



Durham E-Theses

The instability of the demand for money

Gafga, Philip Henry

How to cite:

Gafga, Philip Henry (1987) *The instability of the demand for money*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/6749/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

The copyright of this thesis rests with the author.
No quotation from it should be published without
his prior written consent and information derived
from it should be acknowledged.

THE INSTABILITY OF THE DEMAND FOR MONEY

by

PHILIP HENRY GAFGA

*A thesis submitted to the University of Durham
in candidature for the degree of
Master of Arts*

July 1987



-Z JAN 1988

**In Memory of
My Parents**

ABSTRACT

Author: Philip H. Gafga.

Title of M.A. thesis: The Instability of the Demand for Money.

The demand for money plays an important role in the assessment of the efficacy of monetary policy. Prior to the early 1970s, there was a consensus among the empirical literature that a stable demand for money function existed. During the early 1970s, most empirical studies indicated that the demand for money had shifted about in an unpredictable manner, making the assessment of the efficacy of monetary policy hazardous.

This thesis investigates the causes of the instability of the demand for money by going back to first fundamentals of the theory of the demand for money. Two main possible causes are identified, *viz:* financial innovation, and frequent changes in taxation regimes.

With regard to financial innovation, which may take on the form of lower transaction costs, improved cash-management techniques and the increased proliferation of new substitutes for money, the following propositions are made: that a change in transactions costs affects the demand for money, and that improved cash-management techniques and new substitutes for money will lead to increased interest-elasticities for the demand for money. The use of *divisia* monetary aggregates as a possible replacement for *simple-sum* aggregates is also considered.

With regard to frequent changes in taxation regimes, the theoretical relationship between expected inflation and interest rates as embodied in the Fisher hypothesis is analysed. Two neoclassical monetary growth models are discussed in which each model has a different method of capital financing by the firm, *viz:* all-debt, and debt-equity financing. The Fisher hypothesis is then refined to take into account the different features of each taxation regime. Whilst the original Fisher hypothesis predicts that the nominal interest rate will adjust *pari passau* in response to expected inflation, the refined hypothesis predicts that nominal interest rates will adjust by more than the expected rate of inflation. The refined Fisher hypothesis is then incorporated into the steady-state demand for money, and it is suggested that frequent changes in taxation regimes can lead to the instability of the steady-state demand for money.

CONTENTS

<i>List of figures.</i>	x
<i>List of tables.</i>	xi
<i>Acknowledgements.</i>	xiv
Chapter 1. INTRODUCTION.	
1.1. The purpose of this study.	2
1.2. Plan of discussion.	9
PART ONE: FINANCIAL INNOVATION.	
Chapter 2. REGULATION AND FINANCIAL INNOVATION.	
2.1. Some preliminary remarks.	21
2.1.1. <i>Definitional problems.</i>	21
2.1.2. <i>Taxonomic problems.</i>	25
2.1.3. <i>Inducement to innovation and its diffusion.</i>	35
2.2. Towards a theory of financial innovation.	41
2.2.1. <i>A Schumpeterian approach.</i>	41
2.2.2. <i>The constraint-induced innovation hypothesis.</i>	44
2.2.3. <i>The regulatory dialectic framework.</i>	52
2.2.4. <i>A hybrid theory of financial innovation.</i>	61
2.3. The UK experience of financial innovation.	65
2.3.1. <i>The financial sector during the 1960s.</i>	65
(a) <i>The scenario.</i>	65

	(b) <i>Innovation-inducing developments.</i>	66
	2.3.2. <i>The problem of liability management.</i>	73
	(a) <i>The initial arrangements of Competition and Credit Control (CCC).</i>	73
	(b) <i>The early operational experience of CCC.</i>	76
	(c) <i>The impact of liability management.</i>	78
	(d) <i>The Supplementary Special Deposits Scheme (SSD).</i>	81
	(e) <i>Implications for monetary control.</i>	85
	2.3.3. <i>Other financial innovations.</i>	89
	(a) <i>The switch to variable-rate lending.</i>	89
	(b) <i>Technological change.</i>	94
	2.4. Conclusions.	96
Chapter	3. THE TRANSACTIONS DEMAND FOR MONEY.	98
	3.1. Alternative theories.	100
	3.1.1. <i>Deterministic inventory-theoretic model.</i>	103
	(a) <i>Zero transactions costs.</i>	104
	(b) <i>Fixed transactions costs.</i>	107
	(c) <i>Proportional transactions costs.</i>	113
	(d) <i>Transactions costs as a linear function of transactions.</i>	117
	(e) <i>Income and interest elasticities.</i>	120
	3.1.2. <i>Probabilistic inventory-theoretic model.</i>	126
	(a) <i>The basic concept.</i>	126
	(b) <i>The model.</i>	127
	(c) <i>Properties of the derived demand for money function.</i>	132
	3.1.3. <i>Effects of uncertainty in interest rates.</i>	133
	3.2. Innovations as a time-trend.	143
	3.2.1. <i>'A priori' justifications.</i>	143
	3.2.2. <i>Empirical evidence and analysis.</i>	146
	3.3. Innovations in cash-management.	152

	3.3.1. <i>Outline of theory.</i>	152
	3.3.2. <i>The Porter–Simpson model.</i>	155
	3.3.3. <i>Empirical evidence and analysis.</i>	160
	3.4. The problem of brokerage fees.	172
	3.4.1. <i>The methodology of Porter and Offenbacher (1984).</i>	173
	(a) <i>The derivation of the debits equation.</i>	173
	(b) <i>The derivation of a brokerage fee series.</i>	175
	3.4.2. <i>A critique of Porter and Offenbacher.</i>	176
	3.4.3. <i>Some tentative suggestions for tackling the problem of brokerage fees.</i>	179
	3.5. Conclusions.	181
Chapter	4. MONEY SUBSTITUTES AND AGGREGATION.	184
	4.1. The definition of money.	188
	4.1.1. <i>'A priori' approaches.</i>	189
	(a) <i>Money as a unit of account.</i>	189
	(b) <i>Money as a store of value.</i>	190
	(c) <i>Money as a means of payment or medium of exchange.</i>	191
	(d) <i>The distinction between means of payment and mediums of exchange.</i>	192
	(e) <i>The effects of financial innovation on such a distinction.</i>	193
	(f) <i>Theoretical approaches to the definition of money.</i>	195
	4.1.2. <i>Empirical approaches.</i>	207
	(a) <i>In the context of the causality debate.</i>	208
	(b) <i>Demand for money stability criteria.</i>	210
	4.2. The aggregation problem.	213
	4.2.1. <i>The 'all-or-nothing' approach.</i>	215
	4.2.2. <i>Applications of economic aggregation theory.</i>	221

(a)	<i>Chetty's approach.</i>	222
(b)	<i>The 'dual' problem that wasn't.</i>	231
(c)	<i>Donovan's critique.</i>	240
(d)	<i>The concept of rental prices of money.</i>	241
(e)	<i>Recent developments.</i>	245
4.3.	Divisia aggregation.	246
4.3.1.	<i>The Divisia aggregation procedure.</i>	246
(a)	<i>Basic concepts.</i>	246
(b)	<i>The optimal aggregation procedure.</i>	252
4.3.2.	<i>A critique.</i>	253
4.4.	Empirical evidence and analysis.	255
4.4.1.	<i>A demand for money study using Divisia aggregates.</i>	255
4.4.2.	<i>Other empirical evidence.</i>	259
4.5.	Conclusions.	261

PART TWO: THE FISHER HYPOTHESIS.

Chapter	5. THE FISHER HYPOTHESIS IN HISTORICAL PERSPECTIVE.	264
5.1.	Basic concepts.	267
5.1.1.	<i>Nature of Fisher's original formulation.</i>	267
5.1.2.	<i>Limiting values of variables.</i>	272
5.2.	Interest rates and prices.	276
5.2.1.	<i>Henry Thornton.</i>	276
5.2.2.	<i>Knut Wicksell.</i>	280
5.2.3.	<i>Inflationary expectations and the Wicksellian model.</i>	283
5.3.	The transition period.	289
5.3.1.	<i>Fisher's theory of the transition period.</i>	290
5.3.2.	<i>Fisher's empirical analysis.</i>	293
5.3.3.	<i>A Wicksellian perspective.</i>	298

	5.4. Conclusions.	305
Chapter	6. REFINEMENTS OF THE FISHER HYPOTHESIS.	307
	6.1. The need for refinement.	310
	6.1.1. <i>Bailey's analysis.</i>	311
	6.1.2. <i>Mundell's analysis.</i>	312
	6.2. A neoclassical monetary growth model.	314
	6.2.1. <i>The production function.</i>	314
	6.2.2. <i>The liquidity function.</i>	315
	6.2.3. <i>The consumption function.</i>	316
	6.2.4. <i>The national income identity.</i>	317
	6.3. All-debt financing model.	318
	6.3.1. <i>The firm's demand for capital.</i>	319
	(a) <i>When taxation is absent.</i>	319
	(b) <i>When taxation is present.</i>	322
	6.3.2. <i>The various after-tax rates of return defined.</i>	324
	6.4. Effects of inflation (1).	325
	6.4.1. <i>Effect on capital intensity.</i>	328
	6.4.2. <i>Effect on gross interest rates.</i>	332
	6.4.3. <i>Effect on net interest rates.</i>	334
	6.5. A critique of Feldstein's model.	336
	6.5.1. <i>A re-consideration of the demand for money function.</i>	336
	6.5.2. <i>Relaxation of more assumptions.</i>	339
	6.5.3. <i>Why are liquidity effects a problem?</i>	339
	6.6. Debt-equity financing model.	344
	6.6.1. <i>Initial assumptions.</i>	344
	6.6.2. <i>The installment function.</i>	345
	6.6.3. <i>The firm's optimal behaviour.</i>	346

	6.6.4. <i>Tobin's q-theory of investment under taxation.</i>	351
	6.6.5. <i>The firm's demand for capital in a debt-equity model.</i>	356
	6.7. Effects of inflation (2).	358
	6.7.1. <i>Effect on gross interest rates.</i>	362
	6.7.2. <i>Comparison with the all-debt model.</i>	362
	6.8. The short-run versus the long-run.	363
	6.9. Conclusions.	375
Chapter	7. THE REFINED FISHER HYPOTHESIS AND THE STEADY-STATE DEMAND FOR MONEY.	377
	7.1. The dynamic steady-state demand for money.	380
	7.2. Incorporation of the Fisher hypothesis.	384
	7.3. Empirical evidence and analysis.	389
	7.4. Consistency with theory.	402
	7.5. Conclusions.	406
Appendix	1. A Note on Index Numbers and their Desirable Properties.	409
Appendix	2. A Note on Aggregate Production Functions.	415
	Notes.	419
	References.	441

LIST OF FIGURES

Figure	2.1. A schematic representation of the process of financial innovation within the regulatory dialectic framework.	58
	2.2. A schematic representation of the process of financial innovation under the hybrid theory.	63
	3.1. The Baumol-Tobin 'sawtooth' pattern of disbursements and receipts with different cash-management strategies.	105
	3.2. The Baumol-Tobin model of the transactions demand for money with fixed and proportional transactions costs.	111
	3.3. The Baumol-Tobin model of the transactions demand for money with transactions costs as a linear function of the volume of transactions.	118
	3.4. The demand for money given uncertainty in interest rates and proportional transactions costs.	138
	3.5. Effect of a change in uncertainty of the interest rate on the demand for money.	140
	3.6. Effect of a fall in proportional transactions costs on the demand for money given uncertainty about interest rates.	142
	4.1. The concept of economic aggregation.	226
	5.1. Combination of the analyses by Wicksell and Fisher in the context of a market for loanable funds.	301

LIST OF TABLES

Table	1.1. Regressions based on a specification of the demand for M1 balances for the period 1963-1983 in the United Kingdom.	4
	2.1. Summary of classification schemes used in various studies of financial innovation.	27
	2.2. Time-series of shadow prices for deposits for the First National City Bank for the period 1953-71.	49
	2.3. Price-earnings ratio and price-dividend ratio of large New York City banks and Moody's industrials, and their relationship for the period 1952-72.	51
	2.4. Changes in the distribution of UK residents' sterling bank deposits during the 1960s.	67
	2.5. Distribution of liabilities between sterling and other currencies, 1957-1979.	69
	2.6. The London sterling money markets, 1957-1979.	71
	2.7. Changes in the structure of London clearing bank deposits during the 1970s.	84
	3.1. Summary of selected empirical results that include innovations as a time-trend.	148
	3.2. Summary of selected empirical results that use previous-peak variables as proxies for innovation.	164
	3.3. Post sample errors resulting from dynamic simulations for conventional demand for money functions and those containing previous-peak variables.	165

3.4.	Summary of selected empirical results that use ratchet variables as proxies for innovation.	167
3.5.	Post sample errors resulting from simulations for demand for money functions containing ratchet variables.	169
4.1.	Summary of theoretical definitions of money, and the possible effects of financial innovation on such definitions.	196
4.2.	Estimated elasticities of substitution between money and alternative assets derived from the study by Moroney and Wilbratte (1976).	235
4.3.	Error Summary - Double Logarithmic Money Demand Equations: Divisia Aggregates and Divisia User Costs.	256
4.4.	Error Summar - Double Logarithmic Money Demand Equations: Conventional Aggregates and Interest Rate Opportunity Costs.	257
4.5.	The number of times that each aggregate was best performer in statistical tests of Barnett (1982).	260
6.1.	A neoclassical monetary growth model with all-debt financing.	326
6.2.	Total differentials of all equations in Table 6.1.	327
6.3.	Revised neoclassical monetary growth model with all- debt financing.	340
6.4.	Total differentials of all equations in Table 6.3.	341
6.5.	Revised neoclassical monetary growth model with debt-equity financing.	359
6.6.	Total differentials of all equations in Table 6.5.	360
6.7.	Short-run monetary growth model.	370
6.8.	Total differentials of all equations in Table 6.7.	371

- 7.1. Solved coefficients from equation [12], and summary of calculations leading up to coefficients of equation 1 in Table 7.2. 392
- 7.2. Summary of studies leading up to the derivation of the dynamic steady-state demand for money. 393
- 7.3. Values of μ , Wald test statistics, and elasticities of inverse income velocity with respect to inflation for key corporate tax rates. 400

ACKNOWLEDGEMENTS

The research undertaken which culminated in the writing of this thesis was supported by a grant from the Economic and Social Research Council whose support is gratefully acknowledged here.

I would especially like to thank Professor John Creedy for his numerous helpful comments on earlier drafts of this thesis.

There are many people that I am indebted to for their support and understanding. However, I would like to single out just two for special mention here. I would like to thank Colin Davies, Tutor for Hearing-Impaired Students at the University of Durham, and the Reverend David Jones, Vice-Principal of the College of St. Hild and St. Bede for giving me their encouragement and support, without which this thesis would have never been completed.

Finally, one would expect me to go through the process of acknowledging the services of a typist. However, there is no typist for me to thank because I did the typing myself!

P.H.G.
College of St. Hild and St. Bede.

C H A P T E R O N E
I N T R O D U C T I O N

Section	1.1. The purpose of this study.	2
	1.2. Plan of discussion.	9

CHAPTER ONE

INTRODUCTION

1.1. Purpose of this study

Prior to the early 1970s, it was a commonly held belief that the demand for money was essentially a stable function, and policy-makers came to depend on it for assessing the efficacy of their monetary policies which was, then, a comparatively simple matter since it could be assessed in terms of a basic *IS-LM* model. In that case, a stable demand for money function was linked with a stable *LM* locus.¹

However, during the early 1970s, the picture changed quite dramatically as several econometric studies of the demand for money began to proliferate, indicating that there had been an apparent breakdown in the empirical demand for money function. Perhaps the most typical manifestation of such a breakdown was the tendency for estimated coefficients of the empirical demand for money function to take on nonsensical values which were at variance with theoretically specified values. It was usually the case that the estimated demand for money function exhibited a tendency to make large forecasting errors as the 1970s progressed which essentially meant that the empirical demand for money consistently overpredicted narrowly-defined money balances.

One could enter upon a full survey of all empirical studies of the demand for money that used conventional specifications of the function, and easily come to the conclusion that there has



indeed been a breakdown. Rather than to enter upon a full survey of such studies, it will be sufficient for present purposes to take a look at a typical study by Gafga (1985b) which used a conventional specification of the demand for money which took on the following form:

$$\ln M_t = \beta_0 + \beta_1 \ln Y_t + \beta_2 \ln r_t + \beta_3 M_{t-1} \dots [1]$$

where M denotes nominal M1 balances (consisting of currency and demand deposits), Y denotes GNP at current market prices which is used to proxy the volume of transactions, r denotes the nominal rate of interest which is a proxy for the opportunity cost of holding money balances, and β_i ($i = 0, \dots, 3$) are constants that have to be estimated. The specification contained in equation [1] includes a lagged dependent variable which is used to take account of partial adjustment.

Table 1.1. presents some selected results from the study by Gafga (1985b) which used UK data for the period 1963–1983. The main objective of the study was to examine the stability of the empirical demand for money, and to assess the plausibility of the explanation put forward by Artis and Lewis (1976) that there was a state of disequilibrium in the 'money' market owing to large money-supply shocks of the early 1970s. Two M1 demand for money functions were estimated; in each case, a different rate of interest was used, viz: a short- and long-term rate. In the first regressions for each case, the long-run income- and interest-elasticities are not very far off from theoretically-plausible values for the first part of the sample period. In the second part

TABLE 1.10 Regressions based on a specification of the demand for M1 balances for the period 1963-1983 in the United Kingdom.

Data period	Dependent variable	Constant	Exploratory variables						R ²	D.W.	L-R elasticities	
			M1-1	Y	r ₁	r ₂	Income	Interest				
1963(II) - 1972(IV)	M1	-0.701 (-0.3434)	0.877 (13.559)	0.133 (3.066)	Y	-0.601 (-3.057)	r ₁	0.993	2.19	1.08	-0.50	
1973(II) - 1983(IV)	M1	0.238 (3.434)	0.806 (15.178)	0.168 (3.687)	Y	-0.095 (-3.496)	r ₁	0.998	2.13	0.87	-0.28	
1963(II) - 1983(IV)	M1	0.0918 (1.913)	0.804 (23.431)	0.182 (6.466)	Y	-0.049 (-6.773)	r ₁	0.999	1.97	0.93	-0.25	
1963(II) - 1972(IV)	M1	-0.517 (-2.505)	0.762 (9.849)	0.308 (4.496)	Y	-0.174 (-4.822)	r ₂	0.992	2.03	1.29	-0.73	
1973(II) - 1983(IV)	M1	0.450 (2.629)	0.707 (7.653)	0.245 (3.157)	Y	-0.082 (-2.244)	r ₂	0.997	1.52	0.84	-0.28	
1963(II) - 1983(IV)	M1	0.132 (1.962)	0.798 (13.342)	0.181 (3.484)	Y	-0.034 (-1.944)	r ₂	0.999	1.53	0.90	-0.17	

SOURCE: Goffo (1985b), Table 2.

Table continued on next page...

TABLE 1.1: (continued)

Definition of variables: M1 = nominal money supply on official M1 definition, seasonally adjusted,
 Y = GNP at current market prices, seasonally adjusted,
 r1 = local authority 3-month deposit rate, and
 r2 = consol rate.

Notes: 1. Figures in parentheses denote t statistics.

2. Strictly speaking, the Durbin-Watson statistics are not valid since the regressions include a lagged dependent variable. Therefore the appropriate statistic for serial correlation is the Durbin h statistic which is given by

$$h = [1 - 0.5(D.W.)] \cdot \{[n / (1 - n\text{var}(\alpha_i))]\}$$

where α_i is the parameter for the lagged dependent variable.

of the sample period, the R^2 statistics improve slightly, but that does not necessarily show that the demand for money performs better in that period.² It is the consistency of the estimated specification with the theoretical specification that should serve as a basis on which judgements can be made. It is clear from Table 1.1. that the interest-elasticities have fallen in absolute terms so that they are not consistent with theory which normally specifies a value of -0.5 . Furthermore, the income-elasticities have fallen below unity indicating that there are now economies of scale in holding money balances. However, the long-run elasticities are dependent on the value of the coefficient of the lagged dependent variable which indicates implausibly long adjustment processes so that the results need to be interpreted with some caution. The formal Chow-text procedure revealed that the null hypothesis of parameter stability could be rejected at the 5% significance level for the second specification containing a long-term interest rate, but almost could not be rejected for the one containing the short-term interest rate although it was rejected at the 1% significance level. The conclusions are rather mixed, but if considered in conjunction with other studies, the overall conclusion is that there had been a breakdown in the demand for money during the early 1970s.³ Similar conclusions are also reported for the US, and are discussed fully in Chapter Three of this thesis.

Having ascertained that there indeed exists some form of instability of the demand for money, one would be concerned to explain why an apparently stable demand for money should suddenly turn out to be unstable. Before addressing such a question, it is important to enter upon a brief discourse regarding exactly what is

meant by 'stability'. The most common interpretation of stability is based on parameter stability in which a function exhibits a tendency to shift about rather unpredictably. This is the narrowest concept of stability that have been employed by traditional demand for money studies such as that reported in Table 1.1. The other aspect of stability is *functional* stability in which it is generally assumed that the function need not be static, but should not behave unpredictably or in an erratic manner.⁴ This thesis takes the view that too much emphasis has been placed upon parametrical stability since it is totally unreasonable to assume that functions would remain static over the long-run in spite of many exogenous developments. The idea is that the concept of functional stability should be upheld because theories can be formulated to explain how a function behaves in response to exogenous changes.

In order to investigate the causes of the instability of the demand for money, two possible causes are considered in this thesis. The first one concerns the impact of financial innovation which has taken on various forms such as lower transactions costs, improved cash-management techniques, and the increased proliferation of new substitutes for money. The main object of this thesis is to go back to first fundamentals of the theory of the demand for money, and to try and discover from a theoretical standpoint why there could be instability in the demand for money. A careful and systematic consideration of the theories of the demand for money will indicate that lower transactions costs do have a potentially important role to play in explaining why there has been a reduction in money balances. Furthermore, the effect of improved cash-management techniques is considered so that some *a priori* predictions regarding

the behaviour of the demand for money can be made. Regarding the proliferation of new substitutes for money as a consequence of financial innovation, a criticism is made in this thesis of existing aggregation procedures which implicitly allocate identical weights to each component of the monetary aggregate. Alternative aggregation procedures are considered in which weights are based on the 'moneyness' of an asset so that any monetary aggregates will not tend to overstate the amount of liquidity services available. This is done by going back to first fundamentals of economic aggregation theory. Such an examination of the effects of financial innovation on the demand for money will have served a useful purpose if it sheds more light on the mystery surrounding the instability of the demand for money.

Another possible cause of the instability of the demand for money is that frequent changes in taxation regimes may have exerted their influence on the behaviour of the relationship between expected inflation and nominal interest rates as embodied in the Fisher hypothesis. The basic idea behind the Fisher hypothesis is that the nominal interest rate adjusts *pari passu* in response to changes in expected inflation rates so that the approximate relationship would be

$$R \approx r + \pi \dots [2]$$

where R denotes the nominal rate of interest, r denotes the real rate of interest, and π denotes the expected rate of inflation. According to the analysis in this thesis, the presence of taxation will modify the above relationship such that the nominal interest rate would have to change by about one-and-half times in response to

changes in the expected rate of inflation in the case of a model in which the firm is assumed to finance its capital entirely by issues of debentures, and somewhere in-between in the case of debt-equity financing. This relationship is incorporated into the steady-state demand for money, and an analysis reveals that changes in taxation regimes may even be another factor responsible for the instability of the demand for money.

1.2. Plan of discussion

This thesis is divided into two parts, the first part being concerned with financial innovation, and the second part being concerned with the Fisher hypothesis. With regard to financial innovation, Chapter Two is devoted to a full discussion of the concept of financial innovation as it is rarely defined in any discussion of financial innovation. Several problems of defining and classifying financial innovations are considered, and it does appear that such a definition and a classification scheme is really dependent on the main objective of the study of financial innovation. Several theories of financial innovation are then considered. The first one is essentially a Schumpeterian approach in which changes in the financial sector are predominantly responses to impulses emanating from the real sector. Financial innovation is seen as a way whereby further change in the real sector could be promoted. The second theory concerns those innovations that occur in response to changing constraints imposed upon the financial firm, either externally or internally. A linear-programming model is used to illustrate how rising shadow prices of changing constraints could reflect a rise in compliance costs with a particular constraint. The

third theory is embodied within a regulatory dialectic framework in which financial innovation may occur in response to the growing burdens of regulation imposed in restrictive monetary arrangements. This is quite consistent with the constraint-induced innovation hypothesis, but it goes further in defining time lags in explaining the rate of diffusion of an innovation throughout the financial system. The final theory is what may be termed a hybrid theory in which elements from the previous three theories are drawn together to form a more general theory of financial innovation. The experience of the UK in the realm of financial innovation is then considered. It begins by looking at the scenario of the British banking system in the 1960s, and outlines some innovation-inducing developments that led to a spate of financial innovation in the 1970s and 1980s. Particular emphasis is placed on the problems posed by liability management by the banks for the conduct of monetary policy. A brief review of the effects of high and volatile interest rates on innovative activity then follows. Chapter Two finishes with a review of technological developments that have served to reduce transactions costs.

Chapter Three is mainly concerned with the right-hand side of the transactions demand for money equation. A full examination is undertaken of the theory of the demand for money using both deterministic and probabilistic inventory-theoretic models as originally formulated by Baumol (1952) and Miller and Orr (1966) respectively. The analysis is specifically geared to the consideration of the effects of changes in transactions costs on the demand for money, and particular effort has been made to analyse the effect of a change in the *structure* of transactions costs on the

interest-elasticity of the demand for money. The effect of uncertainty about interest rates is also considered. Several empirical studies are then considered which attempt to capture the effect of financial innovation on the demand for money. Firstly, studies that use a simple time-trend term in addition to the conventional specification are considered. It is seen that the use of time trends is not particularly recommended because of the restrictive assumption that financial innovation takes place at a steady rate over time. Then the use of improved cash-management ratios is analysed using the Miller-Orr model and it is shown that such cash-management techniques will lead to an increase in the interest-elasticity of the demand for money. The various proxies for cash-management techniques used by empirical studies may include previous-peak and ratchet variables for interest rates, and it is shown that the use of such variables lead to an improvement in the performance of the empirical demand for money function after 1973. The explicit use of 'brokerage fees' is strongly recommended by the theory of the demand for money as a means of capturing the effects of lower transactions costs which have normally been incorporated into the constant term. The main difficulty inherent in such a strategy is the paucity of data on brokerage fees. Therefore, Porter and Simpson (1984) consider a highly unorthodox method of deriving a brokerage fee series by indirect means by solving for the brokerage fee in the money demand and debits equations. However, this leads to a serious circularity problem, and the results of Porter and Offenbacher have to be decisively rejected because of this. Some tentative suggestions for overcoming the problem of brokerage fees are then offered. Such suggestions indicate that a study of the term

structure of interest rates may prove insightful in resolving the problem of brokerage fees.

Chapter Four considers money substitutes and their aggregation. Firstly, the definition and identification of money is considered because that is the natural thing to do before one goes on to aggregate over monetary aggregates. A brief survey of *a priori* definitions of money reveals a lack of consensus regarding which assets should be included in the definition of money. The empirical definition of money is then considered with reference to various statistical methods, and it is concluded that such empirical methods are, at best, methodologically unsound. Addressing the aggregation problem, a brief review is made of various methods of measuring the substitutability among various assets since weighted aggregates may use the substitutability among assets as a basis for determining the weights to be used for each asset. It is then argued that simple-sum aggregates are not very appropriate because of their tendency to allocate equal weights to each asset included in the aggregate so that the amount of monetary services expressed by simple-sum aggregates may be seriously overstated. Weighted aggregates attempt to measure the amount of monetary services available by allocating weights which are dependent on the 'moneyness' of an asset. It should be clear that cash and non-interest bearing demand deposits can be allocated weights of unity because their function is wholly monetary and do not function as a store-of-wealth. On the other hand, equities would be allocated weights of zero because their function is entirely as a store-of-wealth. Applications of economic aggregation theory in deriving such weighted monetary aggregates are considered, with

special reference to the seminal work of Chetty (1969). Such aggregates are dependent upon the type of aggregator function (that is, utility functions in this context)⁵. The main drawback is the problem of having to estimate the parameters of arbitrarily-specified utility functions in deriving monetary aggregates. An alternative approach is therefore considered which makes use of index-number theory which only depend on observable prices and quantities for index number aggregates. The use of Divisia monetary aggregation procedures are considered. Empirical results are then considered which compare the relative performance of Divisia aggregates to conventional sum aggregates.

In the second part of this thesis, the Fisher hypothesis is analysed. Chapter Five places the Fisher hypothesis into historical perspective with a view to clarifying some of the concepts inherent in the relationship between inflation rates and interest rates. The basic concepts, as formally formulated by Irving Fisher in 1896, are considered with regard to the behaviour of the theoretical relationship between inflation and interest rates when perfect foresight is assumed. Then a discussion of various early analyses on interest rates and prices is made according to the early works of Thornton (1802) and Wicksell (1896). Fisher's analysis of the transition period is then considered which makes it clear that interest rates do not adjust *pari passu* with inflation rates. Finally, a Wicksellian perspective is introduced which attempts to link Wicksell's natural rate of interest with the nominal and real interest rates.

Chapter Six is largely devoted to a refinement of the Fisher hypothesis in which taxes are introduced. Firstly, some early

attempts at refining the Fisher hypothesis are considered. Taxes then are introduced in two neoclassical monetary growth models which each have different assumptions regarding how the firm finances its capital, viz: all-debt and debt-equity financing. The effects of taxes on the Fisher hypothesis are then analysed in detail. A distinction between the short- and long-run is then made which is important in explaining why short-run relationships between inflation and interest rates may not be clearly defined, but could be more clearly defined in the long-run steady-state.

Chapter Seven derives a steady-state demand for money function, and the modified Fisher hypothesis is then incorporated which indicates that there may be some *a priori* evidence to suggest that there may be some form of parameter instability in the steady-state demand for money. Some empirical results and their consistency with theory are considered.

Finally, there are two appendices at the end of this thesis. The first appendix concerns index numbers and their desirable properties which is essential for the discussion of Divisia index numbers in Chapter Four. The second appendix explains the full derivation of the aggregate production function which is used in Chapter Six. Notes to the chapters are found at the end of this thesis, just before the list of references. All notes start on a fresh page for each chapter.

PART ONE

FINANCIAL INNOVATION

C H A P T E R T W O
REGULATION AND FINANCIAL INNOVATION

Section 2.1.	Some preliminary remarks.	21
2.1.1.	<i>Definitional problems.</i>	21
2.1.2.	<i>Taxonomic problems.</i>	25
2.1.3.	<i>Inducement to innovation and its diffusion.</i>	35
2.2.	Towards a theory of financial innovation.	41
2.2.1.	<i>A Schumpeterian approach.</i>	41
2.2.2.	<i>The constraint-induced innovation hypothesis.</i>	44
2.2.3.	<i>The regulatory dialectic framework.</i>	52
2.2.4.	<i>A hybrid theory of financial innovation.</i>	61
2.3.	The UK experience of financial innovation.	65
2.3.1.	<i>The financial sector during the 1960s.</i>	65
	(a) <i>The scenario.</i>	65
	(b) <i>Innovation-inducing developments.</i>	66
2.3.2.	<i>The problem of liability management.</i>	73
	(a) <i>The initial arrangements of Competition and Credit Control (CCC).</i>	73
	(b) <i>The early operational experience of CCC.</i>	76
	(c) <i>The impact of liability management.</i>	78
	(d) <i>The Supplementary Special Deposits Scheme (SSD).</i>	81
	(e) <i>Implications for monetary control.</i>	85

2.3.3. <i>Other financial innovations.</i>	89
(a) <i>The switch to variable-rate lending.</i>	89
(b) <i>Technological change.</i>	94
2.4. Conclusions.	96

CHAPTER TWO

REGULATION AND FINANCIAL INNOVATION

Until the early 1970s, the idea that the demand-for-money function was inherently stable had come to be taken for granted. However, during the 1970s, it transpired that most monetary relationships, notably the demand for money, exhibited a tendency to disintegrate. This episode served to exacerbate the evident difficulties of the authorities in the conduct of their monetary policies. A vast and prolix literature then emerged, purporting to explain the breakdown in monetary relationships. Among the several hypotheses put forward, it is now becoming quite fashionable to attribute the disintegration of the demand-for-money function to institutional change in the financial sector as exemplified by the process of financial innovation. The study of financial innovation, until recently, has been a relatively neglected subject in monetary economics, and it has now got a well-deserved catalyst when the apparent difficulties of the monetary authorities have made themselves much more manifest.

At the outset, it needs to be stressed that the study of financial innovation is important in itself because, as Kane (1984, p.4) has already suggested, there is a tendency amongst macroeconomic models to treat financial innovation as a purely exogenous development, and 'if policy makers do not incorporate policy-induced innovation into their ex ante planning framework,

their efforts at control will be biased toward shortfall.' This would almost re-echo the work of Lucas in this field who, for example, says that

'...given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of economic models.'
(1981, p.126)

The main point being made in this chapter, which stems from the key phrase 'policy-induced innovation' in the above quotation from Kane, is that all policy-makers need to incorporate the effects of their policies on financial innovative activity into their policy-making framework, that is, to put it another way, to *endogenise* the process of financial innovation.

To tackle the various issues involved, the discussion in this chapter will be organised as follows. Firstly, some preliminary remarks will be made regarding definitional and taxonomic problems of financial innovation. Before any theory of financial innovation can be put forward, it is necessary to define the concept of financial innovation, and to decide on how the various financial innovations may be meaningfully classified. That is the main objective of section 2.1. In the same section, a distinction has to be made between the *inducement* to financial innovation and its *diffusion* as it has been claimed by Podolski (1986) that the latter is often of more economic significance than the former.

Secondly, the various theories of financial innovation are

considered. The first theory (or approach) concerns the inducement to innovate in the financial sector as mainly a response to impulses in the real sector of the economy. Such a notion is implicit in the writings of Schumpeter. The second theory, mainly attributable to Silber (1975), considers how the financial firm might respond to various changes in its constraints which can take various forms, viz: governmental regulations, balance sheet constraints, market-imposed and internally-imposed constraints. Amongst the possible responses of the financial firm, one firm may be encouraged to undertake innovation. The third theory, attributable to the various writings of Kane (for example, Kane (1977, 1981, 1983, 1984)), to be considered purports to explain the rate of diffusion of a financial innovation, and places great emphasis on how governmental regulations may affect the rate of diffusion. Such a theory presumes that the main cause of financial innovation stems from the regulations imposed by the authorities, and many critics have pointed out that there are also other causes. However, in the final sub-section of section 2.2., a hybrid theory will be put forward which should highlight the complex interrelationships between impulses from the real sector, financial innovation, and changing constraints that a financial firm has to face.

Finally, section 2.3. considers the experience of the UK in financial innovation. The main approach is to offer an interpretation of the events in the monetary sector taking place since the 1960s and to show how financial innovative activity undermined the authorities' attempts at regulation.

2.1. Some preliminary remarks

2.1.1. Definitional problems

Many writers have noted a long-run parallelism between developments in the real and financial sectors.¹ As the study of financial innovations, until recently, has been a relatively neglected subject, there is often some difficulty in establishing an adequate analytical framework in which the process of financial evolution can be analysed. One possible starting point would be to note that similarities can be found between technological and financial innovations, and it would therefore be useful to analyse financial innovations along roughly similar lines to those for technological innovations in order to arrive at a working definition of financial innovations.

Technological change, as visualised by Schumpeter in his extensive writings on the subject, may consist of three steps, viz: invention, innovation, and imitation or diffusion.² Invention can be regarded as the very act of conceiving a new product or process and solving the purely technical problems associated with its application. However, as stressed by Schumpeter (1939, p.84), invention does not always induce innovation. For any invention to have any economic significance at all, it is necessary that commercial methods be devised to exploit that invention. Schumpeter used the term 'innovation' in a very special way, relating it to the implementation of a new process or method that alters the production possibilities of a firm. Innovation comprised the entrepreneurial functions required to carry a new technical possibility into economic practice for the first time – identifying the market, raising the necessary funds, building a new organisation,

cultivating the market and so forth. An innovation may not become apparent until imitation or diffusion has taken place on a sufficiently widespread scale. Imitation or diffusion is the stage at which a new product or process comes into widespread use as one producer after another follows the innovating firms's lead. On the basis of the preceding analysis, one may like to conclude that an innovation is deemed to have taken place if, and only if, it leads to a significant change in habits of, say, consumers.³

Some writers have made a distinction between *product* and *process* innovations.⁴ Product innovation is often understood to mean the setting up of a new production function and may be exemplified by the emergence of new products and new markets for them. In contrast, process innovations concern technological advances in production techniques which have the effect of increasing the marginal productivity of either one or some or all of factors of production which then, *ceteris paribus*, leads to an increase in the production of the firm concerned.⁵ Some further distinctions within process innovation may be made, namely that between labour- and capital-deepening. Capital-deepening is a form of process innovation in which the marginal productivities of capital-related inputs increase relative to those of labour-related inputs, leading to an increase in the marginal rate of technical substitution between labour- and capital-related inputs.⁶ As will be argued later on in this sub-section, capital-deepening innovations are potentially relevant for describing some of the more recent financial innovations.

In certain cases, as pointed out by Scherer (1980, pp.409-410), it is possible that the distinction between product and

process innovation may become blurred so that an innovation could simultaneously take the form of a completely new and novel product which may serve to improve the production process so that the marginal productivities of the factors of production may increase. The development of computers is an example that immediately comes to mind.

The various distinctions noted in the preceding paragraphs can be carried over into the realm of financial innovations. Regarding product innovation, typical examples would include the introduction of new financial products such as certificates of deposit in the 1960s, and the setting up of various futures markets. The most prominent example of process innovation would be the computerisation of the customer-bank relationship by the major banks and building societies. It is not too hard to find many examples in which the distinction between product and process innovations becomes blurred. The first example would be the introduction of automated teller machines (ATMs) by banks and building societies since the ATM is a product innovation, but also a process innovation in that it lowers the cost of providing services that would otherwise have been labour-intensive (that is, the ATM may take over some of the functions normally done by bank cashiers). The second example would be the introduction of *dedicated* dealing and quotation systems in the stock exchanges of the world – they are product innovations since the dealing terminals are specifically dedicated for the dealer in securities and could not have been used for other purposes, and they are process innovations since they serve to reduce transactions costs by making it far easier to obtain the prices of securities, for instance.⁷ Such process

innovations are regarded as capital-deepening as it implies a move away from labour-intensive inputs to capital-intensive inputs.

