation systems: a retrospective on a study', *Industrial and Corporate Change*, 1(2), pp. 347-374.
- Nelson, R. R.** (1996) *The Sources of Economic Growth*, Harvard University Press, Cambridge, Massachusetts.
- Nelson, R. R. and S. Winter** (1982) *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, Massachusetts.
- Neumann, G. R.** (1997) 'Search Models and Duration Data', *Working Paper 97-11*, Department of Economics, University of Iowa.
- Neven, D.** (1993) 'Structural adjustment in European retail banking: some views from industrial organisation', in J. Dermine (editor), *European Banking in 1990's*, Basil Blackwell, Oxford.
- Newbold, P.** (1991) *Statistics for Business and Economics*, Third Edition, Prentice-Hall International, London.
- Niehans, N.** (1978) *The Theory of Money*, John Hopkins University Press, Baltimore.
- Niehans, N.** (1982) 'Innovation in monetary policy: challenge and response', *Journal of Banking and Finance*, 6(1), pp. 9-28.
- Niehans, N.** (1983) 'Financial regulation, multinational banking and monetary policy', *Journal of Banking and Finance*, 7(4), pp. 537-551.
- O'Farrell, P. N. and R. P. Oakey** (1992) 'Regional variations in the adoption of computer-numerically-controlled machine tools by small engineering firms: a multivariate analysis', *Environment and Planning*, 24(6), pp. 887-902.
- Oakey, R. R. Rothwell and S. Cooper** (1988) *Management of Innovation in High Technology Small Firms*, Pinter, London.
- Office of Science and Technology** (1996) 'Technology foresight: first progress report', Department of Trade and Industry, London.

- Office of Science and Technology (1997)** 'Action for financial services', Department of Trade and Industry, London.
- Office of Science and Technology (1997)** 'Action for financial services', Department of Trade and Industry, London.
- Oliver, F. R. (1964)** 'Methods for estimating the logistic growth function', *Applied Statistics*, 13, pp. 57-66.
- Oren, S. and S. Smith (1981)** 'Critical mass and tariff structure in electronic communications markets', *Bell Journal of Economics*, 12(2), pp. 467-487.
- Organisation for the Economic Co-operation and Development (OECD) (1992)** *Technology and the Economy: The Key Relationships*, Paris.
- Ostaszewski, A. (1993)** *Mathematics in Economics: Models and Methods*, Blackwell, Oxford.
- Oster, S. (1982)** 'The diffusion of innovation among steel firms: the basic oxygen process', *The Bell Journal of Economics*, 13(1), pp. 45-56.
- Pagan, A. and F. Vella (1989)** 'Diagnostic tests for models based on individual data: a survey', *Journal of Econometrics*, Supplement, 4, pp. 29-59.
- Papaconstantinou, G., N. Sakurai and A. Wyckoff (1996)** 'Embodied technology diffusion: an empirical analysis for 10 OECD countries', *Organisation for Economic Co-operation and Development (OECD) Working Paper 1022-2227*, 4(8), OECD, Paris.
- Parker, P. M. (1994)** 'Aggregate diffusion forecasting models in marketing: a critical review', *International Journal of Forecasting*, 10(2), pp. 353-380.
- Paroush, J. and D. Rosenberg (1986)** 'Automated teller machines and the share of demand deposits in the money supply - the Israeli experience', *European Economic Review*, 30(6), pp. 1207-1215.
- Pawley, M. (1993)** *Financial Innovation and Monetary Policy*, Routledge, London.
- Pawley, M., D. Winstone and P. Bentley (1991)** *UK Financial Institutions and Markets*, Macmillan, Basingstoke.
- Pennings, J. M. and F. Harianto (1992)** 'The diffusion of technological innovation in the commercial banking industry', *Strategic Management Journal*, 13(1), pp. 29-46.