It has been noted by Silber (1975, p.63) and (implicitly by Podolski (1986, p.107)) that there seems to be an objective criterion on which technical innovations may be defined. It is either a product or a process innovation that qualifies for patent protection, and patent data may be the first source of data. However, financial innovations are not subject to patent protection so that it would be very difficult to arrive at some objective criteria on which financial innovations may be defined. Thus, financial innovations may go largely unnoticed at the initial stage, and it is only until wide diffusion has taken place that everyone recognises that a financial innovation has taken place.⁸ The reason for defining the occurrence of an innovation may now become clear when it was argued previously that such innovations can be labelled readily as such when it leads to a significant change in habits. In the realm of financial innovation, the origins of a financial product or service may be very obscure, and such innovations become much more manifest when wide diffusion has taken place. This seems to have been recognised implicitly by Silber (1975, p.64) who suggests a definition of a financial innovation as '[a]n innovation is a change in techniques, institutions or operating policies that have the effect of altering the way an industry functions.' Unless financial innovations are systematically recorded at every stage (and most importantly at the initial stage), it may prove difficult to have any clear concepts about financial innovation.

Podolski (1986, p.108) has drawn attention to another dimension of the definitional problem. It is sometimes the case that

financial innovation does not stem from completely novel ideas or practices. What may appear to be a financial innovation in a market may simply turn out to be that the new product or practice has spilled over from another market into the market. Two examples are cited by Podolski. The first one concerns the adoption of variable interest loans as a means to reducing risk was mainly derived from the practice of UK building societies offering variable interest mortgages. The second one concerns the Eurocurrency market in which the principles involved in Eurocurrency transactions were known before the First World War in the City of London who carried out transactions involving currencies of a third country.

Furthermore, Podolski distinguishes between 'creative responses' and 'adaptive responses'.⁹ Innovation was seen by Schumpeter (1934, pp.65-66) to lead to 'new combinations of productive means', that is to say, a creative response consists of doing something *outside* the range of existing practice. However, if, for example, in response to high and volatile inflation and interest rates, there occurs a change in practices such as shortening the maturity of loans to reflect increased uncertainty about future inflation and interest rates. Such a response may be termed an adaptive response which consists of doing something *within* the existing range of existing practices, and such changes may not be readily labelled as financial innovations. Therefore, a 'precise definition of financial innovation...is likely to be elusive' (Podolski (1986), p.108).

2.1.2. Taxonomic problems

Owing to its varied nature, financial innovations certainly

present some considerable taxonomic problems. Given the existing state of knowledge in this subject, it is not proposed here that a general classification system be devised. Rather, it will be instructive to analyse the various classification systems used by several studies of financial innovation. In particular, it does seem that a suitable classification system is often dependent on what objectives are to be fulfilled by any study on financial innovation. For example, if one were to conduct an investigation into the effects of financial innovations on the conduct and efficacy of monetary policies, a basic classification system would be to divide financial innovations into two broadly-defined categories such as whether or not such innovations have a *direct* influence on the structure and controllability of monetary aggregates, and whether or not they have an *indirect* influence on the monetary aggregates.

Consider Table 2.1. which presents a summary of some studies on financial innovations. In each entry, the author, the main objective of the study, and the classification system used by that study are all given. It should become apparent that the overall picture is that the classification system used is specific to the main objective of the study. Each classification system mentioned in Table 2.1. will be considered in turn now.

Firstly, the study by Silber (1975) introduces a theory of financial innovation in which it is hypothesised that the financial firm seeks to maximise its utility subject to several constraints. When any of these constraints change *substantially*, the firm may respond by innovating a new financial instrument or practice (this hypothesis is discussed further in sub-section 2.2.2.) The firm is capable of making different responses depending on the type of

TABLE 2.1: Summary of classification schemes used in various studies of financial innovation.

Author of study	Remarks	Classification system used
Silber (1975), p.73.	Discussion of the constraint-induced innovation hypothesis of financial product innovation. The classification system is based on the type of response that the financial firm makes in response to changing constraints.	<ul style="list-style-type: none"> A. Endogenising an exogenous item in the firm's balance sheet (e.g. by modifying the instrument or accepted practices with respect to it). B. Introduction of an existing financial instrument (e.g. from another market, or industry, or country). C. Attempting to generate demand for credit or liabilities by modifying an existing asset or liability in the firm's portfolio. D. Introduction of a completely new financial instrument or practice.
Hester (1981)	Discussion of how financial innovations affect the relative efficacy of monetary policies. A suggested classification system for this discussion would be that envisaged by Podolski (1986).	<ul style="list-style-type: none"> A. Innovations that have a direct influence on monetary aggregates. B. Innovations that have an indirect influence on monetary aggregates.
Kane (1981), p.357.	Discussion of how regulatory actions on part of the authorities can lead to financial innovations by regulated firms which take the form of circumventive responses. The classification system would be based on responses made by the financial firm to regulations that restrict expansion in their assets through specified strategies.	<ul style="list-style-type: none"> A. Expansion through geographical diversification strategies. B. Expansion through merger or diversification into other related or unrelated lines of business. C. Expansion by price competition.
Silber (1983), p.91.	An 'informal' test of the constraint-induced innovation hypothesis of financial innovation discussed in Silber (1975) using a wide range of financial innovations as data. There are four broad categories used.	<ul style="list-style-type: none"> A. Cash management. B. Investment contracts. C. Market structures. D. Institutional organisation.
Podolski (1986), p.111.	Discussion of the effects of financial innovations on the structure and controllability of monetary aggregates.	<ul style="list-style-type: none"> A. Innovations that have a direct influence on monetary aggregates. B. Innovations that have an indirect influence on monetary aggregates.

market it operates in, and on other factors. The classification system used is on the basis of the response made by the financial firm in response to a change in any of its constraints. The first type of response to be distinguished by Silber (1975, pp.67 and 69) concerns the endogenising of a previously exogenous item in the balance sheet by the financial firm. In the U.S. at the beginning of the 1960s, there was some part-disintermediation away from commercial banks by large corporations who lent directly to each other in the commercial paper market whose rates of return were more attractive relative to those offered by the commercial banks on their deposits. Because the commercial banks' market share of intermediation among large corporations was threatened, the banks introduced the negotiable certificate of deposit (CDs) as a measure designed to protect their market share. There was also another reason for the introduction of CDs. Banks were subject sometimes to episodes of excess loan demand which drove up interest rates, and therefore could lead to sizeable deposit withdrawals by large corporations. The CD was a financial instrument designed to endogenise the previously exogenous flows of large corporate deposits because it required a deposit for a fixed term so that it could forestall any sizeable deposit withdrawals, and made it easier for banks to bid competitively for funds at a time when their buffer stock of U.S. government securities began to be depleted after the war. In spite of the reserve requirements on CDs, the CDs market outgrew the market for commercial paper because it was more liquid and divisible than commercial paper.¹⁰

The second category to be distinguished by Silber (1975) concerns the introduction of an existing financial instrument or

practice from another market, or industry, or country. A specific example would be the use of repurchase agreements by the commercial banks towards the end of the 1960s. Repurchase agreements (RPs) were transactions in which one party agreed to sell U.S. Treasury securities to another party for a short period of time whilst simultaneously agreeing to buy back the same securities at an agreed price.¹¹ These RPs were originally used by non-bank security dealers as a means of financing their inventories of securities. When Regulation Q became binding in 1968, the commercial banks took on the practice of RPs as a means of circumventing Regulation Q because funds acquired through RP agreements were not subject to Regulation Q.¹² This type of innovation was an introduction of an existing practice from another market.

The third category given by Silber (1975) concerns the modification of an existing asset or liability in a firm's portfolio so as to generate fresh demand for its assets or liabilities. An example cited by Silber (1975, p.67) is the introduction of term loans by commercial banks in the 1930s. The final category noted by Silber (1975, p.73) are completely new items which may include the introduction of computers into banking, and the introduction of credit cards in the 1960s.

There is an inherent difficulty in Silber's classification scheme in view of the discussion towards the end of the last sub-section. Podolski (1986) argued that some financial innovations may not be innovations in the strict sense because they may have spilled over from another market into the market where the 'innovation' is supposed to have taken place. Thus, some of the responses by a

financial firm that were classified by Silber (1975) may not be innovations in the strict sense, especially categories A and B in the relevant entry in Table 2.1.

Secondly, Hester (1981), it ought to be stressed here, does not explicitly use a classification system. When discussing the effects of financial innovations on monetary aggregates, the most sensible approach to classification, as suggested by Podolski (1986, p.111) would be to divide financial innovation into two broadly defined categories which takes into account how financial innovations are likely to influence monetary aggregates and their controllability. The first category concerns those financial innovations that have a direct influence on monetary aggregates. Examples that immediately come to mind would include the creation of new money substitutes, or more specifically, the creation of new assets that are capable of serving as a means of payments, but do not possess the theoretical construction of money as a zero-interest asset. In addition, such new money substitutes may initially lie outside the scope of official definitions of money. That is, until the monetary authorities have more or less fully perceived the effects of new money substitutes on existing monetary aggregates, monetary aggregates will not be re-defined immediately. According to the simple hypothetical example given by Hester (1981, pp.143-146), the introduction of a new financial product or process may alter the signs of money multipliers, and if policymakers do not, at an initial stage, possess sufficient knowledge about the financial innovation that has just taken place, it may very well turn out that the signs of the money multipliers will become indeterminate which has implications for the relative efficacy of monetary policy. The second category given by

Podolski concerns those financial innovations that have an indirect effect on the monetary aggregates. It has been suggested by Podolski (1986, p.111) that such innovations are also relevant because they aid the understanding of influences shaping the structure of monetary aggregates.

The use of a classification system such as that suggested by Podolski will serve as a useful aid in understanding how some of the financial innovations discussed in Hester (1981) and Podolski (1986) affect monetary aggregates. There are several examples of financial innovations that had a direct influence on monetary aggregates. Referring to the earlier example of CDs, the introduction of CDs had the effect of weakening slightly the restrictive effects of monetary policy because the CDs made it easier for the U.S. commercial banks to bid competitively for extra funds by offering higher rates of interest on CDs as long as Regulation Q was not binding. Thus, they were still able to satisfy loan demand, and this may indeed be the very first example of the banking technique of liability management which is discussed in more detail in Section 2.3. A further example is provided by the growth of overnight RPs in the U.S. These were regarded as a part of the transactions medium, but the Federal Reserve failed to appreciate the apparent importance of overnight RPs as these agreements were used by the commercial banks to reduce their deposits at the close of business each day and to get back these deposits at the open of business the following working day. Thus, these overnight RPs had a distortionary effect on narrowly-defined aggregates such as M1 so that it undermined the Federal Reserve's ability to control narrow aggregates. The final example concerns the proliferation of close

substitutes for current accounts in the U.K. It is a well-known fact that when interest rates varied, it tended to distort M1 which was defined so as to include notes and coin in circulation with the public plus sight deposits. To take into account the new interest-bearing substitutes for non-interest bearing transactions balances, the Bank of England set up a new definition which encompassed a wider definition of transactions balances known as M2 (see Bank of England (1982b), pp.224-225). This will be discussed further in Chapter Four which concerns the definition of monetary aggregates.

Several examples can also be given for those financial innovations that have an indirect influence on monetary aggregates. The first example concerns the setting up of one-bank holding companies (OBHCs) in the U.S. when Regulation Q became binding in 1966. The commercial banks underwent congeneric transformations into OBHCs in order to escape the various regulations imposed by the Federal Reserve, especially Regulation Q, and were thus able to issue their own commercial paper at market rates of interest in order to raise funds. Another advantage offered by OBHCs was that these were not subject to the same stringent reporting procedures laid down by the Federal Reserve for commercial banks and thus impaired the Federal Reserve's ability at monetary control. The development of international banking had similar causes.¹³ The final example concerns the development of government-sponsored credit agencies whose primary functions are to intervene in the capital market so as to be able to offer loans to specially designated sectors of the economy (e.g. housing and agriculture) at interest rates below market rates, or just to guarantee

loans. A typical example for the U.K. would be the National Enterprise Board set up in 1975. The main effect of such credit agencies is to blunt the effects of restrictive monetary policies in the specially-designated sectors or 'enterprise zones'. So if monetary policy is to achieve the same *overall* restrictive effects as would have been the case in the absence of such credit agencies, interest rates would have to rise still further.

The third classification system to be considered here is that put forward by Kane (1981). This is mainly based on an analysis of how the banking system has managed to circumvent traditional banking regulations by innovating substitutes for their existing financial instruments and substitutes. The main objective of banking regulation is to restrict to some extent the banks' expansion of their assets so that a single bank, or a small group of banks, may not be seen to be monopolisers. Thus, in the relevant entry of Table 2.1., there are three broadly defined categories of circumventive responses in specific areas of banking regulation. The first category concern regulations designed to restrict the expansion in banks' assets through geographical diversification, and how banks have responded to prohibitions on branching in more than one state, and to restrictions on the number and location of their branches in any one state. Typical examples of such responses might include the establishment of bank-affiliates of the parent bank-holding company, and the provision of automated teller machines (ATMs) at remote sites such as supermarkets and offices. Such responses are designed to undermine the effectiveness of regulations designed to restrict the expansion in banks' assets through geographical diversification. The second category concern responses to regulations that restrict the

expansion in banks' assets by merger activity or *de novo* entry into other related or unrelated lines of business. Mergers are governed in the U.S. by antitrust legislation, and there are regulations which may prohibit the types of activities that any bank-holding company may undertake. Typical examples of circumventive responses may include the process of affiliation between two bank-holding companies rather than an actual merger, and the setting up of non-bank affiliates of the parent bank-holding company. The final category concern regulations that restrict the expansion of banks through price competition. Such responses may usually take the form of non-price competition. An example, with regard to interest payments that were restricted under Regulation Q, would be to offer *implicit* interest payments by way of improved facilities at the bank's branches. Apart from implicit interest payments, other forms of non-price competition also exist in the form of the proliferation of substitutes to traditional financial instruments. For example, substitutes for cheques may take the form of automated electronic transfer schemes; and substitutes for traditional current accounts could take the form of special accounts that circumvent regulations forbidding payment of interest on current accounts such as automated-transfer-of-surplus-funds accounts which automatically transfer surplus funds from current accounts into interest-bearing accounts.¹⁴ Such a classification system is designed to analyse regulation-induced financial innovations by considering how financial firms create substitutes for traditionally-regulated instruments and practices.

The final classification system to be considered here is that used by Silber (1983) in his 'informal' test of the constraint-

induced innovation hypothesis. The data is divided into four broad categories reflecting aspects of financial innovation, viz: cash management, investment contracts, market structures, and institutional organisation. The first category, which has some special relevance to the analysis of the transactions demand for money in Chapter Three, concern innovations that were mainly induced by the historically-high level of interest rates which increased the desire to economise on cash balances. A further change in cash management techniques was also prompted by the availability of technology which lowered transactions costs. Some innovations in investment contracts were caused by the increased volatility of interest rates, such as floating rate loans, and variable rate mortgages. The advent of electronic fund transfer and dealing systems also have affected market structures.

The overall conclusion to be drawn from the preceding analysis of classification systems used by various studies of financial innovations is that it would be very difficult indeed to devise a general classification system that would be capable of fulfilling every objective of each study on financial innovation. The best one can do is to use a classification system that would be specifically suited to the main objective of a study on financial innovation.

2.1.3. Inducement to innovation and its diffusion

A distinction has to be made between the inducement to innovate and the diffusion of an innovation. Such a distinction is necessary to complete any analysis of the process of innovation, and it can be argued that the diffusion of an innovation has much more macroeconomic impact than the mere occurrence of an innovation. Thus, the purpose of this sub-section is to review *briefly* some of

the factors that induce technological innovation and the rate of its diffusion. This will pave the way for the main theoretical discussion of financial innovation in the next section.

For analytical purposes, it will be assumed that the firm is a utility- or profit-maximising entity which accepts several constraints on what it can do in pursuit of its goals. There is some debate as to whether a monopolistic or a competitive market structure is more conducive to innovation. On the one hand, a monopolistic firm, by its very nature, is capable of earning super-normal profits which may create organisational slack in which case, at least, no externally imposed constraints are seen to be binding. The argument, as it goes, is that the monopolistic firm, by earning a rate of profit in excess of that required to cover all its costs, is capable of appropriating some of that profit to research and development activities (R&D) and adopt a more innovative strategy by following a balanced portfolio of R&D so that the cost of failure can be more than offset by its successes. Such innovations resulting from this view of R&D within a monopolistic firm may be termed 'success-slack innovation.'¹⁵ On the other hand, the competitive firm, because of the nature of the market it operates in, is only capable of earning a rate of profit that just about covers its costs so that there is little scope for R&D whose benefits must be weighed carefully against the possible costs of failure which may tend to be higher because the competitive firm may not be able to diversify its R&D to reduce the risks involved.

However, it is important to bear in mind that the distinction between product and process innovation must still be maintained. By their very nature, process innovations have the main effect of

lowering marginal costs such that both the monopolistic and competitive firm are induced to increase their output – the latter by a larger proportion than its monopolistic counterpart. Because the competitive firm is able to expand its output more than its monopolistic counterpart, it is able to realise a higher incremental quasi-rent contribution to its profits which should recoup its R&D outlay.¹⁶ So, if the term 'innovation' is used in its generic sense, then there is some ambiguity as to whether a monopolistic market environment is more conducive to innovation than a competitive market environment. If the term is used in a more precise sense, then the possible conclusions are likely to be more clear-cut, namely that a competitive market environment may be more conducive to *process* innovation whereas a monopolistic market environment may be more conducive to *product* innovation. The preceding discussion has some relevance to the analysis of financial innovation because it will become necessary to decide upon a market environment in which financial firm behaviour may be analysed.

In some situations, one or more of the firm's external constraints may change such that it forces a reduction in its utility or profitability. This may create an atmosphere within the firm in which stress is endemic. In such a situation, extra R&D effort may be undertaken in order to innovate a new product or process which will aim to restore the firm's utility or profitability. Such innovations may be termed 'failure-distress' or 'adversity' innovations since they constitute responses by the firm to adversity. The distinction between 'success-slack' and 'adversity' innovations is an important one for the discussion of the constraint-induced innovation

hypothesis introduced by Silber (1975) in sub-section 2.2.2.

To carry the preceding discussion even further, two main schools of thought on the inducement to innovate can be made out, viz: 'supply-push' or 'technology-push' and 'demand-pull' theories. In order to reinforce the concepts inherent in these theories of innovation, it will be useful to employ the marginal cost-benefit principle. It is assumed that the firm faces a cost-time tradeoff for its R&D programme in which any attempt to accelerate its R&D schedule will incur higher costs. However, if the firm spends too long in its R&D, there will tend to be diminishing returns beyond a point. Furthermore, the firm takes into account the possible benefits from innovating a product. It is assumed that the benefits which would most probably take the form of discounted sales revenue decrease as time passes because if the firm delays its R&D schedule, there is always a risk that it will lose its 'promised' market share if its rivals overtake it in their R&D effort and launch the innovative product before it does so. Indeed, one advantage of being able to accelerate the R&D schedule is that the firm would be able to launch the product earlier so tapping the market profits potential for much longer. Given this scenario, the firm will attempt to choose its optimum pace of R&D, and hence maximum discounted net revenues, by equating the marginal cost of accelerating its R&D schedule to its marginal benefit.

Now, the main concern here is with the *dynamics* of technological innovation.¹⁷ It is a reasonable assumption that as time passes by, there will be a tendency for the stock of technical knowledge to increase. As technological and scientific knowledge advances, what may be impossible today will be feasible but costly

tomorrow and easy the day after tomorrow. The main effect of this advance in technological knowledge, given the preceding analytical framework, is to recalibrate the firm's cost-time tradeoff such that the cost of its R&D schedule will be lower at all points in time, and the *marginal* cost of accelerating its R&D schedule will also become lower. All other things being equal, the firm will now be encouraged to innovate a product earlier. Even if innovating a product was previously deemed to be downright unprofitable, the lowering of the marginal cost of the R&D required may even encourage the firm to start the R&D programme because it is now profitable to do so. Thus, in such 'supply-push' or 'technology-push' theories, innovations are mainly seen to result from autonomous developments in technical knowledge. *Ceteris paribus*, this would suggest that supply-induced innovations are 'forced upon' the consumers by the innovating firms.¹⁸

The second school of thought on factors inducing innovation concern 'demand-pull' theories. Over time, demand conditions in the market will definitely change. Among the possible reasons for such changes would be the extra demand created by a growing population, by growing per-capita incomes and so forth. Such changes would be likely to increase the benefits accruing to a firm who undertakes R&D effort in order to innovate a product. Again, what may have been deemed as being unprofitable to innovate, the firm may now find that the changing demand conditions warrant such an innovation because it is now profitable. So, the main idea in demand-pull hypotheses is that innovation is a response to expanding profit opportunities in growing markets.

Of course, the preceding analysis was dependent on the use

of *ceteris paribus* assumptions. It is indeed most likely that a combination of supply-side and demand-side factors will interact to induce innovations over time. It ought to be stressed that the supply-push and demand-pull theories of innovation are not competing hypotheses, but complementary theories as suggested by Kamien and Schwartz (1982, p.36), namely that the former is useful as an explanation of technological change in the long-run, and the latter is useful in the short-run. The preceding discussion will have an important bearing when examining the process of financial innovation; in particular, the proposition made by Podolski (1986, p.109) that financial innovations are partly in response to changes in the real sector so that such innovations would conform more to demand-pull hypotheses will be examined in sub-section 2.2.1.

Now, Podolski (1986, pp.109-110) has argued that the diffusion of an innovation is more important than the mere occurrence of an innovation, and goes on to suggest that

'[j]ust as the mechanics of the diffusion of technological innovation is potentially relevant to the study of industrial policy, the mechanics of the diffusion of financial innovations may well be relevant to the study of financial and monetary policies.'

There are several factors that determine the rate of diffusion. These may include time lags which may be associated with uncertainty, the cost of adoption of an innovation, and the time taken for information to be systematically gathered on a wide scale; the time taken in learning; and the time taken for other firms to imitate the original innovation. Podolski suggests that, since financial innovations are not subject to patent laws or protection,

imitation may turn out to be easier and cheaper thus leading to 'swarms' or 'epidemics' of secondary financial innovation after the original innovation has taken place. Therefore, it is indeed possible that the diffusion of financial innovations is more rapid than that of technological innovations. The rate of diffusion may be hampered by regulatory forces imposed by regulatory authorities so that there may be a tendency for the rate of diffusion to increase during periods of de-regulation as embodied in the 'regulatory dialectic' framework which will be discussed in sub-section 2.2.3. below.

2.2. Towards a theory of financial innovation

In this section, several hypotheses concerning financial innovation are considered. The first one embodies a Schumpeterian approach in exploring a possible link between changes in the real sector and changes in the financial sector. The second hypothesis concerns what may be termed as 'constraint-induced' innovation as discussed by Silber (1975, 1983). The third one concerns 'circumventive' innovation which is embodied in a 'regulatory dialectic' framework by Kane (1977, 1981, 1983, 1984). At the end of this section, it will be argued that the above hypotheses can be regarded as complementary, and not as competing theories of financial innovation, so as to take into account the various complex interrelationships involved.

2.2.1. A Schumpeterian approach

There have been some occasions on which economists have considered financial innovation to be essentially a reaction to impulses from the real sector. For instance, Silber (1975, p.54) noted that '...the innovation of money responds to a stimulus in the

real sector and in turn influences the potential path of real economic activity.' The origins of such a notion may be traced to the writings of Schumpeter who is especially noted for his analyses of economic evolution instead of states of equilibrium.

In Schumpeter's scheme, the financial sector plays an important role in the evolution of the real sector. Banks are seen to be the dominant financial intermediaries, channelling funds from deposits placed by savers into loans required by firms to carry out their production. In addition, the banks are also seen to be creators of means of payment, and Schumpeter took great pains to distinguish 'credit creation' from savings.¹⁹ Essentially, in a capitalist system, savings, on the one hand, is seen to be the withdrawal of some of the productive resources which are then re-appropriated to their new employment 'through a shifting of means of payment.' (1939, p.111) On the other hand, credit creation is seen to be the process in which new means of payment are created and put at the disposal of the entrepreneurs by the banks, and a further shift in resources may be effected by shifts in the *purchasing power* of means of payment as a consequence of inflationary credit creation. In other words, economic expansion initiated by technological change is financed not by saving, but by credit creation which is considered to be '*...the monetary complement of innovation...*' (1939, p.111)

Now returning to the previous discussion in sub-section 2.1.3. regarding the distinction between supply-push and demand-pull theories of innovation, it should be noted that from the viewpoint of a financial firm, the creation of extra demand for funds by entrepreneurs seeking to finance their innovations would be

represented by increases in marginal benefit which may then induce the financial firm to innovate new sources of funds to satisfy that extra demand. It would seem, therefore, that technological progress in the real sector creates profitable opportunities in the financial sector by creating a demand for funds to support innovating entrepreneurs. This would reinforce Podolski's view, cited above, that financial innovations conform more to demand-pull theories of innovation than to supply-push theories.

It has been argued by Schumpeter (1939, p.111) that the logical relation between technological innovation and credit creation by banks is '...fundamental to the understanding of the capitalist engine...' and '...is at the bottom of all problems of money and credit...' Given the complexities of innovationary processes and the factors which account for financial expansion, the connection between the logical source of credit and actual financial processes might not be apparent and might thus be overlooked. It is 'in no case easy to discern the element of innovation under the mass of induced, derivative, and adventitious phenomena that overlies it. But in the sphere of money and credit the layer is so thick, and the surface so entirely at variance with the processes below, that the first impression of the reader may well be fatal.' (Schumpeter (1939), p.109) Schumpeter, himself, stressed the importance of not losing sight of the fundamental connection between changes in the real sector and its financial consequences by saying that

'...whenever the evolutionary process is in full swing, the bulk of bank credit outstanding at any time finances what has become current business and has lost its original contact with innovation or with the adaptive operations

induced by innovations, although the history of every loan must lead back to the one or the other.'(1939, p.114)

There, however, may be some instances when credit creation may lose its original contact with technological change. After a period of prosperity associated with innovational activity, there may follow a 'secondary wave'. Expectations of continuing prosperity are now no longer justified, but credit creation may continue unabated for some time, resulting in losses which may then put an end to the process of speculative and inflationary credit creation (see Schumpeter (1939), pp.148-149).

2.2.2. Constraint-induced innovation

The process of financial innovation is viewed by Silber (1975, 1983) from a microeconomic point of view of the financial firm. The starting point of Silber's analysis is the assumption that financial firms maximise utility subject to a number of constraints. The most fundamental constraint faced by the financial firm is that its balance sheet identity must hold. There may also be other explicit constraints built into the optimisation problem such as a target rate of growth for total assets, various regulatory requirements, or self-imposed liquidity requirements specifying a desired percentage of the total portfolio in some particular asset.

With respect to utility maximisation by the firm, Podolski (1986, p.185) suggests that 'this is simply another way of saying that, fundamentally, financial firms seek to maximise profits.' There are some grounds for doubting the validity of such a comment here. By assuming that, fundamentally, firms strive to maximise profits, the perspective is made too narrow, and it does seem more realistic

to assume that financial firms pursue a multitude of goals such as achieving a satisfactory rate of growth in its assets, a satisfactory distribution of risk among its assets, and so forth.²⁰ Thus, it may be argued that in maximising its utility subject to several constraints, the firm is striving to do as well as possible in achieving the multitude of goals which it has set for itself; in other words, no generality can be lost in assuming that the firm maximises its utility subject to various constraints, whereas the assumption of profit maximisation can lead to a loss of generality.

Now, the main essence of Silber's 'constraint-induced innovation' hypothesis is that new financial instruments or practices are innovated to lessen the financial constraints imposed on firms. Two types of changes in constraints are distinguished which induce financial firms to undertake the search costs required to modify its traditional policies. The first type concerns exogenous changes in constraints which lead to a reduction in the firm's utility and the firm innovates in an effort to return to its previous level of utility: such innovations are labelled 'adversity innovations' as noted previously in sub-section 2.1.3. above. The second case concerns what may be termed as an increase in compliance costs as a constraint becomes binding, leading to increases in shadow prices of the constraint in a linear programming context. For the purposes of analysis, it would be useful to label such innovations as 'circumventive innovations' since, as will be seen later, they are quite distinct from adversity innovations. The firm will try to remove or modify such constraints. If the constraint is an internally imposed one, the firm can simply revise or suspend that constraint whereas in the case of externally imposed constraints, the firm will

try to circumvent such constraints.

Silber (1975, p.66) argues that if such a hypothesis is to be operational, it requires further specification by distinguishing *normal* stimuli and responses from those of *abnormal* magnitudes. Thus, the approach is to define 'abnormal' magnitudes of change such that an abnormal reduction in the firm's utility or an abnormal increase in the shadow prices of any constraints will lead to innovation. It does seem that the definition of abnormal magnitudes of change is likely to be an arbitrary one as such changes are viewed in historical context *vis-a-vis* normal changes.

It has to be emphasised that there are development costs involved in financial innovations so that an initial rise in the shadow prices of constraints may not necessarily lead to financial innovation. Thus, Silber introduces a time dimension in his linear programming approach to financial innovation by suggesting 'that only a *sustained* increase in shadow prices over time will stimulate new product innovation.' (1983, p.90). Also, as previously discussed in sub-section 2.1.3., lower development costs may encourage even more innovative activity over time.

This hypothesis of financial innovation has been criticised by Podolski (1986, p.186) on the grounds that it '...is both too general and too specific.' Podolski argues that it is too general in the *microeconomic* sense that the firm innovates in order to achieve its goals by circumventing existing external constraints so that the main emphasis is on what may be called 'adversity' innovation owing to a reduction in the firm's utility. Furthermore, it is too specific in the *macroeconomic* sense that it applies to *existing* firms and may, therefore, not explain why new markets, new institutions, and new

monetary standards are set up. However, this is quite correct to some extent in that firms may carry out adversity innovations in response to a reduction in their utility. But, it is important to recognise that there are some cases in which there is no reduction in the firm's utility. It is simply that a constraint may become binding as reflected in rising shadow prices so that the firm's compliance costs increase. That is why, it is felt that 'circumventive innovation' would be a better term to describe those innovations that are in response to rising compliance costs, and to reinforce the distinction between circumventive and adversity innovations. The essential difference is that in adversity innovations, a firm seeks to restore its former utility by removing or modifying any of its constraints whereas in the case of circumventive innovation, the firm seeks to do away with compliance costs by circumventing existing constraints. Such a distinction is implicit in the study carried out by Ben-Horim and Silber (1977) who carry out two separate tests to investigate whether or not adversity innovations have an equal say in the process of financial innovation as circumventive innovations do.

The hypothesis of constraint-induced innovation was formally examined empirically in Ben-Horim and Silber (1977) who investigated the proposition that commercial banks sought to innovate in response to changing constraints. This proposition was investigated with the aid of a linear programming model which consisted of thirteen variables, a profit function, and eighteen constraints. It was assumed that commercial banks maximise their utility subject at least to a balance sheet identity constraint as well as various constraints. Ben-Horim and Silber state that '[t]he

assumption is that banks are profit maximizers. While utility maximization would have been more realistic, for our purposes this seemed to be an unnecessary complication.' (1977, p.282) Further to the criticism of Podolski's comment above, it is clear that the authors of the linear programming study on financial innovation regarded utility maximisation as more realistic, but to implement this model empirically would make the model unnecessarily complicated.

Using historic data for the period 1952-72, the model was solved period by period and the various shadow prices of the constraints were derived as a by-product of optimisation. The hypothesis was that the time series of the shadow prices should rise prior to the introduction of financial innovations, signifying a rising cost of adhering to existing constraints, and then drop after the innovation. Table 2.2. shows a time-series of approximate shadow prices for deposits for the period 1953-71: these figures were derived from one of the time-series plots in Ben-Horim and Silber (1977). The figures do reveal an interesting pattern. Further to the discussion of the introduction of CDs in sub-section 2.1.2. above, in the years up to 1963, prior to the introduction of the negotiable certificate of deposit, the shadow prices exhibited a quite marked increase, and then fell quite sharply thereafter. However, with respect to the development of an Eurodollar market as a means whereby the commercial banks could circumvent the existing regulations by conducting their transactions overseas, there was only a slight increase in the shadow prices. Furthermore, with regard to the introduction of repurchase agreements into the banking system towards the end of the 1960s, the increase in the shadow prices is certainly most marked. On the basis of the evidence just presented,

TABLE 2.2: *Time-series of shadow prices for deposits for the First National City Bank for the period 1953-71.*

<i>Shadow prices for</i>			
<i>Year</i>	<i>Demand deposits</i>	<i>Time deposits</i>	<i>Certificates of deposit</i>
1953	-0.15	0.45	-0.40
1954	0.10	0.40	-0.15
1955	0.30	0.85	0.15
1956	1.00	1.60	0.60
1957	1.35	2.00	1.20
1958	1.60	2.00	1.40
1959	1.60	1.90	1.45
1960	1.35	1.55	1.05
1961	1.25	1.30	1.40
1962	1.10	0.95	1.50
1963	1.20	1.00	1.40
1964	1.10	0.15	0.65
1965	1.30	0.05	0.30
1966	1.40	0.10	0.35
1967	1.70	0.55	0.85
1968	2.15	1.15	1.55
1969	2.55	1.55	1.95
1970	2.05	1.40	1.60
1971	1.35	0.95	1.15

Source: Ben-Horim and Silber (1977), p.288, figure 1a.

it is highly suggestive that the linear programming model was quite capable of identifying the pressures to innovate as exemplified by rising shadow prices of constraints that have changed, and seems to lend some support to the constraint-induced innovation hypothesis.

In an effort to evaluate the extent to which 'adversity' contributed to innovations by New York City commercial banks, Ben-Horim and Silber assumed that a fully specified bank utility function includes stockholders' wealth as a major argument, and that stockholders form their opinion of the value of the bank's stock based on the bank's profit, growth, 'soundness' and so on. The behaviour of bank stock prices can then be regarded as a reflection of investor evaluation of bank utility (1977, p.292). Thus, they take a look at the relationship between price-earnings and price-dividend ratios of commercial banks and industrials which is documented in Table 2.3. The data does suggest that the commercial banks have suffered relative to industrials in investor valuations of the banks' management. Therefore, Ben-Horim and Silber argue that '[t]his seems to qualify as as an adverse experience and helps explain the CD innovation.' This seems to be clear from Table 2.3. since the relative ratio of price-earnings ratios has declined down to 0.69 by 1960, and then increased quite abruptly thereafter. However, such an approach should not be construed as a definitive test on the extent to which adversity explains innovations by financial firms. Much more empirical work would be needed to examine this aspect even further.

Unfortunately, such an approach for trying to identify those times when there is some pressure on financial firms to innovate cannot really be utilised by the monetary authorities for monetary

TABLE 2.3: *Price-earnings ratio and price-dividend ratio of large New York City banks and Moody's industrials, and their relationship for the period 1952-72.*

<i>Year</i>	<i>NYC banks</i>		<i>Industrials</i>		<i>Relative ratios</i>	
	<i>P/E</i>	<i>P/D</i>	<i>P/E</i>	<i>P/D</i>	<i>P/E</i>	<i>P/D</i>
1952	13.10	22.73	10.53	18.02	1.24	1.26
1953	12.90	22.47	9.86	18.83	1.31	1.19
1954	13.35	22.27	11.43	21.46	1.17	1.04
1955	14.93	24.70	12.43	25.45	1.20	0.97
1956	13.13	23.04	14.44	25.71	0.91	0.90
1957	12.01	21.10	13.99	24.33	0.86	0.87
1958	13.31	22.42	18.03	26.04	0.74	0.86
1959	14.25	26.95	18.91	32.05	0.75	0.84
1960	12.43	25.58	18.00	28.74	0.69	0.89
1961	16.77	31.45	20.80	32.89	0.81	0.96
1962	16.29	30.21	17.11	29.50	0.95	1.02
1963	17.29	21.75	17.55	31.25	0.99	1.02
1964	16.95	33.67	18.01	33.56	0.94	1.00
1965	15.00	25.45	17.31	33.56	0.87	0.76
1966	11.76	24.75	15.90	29.07	0.74	0.85
1967	11.93	25.84	18.40	32.15	0.65	0.80
1968	13.24	29.41	17.97	31.85	0.74	0.92
1969	13.03	26.88	17.86	31.85	0.73	0.84
1970	11.38	24.81	17.70	27.72	0.64	0.89
1971	11.52	24.15	18.17	33.56	0.63	0.72
1972	13.62	29.85	17.87	37.74	0.76	0.79

Source: Ben-Horim and Silber (1977), p. 293, Table 4.

control purposes because of the enormous amount of detailed data that would be required – such reporting procedures, if ever at all implemented, would impose great costs on the financial institutions concerned. Whilst this study is a breakthrough in the analysis of financial innovations, its usefulness to the monetary authorities is limited, and certainly cannot predict when new markets, new institutions, and new monetary standards will be set up.

2.2.3. The regulatory dialectic framework

Kane (1977,1981,1983,1984) envisages an economic-political cycle in which the interests of regulators are in direct conflict with those of the regulatees. This is discussed within the 'regulatory dialectic' framework.

Initially, one could envisage a Walrasian-type economy consisting of many markets. It is assumed that only economic forces alone determine the final outcome known as the general equilibrium solution. The *tâtonnement* process is often used to explain how the Walrasian auctioneer guides all the markets towards a set of equilibrium prices in which all markets clear. A new variation of the Walrasian system is then introduced by Kane (1977) in which a 'political market' is added. A political market exists because politicians demand votes and supply regulation services in response to the electorate who supply the votes, and demand regulation services in those markets that they feel most disadvantaged! Kane explains in the following words how the existence of a political market affects the final outcome of the Walrasian economic system:

'Modern economics takes the Walrasian multiple-market auctioneer as its paradigmatic representation of the

contemporary market economy. ...a supplementary "political" market for regulation services opens up for business as soon as the Walrasian auctioneer finishes his work. Transactions in this political market disturb the general economic equilibrium and force the auctioneer back into action again. Continuing interplay between the political and economic markets produces broadly predictable cycles in which controls are set, markets adapt, and controls are re-designed and set for yet another round.' Kane (1977, pp.67-68)

The regulatory dialectic concept embodies an interpretative vision of cyclical interaction between economic and political pressures in regulated markets which is ever-continuing. Political processes of *regulation* and economic processes of *regulatee avoidance* are continually working against each other to determine the final outcome.

Exactly, what makes regulated firms respond more quickly to changes in the economic environment than regulatory agencies? This question may be best answered by considering the objectives of firms and government agencies. Both do have pre-conceived goals which they seek to achieve. Firms usually seek to fulfill a variety of objectives, viz: profit maximisation, market-share, and long-term survival. Government departments usually have the objective of maximising their budget, and their influence over the general decision-making process. The difference between these objectives makes the firm generally more alert to changes in the economic environment, and the firm is likely to have a highly-rationalised structure in which information flows more easily than in a

bureaucratic structure. Firms are then able to evaluate the situation as soon as it occurs, and make decisions on what to do. The marginal cost – marginal benefit principle is a useful tool to apply in deciding if any avoidance action against the regulations will be profitable. Generally, the more burdensome the regulations are, the greater incentive there is to take avoidance action. In contrast, government departments are slow to perceive the latest changes in the environment because of their complex bureaucratic structure which hinders the free flow of information. The previous discussion on the diffusion of financial innovations is particularly relevant here. When the original financial innovation occurs, it will be on a small scale so that it is likely to escape the notice of the regulatory authorities. It is only until wide diffusion has taken place that the regulatory authorities become aware of the new financial innovation and its economic significance. Thus there is an asymmetry in information, which leads to a greater efficiency of adaptability for firms than government departments.

Responses to changes in the environment are usually lagged. Two kinds of lags are distinguished by Kane (1981, p.358). The first lag is called the *innovative* lag and it defines the time taken by all regulatees, on average, to respond to new regulations by devising avoidance schemes in the form of innovations. The other lag, the *regulatory* lag, defines the time taken for the government agencies concerned to perceive a change in the environment and to contemplate the new unexpected problems posed by the new innovations, and to decide on what action (if any) to take. Due to differences in the efficiency of adaptability, these lags can differ in length. Generally, as innovative activity speeds up, the government

agencies are gradually overwhelmed so that the innovation lags get shorter, and the regulatory lags lengthen considerably. Thus, the differential between these two lags is often important enough in explaining the acceleration of financial change in recent times.

It would indeed be possible to shed some more light on the process of financial innovation by dividing the two lags mentioned above into further 'sub-lags'. Kane (1983, p.98) seems to have divided the innovative lag into two sub-lags, viz: the average lag in innovation by less-regulated institutions behind changes in technological and market opportunities, and the average lag in innovation by fully-regulated institutions behind their less-regulated competitors. Kane attributes the first sub-lag, which will be called the *innovative sub-lag*, to project evaluation and gestation. Kane seems to suggest that only the less-regulated institutions are capable of innovating in response to exogenous developments such as technological and market opportunities, and that the fully-regulated institutions are incapable of doing so. However, there may be some instances in which exogenous developments may present the fully-regulated institutions with further economically-feasible opportunities so that they are induced to innovate. As long as the innovation does not lie within the scope of existing regulations, then an innovative sub-lag would be observed. Thus the innovative sub-lag will be taken to refer to the average lag in innovation by *all* institutions behind exogenous developments.

The second sub-lag is attributed by Kane to resistance by fully-regulated institutions to change. But, this is not the sole reason why fully-regulated institutions may lag behind their less-regulated competitors in innovative activity. As will be shown

in the following section on the UK's experience of financial innovation, the London clearing banks could not compete effectively with the unregulated secondary banks in the late 1960s because the clearing banks were seriously handicapped by existing conventions and regulations which were a feature of their cartel. In order to overcome this handicap, the clearing banks set up their own subsidiaries so that they could compete more effectively with the secondary banks. Such a reaction could be termed the *catch-up sub-lag* because the clearing banks had to do two things: first, to overcome their resistance to change, and second, to devise avoidance schemes in order to circumvent the existing conventions and regulations before they could compete with the secondary banks. This catch-up sub-lag may be essentially composed of two sub-lags: the *resistance sub-lag*, and the *avoidance sub-lag*. The resistance sub-lag concerns the average time taken by cartelised or fully-regulated institutions to overcome their resistance to change, and the avoidance sub-lag concerns the average time taken by *any* (i.e. both fully- and less-regulated) institutions to devise avoidance schema in order to circumvent the existing conventions or regulations.

As already argued by Kane (1983) the whole process of financial innovation may become more comprehensible if it is considered in terms of the various lags discussed above. Thus, by dividing Kane's originally defined lags into sub-lags, the process of financial innovation can be made even more comprehensible. This is particularly true if the diffusion of financial innovations is considered. The rate of diffusion of a financial innovation can be determined by both the resistance and avoidance sub-lags. That is,

the rate of diffusion may be lower, the longer are the resistance and avoidance sub-lags, and the converse also holds.

The division by Kane (1983) of the innovative lag into two sub-lags could be carried over into the regulatory lag. It is proposed here to divide the regulatory lag into two sub-lags, viz: *recognition* and *re-regulation* sub-lags. The recognition sub-lag serves as a useful means of emphasising the importance of the diffusion of financial innovations. As pointed out earlier, financial innovations may occur on a small scale so that they may escape the notice of the regulatory authorities (or may even appear to be insignificant and of no consequence). A financial innovation is said to be recognised by the regulatory authorities when the economic significance of the innovation has been perceived. Then there is a further sub-lag in which the authorities contemplate the problems posed by the financial innovation, and they may decide on what action to take. It is important to recognise that 're-regulation' does not necessarily mean that additional regulations will be introduced. Rather, it is intended here that the term should be a generic one, namely that it will embody several possible courses of action, viz: de-regulation, no action, and further regulation. It should be made clear here that the re-regulation sub-lag should not be confused with the definition of the re-regulation lag given by Kane (1984, p.6) since the latter includes the recognition sub-lag.

The process of financial innovation, as embodied within the regulatory dialectic framework, is shown schematically in figure 2.1. There are many ways of interpreting the diagram. If the diagram is interpreted along the lines of the discussion in Kane (1983), exogenous developments such as technological changes induce

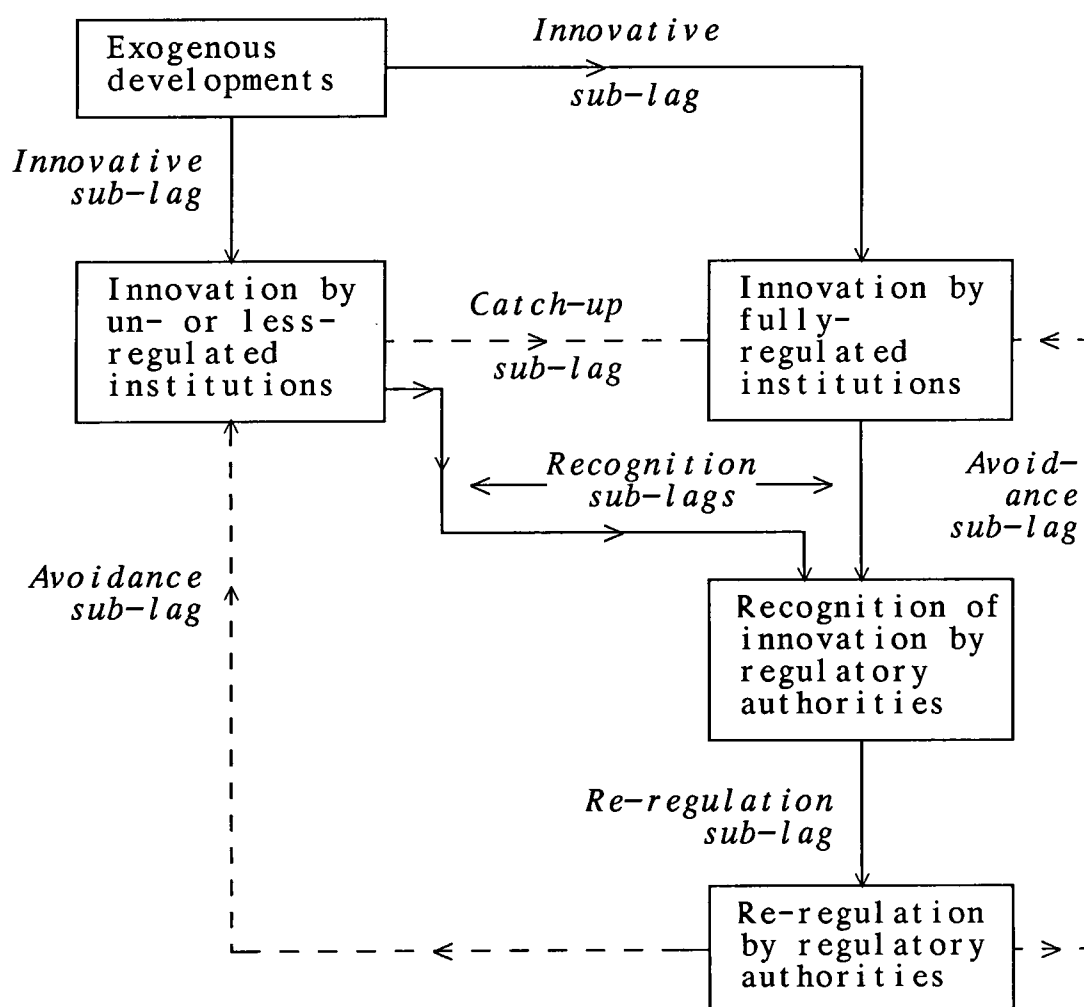


FIGURE 2.1: A schematic representation of the process of financial innovation within the regulatory dialectic framework.

less-regulated institutions to innovate, and the time taken for this innovation to be evaluated may be represented by the innovative sub-lag from 'exogenous developments' to 'innovation by un- or less-regulated institutions'. In order for the fully-regulated institutions to compete effectively against their less-regulated counterparts, they need to overcome their resistance to change, and then devise avoidance schemes which are designed to circumvent the existing conventions and regulations preventing the fully-regulated institutions from taking advantage of the innovation. This would be shown as a catch-up sub-lag from 'innovation by un- or less-regulated institutions' to 'innovation by fully-regulated institutions'. Since the innovation takes time to be recognised by the regulatory authorities, there are recognition sub-lags involved. Once the innovation has been recognised, the regulatory authorities will contemplate the problems posed by the innovation, and this is shown as a re-regulation sub-lag. If they decide to introduce further regulations, this may prompt further innovation as shown by the avoidance sub-lags, and the process known as the regulatory dialectic is initiated, and may repeat itself again. If the authorities decide to take no action at all, or to pursue a course of de-regulation, then there will be no avoidance lags as there are no new regulations to be circumvented - hence the reason for the broken lines showing the relevant avoidance lags. In such an outcome, the regulatory dialectic process may not repeat itself, unless there are further innovation-inducing developments.

However, if exogenous developments induce an innovation that lies well outside the regulatory net, then the fully- and less-regulated institutions are both able to innovate in response to

exogenous developments. Since both types of institutions innovate simultaneously, there is no catch-up sub-lag involved. However, the recognition and re-regulation sub-lags still exist as it takes time for the regulatory authorities to recognise the occurrence of an innovation, and they must decide on what (if any) action to take.

Over time, one can observe some periods in which new regulations are introduced, and some periods of de-regulation. This leads to a distinction between two different types of innovation which are dependent on the prevailing regulatory climate. During periods of regulation, regulated firms become burdened with the new regulations in the form of compliance costs.²¹ This leads them to examine ways of circumventing the new regulations by diversifying out in their product lines or services. This type of innovation is regarded by Kane (1981, p.358) as increasing the *productivity in regulatory avoidance*. In times of de-regulation, there may a tendency for competitive pressures to intensify (say, a squeeze on profit margins) so that there is an incentive for the firm to provide products and services at lower costs. This leads to the second type of innovation which is seen to increase *technical productivity*. A good example of this form of innovation is the growing provision of money transmission services by electronic means. In general, the increasing use of technological advances in communications, and computers opens up a whole new range of economically feasible options for the firm.

Finally, it may be argued that the two hypotheses just discussed above should be regarded as complementary theories of financial innovation, rather than as competing theories. In the regulatory dialectic framework, it was suggested that financial

innovation occurred in response to regulatory and other external constraints, and this invariably re-echoes the hypothesis put forward by Silber (1983), although Kane (1983) either was not aware of Silber's contribution, or did not explicitly acknowledge it. Even during periods of de-regulation, there may still be some constraints (such as market forces) that will encourage financial firms to innovate. But, it can be argued that Kane's regulatory dialectic framework goes much further than Silber's constraint-induced innovation hypothesis since the latter is specifically concerned with the *inducement* to innovation, whereas the regulatory dialectic approach may be concerned both with the inducement to innovation *and its diffusion*.

2.2.4. A hybrid theory of financial innovation

The regulatory dialectic approach to financial innovation has been subject to some criticism on the grounds that it is a far too narrow explanation of the financial innovation process. This is especially true of Silber (1975, p.64) who believes that regulation-induced financial innovation is just a subset of his more general constraint-induced innovation hypothesis. Still, it must be argued here that the constraint-induced innovation hypothesis is still too narrow to explain the full process of financial innovation. It is rather doubtful if a *single* theory could be formulated to explain all aspects of financial innovation. In order to take into account the various complex interrelationships involved, a *hybrid* theory of financial innovation is put forward here which is a combination of all the previous hypotheses discussed in this section. Such a theory will take into account the link between the real and financial sectors of the economy, the changing constraints that induce

innovation, and the lags involved which explain the rate of diffusion of innovations.

It is difficult to say precisely where the process of financial innovation began, but it will be assumed for the sake of argument that in the very long run, there are additions to the stock of knowledge which may then induce technological innovation in the real sector. It has been noted by Schumpeter (see sub-section 2.2.1. above) that innovations are financed by credit-creation rather than by savings. Credit creation may typically take the form of the creation of new means of payment which may then finance further expansion in the real sector; this is a point which Silber (1975, p.54) acknowledged. Thus, the hybrid theory proposes that financial innovations facilitate further expansion in the real sector which may then induce more demand for credit-creation in the very long run. The authorities may seek to keep the rate of expansion in the economy at a respectable pace, that is, to minimise deviations from the target growth rate. It has been argued by Sylla (1982, p.25) that the modern phenomena of regulation-induced financial innovation is a means whereby economic expansion could be accelerated by financial institutions that circumvented government regulations and restrictive monetary arrangements. If figure 2.1. above is re-considered, it will be seen that the real sector of the economy has been 'exogenised' in order to isolate the regulatory dialectic approach to financial innovation from the Schumpeterian approach. The hybrid theory 'endogenises' the real sector so that the complete model may be closed off. Figure 2.2. below shows the modified version of figure 2.1. in which the real sector is now explicitly included. Changes in the real sector may either be

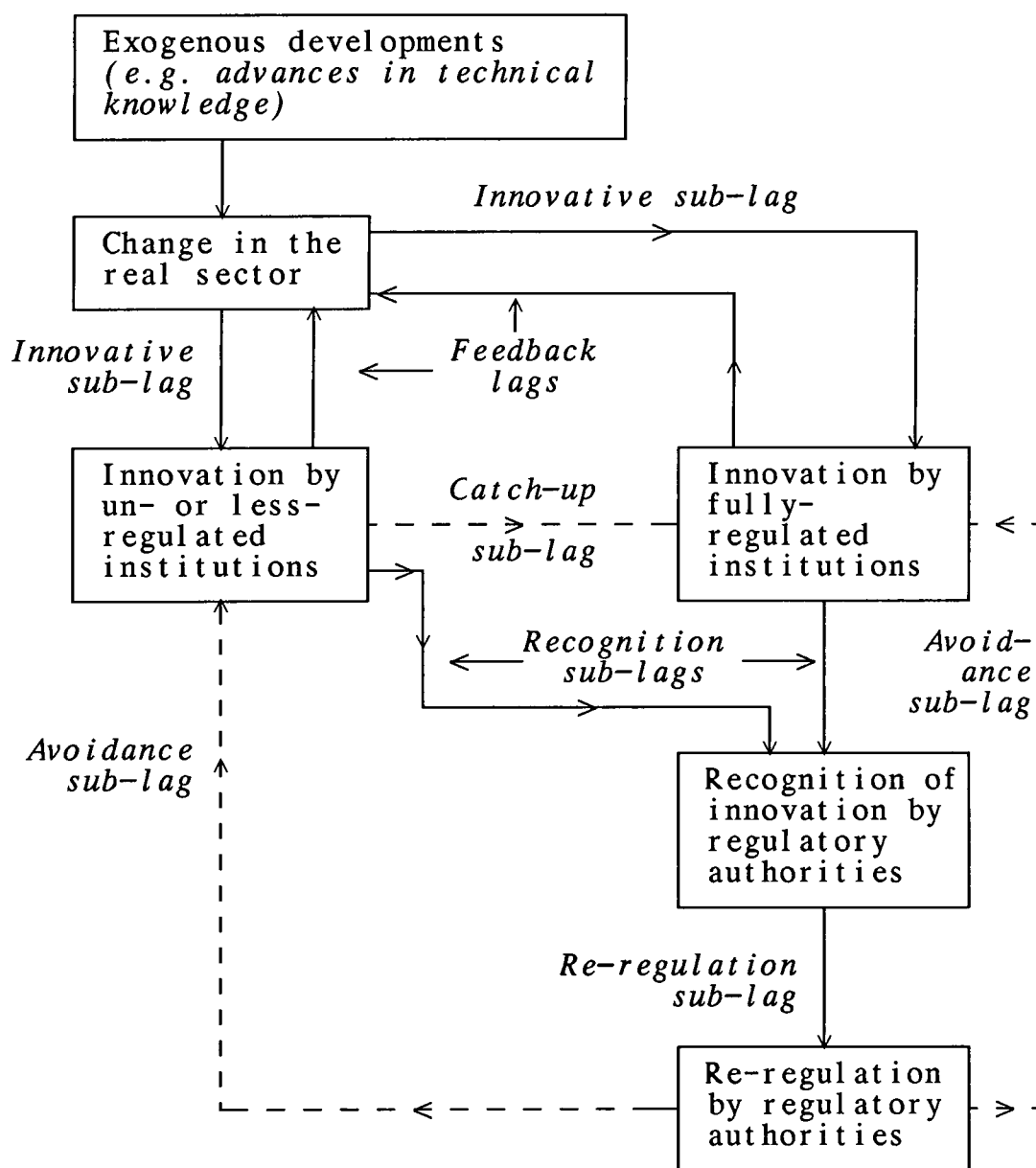


FIGURE 2.2: A schematic representation of the process of financial innovation under the hybrid theory

attributable to exogenous developments such as advances in technical knowledge and population growth, or to 'credit-creation' in the financial sector as shown in figure 2.2. In the latter case, this is shown by what may be tentatively termed as 'feedback lags' from innovations by unregulated or regulated institutions to changes in the real sector. As Schumpeter has already emphasised, it is very important not to lose sight of the fundamental link between credit creation and change in the real sector, so the hybrid theory serves as a useful means in which such a link can be kept in sight in the very long run.

One particularly novel feature of the model outlined in figure 2.2. is that the regulatory dialectic does not necessarily have to come into effect every time an innovation occurs. For instance, changes in the real sector may induce financial innovation which may then feed back directly into the real sector. Or, financial innovations may go through the regulatory dialectic framework, and then feed back into the real sector as a consequence of avoidance action.

Among the possible topics for future research, one could investigate the possibility that government regulations also affect the real sector as well as the financial sector which would be an extension of the regulatory dialectic framework, namely that regulations may have an effect on changes in the real sector which in turn may affect the propensity to innovate in the financial sector.

The preceding paragraphs have tried to explain how financial innovation may occur, and at an accelerating rate. To shed more light on the various hypotheses, the UK experience over the last three decades will now be discussed with reference to monetary

control.

2.3. The UK experience of financial innovation and monetary control

The task of surveying financial innovations that have taken place is not an easy one because regulatory forces are not just the only forces that induce financial innovation. There are also other interrelated factors that are just as important in explaining the process of financial innovation. Amongst such factors are those relating to high and volatile rates of inflation and, therefore, interest rates, and technological advances which all do have important implications for the relative efficacy of monetary policy. Thus, the strategy adopted here is not to consider the regulatory dialectic approach to financial innovation in isolation. Rather, it will be considered in conjunction with other interrelated factors.

2.3.1. The development of the financial sector during the 1960s

(a) *The scenario.* The 1960s saw the beginning of a radical change in the UK money markets and the way the banks went about their business. At the beginning of the decade, the UK monetary scene was dominated by the clearing banks which operated exclusively through their branch networks in a way which had changed little since the beginning of the century. They supplied 95 per cent of the sterling deposits held by UK residents, as Table 2.4. shows. Most of these deposits were still on current account, interest rates on other deposits being linked via the 'Bank Rate' cartel arrangement to the Bank of England's rediscount rate. So, banks could only compete for deposits through non-price means; by expanding the size of their branch networks, for example. These

deposits were supported by bank holdings of cash and liquid assets, Treasury and commercial bills, and call money with the discount market, which were held to ratios set by the Bank of England. There were well-developed secondary markets in Treasury and commercial bills, bank liquidity being ultimately supported by the Bank's operations in these markets and the rediscount facility open to discount houses. Imbalances between different banks were reflected largely in the position with the discount market rather than through inter-bank transactions, and accounts were settled through the Bank of England. Markets in other short-term instruments were still in their infancy. The building societies were collectively small in relation to the clearing banks; according to available figures, in 1955, these building societies numbered nearly 800 of which the top five societies accounted for 40 per cent of the total assets of the building society movement, and the top twenty societies accounted for 65 per cent (Bank of England, 1983, pp.368-369). Their activities were very narrowly defined by the then existing regulations which only permitted them to take deposits exclusively from the personal sector, and to make advances exclusively to finance owner-occupation of houses (i.e. mortgages).

(b) Innovation-inducing developments. The scenario sketched out above had dramatically changed by the end of the decade. As Table 2.4. makes abundantly clear, the clearing banks had lost out significantly to the non-clearers, especially to the subsidiaries of American and other overseas banks in London. This reflected the fact that whilst entry into the cartel was in practical terms hardly possible, entering into competition with it was relatively simple. No licence or other supervisory requirement governed the taking of

TABLE 2.4: Changes in the distribution of UK residents' sterling bank deposits during the 1960s

(£ millions, end December. Figures in parentheses show percentage of total deposits)

	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
(a) London clearing banks	7,168 (84.5)	7,609 (82.8)	8,065 (82.0)	4,477 (80.5)	8,533 (79.6)	9,261 (77.4)	9,766 (75.1)	9,711 (74.5)	10,297 (73.6)	12,201 (74.9)
(b) Other clearing banks	886 (10.4)	921 (10.0)	949 (9.6)	1,006 (9.5)	1,030 (9.6)	1,136 (9.5)	1,208 (9.3)	1,215 (9.3)	1,313 (9.4)	1,438 (8.8)
(c) Other UK banks	431 (5.1)	656 (7.2)	815 (8.4)	1,047 (10.0)	1,159 (10.8)	1,571 (13.1)	1,939 (15.0)	2,103 (16.2)	2,375 (17.0)	2,648 (16.3)
(d) Total	8,485	9,186	9,829	10,530	10,722	11,968	12,907	13,029	13,985	16,287

Source: Wilson Committee (1978), volume V, Table 1, p.112.

deposits from the public by institutions outside the cartel. After the relaxation of foreign exchange restrictions in the late 1950s, these institutions had initially developed alongside the London markets in dollar deposits – the Eurodollar market – but had then found it profitable to take on business in sterling. These banks acted on a wholesale basis in the euromarkets, taking in large denomination deposits for a fixed period of time and lending them on in a similarly structured way, and the growth in foreign currency deposits as a proportion of total deposits is clearly shown in Table 2.5. In contrast to the retail banks, these institutions maintained the viability of their balance sheets by matching the maturity of their liabilities to that of their assets rather than by holding balances of liquid assets.

The overseas banks brought these techniques with them when they moved into their sterling habitat. They proved quite suited to this environment. Together with indigenous secondary banks such as the accepting houses, they helped establish the sterling inter-bank market during the 1960s, and in 1967, they were among the first institutions to issue sterling certificates of deposit (CDs). Together with the discount houses, they developed an active secondary market in CDs. This innovation allowed the issuing banks to take in money at a fixed interest rate and maturity (typically three months) whilst giving the holder the option of liquidating his deposit at any time by selling on to a third party. If the issuers of CDs took the view that interest rates were to rise, it encouraged them to issue CDs of longer fixed maturities so that they expected to deploy the funds provided by CDs profitably by taking advantage of arbitrage opportunities in the money markets.

TABLE 2.5: Distribution of liabilities between sterling and other currencies, 1957-1979

(£ billions, end-December, figures in parentheses show percentage of total liabilities)

	1957	1962	1967	1972	1977	1978	1979
Sterling liabilities	--	10.5	14.9	30.8	56.5	62.8	76.9
		(88.24)	(70.28)	(47.09)	(31.72)	(30.80)	(31.21)
Other currency liabilities	--	1.4	6.3	34.7	121.6	141.2	169.5
		(11.76)	(29.72)	(52.91)	(68.28)	(69.28)	(68.79)
Total liabilities	8.7	11.9	21.2	65.4	178.1	203.9	246.4

Source: Wilson Committee (1980), Appendices, Table 3.11, p. 406.

Meanwhile, an active wholesale market in local authority deposits had grown up. These were held by banks initially, but non-bank holdings grew rapidly, avoiding the need for bank intermediation. These markets grew up alongside the established markets in Treasury and commercial bills and call money with the discount market, and for this reason, such markets are known as 'parallel' markets.^{2 2} Table 2.6. documents the growth of the parallel markets during the pre-CCC years as well as their size relative to existing markets.^{2 3}

The loss of market share experienced by the clearing banks over this period reflected the Radcliffe proposition that the main effect of controlling any set of financial institutions such as the clearing banks would be to cause their business to be lost to competitors. The clearing banks were clearly handicapped by the special-deposit and liquid-asset conventions which they obeyed. Yet the major handicap seems to have been the Bank-rate cartel arrangements which ruled out the issue of wholesale deposits. Under these arrangements, they could only take in 7-day deposits at interest rates linked to the Bank of England's rediscount or Bank Rate. They were, however, able to set up subsidiaries which operated outside these arrangements, and by the early 1970s, most of the major London clearing banks had done this. This illustrates the 'catch-up sub-lag' discussed in the previous section: initially the clearing banks were reluctant to compete with the non-clearing banks by making use of the new money markets because of their resistance to change (the resistance sub-lag)^{2 4} and then when their share of total deposits declined, it encouraged the clearing banks to take action by setting up their own subsidiaries which was a way of

TABLE 2.6: The London sterling money markets, 1957-1979

(£ millions, end year)

	1957	1962	1967	1972	1977	1978	1979
Money at call with discount market	903	1,186	1,662	2,530	3,513	4,004	4,435
Treasury bills	3,388	3,042	3,156	1,719	3,950	2,813	2,480
Commercial bills	250	400	725	1,188	2,169	3,393	5,588
Local authority bills	450	1,071	1,750	240	443	499	599
Other temporary local authority debt				2,145	2,896	3,788	5,135
Deposits with finance houses	99	337	591	437	921	967	1,117
Inter-bank deposits	—	508	1,309	5,068	11,407	13,205	16,433
Sterling certificates of deposit	—	—	—	4,934	4,546	3,678	3,692
Other market loans by the banking sector	—	—	—	—	4,296	4,581	5,314
Total identified	5,090	6,544	9,193	18,261	34,141	36,928	44,793

Source: Wilson Committee (1980), Appendices, Table 3.70, p.510.

operating outside the cartel arrangements (the avoidance sub-lag).

Credit control by the Bank of England during the years up to 1971 rested mainly on two weapons, viz: liquidity controls, and quantitative and qualitative controls. The special deposits scheme was introduced during 1960 in order to control the volume of lending.²⁵ The main idea of the special deposits scheme was to mop up any excess liquidity by calling in special deposits which yielded a rate of interest, but were not included in the liquidity reserve ratio; in effect, it operated rather like a variable reserve ratio, so that the volume of lending could be controlled. The special deposits scheme was largely ineffective because the clearing banks found a way of circumventing it by disposing of government gilts to bring in more liquidity. Thus, the Bank of England resorted to direct quantitative controls which took the form of formal requests to the clearing banks to curb their lending. However, at this point, the Bank of England had not fully appreciated the full importance of the non-clearing banks and the new money markets.²⁶ In response to a higher volume of lending, the Bank cast its regulatory net wider to include the non-clearing banks in 1965 by asking them to limit their lending. It was not until 1967 that the Bank was very concerned about the rapid growth in the non-clearing banks so that it wished to overhaul the existing system of *ad hoc* requests for lending to be curbed in favour of a more comprehensive system. (Grady and Weale, 1986, pp.48-51) This illustrates the operation of the recognition sub-lag in the regulatory dialectic framework: at this point the Bank became more or less fully aware of the economic significance of the non-clearing banks and the new money markets, and therefore wished to

re-regulate. However, at the end of the 1960s, there was a shift in banking, intellectual, and political opinion away from regulation towards a more competitive banking system in which the allocation of credit would become more efficient. In order to design new regulations, there followed a period of drastic de-regulation. In anticipation of the dismantlement of lending ceilings and other restrictions, there was a boom in merger activity in the banking sector. By the end of 1969, the number of clearing banks had been substantially reduced in anticipation of a more competitive banking environment.

2.3.2. The problem of liability management

(a) *The initial arrangements of CCC.* The Bank of England's initial proposals for Competition and Credit Control (CCC) were published in May 1971. The underlying principle was that the Bank would act upon the banks' sterling deposit base rather than by directly guiding their lending to the private sector. In order to provide a 'firm base' for this policy, the banks were to observe a minimum reserve asset-ratio. In addition, they would place special deposits with the Bank when these were called for. This policy implied a greater reliance upon changes in interest rates as a way of controlling private sector lending and in order to facilitate such changes the Bank's tactical support of the gilt-edged market was to be limited. The idea was that this policy would be more flexible than the previous one, putting all banks on a common basis and allowing them to compete freely. In order to make way for this, the clearing banks agreed to abandon the bank rate cartel agreement.

The operational arrangements were agreed with the banks

over the summer months, and in September 1971, the new arrangements took effect, ushering in a new era of competitive banking. The precise details were as follows (Bank of England, 1971b, 1982). Basically, the banks were to maintain a minimum of 12.5 per cent of their 'eligible liabilities' in the form of 'eligible reserve assets'. The definition of eligible liabilities encompassed the following items (Bank of England, 1982, p.76):

- (i) All sterling deposits of an original maturity of two years or under, from UK residents (other than banks) and from overseas residents (other than overseas offices).
- (ii) All sterling deposits, of whatever term, from the UK banking sector *net* of sterling claims (including non-reserve asset lending to listed discount market institutions).
- (iii) All sterling CDs issued, of whatever term, *less* any holdings of such certificates.
- (iv) The bank's net deposit liability, if any, in sterling to its overseas offices.
- (v) The net liability in currencies other than sterling.
- (vi) A reduction equivalent to 60 per cent of the net value of items in transit.

and the eligible reserve assets included

- (i) Balances held at the Bank of England (other than special or supplementary deposits).
- (ii) Treasury Bills.
- (iii) Secured money at call with the London money market.
- (iv) Government stocks with one year or less to maturity.
- (v) Local Authority bills eligible for rediscount at the Bank

of England.

- (vi) Commercial bills eligible for rediscount at the Bank of England up to 2 per cent of eligible liabilities.

Clearing bank holdings of notes and coin were not eligible reserve assets under these arrangements.

It is useful to consider the control mechanism envisaged by the authorities because it serves as a useful yardstick in measuring the circumventive powers of the banking sector in frustrating the initial CCC arrangements. An invaluable insight into the way the authorities expected these arrangements to work in practice is provided by an address by the Governor to a conference of international bankers:

'It is not expected that the mechanism of the minimum asset ratio and Special Deposits can be used to achieve some precise multiple contraction or expansion of bank assets. Rather the intention is to use our control over liquidity, which these instruments will reinforce, to influence the structure of interest rates. The resulting changes in relative rates of return will then induce shifts in the asset portfolios of both the public and the banks. Of course, we do not envisage that there can be a nicely calculated relationship between the size of calls for Special Deposits and the achievement of a desired objective. We expect rather to achieve our objectives through market mechanisms. Special Deposits can be used not only to mop up any abnormal excess liquidity, but also to oblige the banking system to seek to dispose of assets not eligible for the liquidity ratio, for example

gilt-edged stocks of over one year's maturity. By using Special Deposits in this way we shall be able to exert, when appropriate, upward pressure on interest rates - not only rates in the inter-bank market, but also rates in the local authority market and yields on short-term gilt-edged stock.' (Bank of England, 1971a, p.197)

The Bank saw liquidity as being influenced both directly by the supply of reserve assets (via public sector borrowing from the banks) and indirectly by calls for special deposits. The banks would then respond to this reserve-asset pressure by selling secondary reserve assets such as short-term gilts and local authority deposits, increasing yields in the associated market as well as in the inter-bank market. This in turn would put upward pressure on bank lending rates and presumably lead to a reduction in private borrowing.

(b) The early operational experience. The CCC system began its existence facing pressures of a large and unknown magnitude in the form of the frustrated demand for credit hanging over from the previous regime. The scale of these pressures quickly revealed itself as bank lending to the previously restricted sectors accelerated, especially in the commercial property sector.

Another factor put forward for the rapid increase in lending was that the minimum lending rate (MLR) initially failed to keep pace with the money market rates so opening up a differential between the rates of interest. This led to the phenomenon of 'round-tripping' in which corporate treasurers were induced to draw on their overdrafts and lend on the proceeds in the money markets at a higher rate of interest, thus leading to profitable arbitrage.²⁷ The consequence of such developments was an explosive rate of

growth in broad monetary aggregates; for example between 1971 and 1973, the M3 measure grew by 61.5 per cent. This was coupled with the 'reintermediation' process in which the major clearing banks aggressively sought to win back business lost to other banks during the pre-CCC years.

Meanwhile, towards the end of 1973, there was a collapse in the commercial property market which led to a secondary banking crisis. This crisis came about mainly because of the secondary banks' innovative use of the money markets, and their failure to diversify out their asset portfolios which had a disproportionately large proportion in the form of advances to commercial property developers. Thus, when the first defaults came through from the property developers, it had a domino effect on the banking system because of the close interaction of the money markets. The crisis was prevented from spreading out further afield in the banking system by a rescue operation - known as the 'lifeboat' - organised by the Bank of England. In retrospect, hard lessons were to be learnt from this experience. As the Bank of England put it, '[a] principal lesson of these years has been that any system of control sets a premium on avoidance and circumvention; and that prudential regulation is not immune from this rule.' (Bank of England, 1983, p.368), and the Bank resolved to review the prudential arrangements for the banking sector as a whole which culminated in the 1979 Banking Act.

In any event, by the end of 1973 the authorities had decided to retreat from the initial CCC arrangements in favour of a more restrictive regime by the introduction of the Supplementary Special Deposits (SSD) scheme because they were alarmed at the

rapid growth in monetary aggregates. This scheme, known as the 'Corset', operated by requiring banks to make non-interest bearing deposits with the Bank if their interest-bearing eligible liabilities (IBELs) grew at more than a specified rate. Such a deposit imposed a penalty on any bank who took in too much business, and thus tended to restrain periods of rapid monetary growth. Fundamentally, it was a measure designed to counteract the banks' excessive use of 'liability management' techniques in which they meet any demand for credit by bidding for funds from the money markets rather than directly adjusting lending at a *given* level of liabilities as in the old regime; this is now discussed further below.

(c) The impact of liability management. The main reason for considering liability management at some length here is that this new technique in the financial firm does have some serious implications for the relative efficacy of monetary policy. It needs to be stressed straightaway that liability is not just an important financial innovation, but rather the term 'liability management' is used in a very broad sense to describe what Podolski (1986, p.158) calls a 'swarm' of innovations whose cumulative effect is to show a shift in financial management techniques, and it is when the diffusion of such techniques has become sufficiently widespread that the efficacy of monetary policy becomes threatened.

It will be useful at this stage to contrast 'liability management' with its exact opposite 'asset management'. Traditionally, a bank would adjust the asset-side of its balance sheet in response to a change in the liability-side of the balance sheet. So, in effect, the lending policies of the traditional banking sector would have tended to be passive in the sense that the volume of

lending could only be expanded (contracted) if, and only if, the volume of their liabilities (say, deposits) increased (decreased). However, it is the usual practice of banks nowadays to meet any demand for credit by adjusting the volume of their liabilities. Thus, the banks may respond to an increase in the demand for credit by bidding for deposits by offering, for example, attractive interest rates or enhanced withdrawal facilities. It may be noted that the scope for liability management would have been extremely limited for the major clearing banks during the 1960s owing to the restrictions imposed on price-competition by the Bank rate cartel. This is not to say that liability management would have been totally impossible, rather it would have been theoretically possible because the clearing banks would have been capable of competing for deposits by non-price means, but during the 1960s, liability management techniques were still in their infancy, having just been introduced by the more aggressive non-clearing banks as previously mentioned. Wide diffusion of liability management techniques coincided with the beginning of a more competitive banking system.

As previously argued, one of the factors responsible for the rapid growth in broad monetary aggregates was the celebrated 'round-tripping' episode. To understand more fully the developments behind this episode, it is necessary to look in some detail at the role of liability management in the explosive rates of growth in monetary aggregates, and the efforts of the authorities to come to terms with the new techniques in financial management. Basically, the importance of the demand for credit in the determination of the overall size of the banks' balance sheets arises in part because of an observed asymmetry in the flexibility of

interest rates on each side of the banks' balance sheets. This asymmetry mainly stems from the fact that, as a part of the initial CCC arrangements, the Bank Rate was determined by administrative decision so entailing a bias towards delay in its adjustment.²⁸ Now, lending by the banks was done at interest rates which were still linked in some way to the Bank rate, and in order to finance the upsurge in the demand for credit, the banks bid aggressively against each other for deposits (i.e. liability manage) in the money markets. Thus, interest rates in the money markets tended to be more flexible than lending rates, and, as the increase in the demand for credit gathered momentum, interest rates in the money markets were bid up so that there were several occasions on which the Bank rate lagged behind these rates. Thus, the differential between these rates opened up the scope for profitable arbitrage so leading to the 'round-tripping' episode.

The authorities initially attempted to respond to this episode by taking two steps. The first step, taken in October 1972, was to rename the Bank rate as the Minimum Lending Rate (MLR) which was linked to the Treasury Bill rate by a formula.²⁹ This was presumably to make the MLR more 'passive', that is, more responsive to market forces so as to reduce any bias towards delay. The second step was taken in December after the authorities considered that the liquidity position of the banks seemed to be excessive. They called in special deposits in order to mop up any surplus reserve assets so that by the end of 1973, the banks' reserve asset ratios (as a proportion of eligible liabilities) were close to their minimum of 12.5 per cent. However, this was largely circumvented by the banks through skillful use of liability management techniques.