- Perez, C. (1985) 'Microelectronics, long waves and world structural change: new perspectives for developing countries', *World Development*, 13(3), pp. 441-463.
- Petersen, T. (1986a) 'Estimating fully parametric hazard rate models with time-dependent covariates: use of maximum likelihood', *Sociological Methods and Research*, 14(3), pp. 219-246.
- Petersen, T. (1986b) 'Fitting parametric survival models with time-dependent covariates', *Applied Statistics*, 35(3), pp. 281-288.
- Petit, P. (1995) 'Employment and technological change' in P. Stoneman (editor), *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Pindyck, R. S. (1993) 'Investments of uncertain cost', *Journal of Financial Economics*, 34, pp. 53-76.
- Podolski, T. M. (1986) *Financial Innovation and the Money Supply*, Basil Blackwell, Oxford.
- Postrel, S. R. (1990) 'Competing networks and proprietary standards: the case of quadrasonic sound', *Journal of Industrial Economics*, 39, pp. 169-185.
- Primrose, P. L. (1991) *Investment in Manufacturing Technology*, Chapman and Hall, London.
- Quah, D. T. (1997) 'Increasingly weightless economies', *Bank of England Quarterly Bulletin*, February, pp. 49-55, Bank of England, London.
- Quirnbach, H. C. (1986) 'The diffusion of new technology and the market for an innovation', *Rand Journal of Economics*, 17(1), pp. 33-47.
- Rees, J., R. Briggs and R. Oahey (1983) 'The adoption of new technology in the American machinery industry', *Regional Studies*, 18(6), pp. 489-504.
- Reinganum, J. F. (1981a) 'On the diffusion of new technology: a game theoretic approach', *Review of Economic Studies*, 48(3), pp. 395-405.
- Reinganum, J. F. (1981b) 'Market structure and the diffusion of new technology', *The Bell Journal of Economics*, 12(2), pp. 618-624.
- Reinganum, J. F. (1984) 'Practical implications of game theoretic models of R&D', *American Economic Review*, 74, pp. 61-67.

- Reinganum, J. F.** (1985a) 'A two-stage model of research and development with endogenous second mover advantages', *International Journal of Industrial Organisation*, 3, pp. 275-292.
- Reinganum, J. F.** (1985b) 'Innovation and industry evolution', *Quarterly Journal of Economics*, 100(1), pp. 81-99.
- Reinganum, J. F.** (1989) 'The theory of innovation: research, development and diffusion', in R. Schmalensee and R. D. Willig (editors), *Handbook of Industrial Organisation*, North-Holland, Amsterdam.
- Retail Bank Research** (1997) *ATMs and Cash Dispensers*, Retail Bank Research, London.
- Revell, J.** (1986) 'Implications of information technology for financial institutions', in D. E. Fair (editor), *Shifting Frontiers in Financial Markets*, Nijhoff, Dordrecht.
- Rhoades, S.** (1980) 'Monopoly and expense preference behavior: an empirical investigation of a behavioralist hypothesis', *Southern Journal*, 47, pp. 419-432.
- Ridder, G.** (1990) 'The non-parametric identification of generalized accelerated failure-time models', *Review of Economic Studies*, 57, pp. 167-182.
- Riordan, M. H. and D. J. Salant** (1994) 'Preemptive adoptions of an emerging technology', *Journal of Industrial Economics*, 42(3), pp. 247-261.
- Rogers, E.** (1962) *The Diffusion of Innovations*, Free Press, New York.
- Rohlf, J.** (1974) 'A theory of interdependent demand for a communications service', *Bell Journal of Economics*, 5(1), pp. 16-37.
- Romeo, A. A.** (1975) 'Interindustry and interfirm differences in the rate of diffusion of an innovation', *Review of Economics and Statistics*, 57(3), pp. 311-319.
- Romeo, A. A.** (1977) 'The rate of imitation of a capital-embodied process innovation', *Economica*, 44, pp. 63-69.
- Rose, N. L. and P. L. Joskow** (1990) 'The diffusion of new technologies: evidence from the electric utility industry', *Rand Journal of Economics*, 21(3), pp. 354-373.
- Rosegger, G.** (1977) 'Diffusion of technology in industry', in B. Gold (editor), *Research, Technological Change, and Economic Analysis*, Lexington Books, Massachusetts.
- Rosegger, G.** (1996) *The Economics of Production and Innovation*, Butterworth Heinemann, Oxford.