It is important to consider carefully the definitions of 'eligible liabilities' and 'eligible reserve assets'. As previously noted, the former consisted partly of money market deposits, and the latter consisted partly of Treasury Bills. In the traditional system, the response by the banks to pressures on their reserve asset ratios would be to 'asset manage' such that they may run down their liquidity positions, leading to upward pressure on interest rates in these associated assets. Since MLR was linked to the Treasury Bill rate, this would have meant a rise in MLR. This was the mechanism envisaged by the authorities to slow down the growth in the broad monetary aggregates. However, this did not turn out as expected. Instead, the banks responded by bidding for more deposits, driving up the rates of interest in the money markets, and because the emphasis was more on liability management than asset management, there was a tendency for the Treasury Bill rate to adjust more slowly. Thus, 'round-tripping' continued to flourish.

(d) The Supplementary Special Deposits scheme. The SSD scheme was a response to the formidable problems of monetary control posed by liability management and round-tripping. Before the SSD scheme was initiated, the authorities considered possible alternatives. Amongst such alternatives, they considered the possibility of imposing interest rate ceilings like those imposed by *Regulation Q* in the USA. Such ceilings were not adopted because of the ease of substitution between various assets, and would, therefore, not be effective against liability management. Also, it was felt that such ceilings were too distortionary. In the SSD scheme, the size of the deposits required to be placed with the Bank of England varied progressively according to the excess

growth of IBELs. Until November 1974, banks had to make a deposit of 5 per cent of any excess up to 1 per cent of IBELs. A 25 per cent deposit was required on an excess of between 1 and 3 per cent while an excess over 3 per cent required a 50 per cent deposit. After November 1974, the boundaries were raised from 1 to 3 per cent and 3 to 5 per cent but the scheme was otherwise unaltered. By the imposition of such penalties, the scheme forced the banks either to accept lower profits on additional lending, or else to widen the margins they quoted to customers. The cost of placing non-interest-bearing SSDs with the bank was considerably greater than the opportunity cost of acquiring reserve assets, particularly in the second and third penalty tranches, so the financial incentive to widen margins was greatly increased. Thus, the scheme was designed to counteract any form of profitable arbitrage, and to deter the banks' aggressiveness in their competition for deposits.

The efficacy of the SSD scheme, as pointed out in the Bank's obituary notice on the scheme (1982, pp.81-83), is particularly difficult to assess because the imposition of the SSD scheme was usually announced as a part of a package of economic measures so that it is difficult to disentangle the various effects of each measure. However, one thing seems to have been made quite clear: the scheme may have been partly successful in retarding the growth of IBELs, the efficacy of the scheme was seriously undermined by circumvention and avoidance. To make an effective assessment of the efficacy of the scheme, it would be necessary to analyse in detail the changes in the structure of clearing banks' balance sheets occurring between the pre-CCC years and later years.

Unfortunately, the regular banking statistics collected at that time did not offer the kind of detailed information required for such an analysis, distinguishing only between sight deposits, CDs, and other time deposits. However, some useful data can be gleaned from a breakdown of deposits on an annual basis for the years 1971-76 provided by the London clearing banks (LCBS) as part of their evidence to the Wilson Committee. This is summarised in Table 2.7. which shows the strong contrast between the growth of the retail and wholesale elements of the LCBs' deposits during 1972 and 1973. Retail deposits increased by about £2.3 billion (21 per cent) during these two years whilst wholesale deposits, including those taken in through the branch networks, grew by about £5.8 billion (by a factor of about 13.4). These figures are no doubt swollen artificially by the episode of 'round-tripping'. Most interesting of all, after 1974, it does seem that the aggressiveness of the banks in competing for wholesale deposits has been curbed by the imposition of the SSD scheme since wholesale deposits actually declined by about £1.4 billion by the end of the period 1973-75 which coincided with a number of calls for SSDs³⁰ but seem to have risen once 'the corset' was taken off during 1976.

Although the SSD scheme may have retarded the growth of IBELs, there is thought to have been considerable disintermediation; the most important means was the 'bill leak' whereby banks accepted bills of large firms and marketed them with the same ease as certificates of deposit. This leak grew to £2,700 million in the first quarter of 1980 when IBELs were around £35,000 million. There will also have been other forms of disintermediation such as a market in trade bills which had not been accepted by any bank,

TABLE 2.7: Changes in the structure of London clearing bank deposits during the 1970s

(£ millions, end-November)

	1971	1972	1973	1974	1975	1976
A. Branch retail deposits						
(i) current accounts	6,312	6,932	7,179	7,490	8,806	9,499
(ii) 7-day deposits	4,612	4,874	6,060	8,038	8,337	8,238
(Total)	(10,924)	(11,806)	(13,239)	(15,536)	(17,143)	(17,737)
B. Wholesale deposits						
(i) branch wholesale deposits	—	901	2,790	4,041	3,976	4,890
(ii) inter-bank deposits (gross)	287	541	1,357	1,115	716	817
(iii) CD issues	156	1,230	2,112	1,523	581	862
(Total)	(434)	(2,672)	(6,259)	(6,679)	(5,273)	(6,569)
C. Gross deposits (A + B)	11,358	14,532	19,498	22,215	22,419	24,303
D. CD holdings and inter-bank claims	284	1,406	2,486	3,213	2,800	2,757
E. Net deposits (C - D)	11,074	13,126	17,012	19,002	19,619	21,549

Source: Constructed from Wilson Committee (1978), volume V, Table 4, p.119, and Bank of England Quarterly Bulletin, Table 8(2), various issues.

but these were unlikely to have been so important.³¹ The abolition of exchange controls in October 1979 allowed banks to book all excess deposits offshore; it is not clear how much was driven offshore but in any case freedom of capital movement made the SSD scheme unworkable and it was abolished in June 1980.

Thus, to sum up so far, the SSD serves as a useful illustration of the regulatory dialectic approach to financial innovation. Liability management by the regulated financial institutions was, no doubt, important in frustrating the authorities' attempts at monetary control. The SSD scheme was a response by the authorities when traditional tools of monetary policy (i.e. reserve asset ratios, and special deposits) had failed. The growth in wholesale money markets, and even the internationalisation of the British banking system via the relaxation of exchange regulations, will have served to increase the circumventive powers of the financial institutions.

(e) Implications for monetary control. The discussion on the problem of liability management will now be concluded with some comments on the possible effects of liability management on the efficacy of monetary policy. The general picture is that the advent of liability management in wholesale money markets has adversely affected the authorities' ability to control the size of the banks' balance sheets. In the *ancien regime*, where banks were subject to interest rate constraints imposed by the Bank rate cartel, the authorities could enforce a shift in *relative* interest rates by varying the *general* level of market rates. When the authorities wished to be more restrictive, it was comparatively easy to induce a relative shift in interest rates by raising the general level of market rates.

Since rates payable on retail deposits were more or less constrained, this could induce an outflow of funds from the banks into other assets whose rates had risen relative to deposit rates. The banks responded, under asset management, to such outflows by disposing of their surplus reserve assets whereas under liability management, banks could respond by compensatory rises in interest rates on wholesale deposits. Thus, under liability management, the authorities have found it much more difficult to influence relative interest rates. So, as argued by Goodhart (1984, pp.154-155), the demand for funds by the private sector is a key determinant of the response to a general interest rate movement of the whole financial sector, and with liability management, this response is increasingly sluggish.

The main effect of a shift away from asset management towards liability management is to make the demand for money less interest-elastic. Thus, put figuratively in terms of the *IS-LM* framework, the *LM* locus becomes more vertical. In such a case, pursuit of monetary control may lead to considerable interest rate volatility, and the nominal level of interest rates may, therefore, not serve well as an instrument of monetary control. It has been suggested by Goodhart (1984, pp.154-155) that, in spite of the fact that the interest-elasticity of the demand for money has declined, the demand for money has become more sensitive to changes in interest rate *differentials* so that there may be substantial shifts of funds between the various monetary aggregates at a particular general level of interest rates. Thus, one could observe a change in relationships between narrow and broad monetary aggregates. According to Goodhart (1984, pp.155 and 156), prior to liability

management, there tended to be a positive correlation between the growth rates of M1 and sterling M3, whereas under liability management, there either tended to be either no, or even, negative correlation. In particular, it was reported by Goodhart that, for annual rates of growth in the monetary aggregates, the correlation coefficient between M1 and sterling M3 during the 1960s was 0.95 whilst during the 1970s, it was -0.39 .³² Under asset management (i.e. during the 1960s), a rise in the general level of interest rates would tend to lead to a fall in both narrow and broad monetary aggregates (hence positive correlation between M1 and sterling M3), but under liability management, as broad monetary aggregates began to contain more wholesale deposits bearing market-related interest rates, there would now be a tendency for broad monetary aggregates to increase whilst narrow monetary aggregates may decrease (hence negative correlation). On a theoretical note, the greater tendency for funds to be shifted between the various monetary aggregates would lead to a more unstable demand for money which then poses problems for the authorities in terms of greater uncertainty.

The phase of further de-regulation during the late 1970s and early 1980s has tended to increase competitive pressure on the financial sector so that there may be some further scope for more aggressive liability management techniques. This would, of course, exacerbate the difficulties experienced by the authorities in controlling the monetary aggregates. A notable development worthy of special mention here is the clearing banks' entry into the mortgage market as lenders on a large scale in direct competition with the building societies. The banks were unable to enter the mortgage market earlier because of the restrictive supplementary

special deposits. The intensifying of competitive pressure brought about by the erosion of the clearing banks' retail deposit base by the building societies offering better withdrawal facilities at attractive interest rates finally forced the issue of de-regulation. Once released from all restrictions, the clearing banks were able to gain a substantial share of the mortgage market very rapidly. The building societies were often criticised for their varying mortgage queues amongst individual societies because of their cartel arrangements in setting a 'recommended rate'. Their first response was to resort to numerous if expensive ways of offering premia over the cartelised rate for ordinary shares, usually by shortening the required period and easing the notice conditions, and they were then able to eliminate the differential in rates for larger mortgages already offered by the clearing banks. By 1982, when nominal interest rates had fallen, the building societies were now able to maintain their mortgage rates at levels which the clearing banks found it difficult to match profitably.

The implication of increased competition between the banks and building societies is to make the interpretation of changes in monetary aggregates even more difficult, and Goodhart (1984, p.157) states that the overshooting of sterling M3 in 1981 was partly due to the banks' entry in the mortgage market. But, if bank deposits and building society deposits have become such close substitutes, then according to monetary liability interpretations, sterling M3 would have been distorted by the switch in deposits from building societies into banks. Such a distortion led to the adoption by the authorities in 1982 of the PSL2 definition which is a broad measure of private sector liquidity encompassing building society deposits. Thus, it is

particularly important to interpret changes in monetary aggregates with special care, and not in a mechanical manner. This aspect will be discussed further in Chapter Four on monetary aggregates.

2.3.3. Other financial innovations

As previously mentioned, liability management is indeed an important innovation, but, more precisely, it is a swarm of innovations which all reflect a change in traditional management methods of financial institutions. The switch to variable-rate lending, as an innovatory response to the onset of high and volatile interest rates, certainly falls within the purview of liability management in the sense just mentioned. However, as noted in Section 2.1.1. above, the switch to variable-rate lending may not be readily labelled as a financial innovation *per se*, but it needs to be recognised that the trend towards variable-rate lending in the last decade or so has had some important implications for the efficacy of monetary control; indeed one would do well to consider it here. A further aspect is also provided by technological change whose main effects have been to reduce transactions costs generally,

(a) *Effects of switch to variable-rate lending.* In order to understand the importance of high and volatile interest rates in promoting the switch to variable-rate lending, it is necessary to take a look at the concepts underlying interest-rate risk. One of the main functions of financial intermediaries is to hold liabilities to the account of depositors and claims on borrowers in the maturities preferred by the customers. In general, the preferred maturity of depositors and borrowers do not match, so the maturities of liabilities and assets on the balance sheets of financial intermediaries differ, i.e. financial intermediaries undertake maturity transformation.

In the case of the banking institutions, the usual preference of depositors is for the holding of short-term claims, or deposits, on the banking institutions, while borrowers have a preference for loans on more extended terms. In the UK, a large proportion of bank lending is undertaken in overdraft form, which is nominally repayable on call by the banks, but in practice is outstanding for an indefinite period.

Thus, banks, and other non-bank financial intermediaries (NBFIs), engage in maturity transformation. This involves the intermediaries in various kinds of risk, particularly interest-rate risk, if interest-rates on their assets/liabilities are fixed for the duration of the maturity. This risk arises because, with their liabilities generally on a shorter-term basis than their assets, a rise in the general level of interest rates would force them to refinance lending, undertaken earlier at lower fixed rates, on the basis of funds obtained later at higher interest, thus enforcing a running loss. With assets longer than liabilities, and both undertaken on a fixed rate basis, any unforeseen rise in nominal interest rates would bring about an unexpected loss to the banks, while any unforeseen fall in interest rates would lead to a windfall gain. So, generally, an interest-risk arises whenever there is a mismatch in the periodicity over which interest rates on assets and liabilities are fixed. This interest-rate risk is obviously greater when nominal interest rates become more volatile and unpredictable. The volatility of nominal interest rates may tend to rise along with any increase in the volatility of the inflation rate (c.f. the Fisher hypothesis in Chapter Five). Over the last decade or so, the volatility of nominal interest rates has increased, and this has had serious effects on those

institutions which have traditionally used fixed-rate lending.

Responses by financial intermediaries to such volatility in interest rates may be protective. In order to obtain protection from interest-rate risks, it is necessary to adopt some form of variable-rate lending in which lending rates are not fixed but change over time in response to changes in market conditions. There are many ways in which interest rates on longer-term loans can be varied over time in response to such changing conditions. Firstly, the rate can be varied by administrative decision, as in the case of the building societies. However, in some circumstances, interest rates that have been varied by administrative decision may not be wholly acceptable as there would almost certainly be some scope for exploitation of the borrower by lending financial intermediaries. So, in order to be more pragmatic, an alternative could be adopted in which nominal interest rates are related to the rate of inflation plus a margin to allow for a real rate of return. Whilst such a practice has been widely adopted in Latin American countries with phenomenally high rates of inflation, it has not been widely adopted here for reasons that are not yet clear. Instead, it does seem that there is a tendency for nominal interest rates to be more closely related to wholesale market rates, and such lending rates are adjusted periodically.

Such a step, therefore, has tended to reduce the interest-rate risk to lenders in general. Now, with variable-rate lending and given that there is an asymmetry in the adjustment of interest rates on loans in relation to rates on deposits, a rise in interest rates may lead to the 'endowment effect' in which profits of financial intermediaries rise because interest rates on a proportion of deposits

(especially sight deposits) remain fixed whereas lending rates rise. The endowment effect, of course, is subject to erosion over time as fixed-interest deposits are switched in favour of higher-yielding assets. From the viewpoint of borrowers, the switch to variable-rate lending can have a mixed reception. If it turns out that the cash-flow of the borrowers (say, companies) are positively correlated with changes in nominal interest rates, then borrowers stand to benefit. If, on the other hand, changes in cash-flows of borrowers are negatively correlated with changes in nominal interest rates, then it may very well turn out that the risks for borrowers have increased. Such a situation actually came about as a consequence of the worsening of economic conditions in the 1970s, viz: pressure on corporate earnings coupled with high nominal interest rates.

It may, with the benefit of hindsight, have been profitable for companies to continue issuing debentures since nominal interest rates had risen much further after the demise of the debentures market. However, in the early 1970s, the majority of opinion was that nominal interest rates were at an all-time high, and that it might be best to restrain any further issues of debentures until nominal interest rates fell back to 'normal' levels.³³ Such expectations were subsequently proved to be incorrect and, in any event, the rising cost of medium- and long-term borrowing led to the eventual demise of the debentures market.

The demise of the debentures market may have been accelerated in part by the lowering of the cost of bank intermediation. Compared with the enormous transaction costs involved in issuing debt, borrowing on a medium-term basis from

the banks was certainly cheaper and easier. Basically, the cost of bank intermediation refers to the spread between the variable rate charged on loans and the (relatively fixed) rate paid on short-term deposits. Thus, with high nominal interest rates, there was a tendency for the endowment effect to be particularly accentuated so that spreads became smaller. As noted previously, the response of lending to changes in general level of interest rates has tended to become more sluggish, and for interest rate differentials to play an increasingly important role. Thus, it follows that the cost of bank intermediation becomes an increasingly important determinant of the size of balance sheets of banks; as the cost of intermediation falls, the greater will be the volume of lending and deposits. And, this observation would almost certainly be confirmed when one considers the phenomenal increase in bank lending during the early 1970s as a consequence of 'round tripping' in response to *negative* costs of bank intermediation. The implications for the efficacy of monetary policy are similar to those discussed for liability management above.

Now that inflation and nominal interest rates have fallen significantly in the course of the past few years, an interesting question is posed which may serve to shed some more light on the definition of financial innovations discussed in section 2.1.1. above. Recall that the switch to variable-rate lending may have been an *adaptive* response, i.e. not a *creative* response, to the increased volatility of nominal interest rates. It has been suggested by Goodhart (1984, pp.163-164) that if nominal interest rates fall, this could mean that the endowment effect would be eroded, and possibly lead to a higher cost of bank intermediation, thus driving borrowers back into the capital market. It does seem to suggest

that financial innovations could be characterised by their *irreversibility*. That is, according to Schumpeter above, innovations are said to occur if and only if it leads to a once-for-all change in habits. If subsequent experience shows that there is a tendency for banks to revert to fixed-rate lending with higher spreads, and that borrowers were forced back into the capital market, then perhaps the switch to variable-rate lending cannot definitely be labelled as a financial innovation, but as an adaptive response because it lacked the irreversibility characteristic. This question would be well worth pursuing.

(b) *Effects of technological change.* Technological change in the provision of financial intermediation services is certainly one of the most obvious forms that financial innovation can take on. Although it is not too difficult to give a comprehensive catalogue of recent developments in technology applied to the financial sector, it suffices to paint a general picture, and to concentrate more on the relatively neglected question of the implications of technological change in the financial sector for the efficacy of monetary control.

Over the past few years, technological change has been particularly acute in the retail banking sphere, with the growth in ATM (Automated Teller Machine) networks. Such ATMs are now increasingly capable of providing even more basic retail banking services in addition to the basic function of cash dispensing. Such networks are likely to have a profound influence on the supply of financial intermediation services. Firstly, if any institution wishes to enter the market for the provision of financial services, there is less need to undertake highly expensive outlays on setting up a large and diverse branch network since ATMs are capable of being placed

in locations well away from branches. Even so, setting up large ATM networks can still prove to be prohibitively expensive, and there remains some scope for networks to be shared among a group of financial institutions willing to bear a part of the costs involved. Secondly, ATMs are capable of effecting transactions at a much lower unit cost per transaction so that there is a possibility that the cost of bank intermediation will decline so leading to larger balance sheets of financial institutions.

The interpretation of changes in monetary aggregates would have to be carried out with greater care because of the higher proportion of interest-bearing transactions balances in narrower definitions such non-interest-bearing M1 (nibM1) and M1. This aspect will be looked at further in Chapter Four which deals with money substitutes, and aggregation. Furthermore, the higher proportion of interest-bearing transactions balances is likely to lead to a decreased sensitivity of response to changes in general levels of interest rates, and an increased sensitivity of response to changes in relative interest rates since the lower costs associated with supplying financial information is likely to heighten awareness of possible profit opportunities. Thus, the difficulties of the monetary authorities in their conduct of monetary policy may be further exacerbated as the difficulties are similar to those of liability management discussed above. Lower transactions costs are also likely to encourage individuals to hold less non-interest-bearing balances since it is now less costly to switch funds between interest-bearing deposits and their non-interest bearing counterparts. These aspects are discussed further in the next chapter which is primarily devoted to the effects of financial innovation on the transactions demand for money.

2.4. Conclusions

The study of financial innovations, and the forces that bring it about, is hampered by many difficulties which include definitional and taxonomic problems. There is certainly no hard-and-fast classification system for financial innovations, and each financial innovation cannot be considered in isolation because most of the financial innovations are all interrelated with one another. Two hypotheses concerning financial innovation were considered. The first one concerned what may be termed 'constraint-induced' innovation in which financial institutions innovate in response to various constraints. Particular attention was paid to those constraints imposed by regulatory authorities: such innovations may be viewed as attempts to circumvent existing regulations. The second hypothesis embodying the 'regulatory dialectic' framework goes a stage further than the constraint-induced hypothesis in that it also seeks to explain the rate of diffusion of financial innovations. It was particularly noted that if regulatory authorities become overwhelmed by the pace of financial innovation, their responses may turn out to be slower, thus accelerating the diffusion of financial innovations. A hybrid theory of financial innovation was put forward which took into account the link between the real and financial sectors. The experience of the UK in financial innovations was considered. One of the most important financial innovations to be considered is the widespread use of liability management techniques which have tended to undermine the ability of the authorities in controlling broader monetary aggregates. Together with technological change and the increased volatility of nominal interest rates, the authorities have become increasingly unable to

influence relative interest rates due to liability management by financial institutions, and the response to changes in the general level of interest rates has tended to become even more sluggish. The increased sensitivity to changes in relative interest rates has tended to make the interpretation of monetary aggregates even more hazardous in that funds may be shifted between different levels of aggregates.

C H A P T E R T H R E E
THE TRANSACTIONS DEMAND FOR MONEY

Section	3.1. Alternative theories.	102
	3.1.1. <i>Deterministic inventory-theoretic model.</i>	103
	(a) <i>Zero transactions costs.</i>	104
	(b) <i>Fixed transactions costs.</i>	107
	(c) <i>Proportional transactions costs.</i>	113
	(d) <i>Transactions costs as a linear function of transactions.</i>	117
	(e) <i>Income and interest elasticities.</i>	120
	3.1.2. <i>Probabilistic inventory-theoretic model.</i>	126
	(a) <i>The basic concept.</i>	126
	(b) <i>The model.</i>	127
	(c) <i>Properties of the derived demand for money function.</i>	132
	3.1.3. <i>Effects of uncertainty in interest rates.</i>	133
	3.2. Innovations as a time-trend.	143
	3.2.1. <i>'A priori' justifications.</i>	143
	3.2.2. <i>Empirical evidence and analysis.</i>	146
	3.3. Innovations in cash-management.	152
	3.3.1. <i>Outline of theory.</i>	152
	3.3.2. <i>The Porter-Simpson model.</i>	155
	3.3.3. <i>Empirical evidence and analysis.</i>	160
	3.4. The problem of brokerage fees.	172
	3.4.1. <i>The methodology of Porter and Offenbacher.</i>	173
	(a) <i>Derivation of the debits equation.</i>	173
	(b) <i>The derivation of a brokerage fee series.</i>	175

3.4.2. <i>A critique of Porter and Offenbacher.</i>	176
3.4.3. <i>Some tentative suggestions for tackling the problem of brokerage fees.</i>	179
3.5. Conclusions.	181

CHAPTER THREE

THE TRANSACTIONS DEMAND FOR MONEY

The exploration of the transactions demand for money begins here by considering the right-hand side of the demand function in order to examine the effect that each variable has on the demand for money. There are several aspects of financial innovation that need to be borne in mind when analysing the demand for transactions balances because the various changes in variables that stem mainly from financial innovation are likely to have a pronounced effect.

Firstly, as previously seen in the last chapter, financial innovation during the past decade or so has sometimes taken the form of technological change which have the effect of lowering transactions costs. That is the main reason why relatively more emphasis is being placed upon transactions costs and their nature in the theoretical discussion of the demand for money in section 3.1. There are two basic types of inventory-theoretic models of the transactions demand for money. The first one, attributable to Baumol (1952) and Tobin (1956), is a deterministic type in which it is generally assumed that all receipts and disbursements are foreknown with perfect certainty. The second type, attributable mainly to Miller and Orr (1966) and Whalen (1966), is a probabilistic model in which an element of uncertainty concerning one's receipts and disbursements is introduced. The Baumol-Tobin model will be analysed fully in order to highlight the effects of

changes in transactions cost structures on the demand for money. That is, analyses will be carried out for the cases in which transactions costs are fixed, proportional, and linear (i.e. a combination of the last two cases). It is intended to show that the interest-elasticity is indeed sensitive to changes in transactions cost structures. In the final part of section 3.1., uncertainty about future interest rates is introduced according to the analysis carried out by Niehans (1978) which combines parts of Tobin's (1958) portfolio-approach to the demand for money with parts of the inventory-theoretic approach. One implication of the analysis is that as transactions costs approach zero, one must consider the distinction between cash balances that are demanded as an *asset* and those that are demanded as transactions balances.

There have been many attempts to capture the effects of financial innovation on the demand for money by the inclusion of 'innovation' variables as proxies in empirical studies. One view is that financial innovation was essentially a *steady* long-run process of change. The use of time trends as proxies for financial innovation in empirical studies of the demand for money (e.g. Lieberman (1977)) does reflect such a view, and this will be discussed in section 3.2. However, it will be seen that a major objection to such an approach is that consideration of recent economic history reveals that the process of financial innovation is far removed from the concept of steady long-run change, namely that one observes periods of rapid financial change (such as now) and periods in which financial change take place at a more sedate pace.

The second aspect of financial innovation to consider is the presence of high and volatile rates of inflation and interest rates



which have made the opportunity cost of holding cash balances persistently high. Some financial innovations have taken the form of improved cash-management techniques. It is shown in the first part of section 3.3. that the long-run interest-elasticity of the demand for money is likely to be higher than its short-run counterpart. Porter and Simpson (1980) have incorporated a theory of development of cash-management techniques into the Miller-Orr model and it substantiates their claim regarding interest-elasticities made above. The last two parts of section 3.3. will be devoted to a discussion of empirical studies that use different proxies for improved cash-management techniques.

Finally, problems regarding the derivation of a 'brokerage fees' time-series are considered. One particular approach was to estimate two separate equations, one for the demand for money, and one for the volume of debits, and then solve the two equations for the brokerage fee. However, it is pointed out that such a practice can lead to false conclusions in empirical studies such as that of Porter and Offenbacher (1984).

There are also other aspects of financial innovation that affect the left-hand side of the demand for money function (such as liability management), but their discussion is best postponed until the next chapter when monetary aggregates are considered.

3.1. Alternative theories

Modern theories of a transactions demand for money originated in the seminal paper of Baumol (1952), and was also developed independently of Baumol in Tobin (1956). At first sight, the approaches used by the two economists in deriving the

transactions demand for money may look different, but as will be shown later on, these two approaches are certainly equivalent. It often appears to be the case that many researchers, when entering into a theoretical discussion of the transactions demand for money, they often base their discussion on the exposition in Baumol (1952) whereas the exposition in Tobin (1956) has been relatively neglected. In order to do a little more justice to the latter paper, the discussion will be biased more on Tobin's paper than Baumol's paper. Essentially, they both utilised an inventory-theoretic approach which is based on the fact that a time lag exists between disbursements and receipts. It is assumed that there is a two-asset world in which the only two assets are money which earns no interest and bonds which do earn interest. The main objective of these inaugural inventory-theoretic models was to explain why the demand for transactions balances should be related to interest rates and transactions costs. Before proceeding further with the discussion, it will be noted that there are two basic types of inventory-theoretic models. The first type is what may be termed a deterministic model in which the pattern of disbursements and receipts are known with absolute certainty in any time period, whereas the second type, mainly due to Miller and Orr (1966), is a probabilistic model in which the pattern of disbursements and receipts are not known with perfect certainty. The latter model will be considered in sub-section 3.1.2. The following sub-section is devoted to a full discussion of the deterministic inventory-theoretic model of the transactions demand for money.

3.1.1. Deterministic inventory-theoretic model

As previously mentioned, the individual is assumed to know

the pattern of disbursements and receipts with perfect certainty in every time period. At the beginning of each period, the individual is assumed to receive his receipts which is equal to Y . Over the period, the individual will be making disbursements at a steady rate until he has run down his balances completely. This simple case will give a 'sawtooth' appearance to the assets function as shown in figure 3.1. which also shows some cash-management strategies.

(a) *Zero transactions costs.* For the sake of argument, it will be initially assumed that there are no 'transactions' costs and that the number of transactions between cash and bonds is limited to two: the first one being to purchase bonds, and the second one to sell bonds. In the first case, the individual receives his receipts at the beginning of the period, $t = 0$, but does not choose to invest the receipts in bonds immediately. Later on, at time t_1 , B_1 bonds are purchased and then sold off at time t_2 before the individual has even run down his cash balances. The shaded area shows the proportion of the individual's portfolio held in bonds. This is clearly not an optimal strategy because the individual could have earned more interest by buying bonds immediately upon receipt of his receipts. Furthermore, by selling bonds before the individual actually has run down his cash balances, he has incurred an opportunity cost in terms of lost interest that would otherwise have been earned by holding on to bonds a little longer until his cash balances need to be replenished.

Now consider the last two cash-management strategies given in figure 3.1. above. Both of these strategies do conform to the principles of good cash-management laid down in the first case, but the last strategy during the period $t_5 - t_6$ is not quite an optimal

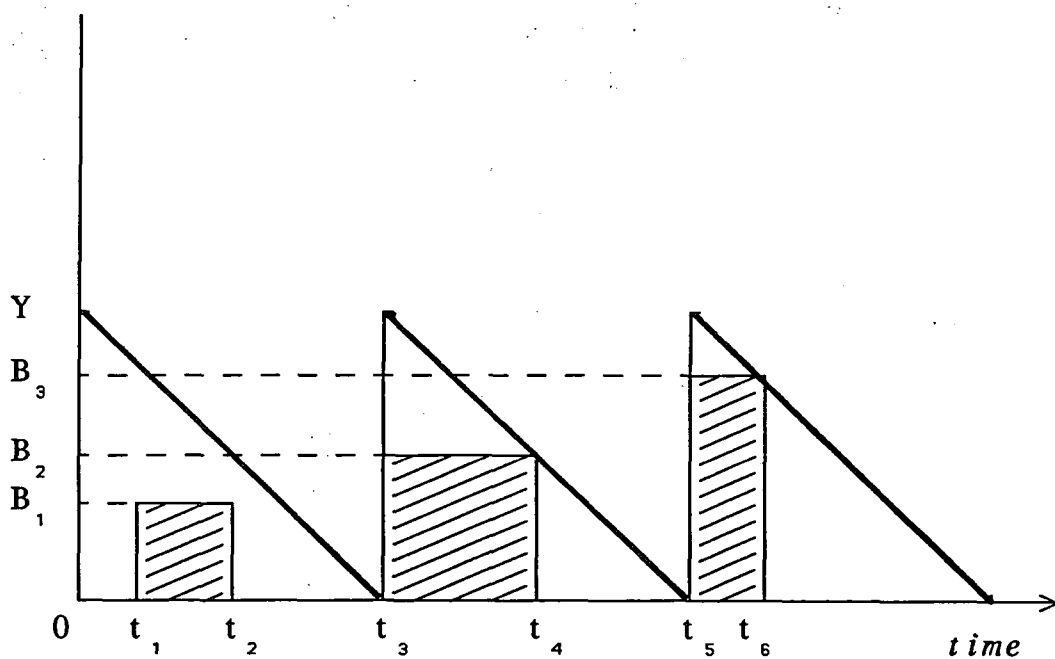


FIGURE 3.1: *The Baumol-Tobin 'sawtooth' pattern of disbursements and receipts with different cash-management strategies. The optimal strategy is during the period $t_3 - t_4$.*

one because too many bonds were bought immediately and therefore the individual's cash balance is run down sooner than necessary so that an opportunity cost is incurred in terms of lost interest by having to sell bonds and holding cash for longer, given that the number of transactions is constrained to only two per period. The second strategy during the period $t_3 - t_4$ is the optimal cash-management strategy to follow because the right amount of bonds $(Y/2)$ were purchased and then held for a period of $(t_4 - t_3)/2$ in order to maximise revenue. Given that the rate of interest on bonds is r and that the average amount of bonds held is $(Y/2)/2 = Y/4$, it follows that revenue is given by $(Y/4)r$.

The argument can be generalised to the case where n transactions are permitted. Let a time period be sub-divided into n sub-periods, $t_1, \dots, t_i, \dots, t_n$ where $t_1 = 0$ and $t_n = 1$. Thus the optimal strategy is to buy $[(n - 1)/n]Y$ bonds at time t_1 and to sell them in equal instalments of Y/n at times $t_i = (i - 1)/n$. Since the average number of bonds held is $[(n - 1)/2n]Y$, it follows that revenue is given by $[(n - 1)/2n]Yr$. From the above reasoning which is due to Tobin (1956, pp.243-244), it can be argued that the average cash balances, denoted by \bar{M} , is given by

$$\bar{M} = \frac{1 - [(n - 1)/n]}{2} \cdot Y = \frac{Y}{2n} \dots [1]$$

It is clear that as the number of transactions allowed becomes infinitely large, it follows that \bar{M} will approach zero and revenue will approach $\frac{1}{2}Yr$, that is

$$\lim_{n \rightarrow \infty} \bar{M} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} R = \frac{1}{2}Yr$$

where R stands for revenue. It should be noted that in the *absolute* absence of transactions costs, the long-run demand for money is totally interest-inelastic because, at any *given* number of transactions, the same cash-management strategy would still be followed, leading to lower revenues though. However, if individuals were allowed to choose their optimum number of transactions, then zero transactions costs would dictate that an infinite number of transactions would take place such that hardly any cash was demanded. The implication of transactions costs approaching zero for the transactions demand for money as a consequence of financial innovation will be discussed later on.

(b) *Fixed transactions costs.* As suggested by Baumol (1952, p.545 and p.546, footnote 5), fixed transactions costs may take the form of a 'brokerage fee' which is payable every time a transaction involving cash and bonds is undertaken. Such a brokerage fee may reflect both objective and subjective costs involved in carrying out a transaction. Typical examples may include the price of a telephone call to a broker to execute an order and the time spent at the bank cashing cheques from the broker. Of course, in the presence of fixed transactions costs, an individual would have to modify his cash-management strategy since an infinite number of transactions would incur disproportionately large costs which may well lead to a net loss. According to the analysis carried out by Baumol (1952), total costs of a particular cash-management strategy are composed of two main elements. The first element is, of

course, the fixed cost of a number of transactions. The second element concerns the opportunity cost of having to hold some cash balances which do not earn any interest at all.

According to Niehans (1978, p.29), any model that exhibits a perpetual regular pattern of disbursements and receipts is called an infinite stationary motion model. It just happens that the Baumol-Tobin 'sawtooth' type model falls into the category of infinite stationary motion models. Thus, for the sake of argument again, given that there are no transactions costs involved and that the individual is allowed an infinite number of transactions, it was previously shown that the maximised revenue would be $\frac{1}{2}Yr$ which would be the same for each period throughout time. Given that transactions costs now exist, it is only sufficient that the individual minimises total costs in order to arrive at an optimal cash-management strategy. This is so because there is really no need to work out the the maximum revenue in each period if it is known that it will be the same for each period. However, one must be warned against generalising too soon from simple cases. If the assumption of infinite stationary motion was to be relaxed, it would follow that maximum revenue would *not* be the same for each period throughout time so that cost minimisation is a necessary but no longer a sufficient condition for an optimal cash-management strategy. In fact, as a little reflection will show, it is now necessary to maximise *net revenue* which should be a necessary and sufficient condition for an optimal strategy. This is presumably the main reason why Tobin (1956, p.247) was critical of Baumol's cost minimisation assumption. So, to avoid any loss of generality here, maximisation of net revenue will be used to arrive at any optimal

cash-management strategy.

Following a reasoning similar to that of Baumol, but embodying Tobin's concept of an optimal number of transactions, it can be shown that net revenue is given by gross revenue less the opportunity cost of holding cash balances and any fixed costs. As previously noted, gross revenue is given by $\frac{1}{2}Yr$, and since the average cash balance held is $Y/2n$, the opportunity cost is $(Y/2n)r$. Given that the fixed brokerage fee is b , it follows that total fixed costs for n transactions equals nb . Hence net revenue for n transactions, denoted by Π_n , is given by

$$\Pi_n = \frac{1}{2}Yr[1 - (1/n)] - nb \quad \dots [2]$$

It is now required to find the optimal number of transactions, n^* , such that net revenue is maximised. This may be found by taking the partial derivative of Π_n with respect to n and setting the resulting expression equal to zero:

$$\frac{\partial \Pi_n}{\partial n} = \frac{1}{2}Yr(1/n^2) - b = 0$$

which may then be solved for n to give

$$n^* = \sqrt{\frac{Yr}{2b}} \quad \dots [3]$$

According to Baumol (1952), the number of transactions is given by Y/M which equals n . Therefore, the optimal demand for money is

given by substituting for n^* from equation [3]:

$$M_f^* = \frac{Y}{n^*} = \sqrt{\frac{2bY}{r}} \quad \dots [4]$$

which is the widely-quoted 'square root' formula given in Baumol (1952, p.547). The optimal cash-management strategy in the case of fixed transactions costs is shown in figure 3.2(a) below. Here, the individual is seen to be holding M_f^* cash balances for an interval of $1/n^*$ which are then run down before selling bonds to replenish his cash balances which are then held for a further interval of $1/n^*$. The process continues until all assets have been used up in the process of transactions. It is interesting to note that the imposition of fixed transactions costs has a similar effect to that of imposing a given number of transactions in the absence of transactions costs which has already been analysed in sub-section 3.1.1(a).

Consider the properties of the demand for money function just derived above. It can be verified that

$$\frac{\partial M_f^*}{\partial r} < 0, \quad \frac{\partial M_f^*}{\partial Y} > 0, \quad \text{and that} \quad \frac{\partial M_f^*}{\partial b} > 0$$

This leads to the implication that if the interest rate increases, the demand for transactions balances decreases. Furthermore, this would be reinforced by a secular decline in brokerage fees brought about by innovations in cash-management techniques. Also, if income or the volume of transactions increases, the demand for money will increase. Since average cash balances are $M^d = M_f^*/2$, it follows

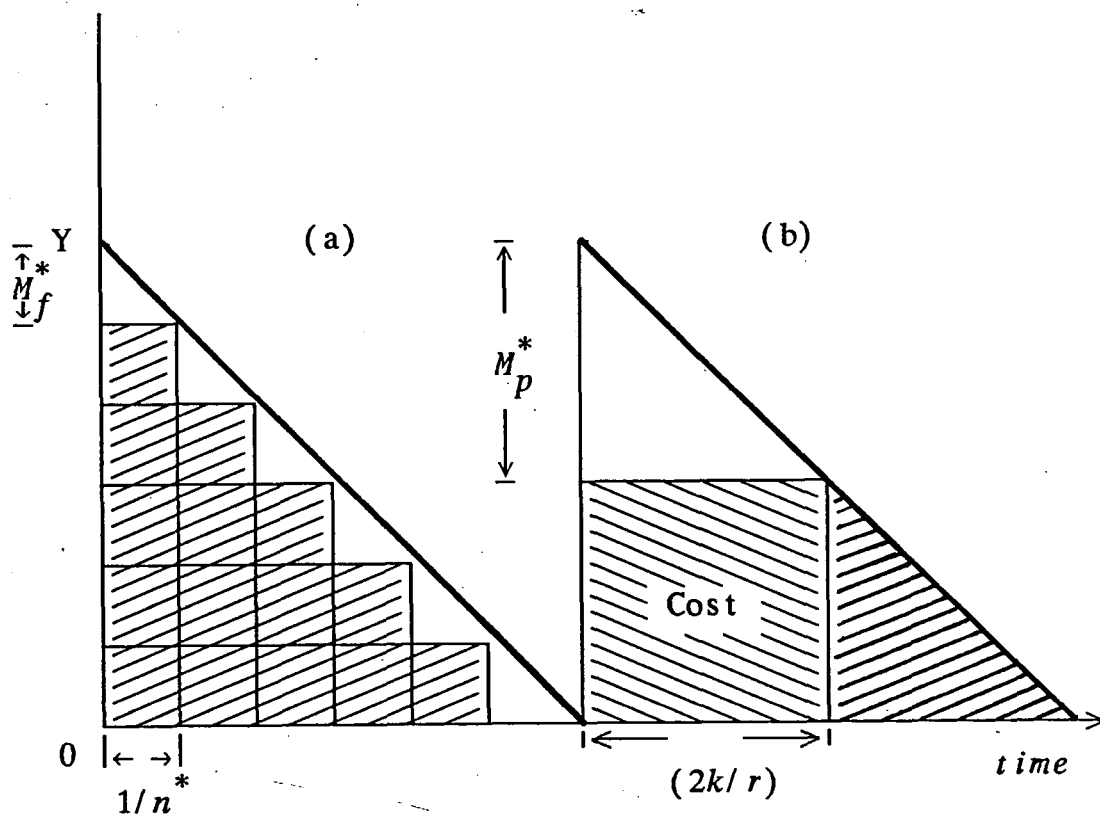


FIGURE 3.2; *The Baumol-Tobin model of the transactions demand for money with fixed and proportional costs.*

that the long-run demand for transactions balances is

$$M_f^d = \sqrt{\frac{bY}{2r}} \quad \dots [5]$$

The demand for real transactions balances (i.e. nominal balances deflated by the price level) is

$$\frac{M_f^d}{P} = \alpha Y^{0.5} r^{-0.5}$$

where $\alpha = \frac{1}{2}(2b)^{\frac{1}{2}}$ which is a 'constant'. The term 'constant' is in inverted commas because transactions can vary over time, leading to shifts in the demand for money function as most empirical studies in the 1970s showed. Taking natural logarithms of the preceding expression, the following equation is the theoretical logarithmic demand for money function:

$$\ln M_f^d = \ln \alpha + 0.5 \ln Y - 0.5 \ln r + \ln P \quad \dots [6]$$

From the above equation, it will be seen that there are economies of scale in holding money balances. This stems from the fact that given a percentage change in income or transactions, the percentage change in money balances will only be half of that for income or transactions. Thus the income-elasticity is equal to +0.5. Furthermore, given a change in interest rates, the demand for money will change less than proportionately in response, implying an interest-elasticity of -0.5. With respect to the price level, there is unit elasticity such that there will be an equiproportionate change in

the demand for money in response to changes in the price level.

(c) *Proportional transactions costs*. What can be said about the demand for money when there are proportional transactions costs instead of fixed costs? In such a case, transactions costs are directly proportional to the value of the transaction involved. First of all, consider, as Niehans (1978, pp.43-47) does, a general situation in which there are some observed periods in which the individual accumulates his assets and other period in which they are decumulated. It is also important to bear in mind that the assumption of perfect foresight still exists so that the individual knows what the pattern of transactions will be. The individual's problem is to decide on how long for should he accumulate his assets in the form of bonds from the beginning of the period, and then accumulate his assets in the form of cash balances. This problem may be tackled by employing the concepts of marginal revenue and marginal costs. When the individual buys and sells bonds, he has to incur the cost of a 'round trip' to the bond market which in effect means that two transactions costs are incurred. Letting k denote the proportional transaction cost, it should be clear that the marginal cost of holding a bond is $2k$.¹ Against such costs, the individual will also take into account the marginal revenue from holding a bond. Letting τ represent the marginal holding period for a bond, it can be shown that marginal revenue is τr . It will be profitable for the individual to accumulate his assets in the form of bonds as long as marginal revenue exceeds marginal cost, that is when $\tau r > 2k$. Given that k and r are constant, further purchases of bonds as time passes by will lead to a lower marginal revenue. The point will come when $\tau r = 2k$ in

which case the individual should stop accumulating his assets in the form of bonds, and accumulate them in the form of cash balances instead as long as $\tau r < 2k$. The individual may continue to accumulate cash balances until he enters a phase in which he has to decumulate his cash balances. Of course, he can decumulate them until they are run down to the point of exhaustion and then continue to finance his transactions by continuous sales of bonds until the next receipt arrives.

Returning to the Baumol-Tobin type pattern of transactions, the individual does not accumulate his receipts over time, but is assumed to have his receipts in one lump at the beginning of each period. In order to decide upon an optimal cash balance at the beginning of the period, the individual will be evaluating the marginal revenues and marginal costs involved by holding bonds. At the point where $\tau r = 2k$, the individual will then know how many bonds to purchase and therefore how much of his receipts should be withheld in the form of cash balances. These cash balances will then be decumulated until they become exhausted, and then the individual can finance the rest of his transactions by continuous sales of bonds. Such a situation is depicted in figure 3.2(b). It should be noted that there is a rectangle marked 'cost' in the figure: it does not really represent the costs of holding bonds, but rather indicates the period in which revenues from holding bonds are more than offset by the transactions costs.

It was shown previously that average cash balances are equivalent to $Y/2n$ where n is the number of transactions. It can be shown that the holding period for cash balances is given by $Y/(Y/n)$ in which case $\tau = 1/n$. Thus, revenue may be defined

as the revenue that would have been obtained if all receipts were immediately ploughed into bonds in the absence of transactions costs less the opportunity cost or loss of revenue caused by the holding of average cash balances which is equal to $\frac{1}{2}(Y/n)\tau r$ or $\frac{1}{2}Yr(1/n^2)$. Since the amount of bonds held is equal to $Y[1 - (1/n)]$, it follows that total transactions costs are $2kY[1 - (1/n)]$. Thus net revenue is given by

$$\Pi_n = \frac{1}{2}Yr[1 - (1/n^2)] - 2kY[1 - (1/n)] \quad \dots [7]$$

The optimal number of transactions and, therefore, the optimal holding period can be obtained by taking the partial derivative of the preceding expression with respect to n and setting it equal to zero:

$$\frac{\partial \Pi_n}{\partial n} = \frac{1}{2}Yr(2/n^3) - 2kY(1/n^2) = 0$$

which is then solved for n to give

$$n^* = r/2k \quad \dots [8]$$

The optimal marginal holding period is therefore $\tau^* = (2k/r)$ and is shown in figure 3.2(b). Since it is known that $n = Y/M$, it follows that the demand for money in the presence of proportional transactions costs is

$$M_p^* = \left[\frac{2k}{r} \right] Y \quad (c.f. \dots [9] \quad [4])$$

where it is required that $2k/r \ll 1$. The reason for such a requirement is that if brokerage costs exceed revenues from holding bonds, it will simply not pay to hold bonds at all. Therefore, all receipts will be held in the form of cash balances which should not be greater than Y . As the rate of interest increases or transactions costs decline, there will be a reduction in receipts that are withheld as cash balances at the beginning of the period. Furthermore, other things being equal, the amount of receipts withheld as cash balances varies in direct proportion to income or the volume of transactions.

Niehans (1978, p.46) has defined the long-run demand for money as $M_p^d = \frac{1}{2} M_p^* \tau^*$ where $\tau^* = (2k/r)$. Thus the long-run demand for money in the case of proportional costs is given by

$$M_p^d = \frac{1}{2} \left[\frac{2k}{r} \right]^2 Y \quad \dots [10]$$

The properties of the above demand for money function may be noted as follows. First, it is proportional to income, and there are no economies of scale in holding cash balances. Furthermore, it varies with the square of transactions costs, and is inversely related to the square of the interest rate whose interest-elasticity is much greater at -2.0 . Such a function is only valid if $r \gg 2k$. This is so because if $r < 2k$ then this would imply that the demand for transactions balances is greater than the total assets available for holding as cash. So if the model is to hold at all, it is vital that the total revenue from interest earnings in any *one* period must be equal to or greater than the transactions costs of a round trip to the bond market.

So far, it has been possible to make statements regarding the various elasticities for the demand for money. On the one hand, it has been shown that, in the case of fixed transactions costs, the interest-elasticity is likely to be -0.5 and, in the case of proportional transactions costs, it is likely to be -2.0 . It was also observed that the income-elasticity is likely to be between $+0.5$ and $+1.0$ in the respective cases. It would seem intuitively plausible that in the presence of linear transactions costs, both elasticities would be somewhere between their two extreme values. This aspect is now the subject of the following paragraphs.

(d) Transactions costs as a linear function of transactions.

Consider the case in which transactions costs are both fixed and proportional, namely that transactions costs are a linear function of the volume of transactions involved, say, $C = b + kY$ where C denotes total transactions costs.

Before proceeding any further with the analysis, it is particularly important to appreciate the notation used for the demand for money so far. It can be seen that the use of subscripts f and p is designed to distinguish between the components of the demand for money that are due to fixed and proportional transactions costs respectively. If no subscript is given, it will be understood that this refers to the demand for money under both fixed and proportional costs.

According to the analysis of Baumol (1952, pp.547-549) and using a similar methodology to that used by Tobin (1956), it can be shown from figure 3.3. that, out of an available asset balance of Y , Y_τ will be withheld from investment at the beginning of the period due to proportional transactions costs where, as before,

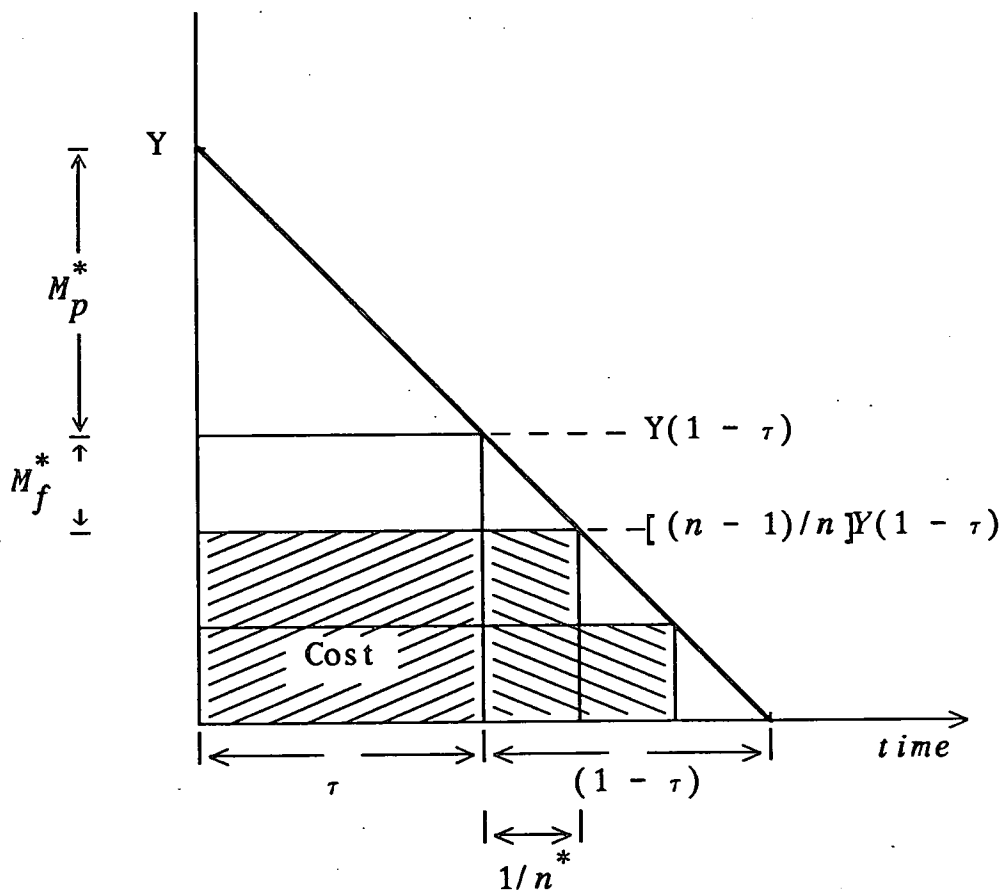


FIGURE 3.3: *The Baumol-Tobin model of the transactions demand for money with transactions costs as a linear function of the volume of transactions.*

$\tau = (2k/r)$. This now leaves an asset balance of $Y(1 - \tau)$ which is available for distribution among bonds and cash. According to Tobin (1956), the optimal strategy would be to allocate $(1/n)Y(1 - \tau)$ as cash and invest the rest into bonds. It has been shown by Tobin (p.245-247) that net revenue in this case is given by

$$\Pi_n = [(n - 1)/n]Yr(1 - \tau)^2 - nb \quad \dots [11]$$

which can be maximised by taking the partial derivative with respect to n and setting it equal to zero:

$$\frac{\partial \Pi_n}{\partial n} = \frac{1}{2n^2} Yr (1 - \tau)^2 - b = 0$$

so that the above equation can be solved to give the optimal number of transactions for the rest of the period $(1 - \tau)$ as follows

$$n^* = \sqrt{\frac{Yr}{2b}} (1 - \tau) \quad \dots [12]$$

It may be recalled that $n = Y/M$ so that

$$(1 - \tau) M_f^* = \sqrt{\frac{2bY}{r}} \quad \dots [13]$$

which may be seen as the demand for money that has been 'scaled down' for a shorter period from $t = \tau$ to the end of the period. It can be shown that the long-run demand for money is given by a

weighted average of the cash balances at times $t = 0$, $t = \tau$, and $t = 1$. Between the beginning of the period and time $t = \tau$, the average cash balance held is $(M_p^* + M_f^*)/2$ which is held for a period of τ ; between time $t = \tau$ and the end of the period, the average cash balance held is only $(M_f^*)/2$ which is held for a period of $(1 - \tau)$. Furthermore, the weights to be allocated are $(M_p^* + M_f^*)/Y$ and $[Y - (M_p^* + M_f^*)]/Y$ respectively so that the long-run demand for money is:

$$M^d = \frac{[M_p^* + M_f^*]^2 + [M_f^*][Y - (M_p^* + M_f^*)]}{2Y}$$

whence

$$M^d = \sqrt{\frac{bY}{2r}} \left[1 + \left[\frac{2k}{r} \right] \right] + \frac{1}{2} \left[\frac{2k}{r} \right]^2 Y \quad \dots [14]$$

which may be written more succinctly as $M^d = M_f^d(1 + \tau) + M_p^d$ after making the necessary substitutions from equations [4] and [10] respectively. This result is the same as shown by Brunner and Meltzer (1967, p.426, equation 4).

(e) *Income and interest elasticities.* It will be of considerable interest to examine the income and interest elasticities for the demand for money under both fixed and proportional costs, and it may even turn out to be possible that some tentative comments could be made regarding the effect of changes in the structure of transactions costs on interest-elasticities.

First consider the interest-elasticity of the demand for money

where transactions costs are assumed to be a linear function of the volume of transactions. Taking the partial derivative of equation [14] with respect to r gives

$$\frac{\partial M^d}{\partial r} = -\frac{M_f^d}{2r} (1 + \tau) + M_f^d \left[-\frac{2k}{r^2} \right] + Y \left[\frac{2k}{r} \right] \left[-\frac{2k}{r^2} \right]$$

and multiplication of the preceding expression by (r/M^d) gives the interest-elasticity of the demand for money

$$\eta_r = -2 \frac{M_p^d}{M^d} - \frac{1}{2} \frac{M_f^d}{M^d} (1 + 3\tau) \quad \dots [15]$$

It is particularly interesting to note that in the absence of proportional transactions costs, the interest-elasticity is equal to $-\frac{1}{2}$ (remember that τ contains the proportional cost term k in which case $\tau = 0$), and in the absence of fixed costs, is -2 . Under proportional and fixed transactions costs that are a linear function of the volume of transactions, the value of the interest-elasticity is a little more complicated to determine because it all depends on the ratio of marginal proportional costs to the interest rate. It will be recalled that such a ratio can only take on certain values, namely that $0 < \tau < 1$. The value of τ cannot be zero under linear transactions costs because, if it were the case, it would imply the absence of proportional costs. The value of τ can exceed unity but if that were the case, there would be no demand for money at all since it would be utterly unprofitable to invest surplus cash balances in bonds. By taking on different values for τ , it would be possible to make some tentative comments on the likely value of the

interest-elasticity. First of all, consider a value of τ equal to unity. In such a case, equation [14] reduces to $-2 (M_b^d + M_f^d)/M^d$. It is clear that $(M_b^d + M_f^d)/M^d$ must equal unity so that the interest-elasticity in the case when marginal proportional costs are equal to the interest rate is equal to -2.0 . Next, consider a value of τ equal to $1/3$ in which case equation [14] reduces to $-2(M_b^d/M^d) - (M_f^d/M^d)$. In such a case where the 'one-way' marginal cost (k) is one-sixth of the interest rate, the interest-elasticity will take on a value somewhere between -2.0 and -0.5 , which is dependent on the proportions of M_b^d and M_f^d to M^d .

To summarise so far, an interest-elasticity of -0.5 may occur in the case where there are only fixed costs, and an elasticity of -2.0 may occur in *two* cases: either when there are only proportional costs or when there are linear costs but the ratio of the marginal proportional cost to the interest rate is unity. In other cases, when there are linear costs, and the ratio of the marginal proportional cost is between zero and unity, an interest-elasticity of somewhere in the range of -2.0 and -0.5 is obtainable. Apart from the emphasis on the value of τ here, such observations are consistent with those made by Niehans (1978, p.51).

It would be most instructive if the behaviour of the interest-elasticity could be analysed as a response to changes in the structure of transactions costs. Such an exercise is attempted here in the hope that some light may be shed on the underlying causes of shifts in interest-elasticities that have been a feature of empirical demand for money studies since the 1970s. A change in fixed transactions costs will be analysed first as it is the simplest case. Taking the partial derivative of equation [15] above with respect to b gives an

expression of the form

$$\frac{\partial \eta_r}{\partial b} = \frac{-\frac{1}{2}(M^d) \left[\frac{\partial M_f^d}{\partial b} \right] (1+3\tau) - \left[-\frac{1}{2} M_f^d (1+3\tau) - 2 M_p^d \right] \left[\frac{\partial M^d}{\partial b} \right]}{(M^d)^2}$$

which is readily simplified to

$$\frac{\partial \eta_r}{\partial b} = \frac{-\frac{1}{2}(Y/2br)^{\frac{1}{2}} \left[\frac{1}{2}(1+3\tau) + (\eta_r)(1+\tau) \right]}{M^d} \quad \dots [16]$$

It is clear that in the absence of proportional transactions costs, $\tau = 0$ and that $\eta_r = -\frac{1}{2}$ so that $\partial \eta_r / \partial b = 0$. Thus a change in fixed transactions costs in the absence of proportional costs will not change the interest-elasticity of the demand for money. Now consider the introduction of proportional transactions costs. The picture is a little more difficult to analyse since the sign of $\partial \eta_r / \partial b$ depends on the ratio of marginal proportional transactions costs to the interest rate, τ . It was previously argued that when $\tau = 1$, $\eta_r = -2$, so that the term within the square brackets in the numerator of equation [16] evaluates to -2 which makes the sign of $\partial \eta_r / \partial b$ unambiguously positive. Thus a fall in fixed transactions costs would increase the interest-elasticity of the demand for money (remember that η_r is negative so a fall in b will lead to a decrease in η_r or an increase in its *absolute* value). This, of course, leads to the happy conclusion that falls in fixed transactions costs, as a consequence of financial innovation, would increase the interest-elasticity of the demand for money which is intuitively plausible.

Considering values of τ between zero and unity, assume that

$\tau = 1/3$ so that the sign of the expression in square brackets in the numerator of equation [16] above will depend on the value of the interest-elasticity. For values of η_r between -0.5 and -1.5 , $\partial \eta_r / \partial b$ will be negative whereas for values less than (greater in absolute value) -1.5 , $\partial \eta_r / \partial b$ will be positive. At an empirical level, one would need detailed knowledge of the structure of transactions costs to be able to make predictions regarding the effect of changes in fixed transactions costs on interest-elasticities. It is not really possible to give general *a priori* predictions.

Changes in proportional costs are now analysed by taking the partial derivative of equation [15] with respect to k which gives

$$\frac{\partial \eta_r}{\partial k} = \frac{-2(M^d)(\partial M_p^d / \partial k) - \frac{3}{2}(M^d)(M_f^d)(2/r) - [-2(M_p^d) - \frac{1}{2}(M_f^d)(1 + 3\tau)](\partial M^d / \partial k)}{(M^d)^2}$$

which after simplification gives

$$\frac{\partial \eta_r}{\partial k} = \frac{-(2 + \eta_r)[(M_p^d)(2/k)] - (\frac{3}{2} + 3\eta_r)[(M_f^d)(2/r)]}{M^d} \dots [17]$$

Here, when fixed transactions costs are absent, it is clear that $M_f^d = 0$ and that $\eta_r = -2$ so that $\partial \eta_r / \partial k = 0$. When there are both fixed and proportional costs and that the ratio of marginal proportional costs to the interest rate is unity (i.e. $\tau = 1$), the interest-elasticity will be -2 so that the first term in the numerator of equation [17] is zero and the second term will be

unambiguously positive. Given such an outcome, it is clear that the sign of $\partial \eta_r / \partial k$ is positive, implying that a fall in proportional costs will lead to an increase in the absolute value of the interest-elasticity of the demand for money. For values of τ between zero and unity, it is not really possible to determine the sign of $\partial \eta_r / \partial k$, but at the *very least*, the interest-elasticity is likely to vary in response to changes in proportional costs.

Thus, one possible interpretation of shifts in interest-elasticities of the demand for money is that there was a underlying shift in the structure of transactions costs brought about by financial innovation.

With regard to the income-elasticity of the demand for money, this can be found by taking the partial derivative of equation [14] above with respect to Y to give

$$\frac{\partial M^d}{\partial Y} = \frac{1}{2}(b/2rY)^{\frac{1}{2}}(1 + \tau) + \frac{1}{2}\tau^2$$

Multiplication throughout by (Y/M^d) gives the income-elasticity of the demand for money under transactions costs that are a linear function of the volume of transactions.

$$\eta_Y = \frac{1}{2}(M_f^d/M^d)(1 + \tau) + (M_p^d/M^d) \quad \dots [18]$$

In the two extreme cases where fixed or proportional costs are only

absent, the income elasticities are equal to +0.5 and +1.0 respectively. Between these two extreme cases, the income-elasticity is likely to be somewhere between +0.5 and +1.0.

How does this analysis compare with the analysis of Brunner and Meltzer (1967, pp.426-427)? The income- or transactions-elasticity derived in equation [18] above is essentially the same as that derived by Brunner and Meltzer (equation 5).² For a *given* transactions costs structure, Brunner and Meltzer argue that as the volume of transactions approach infinity, η_Y will approach unity which substantiates their argument that the demand for money function in equation [14] above is not inconsistent with the quantity theory of money which postulates that there are no economies of scale in cash holdings. This is only valid if the relative importance of fixed transactions costs is comparatively minor.

3.1.2. Probabilistic inventory-theoretic model

(a) *The basic concept.* By relaxing the assumption that the pattern of disbursements and receipts are perfectly foreseen, one has a probabilistic model of the transactions demand for money in which the pattern of disbursements and receipts are subject to uncertainty. It has been argued by Miller and Orr (1966, p.415) that deterministic inventory-theoretic models of the transactions demand for money applies well in the case of households who earn a salary on a regular basis, but is inherently unsatisfactory in the case of professional households and business firms whose cash flows exhibit random behaviour. According to Orr (1970, pp.54-55), the basic idea behind a probabilistic inventory-theoretic model is that in large business firms, the cash balance fluctuates irregularly and unpredictably over time in both directions. An accumulation in

cash balances occurs when receipts exceed disbursements and a decumulation occurs when the reverse is true. When the accumulation in cash balances becomes particularly prolonged, there will come a point when the firm decides that its present level of cash balances is excessive, and chooses to invest a sizeable chunk of the cash balance in interest-bearing assets, or to facilitate loan retirement. Conversely, when a decumulation in cash balances becomes prolonged, there will come a point when the firm will decide that its present level of cash balances are inadequate, and will therefore choose to replenish its cash balances to some acceptable level by liquidating some of its interest-bearing assets or further borrowing.

(b) *The model.* Some of the more trivial assumptions of the Miller-Orr model (1966, pp.417-419) are quite similar to those of the Baumol-Tobin model in that a 'two-asset' world is assumed in which non-interest bearing cash balances and interest-bearing assets form the two main assets. Furthermore, there also exists a fixed brokerage fee, similar in its concept to that employed in Baumol (1952), which is levied on each transaction taking place. It is also assumed that there are negligible delays in the effecting of a transaction, namely that an exchange of interest-bearing assets for cash balances or *vice-versa* takes place simultaneously.

Further assumptions were also made in which the Miller-Orr model becomes substantially different from the Baumol-Tobin model. Regarding the nature of the banking system, it is assumed that overdrafts of any kind are strictly prohibited so that the minimum balance is virtually zero. The next assumption, which is the most important one of the model, is that cash flows are stochastic, and

that during a working day, the cash balance is assumed to change t times by either $+Y$ or $-Y$. The variations in cash balances follow a Bernoulli-type process in which there are two possible outcomes denoted either as a success or failure with probabilities p and $1 - p = q$ respectively. A success is said to occur if the cash balance increases by Y or a failure occurs when it decreases by Y . Miller and Orr (1966, p.419) specifically consider the simplest case of 'zero-drift' in which the probabilities of a success or a failure are equal, that is $p = q = \frac{1}{2}$ so that the distribution of changes in cash balances will have zero mean and a finite variance equal to $\sigma^2 = Y^2 t$. Finally, it is assumed that the firm seeks to minimise its steady-state costs of managing its cash balances. Miller and Orr assume that a firm sets itself upper and lower bounds of cash balances in which cash balances can wander freely within the two limits. However, if the cash balance reaches the upper limit of hY , the firm will initiate a transfer of cash into interest-bearing assets so that the cash balance is reduced by $(h - z)Y$ to $M = zY$. Furthermore, if the cash balance reaches the lower limit of zero, some interest-bearing assets will be liquidated in order to restore the cash balance to M . It is particularly important to stress that, in order to simplify their notation, Miller and Orr (1966, p.422) make the distinction between h and $h' = hY$, and $z = z' = zY = M$. The definition of h' and z' is that these level of cash balances are denominated in *single* currency units whereas h and z are normalised variables such that they denote the level of cash balances denominated in Y currency units. Given the firm's policy of cash-management, the firm aims to minimise its expected average costs. The composition of costs is quite similar to

that of Baumol (1952), namely that there are transactions and opportunity costs. Given that the brokerage fee is set at b and that the probability of a transaction (in either direction) occurring during a day is $P(T)$, it follows that the expected average transactions costs will amount to $bP(T)$. Furthermore, the expected average cash balance is denoted by $E(M)$. Given that the rate of interest on interest-bearing assets is r , the opportunity cost is expected to be $rE(M)$ so that the expected total average cost of the firm's daily cash-management strategy is given by

$$E(c) = bP(T) + rE(M) \quad \dots [19]$$

where $E(c)$ denotes expected average cost as defined in Orr (1970, p.58).

The firm's optimisation problem is to find optimal values of h and z that will serve to minimise expected average costs. Before equation [19] can be optimised to arrive at minimum expected average costs, it is necessary to derive expressions for $P(T)$ and $E(M)$ in terms of h and z . It has been shown by Orr (1970, pp.58-61, equations 9 and 10) that

$$E(M) = (h + z)/3 \quad \dots [20]$$

and that

$$P(T) = t/[z(h - z)] \quad \dots [21]$$

which may be substituted into equation [19] so that

$$E(c) = \frac{bt}{z(h-z)} + \frac{rY[(h-z) + 2z]}{3} \dots [22]$$

remembering that the second term has to be multiplied by Y to reflect the full opportunity cost of holding cash balances. The preceding expression can now be partially differentiated with respect to $(h-z)$ and z , and the resulting expressions set equal to zero as a necessary condition for a minimum in which case

$$\frac{\partial E(c)}{\partial (h-z)} = \frac{-bt}{(h-z)^2 z} + \frac{rY}{3} = 0$$

and

$$\frac{\partial E(c)}{\partial z} = \frac{-bt}{z^2(h-z)} + \frac{2rY}{3} = 0$$

The above two equations can be solved to give the optimal values of h and z as follows

$$z^* = \sqrt[3]{\frac{3bt}{4rY}}$$

and

$$h^* = 3z^*$$

The above two expressions need to be multiplied by Y in order to convert the unit of measurement from Y currency units into *single* currency units:

$$M^* = \sqrt[3]{\frac{3bY^2t}{4r}} \quad \dots [23a]$$

and

$$h^*Y = 3M^* \quad \dots [24]$$

However, from the previous discussion, it can be recalled that the variance of cash balances is $\sigma^2 = Y^2t$ so that equation [23a] becomes

$$M^* = \sqrt[3]{\frac{3b\sigma^2}{4r}} \quad \dots [23b]$$

which is the same as that one derived by Orr (1970, p.62, equation 16). It has been argued by Orr (1970, p.64) that the long-run average demand for transactions balances is given by $(h^*Y + M^*)/3$ so that substitution for h^*Y and M^* from equations [24] and [23b] respectively gives an expression for the long-run demand for transactions balances:

$$M^d = \frac{4}{3} \sqrt[3]{\frac{3b\sigma^2}{4r}} \quad \dots [25]$$

The properties of this demand for money function will now be

discussed.

(c) *Properties of the derived demand for money function.* It is clear that the demand for money function derived above in equation [23b] is an increasing function of transactions costs and a decreasing function of the interest rate. Furthermore, it also exhibits the property that it is an increasing function of the variance in cash balances. Miller and Orr (1966, p.425) have suggested that such a term reflects the degree in which there is a lack of synchronisation between disbursements and receipts. So, if there was an increase in uncertainty regarding the future expected pattern of disbursements and receipts (as reflected in a higher variance of cash balances), the demand for money would increase.

It is easily seen that the interest-elasticity of the demand for money is slightly smaller at $-1/3$, compared with $-1/2$ for the Baumol-Tobin model. One possible reason for this difference is the existence of uncertainty in the pattern of disbursements and receipts so that if a firm unexpectedly found itself short of cash, it would have to go through the process of initiating a transfer of funds between interest-bearing assets and cash which would cost something more than just the brokerage fee *per se*. Such extra costs may be subjective in that being caught out with a lack of cash is likely to cause financial embarrassment for the firm. Thus, there is always a greater incentive to hold on to larger cash balances than in the case of perfect certainty. Given, for example, a rise in interest rates, the firm will be particularly careful not to transfer excessive cash balances into interest-bearing assets simply because the opportunity cost of holding cash balances has risen. Thus, the firm's response to changes in interest rates will tend to be more sluggish in the

case of uncertainty so that the interest-elasticity would tend to be lower.

The presence of a term for the variance of cash balances naturally raises the question of how equation [25] is to be interpreted whenever discussing the question of the income- or transactions-elasticity of the demand for money. Orr (1970, pp.64-65) offers two possible extreme cases in which the rate of transactions may change. First of all, it needs to be recalled that the variance in cash balances is given by $\sigma^2 = Y^2 t$. This means that the variance can change in two ways or a combination of both: either by a change in the size of the average receipt and disbursement (change in Y^2) or by a change in the number of transactions (change in t). On the one hand, if the average size of a transaction changes, then the variance in cash balances will change proportionately in response to a change in Y^2 . In such a case, it is clear that the transactions-elasticity will be equal to $2/3$. On the other hand, if the number of transactions increase, the variance in cash balances will change in response to a change in t so that the transactions elasticity would be $1/3$. Given a combination of such changes, the range of possible values for the transactions-elasticities becomes even larger.

3.1.3. Uncertainty in interest rates

Apart from falling transactions costs brought about by financial innovation, there is also another possible factor that may lead to a decline in the demand for transactions balances. In the last chapter in section 2.3.3(a), it was mentioned that the existence of high and more volatile rates of interest had an important role in the process of financial innovation during the last decade or so.

Would it not be appropriate to treat the existence of more volatile interest rates as an increased uncertainty in the rate of interest?

It is also shown that Niehan's model can serve as a useful vehicle for analysing changes in transactions costs.

Under the influence of the Keynesian liquidity preference theory, the transactions demand is frequently distinguished from the asset demand for money; the former being often presented in the context of inventory-theoretic models as those discussed above, and the latter following Tobin's (1958) mean-variance analysis of portfolio balance. Niehans (1978, pp. 52-59) offers a model for analysing the demand for money by combining elements of inventory-theoretic models that stress transactions costs with Tobin's mean-variance model that stresses uncertainty about interest rates.

Tobin's model did not consider transactions assets, but long-term investment portfolios. The basic idea was that the demand for cash balances was attributed to the risk that the holding of bonds might involve a capital loss in excess of interest income. A common objection raised against Tobin's liquidity preference theory is that there are some interest-bearing assets that are virtually immune from any capital losses so that there would virtually be no demand for cash balances. Niehans (1978) therefore argues that Tobin's theory has little to offer by way of explanation of the demand for cash balances - it only explains the diversification of investment portfolios.

However, such an objection is overruled when transactions costs are introduced into the model so that cash balances will always be held even when there are interest-bearing assets that are immune from capital losses. The following paragraphs will consider

Niehan's analysis of the transactions demand for cash when uncertainty in interest rates is introduced.

It is important to realise at the outset that the analysis is an extension of deterministic inventory-theoretic models, namely that it is still being assumed that the pattern of disbursements and receipts are foreseen with perfect certainty although interest rates are now subject to some uncertainty. This is, of course, a simplifying assumption. It is assumed that future rates of interest are distributed with mean equal to $E(r)$ and standard deviation equal to σ_r . Niehans (1978, p.53) also assumes that the individual attempts to maximise a utility function depending on both the mean and standard deviation of net revenue instead of attempting to maximise expected returns. The main questions that Niehans seeks to answer are: 'What is the effect of introducing uncertainty into the [inventory-theoretic] model...? Will cash balances unambiguously increase or may they conceivably decline?' (1978, p.53)

To set about answering such question, the analysis will be confined to the case of proportional transactions costs, and some tentative comments will be offered on the case when transactions costs are zero. The following definitions are applicable. Let average assets be denoted by A which equals $\frac{1}{2}Y$ and is also equal to the sum of the average cash balance and average bond or interest-bearing asset holding, denoted by \bar{M} and \bar{B} respectively. The following ratios are defined as follows:

$$\mu = \frac{\bar{M}}{\bar{A}} \quad \text{and} \quad \beta = \frac{\bar{B}}{\bar{A}}$$

where $\mu + \beta = 1$. Proportional costs will be considered first of

all.

It can be recalled from equation [7] above that net revenue under proportional transactions costs was defined as

$$\Pi_n = \frac{1}{2}Yr[1 - (1/n^2)] - 2kY[1 - (1/n)] \dots [7 \text{ repeated}]$$

It can be noted that the average cash balance is $\bar{M} = \frac{1}{2}Y(1/n^2)$ and that the average bond holding is $\bar{B} = \frac{1}{2}Y[1 - (1/n^2)]$. The second term in the preceding expression can be re-arranged slightly as $4k(\frac{1}{2}Y)[1 - (1/n)]$. Dividing [7] throughout by A yields the net revenue per currency unit (dollars or whatever) of assets:

$$\frac{\Pi}{A} = \frac{\bar{B}}{A} r - 4k \left[1 - \sqrt{\frac{\bar{M}}{A}} \right]$$

which reduces to

$$\pi = \beta r - 4k[1 - (1 - \beta)^{\frac{1}{2}}] \dots [26]$$

where π denotes net revenue per currency unit of assets. Since the rate of interest is subject to uncertainty, the expected net revenue per currency unit of assets is therefore

$$E(\pi) = \beta E(r) - 4k[1 - (1 - \beta)^{\frac{1}{2}}]$$

while the risk attached to the portfolio with a bond component β is assumed to be a linear function of the standard deviation of the rate of interest, namely that $\sigma_\pi = \beta\sigma_r$ in which case $\beta = \sigma_\pi/\sigma_r$. Thus, the preceding expression, after substitution for β , becomes the

opportunity locus analogous to that of Tobin (1958):

$$E(\pi) = E(r) \left[\frac{\sigma_\pi}{\sigma_r} \right] - 4k \left[1 - \sqrt{1 - \frac{\sigma_\pi}{\sigma_r}} \right] \dots [27]$$

This opportunity-locus shows the combinations of expected return and risk for a given portfolio of cash and bonds. It can be verified that $\partial E(\pi) / \partial \sigma_\pi > 0$ and that $\partial^2 E(\pi) / \partial \sigma_\pi^2 < 0$ which indicates that the opportunity locus is non-linear with a maximum for $E(\pi)$ at, say, σ_π^* . Such a locus is shown by, for example, OP_0 in figure 3.4. below.