- Rosenberg, N. (1976a) 'On technological expectations', *The Economic Journal*, 86, pp. 523-535.
- Rosenberg, N. (1976b) *Perspective on Technology*, Cambridge University Press, New York.
- Rosenberg, N. (1982) *Inside the Black Box*, Cambridge University Press, Cambridge.
- Ryan, B. and N. C. Gross (1943) 'The diffusion of hybrid corn in two Iowa communities', *Rural Sociology*, 8, pp. 15-24.
- Rymes, T. K. (1971) *On Concepts of Capital and Technical Change*, Cambridge University Press, Cambridge.
- Sadanand, V. (1989) 'Endogenous diffusion of technology', *International Journal of Industrial Organisation*, 7, pp. 471-487.
- Sakurai, N., E. Ioannidis and G. Papaconstantinou (1996) 'The impact of R&D and technology diffusion on productivity growth: evidence for 10 OECD countries in the 1970s and 1980s', *Organisation for Economic Co-operation and Development (OECD) Working Paper 1022-2227*, 4(12), OECD, Paris.
- Saloner, G. and A. Shepard (1995) 'Adoption of technologies with network effects: an empirical examination of the adoption of automated teller machines', *Rand Journal of Economics*, 26(3), pp. 479-501.
- Salop, S. (1979) 'Monopolistic competition with outside goods', *Bell Journal of Economics*, 10, pp. 141-156.
- Salter, W. E. G. (1966) *Productivity and Technical Change*, Second Edition, Department of Applied Economics, Monographs No. 6, University of Cambridge Press, Cambridge.
- Saracho, A. I. (1997) 'The diffusion of a durable embodied capital innovation', *Economics Letters*, 54(1), pp. 45-50.
- Sarkar, J. (1998) 'Technological diffusion: alternative theories and historical evidence', *Journal of Economic Surveys*, 12(2), pp. 131-176.
- Satchell, P. (1998) *Innovation and Automation*, Ashgate, Aldershot.
- Saving, T. R. (1971) 'Transactions costs and the demand for money', *American Economic Review*, 3(1), pp. 407-420.
- Saviotti, P. P. (1997) 'Innovation systems and evolutionary theories', in C. Edquist (editor), *Systems of Innovation: Technologies, Institutions and Organisations*, Pinter, London.

- Saxonhouse, G. R.** (1977) 'Regressions for samples having different characteristics', *Review of Economics and Statistics*, **59**, pp. 234-237.
- Scarborough, H. and R. Lannon** (1988) 'The successful exploitation of new technology in banking', *Journal of General Management*, **13(3)**, pp. 38-51.
- Scherer, F. M.** (1965) 'Firm size, market structure, opportunity and the output of patented inventions', *American Economic Review*, **55**, pp. 1097-1125.
- Schmidt, G. and A. D. Witte** (1988) *Predicting Recidivism Using Survival Models*, Springer-Verlag, New York.
- Schmidt, P. and A. D. Witte** (1989) 'Predicting criminal recidivism using split population survival-time models', *Journal of Econometrics*, **40(1)**, pp. 141-159.
- Schmookler, J.** (1966) *Invention and Economic Growth*, Harvard University Press, Cambridge, Massachusetts.
- Schumpeter, J. A.** (1934) *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*, Harvard University Press, Cambridge, Massachusetts.
- Schumpeter, J. A.** (1939) *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*, Two Volumes, McGraw-Hill, New York.
- Schumpeter, J. A.** (1942) *Capitalism, Socialism, and Democracy*, Harper and Row, New York.
- Schumpeter, J. A.** (1947) 'The creative response in economic history', *Journal of Economic History*, **7**, pp. 149-159.
- Seaton, J. S.** (1996) 'What are the differences between finance portfolios of large R&D and non-R&D quoted companies?', *Economic Research Paper No 96/8*, Department of Economics, Loughborough University.
- Shleifer, A.** (1986) 'Implementation cycles', *Journal of Political Economy*, **10**, pp. 101-112.
- Shy, O.** (1996) *Industrial Organization: Theory and Applications*, MIT Press, Cambridge, Massachusetts.
- Silber, W. L.** (1975) *Financial Innovation*, Lexington Books, Farnborough.
- Silber, W. L.** (1983) 'The process of financial innovation', *American Economic Review*, **73(2)**, pp. 89-95.
- Silverberg, G.** (1988) 'Modelling economic dynamics and technical change: mathematical approaches to self-organization and evolution', in G. Dosi, C.