Consider now the indifference map as exemplified by the indifference curves I_0 and I_1 which are drawn such that the individual is assumed to be a 'risk-averter'. The reason for the upward sloping indifference curves is that the individual's marginal risk premium tends to rise as risk increases. There are also other cases in which an individual may either be a 'risk-lover' with downward sloping indifference curves or be 'risk-neutral' with horizontal indifference curves. In the absence of uncertainty, the individual (in any risk category) will choose a portfolio that maximises expected net revenue such as $E(\pi)_{\max}$ in figure 3.4. above. When uncertainty is introduced, a risk-averting individual will choose his optimum portfolio such that a proportion β^* will be held in the form of bonds whereas μ^* will be held in the form of cash balances - this is shown by point E_0 in figure 3.4. It does follow that a risk-averter will hold more cash balances in the presence of uncertainty than would have been the case in the

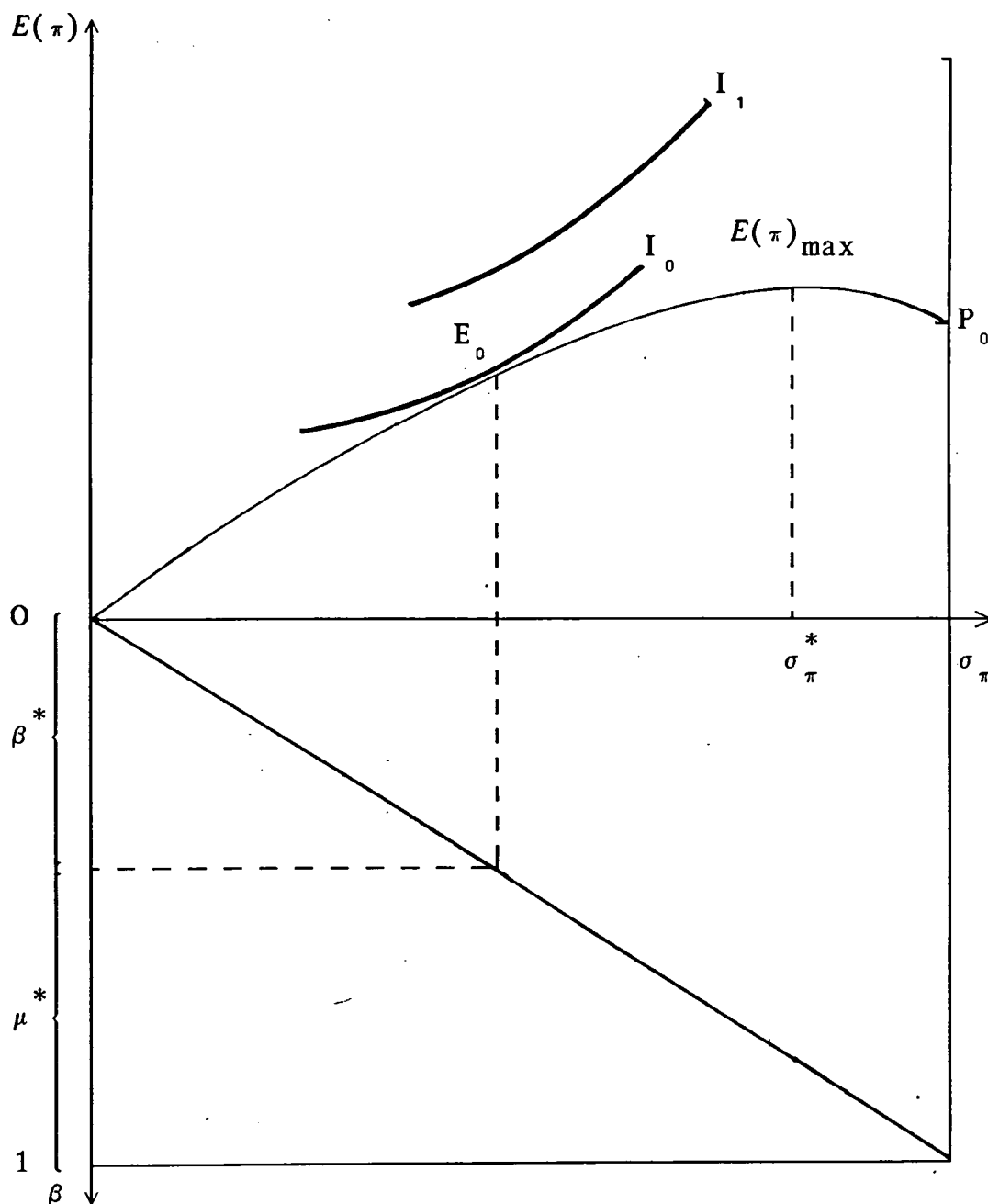


FIGURE 3.4: *The demand for money given uncertainty in interest rates and proportional transactions costs.*

absence of uncertainty. By analogous reasoning, a risk-loving individual would hold less cash balances in the presence of risk whereas a risk-neutral individual will not adjust his cash balances in any way.

Two changes will now be considered, the first change being an increase in uncertainty, and the second being a fall in transactions costs. When uncertainty about the future rate of interest increases, this is reflected by an increase in σ_r so that for any given level of $E(\pi)$, σ_π will be larger.³ This would be represented by a pivoting of the opportunity locus from OP_0 to OP_1 such that $E(\pi)_{\max}$ is unchanged, but at a higher σ_π as shown in figure 3.5. below. It has to be stressed that the effect on the demand for money is not at all that unambiguous because it essentially depends on the income and substitution effects. Given the individual's indifference map, the demand for money could either increase or decrease. If equilibrium moves from point E_0 in figure 3.5. to point E_1 , this reflects a substitution effect from point E_0 to, say, point S on indifference curve I_0 and then a weak income effect from point S to point E_1 . Alternatively, there could be a strong income effect from S to E_2 . In the former case, an increase in uncertainty actually leads to an increase in the proportion of the portfolio held as cash, whereas in the latter case, it could lead to a decrease. These changes are equivalent to a fall and a rise in the proportion of the portfolio held as bonds, denoted by β_1^* and β_2^* respectively. Thus there is some ambiguity regarding the effect on the demand for money of an increase in uncertainty about future interest rates although there is certainly a reduction in the individual's utility.

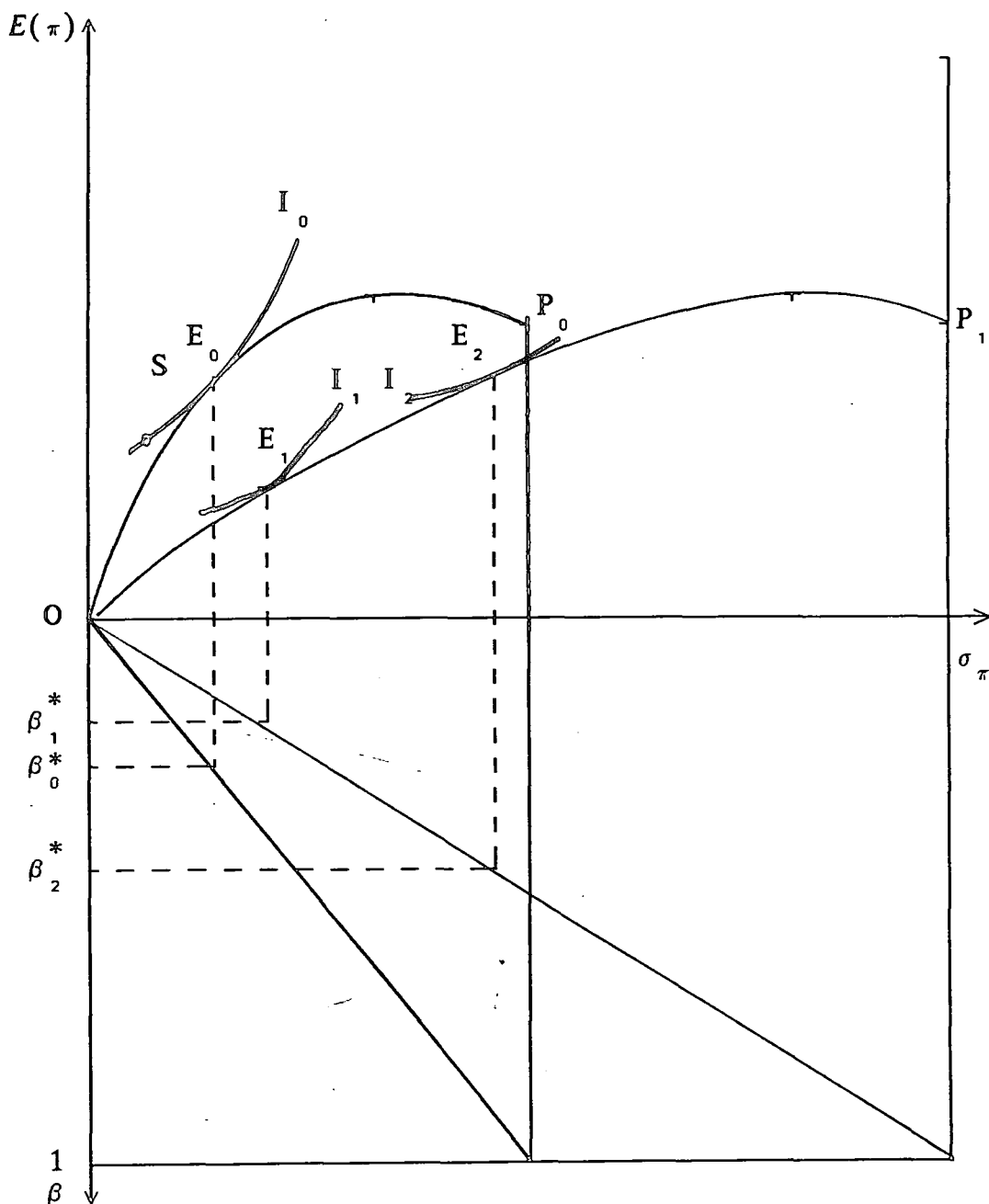


FIGURE 3.5: *Effect of a change in uncertainty of the interest rate on the demand for money.*

Now consider a change in transactions costs whose effect is to move the opportunity locus towards a straight line such as OP_2 in figure 3.6. below. The reason for this is quite clear: consider equation [27] and let proportional transactions costs approach zero. The opportunity locus will then reduce to a linear form such as

$$E(\pi) = E(r)(\sigma_\pi / \sigma_r)$$

Furthermore, σ_π will also increase in response to falling transactions costs, approaching σ_r as k approaches zero.⁴ In figure 3.6., the initial opportunity locus is given by OP_0 , with $\sigma_\pi = \sigma_0$, and the individual is at an optimum at point E_0 such that the individual will hold a proportion β_0^* in bonds. When proportional transactions costs fall, the opportunity locus will pivot towards the straight opportunity locus - this is shown in figure 3.6. by a shift of OP_0 to OP_1 , and σ_π will increase to σ_1 . Here, it is being assumed that a strong income effect is in operation so that the individual is now optimising at point E_1 where the proportion held as bonds has risen from β_0^* to β_1^* . If proportional transactions fall to zero (if ever), then the opportunity locus will become OP_2 which is linear itself, and σ_π will be equal to σ_r . Again, assuming a strong income effect, the individual will now optimise at point E_2 and the proportion held as bonds will rise further to β_2^* . As figure 3.6. is drawn, there is a decline in the demand for transactions balances with an increase in utility for the individual.

In the last case, it will be seen that the opportunity locus OP_2 is indistinguishable from Tobin's opportunity locus. The implication would be that cash balances would only be held as an

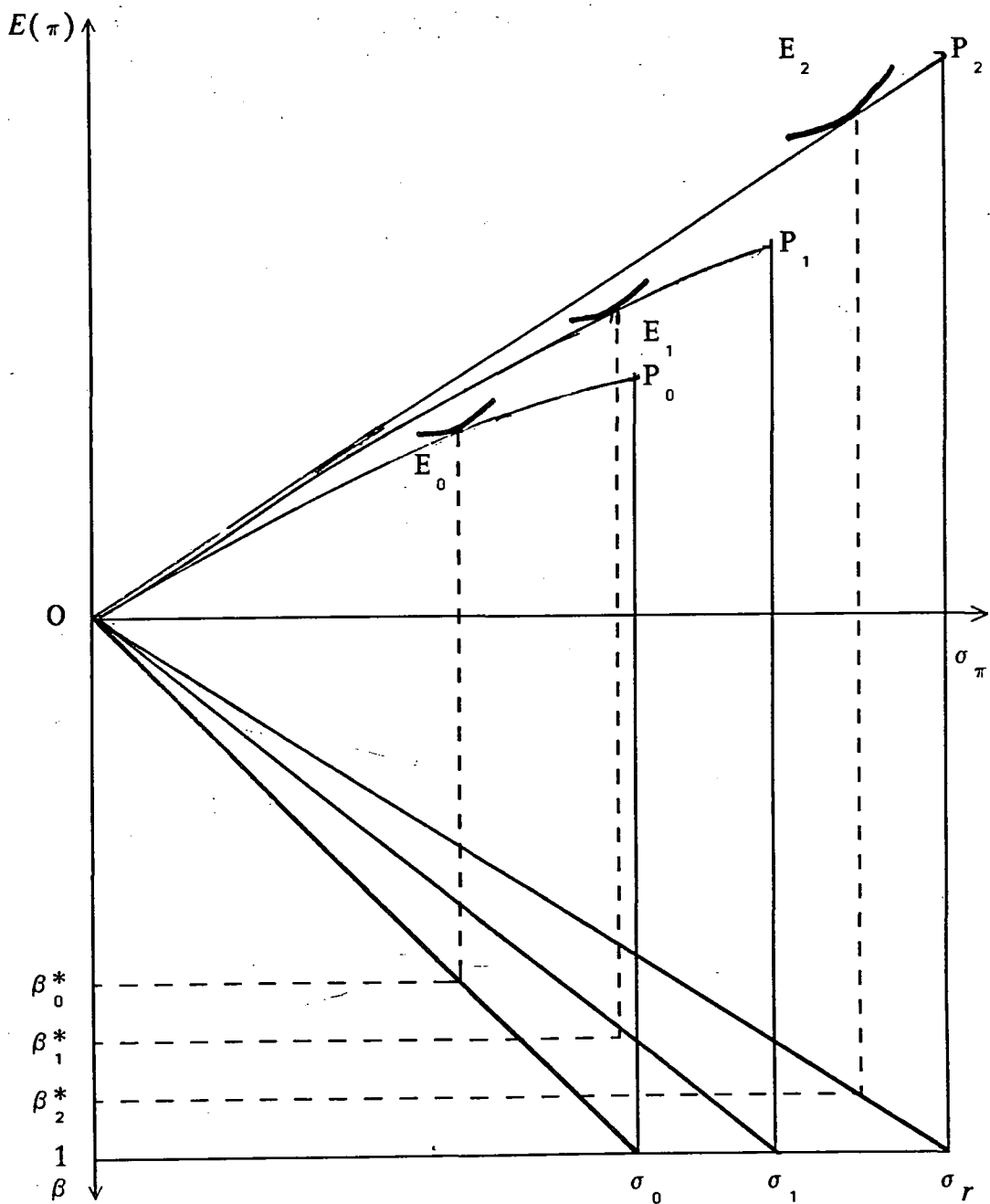


FIGURE 3.6: *Effect of a fall in proportional transactions costs on the demand for money given uncertainty about interest rates.*

asset and not for transactions balances. Given that there were interest-bearing assets in existence that were immune from capital losses, one would have to consider carefully if any cash balances would be held ever at all (Podolski (1986), pp.205–207).

Having discussed the alternative theories of the demand for money which have been extended to include uncertainty in the pattern of disbursements and receipts, and uncertainty in interest rates, the overall picture seems to be that falling transactions costs and increased uncertainty in interest rates may lead to a reduction in the demand for transactions balances, although such an outcome is of course dependent on the relative strengths of the substitution and income effects as the analysis in the last few paragraphs has already demonstrated. The rest of this chapter will now be mainly concerned with the empirical work that has already been undertaken in order to investigate whether or not financial innovation has been largely responsible for the instability of the demand for money.

3.2. Innovations as a time trend

3.2.1. A priori justification

It has been argued by Lieberman (1977, p.308) that it is possible to mis-specify a standard empirical demand-for-money function simply by ignoring the effects of technological change. Even if interest rates and the volume of transactions were held constant, there would still be strong *a priori* reasons why the demand for money would decline over time. Consider again equation [6] which was given in sub-section 3.1.1(b) above. This equation depicts a theoretical demand for money function in which α denotes the constant term. On closer inspection, it will be seen

that α is equal to $\frac{1}{2}(2b)^{\frac{1}{2}}$, and one can hardly call the term α a 'constant' if brokerage fees are allowed to change over time. When brokerage fees decline, this, *ceteris paribus*, would be reflected in a downward shift in the demand for money function. By the explicit exclusion of brokerage fees in the empirical specification of the demand for money function, one is likely to come up with biased estimates of its coefficients.

In order to overcome such difficulties, several approaches have been proposed in order to try and capture the effects of innovation on the demand for money. One possible approach would have been to include brokerage fees explicitly in the specification, but, as will be seen in section 3.5. below, such an approach suffers from the major drawback that there is a relative paucity of data on brokerage fees.

One alternative approach suggested by Lieberman (1977, p.309) would have been to include variables that measure the level of activity in an innovative process or technique. Thus, for example, one could use the volume of credit card credit to reflect the trend away from making payments with conventional transactions balances into payments by credit cards. This was an approach that was utilised by Johnston (1984) for the United Kingdom. Such variables included the number of bank current accounts per head of population, the number of building society accounts per head of population, the total number of credit cards issued, and the number of ATMs in operation. Unfortunately, such variables only reflect a small sub-set of the vast range of financial innovations that have taken place so far. Even if one were to include all variables showing the level of activity in the most significant financial

innovations, there are many problems that have to be contended with. Firstly, how are significant financial innovations to be defined? The difficulties inherent in such a definition are quite clear. Secondly, there is the problem that a large number of financial innovations would not be very amenable to econometric analysis since a large number of variables in an empirical specification of the demand for money would be likely to reduce the number of degrees of freedom on which to base statistical tests. The inclusion of a large number of variables, according to Judd and Scadding (1982, p.993), would be likely to violate the criterion of a stable demand for money function on the grounds that a relationship that requires a large number of variables to order to pin it down is, in effect, not predictable.

The approach suggested by Lieberman (1977, p.309) is to use a simple linear time trend as an additional variable. In justifying such an approach, Lieberman has made a subtle distinction between endogenous and exogenous innovations. He argues that interest rates not only reflects the opportunity cost of holding conventional transactions balances, but also the 'induced improvements in technology which tend to reduce money demand.' As will be seen later on in section 3.3., this is an idea not far removed from that of Porter and Simpson (1980) regarding the effect of high interest rates on the rate of investment in new improved money-management techniques which are designed to economise on conventional transactions balances in face of persistently high opportunity costs. Lieberman goes on to argue that 'a separate technological change variable is necessary to measure the effects of exogenously produced technological change.' One of the simplest ways of capturing the

effect of technological change on the demand for money would be to include an exponential decay term whose parameter measures the mean rate of technological change per annum. Such a variable, of course, reflects the view that technological change over the post-war period has been characterised by a *steady* process of change.

Thus, in addition to conventional specifications of empirical demand for money functions given in Chapter One above, a time trend variable may be included such that

$$(M^d/P) = \beta_0 Y^{\beta_1} r^{\beta_2} e^{\beta_3 \text{time}} \epsilon \quad \dots [28]$$

As mentioned in Chapter One, a partial adjustment process of the form $\ln(M/P)_t - \ln(M/P)_{t-1} = \gamma[\ln(M^d/P)_t - \ln(M/P)_{t-1}]$ may be included after taking the logarithms of equation [28] above, in which case it becomes:

$$\begin{aligned} \ln(M/P)_t = \gamma \ln \beta_0 + \gamma \beta_1 \ln Y_t + \gamma \beta_2 \ln r_t + \gamma \beta_3 \cdot t \\ (1 - \gamma) \ln(M/P)_{t-1} + \epsilon_t \quad \dots [29] \end{aligned}$$

where ϵ_t is a stochastic term which is equal to $\ln \epsilon_t$, and each variable as used by various empirical studies are defined in Table 3.1. which will be discussed in detail in the next sub-section.

3.2.2. Empirical evidence and analysis

Table 3.1. presents a selection of empirical results derived from a comparison of two studies that actually used a time-trend variable to represent technological innovation. To facilitate comparison with the standard specification of the empirical demand for money function, results of regressions for the standard

specification are also shown. The first two equations of Table 3.1. show the results of regressions run by Lieberman (1977) using annual data for the U.S. from 1947 until 1973.⁵ Equation 3.1.1 shows a conventional specification with a lagged dependent variable to take account of the fact that money balances do not adjust fully to desired levels within a year. As the coefficient to the lagged dependent variable shows, it implies an implausibly long adjustment period at a rate of about 18.3% per annum. The long-run elasticities of the demand for money can be calculated as 0.869 and -0.415 for real GNP and the interest rate respectively. The value of the income-elasticity implies that there would be no economies of scale in holding transactions balances, and the interest-elasticity seems to fall fairly close to the accepted theoretical value of -0.5.

When a time-trend term is added, as shown in equation 3.1.2 of Table 3.1., there is some slight improvement in that the adjustment period for money balances is now shorter at a rate of about 22.8% per annum. However, all but the real GNP and lagged dependent variables are now statistically insignificant at the 5% significance level. Thus, the addition of a time-trend variable does not have much of an effect.

Equations 3.1.3 to 3.1.7. of Table 3.1. give some of the results reported by Porter and Simpson (1980) which used quarterly data from 1959 to 1980. Equations 3.1.3 and 3.1.4 show regressions for a standard specification of the demand for money function. Two regressions were carried out; the first being for the period 1959:4-1974:2 and the second being for the period 1959:4-1980:2. The results are a typical example of the story of the breakdown in the demand for money function during the early

TABLE 3.1: Summary of selected empirical results that include innovations as a time trend.

Study	Eq. no.	Specification and regression results	Summary statistics	
Lieberman (1977) 1947-73[A]	3.1.1.	$(M1/P) = 0.04 + 0.1596NP - 0.076r_1 + 0.817(M1/P)_{-1}$ (0.77) (3.62) (-2.11) (8.21)	$R^2 = 0.935$ D.W. = 2.00 S.E. = 0.016	
	3.1.2.	$(M1/P) = -0.29 + 0.2546NP - 0.051r_1 - 0.004T + 0.772(M1/P)_{-1}$ (-0.49) (2.12) (-1.14) (-0.86) (6.79)	$R^2 = 0.943$ D.W. = 1.97 S.E. = 0.016	
Porter & Simpson (1980)	3.1.3.	$(M1A/P) = 0.788 + 0.1616NP - 0.020r_2 - 0.018r_3 + 0.660(M1A/P)_{-1}$ (1.93) (3.31) (-3.59) (-1.05) (5.15)	$\bar{R}^2 = 0.992$ D.W. = 1.58 S.E. = 0.0049	
	3.1.4.	$(M1A/P) = -0.383 + 0.0206NP - 0.023r_2 - 0.003r_3 + 1.05(M1A/P)_{-1}$ (-2.28) (1.18) (-4.93) (-0.18) (36.45)	$\bar{R}^2 = 0.982$ D.W. = 1.71 S.E. = 0.0064	
	3.1.5.	$(M1A/P) = -0.060 + 0.2526NP - 0.023r_2 - 0.013r_3 - 0.001T + 0.707(M1A/P)_{-1}$ (-0.14) (4.12) (-4.83) (-0.90) (-2.74) (6.00)	$\bar{R}^2 = 0.993$ D.W. = 1.71 S.E. = 0.0047	
	3.1.6.	$(M1A/P) = -0.365 + 0.6046NP - 0.011r_2 - 0.072r_3 - 0.0067T + 0.38(M1A/P)_{-1}$ (-0.32) (4.43) (-1.16) (-0.39) (-3.97) (1.85)	$\bar{R}^2 = 0.918$ D.W. = 1.63 S.E. = 0.0065	

Table continued on next page....

TABLE 3.1: (continued)

Study	Eq. no.	Specification and regression results	Summary statistics
Porter & Simpson (1980) continued...			
1959:4-80:2	3.1.7.	$(MIA/P) = -1.03 + 0.1996NP - 0.026r_2 + 0.004r_3 - 0.0014T + 0.953(MIA/P)_{-1}$	$R^2 = 0.986$ $D.W. = 1.73$ $S.E. = 0.0057$
		(-5.56) (4.96) (-7.17) (0.34) (-4.74) (31.33)	

Source: Tables 1 and 2 in Lieberman (1977, pp.311 & 312),
 Tables R-1 and R-3 in Porter and Simpson (1980, pp.206 & 212).

Definition of variables:

M1 = Equivalent to M1A prior to 1979 in the U.S.
 M1A = Demand deposits at all commercial banks other than those due to domestic banks, the U.S. government, and foreign banks and official institutions, less cash items in the process of collection and Federal Reserve float, plus currency outside the Treasury, Federal Reserve banks, and the vaults of commercial banks, plus travellers' checks of all non-bank issuers.

GNP = Real gross national product.

F = Implicit GNP deflator.

r1 = Moody's AAA corporate bond rate.

r2 = Treasury Bill rate.

r3 = Commercial bank passbook savings rate.

T = Linear time trend.

Key to table: [A] = Annual data.

[Q] = Quarterly data.

Figures in parentheses denote 't' statistics.

1970s. Equation 3.1.3 shows that the long-run income elasticity of the demand for money is 0.474 which is not far off from the theoretical prediction of 0.5. A further feature may also be noted in which two interest rates are included in the empirical specification; one reflecting the Treasury Bill rate, and the second reflecting the commercial bank passbook savings rate. It is not clear from Porter and Simpson (1980) why two interest rates were included, however, it would probably reflect the nature of U.S. banking regulation up to 1980 in which there were a set of regulated interest rates that could be paid by commercial banks, and another set of open-market interest rates on short-term instruments which were unregulated. Thus, the long-run elasticity would be given by the sum of the two short-run elasticities divided by the coefficient to the lagged dependent variable: the calculations show a value of -0.112 which seems rather small considering the theoretical prediction of -0.5. The coefficient to the lagged dependent variable shows an adjustment rate of about 34% per quarter which indicates that complete adjustment will not take place within a year. Equation 3.1.4. shows the second regression for the period after 1974, and it does indicate that this estimated demand for money function is very far off from its theoretical counterpart in that there is a nonsensical coefficient to the lagged dependent variable in excess of unity, leading to nonsensical values for the long-run elasticities.

Equations 3.1.5 to 3.1.7 show the results of regressions that include a time-trend variable. The most interesting result to come from equation 3.1.5 is that the coefficient for the time trend variable is actually significantly different from zero at the 5%

significance level, and it implies that the demand for M1A balances during the period 1959–1974 declined at a rate of 0.34 per cent per quarter. The coefficient to the lagged dependent variable has now increased in comparison with equation 3.1.1. implying that the period of complete adjustment is now longer and still over a year. Another interesting feature to note is that the long-run income-elasticity has now risen to 0.86 which indicates that there are now less economies of scale in holding conventional transactions balances. However, the long-run interest elasticity is still rather small at -0.123 .

Equation 3.1.7 shows the same period of fit as for equation 3.1.4 of Table 3.1. It will be noted that the coefficient to the lagged dependent variable implies an implausibly long adjustment period, and this is reflected in rather dubious values for the income- and interest-elasticities of the demand for money. The coefficient to the interest rate on passbook savings has actually become positive, although it is insignificantly different from zero at the 5% significance level. Thus, overall, it is concluded that the addition of a time trend does not have much improvement on the empirical demand for money function.

A major objection against the use of time-trend variables to represent technological change has been raised by Porter and Simpson (1980, p.176) who say that the use of such variables reflects the view that innovations in cash-management occur at a steady rate over time. Such a view is difficult to justify when recent economic history is taken into account which indicates that there have been observed periods in which innovations have occurred at an accelerated pace. There is certainly a case for trying to

endogenise innovations in cash-management techniques as they are very much influenced by the behaviour of interest rates. Thus the following section considers in some detail how cash-management innovations have been endogenised within the demand for money function, and empirical results will, of course, be presented.

3.3. Innovations in cash-management

3.3.1. Outline of theory

The main proposition put forward by Porter and Simpson (1980, p.165) is that, in the short run, '...the demand for traditional monetary assets is somewhat insensitive to changes in opportunity costs but in the long run the response tends to be much stronger as more substitutes are developed and used.' Porter and Simpson argue that the presence of record-high interest rates in the late 1970s led to record-high opportunity costs for holding ordinary transactions balances. The public did not only try to economise on existing cash balances, but they they invested in new money-management techniques that were designed to lower the amount of transactions balances required for a given amount of spending in a climate in which it was generally expected that high opportunity costs would persist.

The main reason for the distinction between short-run and long-run responses to high opportunity costs of holding transactions balances is that when individuals expect *persistently* high opportunity costs, they have an incentive to actively seek ways to modify their cash-management systems in order to reduce their conventional transactions balances permanently. Porter and Simpson (1980, p.166) contrast this long-run response with the short-run

response in which an individual is confronted with a set of *given* cash management techniques by which the individual can economise on his conventional transactions balances. Thus, it is contended that the short-run demand for money schedule is drawn for a set of given money-management techniques whereas the long-run demand schedule is drawn so as to allow for variations in cash-management techniques. Owing to the relative unpredictability of financial innovations, the short-run demand for money would be relatively more predictable than its long-run counterpart.

It needs to be emphasised that the preceding analysis was carried out in the context of a monetary system in which explicit interest payments on demand deposits was expressly prohibited (such as was the case in the U.S. prior to 1980). If explicit interest payments on demand deposits were allowed, would the above analysis hold true? An answer can be provided if one considers a banking system in which reserve requirements are operational. From the experience of the banks, it will be seen that individual customer accounts are subject to a uncertain cash-flow pattern, and one advantage of the bank is the ability to pool such cash-flow disturbances such that the aggregate cash-flow disturbance is minimised. If it is assumed that such cash-flow disturbances were exactly offsetting each other in the aggregate, there would really be no need for the bank to undertake management of its reserve assets in order to fulfil reserve requirements. Thus, Porter and Simpson (1980, p.168) argue that in the special case where reserve positions do not fluctuate, the costs of reserve management by the bank will approach zero. Thus, a bank would be able to offer rates of return on its demand deposits equivalent to those offered on short-

term market instruments. However, in spite of much diversification by the bank, there always exists the possibility that the bank's reserve position will be subject to some form of fluctuation because cash-flow disturbances for each of the bank's customers do not exactly offset each other in the aggregate. In such a case, the bank has to undertake the costs of managing its reserve position, and such costs will be passed on to its demand deposit customers in the form of lower yields. Thus, there will exist a differential between rates of return on demand deposits and those on short-term market instruments so that such a differential would be equivalent to the opportunity cost of holding conventional transactions balances. Of course, such opportunity costs would be lower in this case where interest is being paid on demand deposits than in the case where no interest is being paid, but the incentive to invest in improved cash-management techniques is still there, albeit in a weakened form. Even if one were to consider a banking system in which no reserve requirements are operational, the bank still has to maintain some reserves as a prudential measure in meeting some of its customers' daily withdrawal patterns so that the bank still incurs costs of managing its reserves. Thus, the higher the opportunity cost is, and the longer it is expected to persist, the greater will be the incentive to invest in new cash-management techniques.

There are also other factors which may serve to strengthen the incentive to invest in new cash-management techniques in face of persistently high opportunity costs. The increasing proliferation of close substitutes to conventional transactions balances is likely to increase the incentive to invest in cash-management techniques that are designed to transfer surplus funds from those demand deposits

that have a relatively high opportunity cost to those short-term liquid assets that bear a relatively lower or even zero opportunity costs. The ever-continuing pace of technological innovations will also serve to reduce even further the transactions costs involved in switches between conventional transactions balances and short-term liquid assets.

3.3.2. *The Porter-Simpson model*

In order to give their theory of improved cash-management techniques affecting the demand for money a firmer foundation, Porter and Simpson (1980, pp.193-198) have developed a model that endogenises cash-management techniques into the Miller-Orr model of the demand for money. It was seen in sub-section 3.1.2. that the main assumption of the Miller-Orr model is that cash flows of the firm are subject to some uncertainty, and Porter and Simpson argue that this is tantamount to assuming that such cash flows are exogenous. They point out that by adopting improved cash-management techniques, the firm is able to reduce the uncertainty about its future cash flows. In other words, Porter and Simpson have partly endogenised the firm's cash flow by endogenising cash-management techniques.

Porter and Simpson denote a unit of cash-management services by the variable λ , and they assume that the cost of such services is fully variable such that the cost of cash management services is equal to λe where e denotes the variable cost of cash-management services. However, they still assume that the 'brokerage fee' is still a fixed cost. Furthermore, a function $g(\lambda)$ is defined by Porter and Simpson which reflects a factor, taking on a value between zero and unity, that reduces the variance of the

firm's cash flow. Thus, if a firm purchases λ units of cash-management services, then the variance of its future cash flows is reduced from σ^2 to $\sigma^2 g(\lambda)$. It is quite clear that if the firm does not invest in any cash-management services, then the variance of the firm's future cash flows remains unchanged, namely that $g(0) = 1$. As the firm purchases even more cash-management services, $g(\lambda)$ will fall at a decelerating rate so that beyond a certain point, diminishing marginal returns will set in, that is, $g'(\lambda) < 0$ and $g''(\lambda) > 0$.⁶

Following the same exposition as given in Miller and Orr (1966, p.423), Porter and Simpson (1980, p.193) have shown that the expected cost function is given by

$$E(c) = \frac{b\sigma^2 g(\lambda)}{z'(h' - z')} + \frac{r(h' + z')}{3} + \lambda e$$

where all the variables have the same definitions as those given prior to equation [22] in sub-section 3.1.2. above, except for λ and e which were defined in the preceding paragraph. Note that primes have been added to the variables h and z to make it clear that they are denominated in *single* currency units where, as before, $h' = hY$ and $z' = zY$. Recalling that $\sigma^2 = Y^2 t$, the preceding expression can be re-arranged into a form that will be directly comparable with equation [22]:

$$E(c) = \frac{btg(\lambda)}{z(h-z)} + \frac{rY[(h-z) + 2z]}{3} + \lambda e \quad \dots [30]$$

which is a form based on the exposition of Orr (1970). As can be

seen quite readily, the only difference between equations [22] and [30] is the presence of the factor $g(\lambda)$ and the cost of cash-management services, λe . As before, the firm is assumed to minimise its expected costs so that the necessary conditions for a minimum are that

$$\frac{\partial E(c)}{\partial (h - z)} = \frac{-btg(\lambda)}{(h - z)^2 z} + \frac{rY}{3} = 0 \quad \dots [31]$$

and

$$\frac{\partial E(c)}{\partial z} = \frac{-btg(\lambda)}{z^2(h - z)} + \frac{2rY}{3} = 0 \quad \dots [32]$$

There is also a further condition to be satisfied if λ is to be chosen so as to minimise $E(c)$:

$$\frac{\partial E(c)}{\partial \lambda} = \frac{btg'(\lambda)}{z(h - z)} + e = 0 \quad \dots [33]$$

The above equations can be solved to give the following expressions, after remembering to convert back z and h into single currency units:

$$hY = 3M \quad \dots [34]$$

where it is to be recalled that $zY = M$, and

$$-M \frac{g'(\lambda)}{g(\lambda)} = \frac{3e}{2r} \quad \dots [35]$$

and

$$M^3 = \frac{3b\sigma^2 g(\lambda)}{4r} \quad \dots [36]$$

It is now required to show that under cash-management innovation, the interest-elasticity for the demand for money will be greater in absolute value than $-1/3$, namely that

$$\frac{\partial \ln M}{\partial \ln r} < -\frac{1}{3}$$

It is first necessary to take logarithms of equation [36] above, and then differentiate it with respect to $\ln r$:

$$\frac{\partial \ln M}{\partial \ln r} = -\frac{1}{3} + \frac{1}{3} \cdot \frac{\partial \ln g(\lambda)}{\partial \ln r}$$

The reason for the second term in the above expression is that as interest rates change, $g(\lambda)$ will change via a change in λ . The interest-elasticity of the demand for money is therefore

$$\frac{\partial \ln M}{\partial \ln r} = -\frac{1}{3} + \frac{1}{3} \cdot \frac{\partial \ln g(\lambda)}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial \ln r} \quad \dots [37]$$

According to the definitions given by Porter and Simpson (discussed above) of the function $g(\lambda)$, it can be seen that $g'(\lambda) < 0$, and the second term will be negative if and only if $\partial \lambda / \partial \ln r > 0$. Such a requirement is not too difficult to justify on *a priori* grounds if one considers that the firm varies its investment in cash-management techniques in direct response to changes in interest

rates. Thus, from equation [35], the interest-elasticity is unambiguously greater in absolute value under conditions when cash-management techniques are allowed to vary rather than for *given* cash-management techniques.

Porter and Simpson (1980, p.193, equation A-4) have derived a 'fourth-root' demand for money function for the case where cash-management techniques are allowed to vary for a specific form of the function $g(\lambda)$. Their procedure, if carried over to the general form of the function $g(\lambda)$, would have given

$$M^4 = \frac{-9b\sigma^2 e [g(\lambda)]^2}{8r^2 g'(\lambda)} \quad \dots [38]$$

which is achieved by multiplying equation [36] above throughout by M and substituting for M on the right hand side from equation [37]. As argued in sub-section 3.1.2(c) above, the long-run demand for money is given by $M^d = (hY + M)/3 = 4M/3$ so that from equation [36] above, the long-run demand for money when cash-management innovations have been endogenised is

$$M^d = \frac{4}{3} \left[\frac{-9b\sigma^2 e [g(\lambda)]^2}{8r^2 g'(\lambda)} \right]^{\frac{1}{4}} \quad \dots [39]$$

This expression will only be valid if $g'(\lambda) < 0$. Consider for example a specific form of the function $g(\lambda) = 1/(1 + \lambda)$ where $g'(\lambda) = -1/(1 + \lambda)^2$. Therefore the long run demand for money would be given by $M^d = \alpha [(b\sigma^2 e)/8r^2]^{\frac{1}{4}}$ where $K = (4/3)(9/8)^{\frac{1}{4}}$.

There are two ways of capturing the effects of high interest

rates on cash-management innovations; firstly by using past-peak variables, and secondly, by using ratchet variables. Their *a priori* justifications will be discussed in the next sub-section along with empirical results.

3.3.3. The use of ratchet variables as proxies

One justification for the use of previous-peak interest rates in empirical demand for money specifications is that there may be termed as an 'awareness threshold' in which a greater awareness of higher opportunity costs of holding conventional transactions balances may come about as a result of interest rates surpassing their previous peak. (Porter and Simpson (1980), pp.179-180) When awareness of higher opportunity costs has become accentuated in times of high interest rates, it creates a favourable climate for arousing greater interest in new cash-management techniques in the anticipation that interest rates will continue to be higher in the future. So, what was originally deemed to be unprofitable, new cash-management techniques will now become profitable, and once investment in such techniques has already taken place, it tends to have a permanent effect on the demand for money, even after interest rates have fallen beyond their previous peak. Thus, previous peak interest rates that are included in empirical demand for money functions may serve well to capture the effects of innovation.

However, it has to be recognised that investment in new cash-management techniques does not simply take place overnight; in fact, it would be more reasonable to postulate that such investments take time to implement. As previously discussed in Chapter Two, there are numerous factors that determine the rate of diffusion of

financial innovations. For instance, the technology required to implement the new cash-management techniques may have to be developed, and there are also 'learning periods' in which firms and individuals seek to familiarise themselves with the new technology before adapting it on a wider scale. Given that there are delays inherent in the implementation of new cash-management techniques, it is reasonable to take the view that such innovations will take time to have a significant effect on the demand for money. Thus, Porter and Simpson (1980) suggest the use of a ratchet variable (as opposed to previous-peak variables) would be more appropriate under the circumstances.

Porter and Simpson (1980) then go on to define the ratchet variable which is the cumulative sum of positive terms, each term being the differences between an opportunity cost variable, v_j , and a n -period moving average of the most recent opportunity cost variables, v_i for $i = j - (n - 1), \dots, j$ so that

$$s_t = \sum_{j=1}^t \left[v_j - \frac{1}{n} \sum_{i=j-n+1}^j v_i \right]^+ \quad \dots [40]$$

The notation $()^+$ is used by Porter and Simpson to denote the fact that if the current opportunity cost variable is greater than the moving average, then the positive difference is added to the cumulative sum; otherwise a negative difference adds nothing.

As previously discussed in the last sub-section, if interest rates rise beyond a certain threshold effect, then the Porter-Simpson model predicts a rise in the interest-elasticity of the demand for money during the long run. Thus, Porter and Simpson (1980, p.183) consider different functional forms for the ratchet variable in

order to capture the variability in interest-elasticities as interest rates rise beyond a threshold level. In particular, they consider three functional forms, viz: linear, linear times logarithm, and power transformation. The linear functional form, s_t , will give an interest elasticity of cs_t where c is the coefficient to the ratchet variable in the regression, and the linear times logarithmic functional form, $s_t \times \ln(s_t)$, will give an interest-elasticity of $cs_t[1 + \ln(s_t)]$. As will be seen in the following analysis of empirical results, it seems that Porter and Simpson (1980) regard this functional form for the ratchet variable as giving the best overall performance, and this is confirmed by Porter and Offenbacher (1984, p.92, footnote 5) who say that this form has been 'used exclusively'. The final functional form to be considered is the power transformation form, s_t^λ . It is not explicitly clear from Porter and Simpson what the parameter λ does represent. It would be plausible to assume that this stood for the number of units of cash-management services purchased, as previously defined in the last sub-section. Thus, if λ were to increase, then s_t^λ would increase, leading to a rise in the interest-elasticity of $c\lambda s_t^\lambda$. Porter and Simpson also included another functional form of the ratchet variable which is $\ln(s_t)$, that has the property of constant elasticity.

Porter and Simpson used a Shiller distributed lag estimation technique in which they use a four-quarter lag for the Treasury Bill rate, a three-quarter lag for real GNP, a six-quarter lag for the money management ratchet variable, and the passbook savings rate entered the regression contemporaneously. Except for the ratchet variable, the regression equation was entered in double logarithmic form:

$$\begin{aligned} \ln(M/P)_t &= \beta_0 + \sum_{j=0}^3 \beta_{1,j} r_{1,t-j} + \beta_2 r_{2,t} \\ &+ \sum_{j=0}^2 \beta_{3,j} (Y/P)_{t-j} + \sum_{j=0}^5 \beta_{4,j} g(s_{t-j}) \\ &\dots [41] \end{aligned}$$

No lagged dependent variable was specified. Some selected results of the above regression are shown in Table 3.4., and will be discussed later on.

Consider Tables 3.2. and 3.3. which summarise some selected empirical results from regressions that include previous-peak (*not* ratchet) variables as proxies for innovation. Table 3.2. shows two regressions selected from Goldfeld (1976) which show the effect of the addition of the previous-peak in commercial paper rates, r_p . Equation 3.2.1. of Table 3.2. shows a typical regression for a standard specification of the demand for money. Whilst it is not possible to say anything about the overall improvement in the fit to the data resulting from the addition of the previous-peak variable in equation 3.2.2. as figures for R^2 were not reported by Goldfeld (1976), it is possible to discern some improvement in the properties of the specification contained in equation 3.2.2 from those of the standard specification. For example, the coefficient to the lagged dependent variable has decreased from 0.822 to 0.767, implying that the adjustment period has shortened somewhat, but still rather long. Furthermore, a test of the null hypothesis that the coefficient to the previous-peak variable is insignificantly different from zero indicates that the null hypothesis can be rejected at the 5% significance level, implying that the previous-peak variable is

TABLE 3.2: Summary of selected empirical results that use past-peak variables as proxies for innovation.

Study	Eq. no.	Specification and regression results	Summary statistics
Goldfeld (1976)	1952:2-73:4[Q]	3.2.1. $(M1/P) = \beta_0 + 0.112GNP - 0.035r_1 - 0.010r_2 + 0.822(M1/P)_{-1}$ (5.4) (-2.9) (-4.2) (14.1)	$R^2 = \text{n.a.}$ $D.W. = \text{n.a.}$ $S.E. = \text{n.a.}$
	1952:2-73:4[Q]	3.2.2. $(M1/P) = \beta_0 + 0.154GNP - 0.037r_1 - 0.009r_2 - 0.020r_p + 0.767(M1/P)_{-1}$ (6.6) (-3.3) (-4.9) (-3.3) (13.4)	$R^2 = \text{n.a.}$ $D.W. = \text{n.a.}$ $S.E. = \text{n.a.}$

Source: Tables 5 and 6 in Goldfeld (1976, pp.698 & 701)

Definition of variables:

M1 = Equivalent to M1A prior to 1979 in the U.S.
M1A = Demand deposits at all commercial banks other than those due to domestic banks, the U.S. government, and foreign banks and official institutions, less cash items in the process of collection and Federal Reserve float, plus currency outside the Treasury, Federal Reserve banks, and the vaults of commercial banks, plus travellers' checks of all non-bank issuers.

GNP = Real gross national product.

P = Implicit GNP deflator.

r_1 = Interest rate on time deposits.

r_2 = Treasury Bill rate.

r_p = Previous-peak commercial paper rate.

Key to table: [Q] = Quarterly data.

β_0 = Constant term, but value was not reported.

n.a. = Either not available or not reported.

Figures in parentheses denote 't' statistics.

TABLE 3.3: *Post sample errors resulting from dynamic simulations for conventional demand for money functions and those containing previous-peak variables.*

<i>Aggregate</i>	<i>Specific- ation</i>	<i>Sample period</i>	<i>Simulation period</i>	<i>Annual errors</i>	
				<i>Mean</i>	<i>RMSE</i>
GM1	Conv.	52: 2- 73: 4	74: 1 76: 2	n. a.	4. 8
GM1	P. P.	52: 2- 73: 4	74: 1- 76: 2	n. a.	3. 6
DD	Conv.	55: 4- 74: 2	74: 3- 80: 2	-4. 58	6. 31
DD	P. P.	55: 4- 74: 2	74: 3- 80: 2	-3. 76	4. 20
M1B	Conv.	59: 4- 74: 2	74: 3- 80: 2	-2. 56	4. 27
M1B	P. P.	55: 4- 74: 2	74: 3- 80: 2	-1. 68	2. 43

Source: Tables 5 & 6 in Goldfeld (1976, pp.696 & 701), and Tables B-1, B-2, B-6, & B-7 in Porter and Simpson (1980, pp.206-207, 208-209, 224-225, & 226).

Notes: Conv. = Conventional specification, P.P. = Specification that includes previous-peak interest rate, GM1 = M1 definition used in Goldfeld (1976) but is equivalent to M1A, DD = Demand deposits in Porter and Simpson (1980), and M1A and M1B are those aggregates as defined in Table 3.4. below for Porter and Simpson (1980).

Figures for annual mean errors and RMSEs are given in percentages.

another important explanatory variable. Further insight into the empirical results can be obtained by considering Table 3.3. which presents a summary of selected results showing mean annual errors and root mean square errors arising from dynamic simulations of demand for money functions based on equations 3.2.1 and 3.2.2. of Table 3.2. for Goldfeld (1976), and similar specifications for Porter and Simpson (1980).⁷ The results do indicate an overall reduction in both annual mean errors and root mean square errors for those specifications that include previous-peak variables.

Consider now Tables 3.4. and 3.5. which show some selected results of the empirical work of Porter and Simpson (1980). The period of simulation was 1974:3-1980:2 which is the same simulation period for the conventional specifications of the demand for money function and those containing a previous-peak variable.⁸ It will be seen from Table 3.5. that the demand for money function that contains a logarithmic-times-linear functional form for the ratchet variable seems to have the best overall performance in terms of the smallest RMSE for both quarterly and annual errors. When the annual RMSEs are compared with those given in the last four rows of Table 3.3., it will be seen that the use of a ratchet variable instead of a previous-peak variables gives superior results, except for the case of constant-elasticity ratchet variables which seems to substantiate the theory put forward by Porter and Simpson that the interest-elasticity increases over the long-run when cash-management techniques are allowed to vary.

Having established that Porter and Simpson arrived at the 'best' demand for money which includes the linear-times-log specification for the ratchet variable, Table 3.4. shows a summary

TABLE 3.4: Summary of selected empirical results that use ratchet variables as proxies for innovation.

Study	Eq. no.	Specification and regression results	Summary statistics
Porter and Simpson (1980)			
1955:1-74:2[0]	3.4.1.	$(DD/P) = 1.85 + 0.506GMP - 0.044r_1 - 0.030r_2 - 0.0010r_1$ (4.45) (7.82) (-2.15) (-2.40) (-3.30)	$\bar{R}^2 = 0.985$ D.W. = 1.32 S.E. = 0.0054
1955:1-80:2[0]	3.4.2.	$(DD/P) = 2.20 + 0.448GMP - 0.0010r_1 - 0.027r_2 - 0.0018r_1$ (3.09) (4.55) (-1.27) (-2.03) (-5.67)	$\bar{R}^2 = 0.983$ D.W. = 1.31 S.E. = 0.0067
1955:1-74:2[0]	3.4.3.	$(M1B/P) = 1.93 + 0.527GMP - 0.039r_1 - 0.027r_2 - 0.0007r_1$ (5.35) (9.44) (-2.23) (-2.59) (-2.83)	$\bar{R}^2 = 0.993$ D.W. = 1.37 S.E. = 0.0046
1955:1-80:2[0]	3.4.4.	$(M1B/P) = 1.99 + 0.512GMP - 0.008r_1 - 0.026r_2 - 0.0011r_1$ (5.13) (8.78) (-1.43) (-2.47) (-5.47)	$\bar{R}^2 = 0.989$ D.W. = 1.33 S.E. = 0.0052

Table continued on next page....

TABLE 3.4: (continued)

Source: Table B-4 in Porter and Simpson (1980, pp.219-220).

Definition of variables:

- DD = Demand deposits.
 M1A = Demand deposits at all commercial banks other than those due to domestic banks, the U.S. government, and foreign banks and official institutions, less cash items in the process of collection and Federal Reserve float, plus currency outside the Treasury, Federal Reserve banks, and the vaults of commercial banks, plus travellers' checks of all non-bank issuers.
 M1B = M1A plus negotiable order of withdrawal (NOW) and automatic transfer service (ATS) accounts at banks and thrift institutions, credit union share draft accounts, and demand deposits at mutual savings banks.
 GNP = Real gross national product.
 P = Implicit GNP deflator.
 r₁ = Passbook savings rate.
 r₂ = Treasury Bill rate.
 r_r = Ratchet variable which is defined as the cumulative sum of the positive elements of the difference between the current 5-year rate on U.S. government notes and a 12-quarter moving average of recent yields on these securities. Functional form is linear times log.

Key to table: (Q) = Quarterly data.

Figures in parentheses denote 't' statistics.

Note: No coefficients for the lagged dependent variable were reported.

TABLE 3.5: *Post sample errors resulting from simulations for demand for money functions containing ratchet variables.*

<i>Aggreg- ate.</i>	<i>Specific- ation of ratchet variable.</i>	<i>Quarterly errors</i>		<i>Annual errors</i>	
		<i>Mean</i>	<i>RMSE</i>	<i>Mean</i>	<i>RMSE</i>
DD	Linear	-2.73	5.91	-2.84	3.46
DD	Log	-3.98	6.32	-4.16	4.81
DD	Log×linear	-2.40	5.69	-2.49	3.22
DD	Power trans- formation	-0.05	6.46	-0.11	3.68
M1B	Linear	-0.95	4.44	-1.00	2.30
M1B	Log	-1.76	4.75	-1.86	2.97
M1B	Log×linear	-0.73	4.43	-0.76	2.31
M1B	Power trans- formation	-0.95	4.45	-1.00	2.31

Source: Table 5 in Porter and Simpson (1980, p.184).

Notes: Sample period is 1955:1-74:2, and period of simulation is 1974:3-80:2.

Specifications of ratchet variables: Linear = s_t , Log = $\ln(s_t)$, Log×linear = $s_t \times \ln(s_t)$, and Power transformation = s_t^λ .

DD = Demand deposits in Porter and Simpson (1980), and M1B is the aggregate as defined in Table 3.4. above for Porter and Simpson (1980).

Figures for quarterly and annual mean errors and RMSEs are given in percentages.

of the regression results for the linear-times-log specification. A study of the 't'-ratios in Table 3.4. for the ratchet variables in all equations indicate that the inclusion of such variables should be encouraged on the grounds of their statistical significance at the 5% significance level, and it does seem that their statistical significance increases as the sample period is extended beyond 1974 to 1980.

In judging the stability of their demand for money functions for various aggregates, Porter and Simpson (1980, p.202) compare income- and interest-elasticities for each specification over two sample periods. Thus, for example, equations 3.4.1 and 3.4.2 of Table 3.4. indicate that the long-run income-elasticity has fallen from 0.506 to 0.448 for the demand deposits aggregate, whereas for the M1B aggregate, it has fallen from 0.527 to 0.512 as the sample period is extended to 1980 from 1974. Furthermore, the elasticity for the Treasury Bill rate with respect to demand deposits has fallen in absolute value from 0.030 to 0.027, whereas for M1B, it has fallen in absolute value from 0.027 to 0.026. Thus, it could have been concluded that such demand for money functions exhibited remarkable stability. However, the picture becomes rather doubtful when the passbook savings rate elasticities are considered: all equations in Table 3.4. indicate a sharp fall in absolute value. It has been suggested by Porter and Simpson (1980, pp.201-202) that this may be due to the fact that the passbook savings rate is a primitive form of proxy for cash-management techniques, and that the sharp fall in the passbook savings rate elasticity may be more than offset by the change in the elasticity for the ratchet variable so that there would tend to be a small fall in the absolute value of the combined 'cash-management impacts' elasticity. On the face of

such evidence, Porter and Simpson conclude that such equations for the M1B aggregate are 'remarkably stable.'

However, in a later study, Porter and Offenbacher (1984, pp.54-55) admit that such a specification containing ratchet variables cannot be expected to hold up well. The reason is that, even increasing interest-elasticities caused by innovations in cash-management has a basis in theory, in practice, there are any number of functional forms for the ratchet variable which would exhibit increasing elasticities so that the choice of the best functional form for the ratchet variable is essentially an arbitrary one. Secondly, the specification that includes ratchet variables assumes that all cash-management innovations are purely endogenous. It is not too difficult to recognise that there are also other forms of cash-management innovations which are exogenous, namely that pure technological change (such as improvements in information technology) have no relationship to interest rates when they rise beyond a threshold level. In light of the discussion on the distinction made by Lieberman (1977) between endogenous and exogenous innovations (discussed in Section 3.2. above), it would be well worth while to try the inclusion of a time-trend variable in Porter-Simpson equation in future empirical work, in spite of the fact that exogenous technological innovations are assumed to occur smoothly over time. If such an approach fails, then one must try to consider new ways of allowing for the effects of exogenous technological change on the demand for money.

3.4. The problem of brokerage fees

A considerable part of this chapter was devoted to the discussion of the importance of transactions costs in the first section. It was clear that a fall in transactions costs, however defined, will invariably lead to a decline in the demand for cash balances. Virtually all empirical studies of the demand for money were unable to include any form of 'brokerage fee' variable in their specifications of the demand for money simply because there was no brokerage fee series available. Even if a brokerage fee series existed at all, there still would be some formidable difficulties of trying to quantify such transactions costs, because, as pointed out in sub-section 3.1.1(b) above, the brokerage fee also includes some subjective costs which may include, for instance, the premium on time an individual may place by having to queue up at the cashier's window in a bank, or even the time and effort required by the individual in communicating with his stockbroker. Even if there was no subjective component in the brokerage fee, there is still the problem of finding a 'representative' asset on which to base a brokerage fee series: this problem is not too far removed from that one of deciding upon the appropriate interest rate to be included in an empirical specification of the demand for money.

In order to overcome such problems, Porter and Offenbacher (1984) present a highly unorthodox way of deriving a brokerage fee series, which if taken too literally, would have led to some highly spurious empirical results when analysing the stability of the demand for money. The reason why such an unorthodox method of deriving a brokerage fee series is being discussed here is to warn against the adoption of such unorthodox methodology.

3.4.1. The methodology of Porter and Offenbacher (1984)

(a) *The derivation of the debits equation.* Porter and Offenbacher (1984, pp.55-66 and pp.92-93, footnote 8) have shown how a brokerage fee series could be derived by solving for the brokerage fee, b , from the money demand and debits equations. In order to derive the debits equation, Porter and Offenbacher use the Miller-Orr probabilistic inventory-theoretic model, discussed in sub-section 3.1.2. above. First, it can be recalled from equation [21] in sub-section 3.1.2. above that the probability of a transaction occurring *in either direction* was defined to be:⁹

$$P(T) = t/[z(h - z)] \quad \dots [21 \text{ repeated}]$$

where h denotes the 'ceiling' cash balance which triggers off a transfer of $(h - z)$ into interest-bearing assets, and z denotes the level of the cash balance that the firm will return to after a transfer, and t denotes the number of times which the cash balance changes either by $+Y$ or $-Y$; in other words, it may be regarded as the 'turnover' rate. However, Porter and Offenbacher (1984, p.92) argue that the probability of *debiting the current account* (i.e. cash balance) is quite distinct from the probability of a transaction occurring in either direction. Thus, they define the probability of a debit occurring, $P(D)$, as:

$$P(D) = t/[h(h - z)] \quad \dots [42]$$

and it is argued that expected debits, $E(D)$ are given by $P(D) \cdot (h - z)$ so that

$$E(D) = t/h = t/3z$$

after recalling from sub-section 3.1.2(a) that $h = 3z$. Now, the preceding expression is denominated in Y currency units, so in order to convert to single currency units, the preceding expression has to be multiplied throughout by Y to get

$$d = E(D) \cdot Y = tY/3z$$

It will be recalled from the discussion of the Miller-Orr model that the optimal demand for money was given by

$$z = \sqrt[3]{\frac{3bt}{4rY}}$$

so that when substituted into the preceding expression for expected debits,

$$(3d)^3 = \frac{t^3 Y^3}{(3bt/4rY)} = \frac{4Y^4 t^2 r}{3b} = \frac{4\sigma^4 r}{3b}$$

after noting that $(Y^2 t)^2 = (\sigma^2)^2 = \sigma^4$. A final re-arrangement gives the debits equation:¹⁰

$$d = \frac{1}{3} \sqrt[3]{\frac{4\sigma^4 r}{3b}} \quad \dots [43]$$

It is interesting to note that as the interest rate rises, the number of debits will increase, and as brokerage fees decline, the number of debits will also increase.

(b) *The derivation of the brokerage fee series.* Having derived the debits equation, Porter and Offenbacher proceed to derive an indirect brokerage fee series. They show that the turnover of non-financial debits is given by $t = d/M$ so that

$$t = \frac{1}{3} \left[\frac{4\sigma r}{3b} \right]^{2/3} \dots [44]$$

where M has been substituted from equation [23b] in sub-section 3.1.2(b) above. The preceding expression is then solved for the brokerage fee:

$$b^{(t)} = \frac{4}{3} \cdot \left[\frac{1}{3} \right]^{2/3} \sqrt{\frac{\sigma^2 r^2}{t^3}} \dots [45]$$

where the superscript (t) on the brokerage fee indicates one of the methods of deriving a brokerage fee series, namely through equation [44]. By using debits data, Porter and Offenbacher estimated equations for d and M , and then solved them for the brokerage fee. A brokerage fee series was then constructed by substituting for the transactions and interest rate variables. The resulting brokerage fee was then plugged back into the empirical demand for money equation which takes the form:

$$\begin{aligned} \ln(M/P) = \beta_0 + \beta_1 \ln(Y/P) + \beta_2 \ln(r) + \beta_3 \ln(b^{(t)}/P) \\ + \ln(M/P)_{-1} \end{aligned}$$

Porter and Offenbacher then presented a large amount of statistical evidence purporting to show that brokerage fees were responsible for

shifts in the demand for money function. However, it would be meaningless to analyse their empirical results as the following critique of their methodology will indicate.

3.4.2. A critique of Porter and Offenbacher

Porter and Offenbacher provide extensive evidence that the coefficient to $b(t)$ is always positive and highly significant. However, their empirical results for the US cannot be commented upon in detail simply for the reason that these results are suspect to circularity. Let equation [23b] in sub-section 3.1.2(b) above re-written so that

$$M = \alpha_1 \sqrt[3]{\frac{\sigma^2 b}{r}} \quad \dots [46]$$

where $\alpha_1 = (3/4)^{1/3}$. To see how the possibility of circularity could come about, suppose that, according to Hein (1984), there was an *exogenous* downward shift in the demand for money so that it leads to a reduction in the coefficient α_1 . Again, let equation [42] be re-written so that

$$d = \alpha_2 \sqrt[3]{\frac{\sigma^4 r}{b}} \quad \dots [47]$$

where $\alpha_2 = (1/3)(4/3)^{1/3}$. If the Porter-Offenbacher procedure of estimating the brokerage fee is followed, the brokerage series so derived would show a shift at that point where there was an exogenous downward shift in the demand for money. This is clearly shown if equation [45] is re-arranged to show the coefficients α_1 and α_2 :

$$b(t) = \left[\frac{\alpha_2}{\alpha_1} \right]^{2/3} \sqrt{\frac{\sigma^2 r^2}{t^3}} \quad \dots [48]$$

It is therefore clear that as α_1 changes, $b(t)$ will shift. Thus, it would have been misleading to conclude that the brokerage fee was responsible for the money demand shift, when in fact, it was due to an exogenous shift. The results presented by Porter and Offenbacher therefore force the conclusion that brokerage fees explain the shift in money demand.

To overcome the circularity problem, Porter and Offenbacher (1984, p.63) suggest the inclusion of financial debits as a proxy for the brokerage fee which the Miller-Orr model suggests is inversely related to the brokerage fee - this may be seen if equation [47] is solved for b to give

$$b(d) = \left[\frac{\alpha_2}{d} \right]^3 \sigma^4 r \quad \dots [49]$$

where the superscript (d) on the brokerage fee denotes that this is the brokerage fee derived from a debits equation. So, it is seen that Porter and Simpson include financial debits as an additional explanatory variable in the empirical demand for money equation. However, Hein (1984) has pointed out another weakness in such an approach. When the conventional demand for money is estimated again with financial debits as an additional variable, the coefficient to this variable has negative sign which could lead to further misleading conclusions. Consider writing equations [46] and [47] in

logarithmic form so that

$$\ln M = \ln \alpha_1 + (1/3)\ln b + (2/3)\ln \sigma - (1/3)\ln r$$

and

$$\ln d = \ln \alpha_2 + (1/3)\ln r + (4/3)\ln \sigma - (1/3)\ln b$$

The logarithmic debits equation can then be solved for the brokerage fee so that

$$\ln b = -3\ln d + 3\ln \alpha_2 + \ln r + 4\ln \sigma$$

which is then substituted into the logarithmic demand for money equation to give

$$\ln M = (\ln \alpha_1 + \ln \alpha_2) + 2\ln \sigma - \ln d$$

From this equation, it should be clear that if financial debits are included on the right-hand side of a money demand equation, then the specification suggested by Porter and Offenbacher would say that interest rates have *no effect*, and that the long-run elasticity with regard to financial debits would be unity. Their results reject the above conclusions suggested by the above equation, and it is on these two points that their model has to be rejected. Therefore, the hypothesis that lower brokerage fees cause a shift in money demand remains unverified.

3.4.3. Some tentative suggestions for tackling the problem of brokerage fees

In spite of the efforts of Porter and Offenbacher (1984) to overcome difficulties posed by the paucity of data on brokerage fees, the hypothesis that the demand for money shifted in response to a fall in brokerage fees still remains untested. However, there is an interesting question regarding the comparative performance of different specifications of the demand for money function. As the preceding discussion in sections 3.2 and 3.3. has made it clear, the conventional specifications exhibited a tendency to break down when the sample period was extended beyond the early 1970s. In contrast, a specification that excluded short-run interest rates, but included long-run interest rates, such as that given in Hamburger (1977) showed a tendency to perform very well in dynamic simulations. The main difference was that whilst conventional specifications consistently overpredicted money demand, Hamburger's specification either tended to overpredict or *underpredict* money demand with only relatively small errors. This question was addressed in Hamburger (1984, pp.112-114) who attempted to explain why his specification held up well whilst others had failed.

Hamburger (1984), in explaining why the conventional specification broke down, suggests that financial innovation taking the form of lower transactions costs for short-term assets have tended to increase the *net* rate of return on such assets so that less money was being held. This is a view of financial innovation shared by many researchers on the demand for money, but, unfortunately, Hamburger does not explain exactly why his particular specification, containing long-term instead of short-term interest

rates, did not break down. A highly tentative explanation was put forward by Meyer (1984, pp.122-125) in his discussion of Hamburger (1984) which almost certainly gives the idea that a careful study of the term structure of interest rates may prove to be useful in analysing financial innovation and its effect on the demand for money, and may even turn out to be a fruitful avenue for future empirical research.

Meyer (1984), for the *sake of argument*, introduces a very simple term structure equation relating net long-term interest rates to net short-term interest rates. Such a relationship, in its most basic form, may take the following form:

$$(r_l - b_l) = \alpha(r_s - b_s) \quad \dots [50]$$

where the subscripts, l and s , to the interest rates and 'brokerage fees' indicate long-term and short-term respectively, and α is a constant of proportionality. The above term structure equation is seen by Meyer (1984) to be the link between conventional specifications of the demand for money and that of Hamburger (1977, 1984). In periods prior to the early 1970s, it was a reasonable assumption that transactions costs remained constant (or at least changed in the same proportion) so that the same relationship could exist between *gross* long-term and short-term interest rates, namely that $r_l = \alpha r_s$ without actually having to include transactions costs in the term structure equation. Thus, it explains why both specifications of the demand for money performed equally well in the period prior to the early 1970s, namely that there would not have been much difference if gross long-term interest rates were

included in the empirical demand for money equation as proxies for gross short-term interest rates instead of including the latter rates explicitly.

The picture changes considerably when one relaxes the assumption of constant transactions costs. If it is supposed that short-term transactions costs have fallen relative to long-term transactions costs, then short-term assets would be relatively more attractive to hold so that long-term gross and net interest rates must rise relative to short-term gross interest rates. Therefore, Meyer argues that the explicit inclusion of short-term gross interest rates in the empirical demand for money equation will fail to capture the effects of financial innovation whereas the inclusion of the long-term gross interest rate as a proxy for short-term *net* interest rates will capture such effects, leading to the superior performance of Hamburger's money demand equation in the period after the early 1970s.

Thus, in conclusion, a careful study of the term structure of interest rates may provide some further insights regarding the effects of financial innovation on relative net interest rates, and there is really no need to base such a study on equation [50] above, as there are other term structure equations that are much less restrictive.

3.5. Conclusions

This chapter has examined in some detail the right hand side of the demand for money equation in order to ascertain that financial innovation may be responsible for a reduction in the transactions demand for money. In the first section, it was shown

that if financial innovation was exemplified by lower transactions costs, the demand for money would decline. Given that transactions costs are a linear function of the volume of transactions (as discussed in sub-section 3.1.1(d)), it is possible to argue that a change in the structure of transactions costs is likely to affect the interest-elasticity, giving the impression of a unstable demand for money function. Various ways of capturing the effects of financial innovation were considered. The simplest approach was to use a time-trend variable to represent a steady decline in transactions costs. However, when one considers how the pace of financial innovation has become more rapid in recent years, the use of time-trend variables has to be ruled out. The second approach considered was the inclusion of previous-peak and ratchet variables in empirical demand for money equations. Their inclusion was justified on the grounds that there may exist a threshold level beyond which interest rates may rise, which may then induce investment in cash-management techniques in order to economise on cash balances in face of persistently high opportunity costs, and such investments are irreversible. The empirical evidence presented indicated that a specification including ratchet variables was superior to those that proxied innovations by either previous-peak variables or time-trends. However, the use of ratchet variables reflects the assumption that all innovations were endogenous. It is not too difficult to find examples of exogenous innovations that will also affect the demand for money so that some effort must be devoted to finding ways in which the effects of exogenous innovations may be captured in addition to those of endogenous innovations. A controversial method of explicitly including brokerage fees in the

empirical demand for money was considered. The main essence of the approach was to derive an indirect brokerage fee series by solving for the brokerage fee from the money demand and debits equations. However, such a specification will have circularity problems in that an exogenous shift in the demand for money would also be reflected in a shift in the brokerage fee series so forcing the conclusion that brokerage fees were responsible for the shift in the demand for money. When an attempt was made to overcome the circularity problem by including financial debits as a proxy for the brokerage fee, the interest rate disappeared from the demand for money equation which is inconsistent with the theory of the transactions demand for money. As long as there is a paucity of data on brokerage fees, the problem of capturing adequately the effects of financial innovation on the demand for money are likely to persist. Finally, a potentially promising avenue for future research was briefly considered in which shifts in the term structure of interest rates may reflect a underlying change in relative transactions costs of long- and short-term assets. It is suggested that a careful study of the term structure of interest rates may provide some further insights into the effects of financial innovation on the demand for money.

CHAPTER FOUR
MONEY SUBSTITUTES AND AGGREGATION

Section	4.1. The definition of money.	188
	4.1.1. <i>'A priori' approaches.</i>	188
	(a) <i>Money as a unit of account.</i>	189
	(b) <i>Money as a store of value.</i>	190
	(c) <i>Money as a means of payment or medium of exchange.</i>	191
	(d) <i>The distinction between means of payment and mediums of exchange.</i>	192
	(e) <i>The effects of financial innovation on such a distinction.</i>	193
	(f) <i>Theoretical approaches to the definition of money.</i>	195
	4.1.2. <i>Empirical approaches.</i>	204
	(a) <i>In the context of the causality debate.</i>	208
	(b) <i>Demand for money stability criteria.</i>	210
	4.2. The aggregation problem.	213
	4.2.1. <i>The 'all-or-nothing' approach.</i>	215
	4.2.2. <i>Applications of economic aggregation theory.</i>	221
	(a) <i>Chetty's approach.</i>	222
	(b) <i>The 'dual' problem that wasn't.</i>	231
	(c) <i>Donovan's critique.</i>	240
	(d) <i>The concept of rental prices of money.</i>	241
	(e) <i>Recent developments.</i>	245
	4.3. Divisia aggregation.	246
	4.3.1. <i>The Divisia aggregation procedure.</i>	246
	(a) <i>Basic concepts.</i>	247
	(b) <i>The optimal aggregation procedure.</i>	252
	4.3.2. <i>A critique.</i>	253

4.4. Empirical evidence and analysis.	255
4.4.1. <i>A demand for money study using Divisia aggregates.</i>	255
4.4.2. <i>Other empirical evidence.</i>	259
4.5. Conclusions.	261

CHAPTER FOUR

MONEY SUBSTITUTES AND AGGREGATION

The previous chapter considered the effects of financial innovation on the right hand side of the demand for money equation, namely, on transactions costs, and cash-management techniques. There is, however, still another aspect of financial innovation that needs to be considered. It is the increasing proliferation of new close substitutes for traditional forms of money that has had an important influence on the behaviour of monetary aggregates which are often used as dependent variables on the left hand side of the demand for money equation.

In recent times, concern has been expressed regarding the validity of conventional simple-sum aggregation procedures for the aggregation of monetary assets since such procedures implicitly allocate equal weights to each component of the monetary aggregate, which often implies that there exists perfect substitutability among all assets contained in that aggregate. Especially in the case of broad aggregates, such a presumption is hardly justified. Before addressing the aggregation problem, it is most natural that the definition of money be considered first in section 4.1. This discussion will draw upon the extensive literature of *a priori* and empirical definitions of money.

After having discussed the identification and definition of money, the aggregation problem, in its most fundamental form, involves deciding upon which monetary assets are to be included in

the monetary aggregate. This is usually done with reference to the substitutability among assets by the use of conventional demand for money studies. However, such approaches are rather unsatisfactory because it can lead to an 'all-or-nothing' criterion in the case of simple-sum aggregation because, if on the one hand, such assets were not deemed to be sufficiently close substitutes to those assets contained in the aggregate, then they would be excluded. Thus, there is always a possibility of understating the amount of monetary services which is the information that a monetary aggregate is supposed to provide. On the other hand, if such assets were found to be sufficiently 'close' substitutes, then their inclusion in the monetary aggregate would be 'justified', but only at the peril of overstating the amount of monetary services available. Thus, the second part of section 4.2. considers the derivation of weighted monetary aggregates by the application of economic aggregation theory which fundamentally involves the specification and estimation of utility functions.

In spite of the many attractions that economic aggregation theory has to offer, there still exists an element of arbitrariness in the specification of utility functions and 'budget' constraints and their estimation. To overcome such difficulties, an alternative is proposed which utilises index-number theory. This latter approach has its merits because the construction of quantity indices only depends on the existence of *observable* prices and quantities. The Divisia quantity index has been proposed as the basis on which monetary quantity indices can be constructed. Such an approach is considered in section 4.3. which also presents a critique of monetary quantity indices. One of the biggest problems likely to be faced by

the monetary authorities in the use of Divisia monetary indices is that there is a need to forecast *each separate* quantity and price of each component of the index which may pose some particular difficulties such as information-processing capabilities, and so forth. Section 4.4. considers some empirical evidence which looks at the relative performance of Divisia aggregates *vis-a-vis* conventional sum aggregates.

4.1. The definition of money

Before considering the problem of aggregating over monetary aggregates, it is first necessary to consider some approaches to the definition of money in order to see whether or not any generally-accepted definition of money exists. To attempt an aggregation of money, however defined, before the conceptualisation and definition of money takes place would be tantamount to 'putting the cart before the horse.' Fundamentally, there are two basic approaches to the definition of money, viz: *a priori* and empirical approaches. The former approach strives to arrive at a working definition of money by resorting to *a priori* considerations of its essential functions and qualities, and the latter approach attempts to do the same but by resorting to empirical means such as dual correlation and stability of demand for money criteria.

4.1.1. 'A priori' approaches

Money is usually viewed in terms of its functions, viz: a unit of account, a store of value, a means of payment or medium of exchange, and a standard of deferred payment. But, for the purposes of aggregation over monetary assets, the main emphasis is on money functioning either as a means of payment or medium of

exchange since a monetary aggregate serves to measure the amount of monetary services available. The term 'monetary services' is taken here to refer to the main function of money as a means of payment or medium of exchange, and special care has to be taken to separate this function of money from its store of value function because, as will be seen later, this function is what makes money into an *asset in the conventional sense* which is quite distinct from a *monetary asset*, namely that the latter offers a flow of liquidity or monetary services whereas the former functions as a store of wealth.

The following theoretical discussion is, therefore, organised as follows. Firstly, the function of money as a unit of account and as a store of value will be briefly touched upon. Then, the rest of this sub-section will be devoted to money functioning either as a means of payment or as a medium of exchange, with some reference being made to the effects of financial innovation on the distinction between a means of payment and a medium of exchange.

(a) *Money as a unit of account.* Money functioning as a unit of account is an abstract form of money which serves as a common denominator, in terms of which the exchange value of all other goods and services can be expressed. It has been argued by Brunner and Meltzer (1971, p.787) that the introduction of a unit of account reduces the number of exchange ratios that need to be known to only $N-1$ ratios, where N is the number of commodities in existence in the economy.¹ This may be contrasted with the case of a barter economy in which it is necessary to have a knowledge of the exchange ratio for each separate pair of commodities so that the total number of exchange ratios is equal to

$[(N - 1)N]/2$. Thus, the introduction of a unit of account serves to reduce the costs involved in collecting and processing the information contained in the exchange ratios. Money, as a unit of account, must have the essential quality that each unit is identical to each other in terms of quality.

(b) Money as a store of value. The means of payment represents generalised purchasing power, so that it may be held and act as a store of value or wealth until the point in time at which the individual wishes to exercise his purchasing power. Thus, it could be argued, for analytical purposes in this chapter, that such assets may be seen to have two main attributes, viz: means of payment, and store of wealth. Of course, such attributes can come in varying proportions for each different asset. At the one end of the spectrum, currency possesses the full attribute of means of payment, whereas a very negligible proportion may be a store of wealth attribute.² At the other end of the spectrum, there exists some assets that possess the full attribute of store of wealth (such as equities, or to take an even more extreme example, residential properties), but hardly fulfil the function of a means of payment. Somewhere along the spectrum of assets, there will exist a certain class of assets that are capable of being realised as means of payment at relatively small cost, such as inconvenience, the levying of transactions costs, the loss of interest if no notice of withdrawal is given, and so forth. Whether or not such assets can be readily included in the definition of money is dependent on the relative importance of such conversion costs. Thus, the higher such conversion costs are, the less likely that such assets will be included in the definition of money.

It is precisely the store of value function which is emphasised by Friedman (1964) in his definition of money. He argues that money is a 'temporary abode of purchasing power' because it enables individuals to separate the act of purchase from the act of sale. However, the time period to which the term 'temporary' applies is not fixed, and consequently a range of assets, and not merely the means of payment, may act as temporary abodes of purchasing power. If such assets are to be included in the monetary aggregate, then it is imperative that the store of value function be excluded as far as possible in order to avoid the risk of overstating the amount of 'monetary services' available.