- Freeman, R. Nelson, G. Silverberg and L. Soete (editors), *Technical Change and Economic Theory*, Pinter, London.
- Silverberg, G. (1991) 'Adoption and diffusion of a technology as a collective evolutionary process', *Technological Forecasting and Social Change*, 39(1-2), pp. 67-80.
- Silverberg, G., G. Dosi and L. Orsenigo (1988) 'Innovation, diversity and diffusion: a self-organization model', *The Economic Journal*, 98, pp. 1032-1054.
- Sinha, R. K. and M. Chandrashekar (1992) 'A split hazard model for analysing the diffusion of innovations', *Journal of Marketing*, 29(1), pp. 116-127
- Soete, L. and R. Turner (1984) 'Technology diffusion and the rate of technical change', *The Economic Journal*, 94, pp. 612-623.
- Solow, R. M. (1960) 'Investment and technical progress', in , K. J. Arrow, S. Karlin, and P. Suppes (editors), *Mathematical Methods in the Social Sciences*, Stanford University Press, Stanford.
- Sommers, P (1980) 'The adoption of nuclear power generation', *The Bell Journal of Economics*, 11(1), pp. 283-291.
- Spence, A. M. (1981) 'The learning curve and competition', *Bell Journal*, 40, pp. 49-70.
- Spencer, P. D. (1986) *Financial Innovation, Efficiency and Disequilibrium: Problems of Monetary Management in the UK 1971-81*, Clarendon Press, Oxford.
- Stigler, G. J. (1987) *The Theory of Price*, Fourth Edition, Macmillan, New York.
- Stoneman, P. (1976) *Technological Diffusion and the Computer Revolution: The UK Experience*, Cambridge University Press, Cambridge.
- Stoneman, P. (1981) 'Intra-firm diffusion, bayesian learning and profitability', *The Economic Journal*, 91, pp. 375-388.
- Stoneman, P. (1983) *The Economic Analysis of Technological Change*, Oxford University Press, Oxford.
- Stoneman, P. (1986) 'Technological diffusion: the viewpoint of economic theory', *Ricerche Economiche*, 40(4), pp. 585-606.
- Stoneman, P. (1987) *The Economic Analysis of Technology Policy*, Clarendon Press, Oxford.
- Stoneman, P. (1989) 'Technological diffusion, vertical product differentiation and quality improvement', *Economics Letters*, 31(3), pp. 277-280.

- Stoneman, P.** (1990a) 'Technological diffusion, horizontal product differentiation and adaptation costs', *Economica*, 57, pp. 49-62.
- Stoneman, P.** (1990b) 'The intertemporal demand for consumer technologies requiring hardware and software inputs', *Warwick Economic Research Paper No 355*, Department of Economics, University of Warwick.
- Stoneman, P.** (1991) 'Copying capabilities and intertemporal competition between joint input technologies: CD vs DAT', *Economics of Innovation and New Technology*, 1, pp. 233-241.
- Stoneman, P.** (1995) *Handbook of the Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Stoneman, P. and M J. Kwon** (1996) 'Technology adoption and firm profitability', *The Economic Journal*, 106, pp. 952-962.
- Stoneman, P. and M. J. Kwon** (1994) 'The diffusion of multiple process technologies', *The Economic Journal*, 104, pp. 420-431.
- Stoneman, P. and M. Karshenas** (1993) 'The diffusion of new technology: extensions to theory and evidence', in P. Swann (editor), *New Technologies and the Firm*, Routledge, London.
- Stoneman, P. and N. J. Ireland** (1983) 'The role of supply and demand factors in the diffusion of new process technology', *The Economic Journal*, Supplement, 93, pp. 66-78.
- Stoneman, P. and P. A. David** (1986) 'Adoption subsidies vs. information provision as instruments of technology policy', *The Economic Journal*, 96, Supplement (March), pp. 142-151.
- Stoneman, P. and P. Diederer** (1994) 'Technology diffusion and public policy', *The Economic Journal*, 104, pp. 918-930.
- Sugden, R.** (1991) 'Rational choice: a survey on contributions from economics and philosophy', *The Economic Journal*, 101, pp. 751-785.
- Swann, P.** (1992) 'Rapid technological change, technological visions, corporate organisation and market structure', *Economics of Innovation and New Technology*, 2, pp. 3-25.
- Swann, P.** (1993) *New Technologies and the Firm*, Routledge, London.
- Tanny, S. M. and N. A. Derzko** (1988) 'Innovators and imitators in innovation diffusion modelling', *Journal of Forecasting*, 7(4), pp. 225-234.

- Tarde, G.** (1903) *The Laws of Imitation*, Rienhart and Winston, New York.
- Taylor, M. P.** (1987) 'Financial innovation, inflation and the stability of the demand for broad money in the united kingdom', *Bulletin of Economic Research*, 39(3), pp. 225-233.
- The Bankers' Magazine** (1982) 'Automation opens new opportunities for managers', *The Bankers' Magazine*, November, pp. 8-13.
- The Guardian** (1994) 'Customers in crossfire as banks go to war over Visa', *Guardian*, October 22, p. 33.
- The Sunday Times** (1994) 'Access to cash machines curbed', *The Sunday Times*, July 24, p. 1.
- The Tower Group** (1997) *The Future of Automated Teller Machines (ATMs)*, The Tower Group Consultancy, Netwon, Massachusetts.
- Thirtle, C. G. and V. W. Ruttan** (1987) *The Role of Demand and Supply in the Generation and Diffusion of Technical Change*, Harwood, London.
- Tirole, J.** (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge, Massachusetts.
- Tobin, J.** (1956) 'The interest elasticity of the transactions demand for cash', *Review of Economics and Statistics*, 38(3), pp. 241-247.
- Trajtenberg, M. and S. Yitzhaki** (1989) 'The diffusion of innovations: a methodological reappraisal', *Journal of Business and Economics*, 7(1), pp. 35-47.
- Tsur, Y., M. Sternberg and E. Hochman** (1990) 'Dynamic modelling of innovation process adoption with risk aversion and learning', *Oxford Economic Papers*, 42(2), pp. 336-355.
- Usher, A. P.** (1954) *A History of Mechanical Inventions*, Harvard University Press, Cambridge, Massachusetts.
- Van Reenan, J.** (1997) 'Employment and technical innovation: evidence from UK manufacturing firms', *Journal of Labour Economics*, 15(2), pp. 255-284.
- Vesala, J.** (1993) 'Retail banking in European financial integration', *Discussion Paper No 77*, Bank of Finland, Helsinki.
- Vesala, J.** (1994) 'Banking industry performance in Europe: trends and issues', Paper presented at the OECD Financial Experts' Meeting, Paris.

- Vettas, N. (1998) 'Demand and supply in new markets: diffusion with bilateral learning', *Rand Journal of Economics*, 29(1), pp. 215-233.
- Vickers, J. (1986) 'The evolution of market structure when there is a sequence of innovations', *The Journal of Industrial Economics*, 35, pp. 1-12.
- Viren, M. (1992) 'Financial innovations and currency demand: some new evidence', *Empirical Economics*, 14, pp. 451-461.
- von Hippel, E. (1988) *The Sources of Innovation*, Oxford University Press, Oxford.
- von Weizsacker, C. C. (1966) 'Tentative notes on a two step model with induced technical progress', *Review of Economic Studies*, 33, pp. 245-251.
- Walker, D. (1978) 'Economies of scale in electronic funds transfer systems', *Journal of Banking and Finance*, 2(6), pp. 65-78.
- Weiss, A.M. (1994) 'The effects of expectations on technology adoption: some empirical evidence', *The Journal of Industrial Economics*, 42(4), pp. 341-360.
- Westaway, P. and D. Walton (1991) 'Endogenous financial innovation and the demand for M0', in M. P. Taylor (editor), *Money and Financial Markets*, Basil Blackwell, Oxford.
- White, H. (1982) 'Maximum likelihood estimation of mis-specified models', *Econometrica*, 50(1), pp. 1-24.
- Williams, M. R. (1985) *A History of Computing Technology*, Prentice-Hall, London.
- Wind, Y, V. Mahajan, R. N. Cardozo (1984) *New-Product Forecasting: Models and Applications*, Lexington Books, Lexington.
- Witt, U. (1986) 'Firm's market behaviour under imperfect information and economic natural selection', *Journal of Economic Behaviour and Organization*, 7, pp. 265-290.
- Witt, U. (1991) 'Economics, sociobiology and behavioral psychology on preferences', *Journal of Economic Psychology*, 12(4), pp. 557-573.
- Witt, U. (1997) 'Lock-in vs. critical masses - industrial change under network externalities', *International Journal of Industrial Organization*, 15, pp. 753-773.
- Wyatt, G. (1986) *The Economics of Invention: A Study of the Determinants of Inventive Activity*, Harvester, Oxford.
- Zilberfarb, B. Z. (1989) 'The effect of automated teller machines on demand deposits - an empirical analysis', *Journal of Financial Services Research*, 2(1), pp. 49-57.

Zimmerman, M. B. (1982) 'Learning effects and the commercialization of new energy technologies: the case of nuclear power', *Bell Journal of Economics*, 13(2), pp. 297-310.