(c) Money as a means of payment or medium of exchange. The unit of account may have a physical counterpart which is money in its more 'concrete' or tangible form. By 'concrete', it is not meant that the money necessarily exists in a physical form (though it may do so), but that ownership of it is capable of being transferred and that there is a supply of it which, to a greater or lesser extent, is capable of being quantified. This is money acting as a means of payment and, as such, money is also a medium of exchange. That is, it is an intermediary that comes between final exchanges and thereby obviates the need for establishing a 'double coincidence of wants' before an exchange can take place. The means of payment is accepted by in return for goods and services because the recipient knows that it can, in turn, be used in exchange for the goods and services that the recipient requires. The essential characteristic of the means of payment is that it is generally acceptable and re-usable almost immediately. Currency or legal tender is an example that immediately comes to mind.

(d) *The distinction between means of payment and mediums of exchange.* So far, it has been assumed that the terms 'means of payment' and 'medium of exchange' can be used interchangeably, as is certainly the case of currency, but there are certain classes of assets in which such terms cannot be used interchangeably. Some researchers have defined the means of payment to be anything that enables goods and services to be acquired without the need to supply other goods and services in exchange. For example, Clower has argued that

'[t]he essential issue here is whether the tender of any given financial instrument permits a buyer to take delivery of a a commodity from a seller. On this criterion, trade credit qualifies as money - trade credit being interpreted to include credit card and overdraft facilities, department store credit and travellers' cheques, as well as commercial paper and book credits.' (1971, p.21)

However, it has been pointed out by Shackle (1971) that while a means of payment is also a medium of exchange, it does not necessarily follow that all mediums of exchange are means of payment. A medium of exchange is anything that enables a transaction to take place in the absence of a 'double coincidence of wants', but the receipt of a medium of exchange does not necessarily mean that it can be used immediately by the recipient in return for other goods and services. Therefore, there is a time-lag involved between the receipt of a medium of exchange and the *effective* settlement of the associated debt. Shackle says that

'[p]ayment has been made when a sale has been

completed. Payment has been made when the creditor has no further claim. Payment is in some sense final.' (1971, p.33)

On this basis, it can be argued on the one hand that currency or legal tender is both a means of payment and a medium of exchange because the time-lag involved between the receipt of the currency and the effective settlement of the associated debt is zero. In other words, the recipient of the currency can use it immediately if desired in exchange for other goods and services. On the other hand, demand deposits at clearing banks operated by cheque are certainly not a means of payment, but rather a medium of exchange since the recipient of a cheque drawn on a demand deposit has to present it for clearing first, and cannot immediately use it in exchange for other goods and services. Thus, there is a time-lag involved between the receipt of the medium of exchange and the effective settlement of debt in which the account of the payer is debited in favour of the payee. Certainly, there are also other assets in existence that fulfil the function of a medium of exchange but not of a means of payment. Typical examples include credit cards, and in more recent times, certain categories of building society deposits.

(e) The effects of financial innovation on such a distinction.

There are good *a priori* grounds to argue that the definition of money, based on its means of payment function, is capable of changing during times of financial innovation. One of the most manifest forms that financial innovation has taken on is technological change in the way payments are being processed. The bank-

customer relationship has undergone somewhat of a revolution in that increasing use of technology is being made to improve the efficiency in which banks relate to their customers at the most basic level of services. Consider the example of the possible widespread use of EFTPOS (Electronic Funds Transfer at Point Of Sale) terminals operated by debit cards as a possible replacement for cheques.³ It has been envisaged that the use of EFTPOS terminals will reduce the time-lag between the receipt of the medium of exchange (in the form of the debit card) and the effective settlement of the associated debt which should take a few minutes rather than days. Thus, an interesting dimension is added to the definition of money. Would the use of EFT technology change the definition of money such that demand deposits operated by debit cards, instead of cheques, could be eligible for a classification as means of payment? Of course, currency is the perfect theoretical construct of a means of payment, but whether or not demand deposits can be regarded as a means of payment in the future owing to EFT technology is essentially dependent on the time-lag involved between the receipt of the medium of exchange and the effective settlement of the debt. Even if this time-lag is only a few minutes, there are good grounds for regarding that time-lag as being negligible, and the distinction between means of payment and mediums of exchange would then collapse.

It must be stressed that the use of automated teller machines (ATMs) as quite distinct from EFTPOS terminals does *not* make demand deposits or even building society deposits means of payments because the act of withdrawing cash is essentially a conversion process from a medium of exchange to a means of

payment. If money is regarded in terms of its function as a store of value, then the existence of ATMs must make demand deposits, and even building society deposits, temporary abodes of purchasing power as a consequence of financial innovation.

(f) *Theoretical approaches to the definition of money.* In the final part of this sub-section, various approaches to the theoretical definition of money will be considered, and if it is deemed that financial innovation will have an effect on such a definition, they will also be discussed. The discussion is organised such that the narrowest definitions of money are considered first before moving on to consider the next component in the definition of money. To help with the discussion, the various theoretical definitions of money put forward by several studies are summarised in Table 4.1. below. The inclusion of assets in the *original* definition of money will be indicated with a tick (✓), whereas the suggested inclusion of additional assets to the original definition of money as a consequence of financial innovation will be indicated by a plus (+).

Currency and demand deposits can both be eligible for inclusion in the *traditional* definition of money on the grounds that they both serve as mediums of exchange since other assets may not serve as a medium of exchange equally well as currency and demand deposits. Pesek and Saving (1967, 1968)

TABLE 4.1: *Summary of theoretical definitions of money and the possible effects of financial innovation on such definitions.*

<i>Study</i>	<i>Included assets</i>					
	<i>C</i>	<i>DD</i>	<i>TD</i>	<i>NBCD</i>	<i>NBFID</i>	<i>ACL</i>
Pesek and Saving (1967, 1968)	✓	✓		+	+	
Morgan (1969)	✓	✓	✓			
Friedman (1964)	✓	✓	✓	✓	+	
Newlyn (1964)	✓	✓	✓	✓	+	
Yeager (1968)	✓	✓	✓	✓		
Gurley and Shaw (1955, 1960)	✓	✓	✓	✓	✓	
Radcliffe Committee (1959)	✓	✓	✓	✓	✓	✓

Notes: *C* = currency, *DD* = demand deposits with clearing banks, *TD* = time deposits with clearing banks, *NBCD* = non-clearing bank deposits, *NBFID* = non-bank financial intermediary deposits, and *ACL* = all credit lines.

A tick (✓) denotes an asset that is included in the original definition of money suggested by that study, whereas a plus (+) denotes those additional assets that could be included in the definition of money as a consequence of financial innovation.

agree with the inclusion of currency and demand deposits in the definition of money and the exclusion of all other forms of assets, but for very different and controversial reasons.

Pesek and Saving attempt to distinguish between money and other assets by using a net wealth principle. Money consists of items used as means of payment which are assets to their holders but are not a liability to others.

Money, Pesek and Saving argue, is a net resource of the community, a constituent part of the net wealth of that community. All money renders services in facilitating the exchange of goods, and thereby promoting the division of labour and increases output and productivity. Money is not a debt of its issuer, but a service-providing product which is produced and sold by the money and banking industry. The criterion used by Pesek and Saving to establish whether an item is money or just a debt is the absence or presence of interest: '[i]n any business transaction, if a loan exists, the lender will demand interest from the borrower: if production and sale exist there will be no such payments.' (1967, p.173) State-issued fiat currency, for example, given the large difference between its exchange value and costs of production (seignorage), is clearly part of the community's net wealth; it is an asset to its holders without being a liability to its producers.

In the case of bank deposits, Pesek and Saving argue that there is a clear theoretical difference between the role of banks as producers of demand deposits transferable by cheque and their role as financial intermediaries borrowing funds at one rate of interest and lending at another. Demand deposits are regarded as a product of the banking industry, sold by the banks for currency or

for financial claims (e.g. government securities), or sold for credit. Bank money, like fiat currency, is seen as being resource-cheap in the sense that the real resources used to produce a unit of bank money are extremely small; in other words, bank money has low costs of production. Because of this, bank money cannot be produced under conditions of free entry into the industry. Production is restricted to a limited number of producers so that the price per unit of bank money is kept appreciably above the cost per unit of producing it. According to Pesek and Saving, the fact that banks do produce and sell demand deposits is quite clear because no interest is given on them:

'If bank money were a debt of the banks and not a product sold by the banks, we would see the borrower (the banks) paying interest to the holders of bank money. If the bank money was purchased for cash we do not see any interest payments: on the contrary, many of us pay service charges to the banks.' (1968, p.184)

Time deposits do bear interest, and are, therefore, a debt of the bank and do not add to the community's net wealth. So according to Pesek and Saving, there is a clear theoretical demarcation between the means of payment and other items, the distinction being whether interest is or is not paid. Thus, on the basis of the criteria proposed by Pesek and Saving, only currency and demand deposits are eligible for inclusion in the definition of money. This is shown in the relevant entry of Table 4.1.

One problem arising from the approach of Pesek and Saving concerns the case of payment of interest on demand deposits. If some demand deposits bear some interest, they are considered by

Pesek and Saving to be joint products, part money and part a liability of the bank issuing them. This problem could be further exacerbated by the payment of implicit interest on demand deposits in order to offset some bank service charges. What proportion of interest-bearing demand deposits constitute money and what proportion debt depends, according to Pesek and Saving, on the ratio of the demand deposit rate (or notional interest rate in the case of implicit interest payments) to the market rate of interest. The 'moneyness' of demand deposits declines as the interest rate paid approaches the market rate of interest, until the point where the deposit equals the market rate whereupon the demand deposit ceases to be money and is wholly debt. But, there still exists the problem of defining the appropriate market interest rate. The absence of presence of interest on an asset does not appear to provide a sufficient means of distinguishing between money and other assets, particularly when some items both circulate as means of payment and pay interest so that the problem is that of deciding the amount of 'moneyness' they comprise.

A further problem with the approach of Pesek and Saving was pointed out by Friedman and Schwartz (1969) who say that the analysis confuses price with quantity and marginal with average concepts. Pesek and Saving argue that if bank demand deposits pay interest at the market rate, their value as money must be zero if there is to be equilibrium on the demand side. That is, the means of payment services provided by the deposits are in effect available as a free good in the sense that no interest has to be foregone in order to enjoy them: their price is zero. But the value to which Pesek and Saving refer must be the *marginal value* of the money

services provided by the deposits and not the average or total value. Though the marginal unit of deposits provides no non-pecuniary services, each of the intramarginal units may well do so as Friedman and Schwartz point out:

'A zero *price* for the transactions services of demand deposits does not mean that the *quantity* of money in the form of demand deposits is zero. Alternatively, a marginal yield of transactions services of zero does not mean that the *average* yield is zero.' (1969, p.5)

This argument is certainly repeated by Laidler (1969) who argues that

'In the first place, it seems wrong to argue, as Pesek and Saving implicitly do, that the introduction of time deposits and other financial assets into an economy has no net effect on its welfare. If this were the case it is hard to see how these assets would ever come into being...it is only from the marginal unit of such assets that there is no net gain, for it is only the marginal unit that is held solely for the interest it bears.' (1969, p.513)

Thus, the criteria proposed by Pesek and Saving have to be rejected. It should also be noted that the problems posed by the application of their criteria are likely to be exacerbated by financial innovation which has produced a bewildering array of interest-bearing deposits that have the ability to circulate as a medium of exchange. This would be true of some deposits offered by non-clearing banks and building societies (which fall into the category of non-bank financial intermediaries). Thus, such assets would have to be incorporated into the definition of money, not to

mention the difficulties of deciding upon an appropriate 'benchmark' market rate of interest by which their 'moneyness' could be measured. This is shown by the pluses in Table 4.1. which excludes time deposits.

The next stage of the discussion is to determine whether time deposits should be included in the definition of money. Morgan (1969) argues that currency and all clearing bank deposits should be included in the money definition. The classification of money is based on responses to excess demand/supply of assets. According to Morgan, it suggests that the essential characteristic of money is that the response to excess supply/demand for it will manifest itself as an excess demand/supply respectively of all other assets which, assuming their prices are flexible, would imply an increase/decrease respectively in the prices of these assets. Morgan suggests that two conditions are necessary for this response to occur. Firstly, the price of the asset must be fixed in terms of the unit of account, so that an excess demand/supply is not reflected in a change in the price of that asset. The second condition is that the asset supply should be exogenous:

'[i]n the sense that the amount issued by any one issuer is not affected by the transactions of any transactor that is not itself an issuer of an asset qualifying as money.'
(1969, p.242)

The strict application of such criteria leads to a definition of money in which currency is only eligible for inclusion as it is not reasonable to assume that bank deposits are exogenous because the volume of such deposits could be changed by the actions of depositors. In order, therefore, to include bank deposits in a

money definition, Morgan is forced to introduce an additional condition, that there exists a mechanism by which the monetary authorities could offset the effects of actions by depositors on the volume of bank deposits. This condition would be met only by banks who keep operational reserves with the central bank, with non-clearing banks keeping their reserves with the clearing banks themselves. Therefore, Morgan's application of the above criteria leads to a definition of money in which all deposits of the clearing banks only are included, and is shown in Table 4.1. However, this definition is crucially dependent on the restrictive assumption that the monetary authorities can, and choose to, control the volume of bank deposits in this way.

As previously seen, the function of money as a store of value leads Friedman (1964) to see money as a temporary abode of purchasing power. Such assets that are to be included in this definition of money must be capable of being converted into the medium of exchange itself at relatively little inconvenience to the holder of such assets. Such assets may include currency, demand and time deposits with the clearing banks, and all non-clearing bank deposits. In view of the discussion in section 4.1.1(e) above, it may be that the deposits of nbfi's may also have to be included in the definition of money as they are more capable now of serving as temporary abodes of purchasing power as a consequence of financial innovation. This is reflected in Table 4.1. by the suggested inclusion of such deposits to the original definition.

There is also another argument for the inclusion of non-clearing bank deposits, but for the exclusion of nbfi deposits which have been advanced by Newlyn (1964), and Newlyn and Bootle

(1978). Newlyn and Bootle identify two criteria for distinguishing a means of payment from other assets. Firstly, an asset used to finance payment is a means of payment if

'a payment that will not involve any change in the asset/liability complex of the public other than that between the payer and payee.' (Newlyn, 1964, p.336)

This is termed as the 'neutrality' criterion. The second criterion is that the payment should leave the aggregate of the asset unchanged.

On the basis of their two criteria, Newlyn and Bootle classify as money currency and bank deposits (including those of non-clearing banks). Currency is seen as a means of payment according to the above two criteria since the transfer of currency in an exchange transaction leaves the total unchanged and has no further repercussions, the effects of the exchange are confined only to the two individual parties concerned. Both bank demand and time deposits also qualify as means of payment since the financing of payments using these deposits would only affect the payer's and payee's individual deposit totals; the total of bank deposits would remain the same, and there would be no changes on the asset side of the bank's balance sheet.

The deposits of nbfi's do not, according to Newlyn and Bootle, satisfy the criteria stated above. If one were to take building societies as forming part of the nbfi category, it is possible to argue that such building societies hold their deposits with the banks. If it is assumed, for example, that building society deposits are withdrawn, this will only involve a transfer of bank deposits from the building society to the individual that is withdrawing the funds. Thus, the aggregate of bank deposits would remain

unchanged whereas the aggregate that includes nbfi deposits would fall. On the basis of the neutrality criterion, it would follow that nbfi deposits are ineligible for inclusion in the definition of money.

However, a major flaw in these criteria ensues when the individual concerned in the above example chooses to deposit his withdrawn funds in another building society account. In that case, the aggregate containing nbfi deposits would remain unchanged, and therefore, nbfi deposits would now be eligible for inclusion in the definition of money. Newlyn and Bootle regard such instances as being insignificant, but it is very doubtful indeed that such a view could persist in times of rapid financial innovation. In recent years, there has been an increased proliferation of financial instruments offered by the nbfi's with interest and withdrawal options in varying combinations. It is surely reasonable to see that, in view of increased competition among the nbfi's for deposits, there would be an increasing tendency for funds to be transferred among nbfi deposits in response to competitive bidding for funds. Thus such transfers would now have to be seen as being more significant such that one would have to seriously contemplate the inclusion of nbfi deposits in the definition of money. Thus, Table 4.1. shows the suggested inclusion of nbfi deposits in the original definition of money proposed by Newlyn and Bootle.

Yet another argument for limiting the definition of money so as to include currency and all bank deposits was introduced by Yeager (1968). Yeager argues that there is an asymmetry in the outcome of changes in the public's asset preference between assets used as means of payment and non-money assets. It is this asymmetry which can be used to distinguish between money and

other assets. The means of payment does not have a specific market of its own, and its accounting price is set permanently to unity. As individuals can change the holdings of the means of payment by adjusting their expenditure relative to income, any general excess demand or supply will be felt as a deficiency of demand or excess of demand respectively for other goods and services. An excess demand/supply for the means of payment thus has widespread repercussions, affecting prices in the economy.

With non-money assets, on the other hand, individuals can only change their holdings by entering the specific market in which that asset is traded and then either purchasing or selling it. In the case of a market-clearing non-money asset market, the main impact of excess demand or supply for the asset would, according to Yeager, be largely confined to that specific asset market, producing a change in price or supply of that asset. In the case of non-money asset traded in a market which does not clear, the excess demand or supply would be diverted to other markets, but these repercussions are likely neither to be widespread or substantial. Thus, Yeager considers that an excess demand/supply for the means of payment asset has widespread repercussions in a way that an excess demand/supply for anything else does not have.

Yeager's approach, however, does not provide a sharp line of demarcation between a means of payment and other assets. It should be seen that portfolio balance theory shows that adjustments to excess demand in one market are bound to have repercussions in other asset markets. Thus, the difference between means of payment and other assets would appear to be a matter of degree rather than of kind. At least, the general effects postulated by

Yeager can flow from currency and bank deposits.

The final stage of the discussion concerns those arguments that favour a broadening of the definition of money to include other assets other than currency and bank deposits. In recent times, the role of non-bank financial intermediaries has grown increasingly important such that there existed serious problems for the monetary authorities because of the widespread practice of liability management as a means of circumventing restrictive monetary arrangements (see Chapter Two). According to Gurley and Shaw (1955, 1960), it would mean that the traditional demand for money was now getting more interest-elastic than formerly. Thus Gurley and Shaw favour a broad definition of money which incorporates deposits held with nbfi's. This approach does have its merits because by broadening the scope of monetary aggregates, one internalises substitution effects amongst a wider range of assets. This aspect is discussed further in sections 4.2. and 4.3. below. Finally, a very broad definition of money is based on the Radcliffe Committee's concept of liquidity which is expressed as follows:

'A decision to spend depends not simply on whether a would-be-spender has cash or "money in the bank" although that maximum liquidity is obviously the most favourable springboard. There is the alternative of raising funds by selling an asset or by borrowing; and the prospect of a cash flow from future sales of a product both encourages committment beyond immediately available cash and makes borrowing easier.' (1959, p.132)

Such concept would have led to the inclusion of all lines of credit in the definition of money, but the major problem is that some of

it is unquantifiable, and probably of little practical use from the viewpoint of the monetary authorities.

What can be concluded from this survey of theoretical definitions of money? First and foremost, it is clear that there exists no clear consensus on what constitutes an acceptable definition of money, as Table 4.1. makes abundantly clear. It seems that the *traditional* definition of money would comprise mostly of currency, and demand and time deposits held with the clearing banks. However, as has already been pointed out on several occasions in the preceding paragraphs, there are good *a priori* grounds for supposing that financial innovation will lead to a change in the definition of money. From Table 4.1., it would seem that a suggested *modern* definition of money would incorporate certain deposits held with the non-clearing banks and non-bank financial intermediaries. By 'certain deposits', it is meant that only those deposits that are capable of serving as a medium of exchange should be included in the definition of money, such as in the case of the aggregate M2 recently introduced by the Bank of England (discussed in section 4.6. below), and not just every financial instrument that can be put under the regulatory net.

4.1.2. Empirical approaches

In this sub-section, an alternative approach to the theoretical definition of money is considered briefly. It is not intended here to embark upon a full and comprehensive survey of the empirical literature regarding the definition of money because, to date, several empirical approaches have been shown to be methodologically unsound in that they attempt to arrive at a working definition of money *before* any form of conceptualisation can be made regarding

the meaning of money. Two main empirical approaches to the definition of money are considered here in two contexts, viz: the causality debate and demand for money stability. Fundamentally, whilst these two approaches attempt to contribute to the causality debate and the debate on the stability of the demand for money respectively, it is quite common to see the definition of money by 'best' results as a by-product of such studies.

(a) *In context of the causality debate.* In a well-known study of the relative performance of a simple Keynesian macroeconomic model with a simple monetarist model, Friedman and Meiselman (1963) suggested that 'the precise empirical definition of variables should be selected so as to put the theory in question in its best light.' (p.181) In outline, the study was mainly concerned about the causality between money and nominal GNP, and Friedman and Meiselman regressed the level of consumption (representing endogenous income) on the stock of money and the level of autonomous expenditure over a long period of history, interpreting the coefficients as money and expenditure multipliers. The regression equation took on the following form:

$$C = \alpha + VM + kA \quad \dots [1]$$

where C denotes the level of consumption, M denotes the stock of money, A denotes autonomous expenditure, α denotes a constant term, and V and k denote the money and expenditure multipliers respectively. The product of the regression was that Friedman and Meiselman could obtain the partial correlations between C and M , and C and A taking account of the correlation between M and A . The results suggested that the correlation between C and M is

greater than that between *C* and *A* with the single exception of the period 1929–1936.

In order to select the 'best' definition of money which would give the best form of corroboration for the results of Friedman and Meiselman, two criteria were used. The first criterion was that money was to be that aggregate which had the highest correlation with money income, and the second criterion was that income must be more highly correlated with the aggregate than with each component of the aggregate. Friedman and Meiselman took into account only three components of the money supply (currency, demand deposits, and time deposits) and tried to determine whether time deposits should be included in the money aggregate. If the aggregate inclusive of time deposits were more highly correlated with income than that excluding it, and if the correlation between time deposits alone and income were less than the correlation with the aggregate, then time deposits would be considered to be a close substitute for other components and would have to be included in the aggregate selected. The choice fell on the broad aggregate embodying all three components. This aggregate was then used to show that there was a close and consistent relation between the stock of money and income (or aggregate consumption).⁴

This approach was subjected to criticism by DePrano and Mayer (1965) who argued that the selection of data in this manner resulted in the circularity of argument: 'this practice of using the same data or roughly similar data, both to choose the definitions of variables (the definition being, of course, really part of the overall hypothesis) and to test the hypothesis is particularly suspect.' (p.532) This reiterates the comment made at the beginning of this

sub-section that such approaches are methodologically unsound in that it attempts to find the 'best' definition of money before actually conceptualising the meaning of money. The correct procedure, of course, would be to arrive at some preconceptions about the definition of money, and then to use data based on this preconception to test any hypotheses.⁵

(b) Stability of demand for money criteria. An alternative approach to the empirical definition of money is to use demand for money stability criteria. As previously seen in the context of the causality debate, this approach seeks to put the theory of the demand for money in its best light by selecting the most appropriate empirical definitions of money. The notion of a stable demand for money is usually seen in terms of a monetary aggregate whose real value bears a relatively stable relation to a small number of variables, such as wealth or income, interest rates, and the rate of change in the prices.⁶ Meltzer (1963) took this issue even further when he suggested that

'[t]he problem is that of defining money so that a stable demand function can be shown to have existed under differing institutional arrangements, changes in social and political environment, and changes in economic conditions, or to explain the effects of such changes on the function.' (p.222)

The main difficulty inherent in such an approach is how to define stability. As previously discussed in Chapter One, the term 'stability' is capable of having different interpretations. It is often the case in the empirical literature using traditional econometric methodology that the concept of stability is narrowly defined, that

is, it refers to *parametric* stability. Thus, some of the statistical tests for stability are geared specifically for parametric stability over different sample periods. Some tests for parametric stability are now discussed.

The first, and most naive, approach is to base one's conclusions about the stability of the demand function on the basis of R^2 statistics alone. Strictly speaking, such statistics only measure the 'goodness of fit' of a particular specification as the sample period changes, and are not indicative of relative performance over different specifications using different dependent variables. This point was argued by Goldfeld (1973, pp.585-589) who demonstrated that constraining the income elasticity to some specified value led to an increase in the R^2 statistics. Basically, his approach involved estimating an unconstrained form of the empirical demand for money function, and the resulting estimate of the income-elasticity, say, $\hat{\beta}$ was then used as a basis for constraining the value of the income elasticity such that the dependent variable was $(M/\hat{\beta}Y)$.⁷ The constrained form of the demand for money function produced essentially the same results as the unconstrained form, except that the R^2 statistic changed. Given that both forms were estimated over the same sample period, one would have expected the R^2 statistics to remain unchanged. Thus, at best, the R^2 statistics should not be used as the sole criterion for selecting the 'best' specifications from a range of specifications using different dependent variables.

The second possible approach that could be utilised is to examine the behaviour of the estimated coefficients as the sample period is changed. What may have turned out to be a set of

theoretically-plausible estimated coefficients could change quite suddenly as the period changes so that one is faced with a set of estimated coefficients that is totally at variance with theoretical considerations. This problem is particularly apparent in those specifications that incorporate a partial adjustment process for money balances. There even have been recorded instances when the coefficient to the lagged dependent variable has exceeded unity so that one gets a set of 'nonsense' long-run elasticities which give the impression that the results are in total disagreement with theoretical predictions regarding the direction of change.⁸ A formal test procedure is sometimes used which involves splitting the data into sub-periods, and the null hypothesis of parameteric stability can then be tested.⁹

The final approach, as previously discussed in the last chapter, is to run some dynamic simulations to determine whether or not a specification continues to 'behave well' in that it does not systematically make forecasting errors. The basic approach is to run a simulation of the estimated demand for money function for the post-sample period. This criterion uses mean errors and root mean square errors, and the 'best' specification may be chosen on those errors that have the smallest magnitude.

In a survey of the empirical literature on the demand for money prior to the 1970s, Boorman (1985) concluded that '...the evidence supporting the existence of a reasonably stable demand for money function would seem to be overwhelming.' (p.79) Given such a conclusion that the demand for money function is reasonably stable for most of the monetary aggregates, this criteria would be rather inconclusive in pointing towards the 'best' definition of

money. It has been suggested by Laidler (1969, pp.523-524) that, in spite of the inconclusiveness of the empirical evidence prior to the 1970s, the 'best' definition of money should be that aggregate which was the easiest to control. However, it did become very apparent that the demand for money function had broken down during the 1970s, and one should now seriously contemplate the possibility that the definition of money by best results is ceasing to be a feasible option.¹⁰

4.2. The aggregation problem

Having discussed the definition of money which was favoured to include certain deposits of non-clearing banks and nbfi's in addition to the traditional definition of money, the main problem is now that of aggregating over various components to form a monetary aggregate which should convey information regarding the amount of monetary services available in that aggregate. The main difficulty is to try and measure the degree of 'moneyness' in an asset. As previously argued at the beginning of this chapter, it is useful for present purposes to view each asset as fulfilling two main functions (or having two main attributes), viz: monetary and store-of-value or wealth functions. The 'moneyness' of an asset depends primarily on the relative importance of its monetary function to its wealth function. The monetary function of an asset is essentially to provide a flow of liquidity or monetary services whereas the wealth function may be seen to provide an abode of purchasing power that would be capable of attracting a rate of return to the holder of the asset. It should be especially noted that the term 'abode of purchasing power' used here is quite distinct

from the term 'temporary abode of purchasing power' as used by Friedman (1964). The latter term, as previously seen in sub-section 4.1.1(b) refers to a special class of assets that would be capable of being converted into the medium of exchange itself at relatively inconvenience or cost, whereas the former term is specifically designed as a *generic* term to refer to an even wider range of assets (from currency to the most illiquid assets) that would be capable of being converted into the medium of exchange itself, but there exists a class of assets which do not fall within the category of 'temporary abodes of purchasing power' because of the sheer inconvenience and high cost involved in such conversions. Of course, each asset can fulfill both functions but in varying proportions. The aggregation problem is stated as follows: it is desired to aggregate over a variety of assets that are capable of fulfilling the monetary function to a more or less significant extent such that only the monetary component of each asset is only included in the final monetary aggregate.

There are two main approaches to the aggregation of monetary assets. The first one is based on traditional demand for money studies which use different interest rates on alternative assets in order to arrive at a set of interest cross-elasticities. The main idea behind this approach is to use the cross-elasticity estimates as a basis for deciding upon which assets can be regarded as sufficiently close substitutes for money; if such assets are deemed to be 'close' substitutes for money, then these assets would be regarded as being eligible for inclusion in the definition of money. However, as will be seen below, this leads to an 'all-or-nothing' approach so that certain assets could either be included or excluded from the

monetary aggregate, but only at the peril of overstating or understating respectively the amount of monetary services available. Thus, the second approach is considered which applies economic aggregation theory to the aggregation of monetary assets. This approach involves specifying a consumer utility function and a 'budget' constraint, and the constrained optimisation problem is solved to obtain a set of asset demand equations whose parameters have to be estimated in order to arrive at the parameters of the utility function. In effect, this approach is equivalent to deriving a *weighted* monetary aggregate whose weights depend on the 'moneyness' of each asset included.

4.2.1. The 'all-or-nothing' approach

In order to measure the degree of substitutability among alternative liquid assets, the most common approach that has been utilised is to specify and estimate a demand function for narrowly defined money (say, currency and demand deposits) with the rates of return on one or more 'near-money' assets, plus an income or wealth measure, as explanatory variables. Interest cross-elasticity estimates, derived from the regression coefficients, are then used as a measure of the degree of substitutability among alternative liquid assets. There is, however, no intention here of embarking on a full survey of the empirical literature in this area as such a survey has already been undertaken by Feige and Pearce (1977). It will be sufficient for present purposes to point out some of the pitfalls, apart from those general methodological problems of demand for money studies, inherent in such an approach.

Firstly, the use of simple-sum monetary aggregates as a dependent variable in a demand for money study is highly suggestive

of the implicit assumption that the components of that aggregate are perfect substitutes for each other. This is so because the simple-sum aggregation procedure allocates equal weights to each component of the aggregate. In the particular case of narrowly defined monetary aggregates that only include currency and demand deposits, such an assumption of perfect substitutability between currency and demand deposits would appear to be only justified by the additional assumption that the rate of interest on demand deposits is identical to zero. This is a commonly-used device of doing away with some of the problems posed by money having its 'own' rate; in other words, the 'own' rate of money is quietly disregarded from the analysis, or the observations are simply relegated to a footnote. However, there are indeed some serious grounds for questioning such an assumption because of what may be termed as 'implicit' interest payments on demand deposits as a means in which, say, bank service charges could be partly or wholly offset, or as a means whereby additional benefits of being a customer of the bank could be passed on in the form of, say, concessionary interest rates on overdrafts or loans. It is particularly important to recognise the importance of 'own' rates because they are capable of affecting, to a greater or lesser extent, the degree of substitutability between money and 'near-money' assets. In order to find a proxy for the own rate, some studies have used the negative of the ratio of service charges to demand deposit totals, and others have experimented with the ratio of the differential between operating costs of the bank attributable to demand deposits and service charges to the ratio of demand deposit totals. It has, however, been argued by Boyd (1976) that such proxies are

determined by the forces of demand and supply for demand deposits, and the exogeneity of such variables would be put into considerable doubt since a condition for unbiased OLS (ordinary least squares) estimates is that the explanatory variables should be exogenously determined.

Digressing briefly from the discussion, it should also be noted that as a monetary aggregate is gradually broadened to include, say, time deposits, it becomes even more important to include an own rate. Artis and Lewis (1976, pp.150-151) suggest the use of a weighted average of interest rates which is based on the rate of interest for each component of the aggregate, and the weights are determined primarily by the relative value of each component in that aggregate. This is particularly seen as a device for overcoming the problems of multicollinearity (to be discussed below), but the particular interest rate for demand deposits is assumed to be zero, thus avoiding some of the difficult problems of measuring the implicit rate on demand deposits. Artis and Lewis (1976) do admit that '[t]he procedure implies that changes in the service yield and the rebating rates to defray charges on current account are ignored and is subject to further error where the proxy rates give an inefficient reading of changes in the true rates of interest offered.'(pp.150-151)

Returning to the original discussion, once a decision has been reached on a measure of the rate of return on demand deposits, an attempt must then be made to select suitable interest rates on 'near-money' assets so that the estimated coefficients to such variables may serve to measure the degree of substitutability between money and 'near-money' assets. On theoretical grounds,

one would like to include all relevant interest rates such as those on time deposits, savings and loan shares, etc. But the introduction of a large number of rates may pose some serious problems of multicollinearity as interest rates on various assets may be related to each other. The effect of multicollinearity is essentially to increase the sample variance, or equivalently, to reduce the precision of the estimated coefficients, leading to a higher likelihood of misspecification through the discarding of relevant explanatory variables. At best, only a limited number of interest rates could be included, and some studies have experimented with the use of interest-rate differentials as a way of reducing multicollinearity.¹¹

Finally, the most serious problem is the apparent lack of agreement as to what value of cross-elasticity is to be taken as indicating a close substitutability. One can, of course, adopt the price-theoretic convention that those elasticities which have an absolute value of less than unity should be regarded as 'inelastic', and those that are greater than unity, as 'elastic'. It would then seem that the issue of measuring the degree of substitutability between money and 'near-money' assets is a comparatively straightforward one. However, as will be seen below, this convention has not been systematically adhered to.

Commenting on the results of their survey, Feige and Pearce (1977) say that '...linguistic characterizations of empirical findings often convey the impression of dramatically diverse and conflicting results, even when the underlying arithmetic magnitudes described are statistically indistinguishable from one another.' (p.443) The following exchange between Lee and Hamburger is a particularly illustrative example of the 'disagreement' that exists in interpreting

the results. Both Hamburger (1969) and Lee (1967, 1969) use a narrow definition of money which encapsulates currency and demand deposits only, and the dependent variable is per-capita real balances. Lee (1967) included a variety of interest rate variables in his specification and, according to Feige and Pearce (1977, Table 1(b), p.453), the cross-elasticities for savings and loan shares were in the region of -0.371 to -0.637 .¹² On the basis of his evidence, Lee (1967) argued that savings and loan shares were close substitutes for money, in spite of contravening the price-theoretic convention regarding the labelling of elasticities as 'inelastic' or 'elastic'.¹³ In a comment by Hamburger (1969) on Lee (1967), he argued that the use of interest rate *level* rather than differential variables led to lower cross-elasticities, and estimated cross-elasticities for savings and loan shares to be in the range of -0.137 to -0.374 . This was interpreted by Hamburger (1969) as evidence that savings and loan shares were not good substitutes for money, and saw no reason why there should be concern over the effects on monetary policy of the growth of non-bank financial intermediaries.¹⁴ In his reply to Hamburger, Lee (1969) essentially confirmed Hamburger's findings that the use of interest rate level variables led to lower cross-elasticity estimates. For savings and loan deposits, these estimates were in the region of -0.185 to -0.517 . However, Lee continued to maintain that this was still evidence that non-bank financial intermediary deposits were good substitutes for money, thus lending some support the Gurley-Shaw hypothesis.¹⁵ The study by Cagan and Schwartz (1975), designed to test the Gurley-Shaw hypothesis that the demand for money was getting more interest-elastic as a result of an increasing number of money substitutes, found that, on

the contrary, that it was becoming less interest-elastic, which would have implied an even greater impact of monetary policy on economic activity.

In concluding their survey, Feige and Pearce (1977) state that
'...point estimates of cross-elasticities between money and near-monies are surprisingly consistent and display relatively weak substitution relationships. This conclusion emerges despite the fact that the studies surveyed often reveal heated semantic differences, which on the surface give the impression that the studies are grossly at odds with one another....it is difficult to escape the conclusion that there does indeed exist an unacknowledged empirical consensus on the inelasticity of responses of the demand for money to changes in the rates of return on "money substitutes".' (p.463)

However, one should be warned against interpreting the above findings in terms of the insignificance of the effects of financial innovations on either the demand for money or monetary policy. Multicollinearity amongst interest rates, and the incorporation of only the readily available interest rates in empirical demand for money functions might understate the impact of financial changes or factors such as transactions costs, on which there are no readily available data. Feige and Pearce argue that

'[f]uture research on the issue of substitutability will therefore require not only the creation of more relevant data bases, but also a growing attention to institutional detail, which will hopefully enable us to take account of qualitative changes in asset characteristics in addition to

our current measures of quantitative changes in asset holdings.' (p.464)

However, this approach has been criticised by Chetty (1969, p.271) on the grounds that it leads to an 'all-or-nothing' criteria (hence the title of this sub-section) because it is exceedingly difficult to find assets that would be perfect substitutes for currency and demand deposits, and yet there are very few assets that would be totally unrelated to money. Thus, the 'all-or-nothing' approach is capable of producing simple-sum aggregates which may not accurately convey the information regarding the amount of monetary services available. The application of economic theory to the derivation of weighted monetary aggregates is now the subject of the following sub-section.

4.2.2. Applications of economic aggregation theory

This sub-section will discuss some applications of economic aggregation theory to the derivation of economic monetary aggregates which may be regarded primarily as a set of weighted monetary aggregates whose weights are dependent upon the degree of 'moneyness' that each component asset has. Fundamentally, this approach involves the explicit specification of a utility function and a 'budget' constraint. The constrained optimisation problem is then solved to derive a set of asset demand functions which form the basis of a regression model whereby the parameters can be estimated. The estimated coefficients form the basis of the weights used in aggregator functions.¹⁶ The literature is technically demanding, and the discussion is organised as follows.

Firstly, the seminal work of Chetty (1969) is considered, and then the study by Moroney and Wilbratte (1976) is discussed. The

latter study is mainly based on concepts used by Chetty, but supposedly uses 'duality theory' in which the utility function is used as a 'technological transactions' constraint, and households are assumed to *maximise* wealth subject to this constraint. A critique of both studies is then given which is based mainly on Donovan (1978). Finally, some recent developments in this literature are considered which includes the specification of a GES (generalised elasticity of substitution) utility function by Boughton (1981), and the translog (transcendental logarithmic) utility function by Ewis and Fisher (1984).

(a) *Chetty's model.* In an innovative paper, Chetty (1969) suggested a more direct measure of substitution between financial assets than the interest cross-elasticity discussed in the previous sub-section. His general approach was to regard money as a weighted average of monetary assets with weights being related to the substitution parameters. It will be useful to consider the two-asset case first because some useful insights can be gained into the most fundamental parts of an economic monetary aggregate before going on to consider more general cases. Chetty initially started his analysis by considering just two assets taking the form of money (defined to be currency and demand deposits) and time deposits. In order to derive the asset demand equations, a CES (constant elasticity of substitution) utility function is specified such that

$$U = [\beta_0 M^{-\rho} + \beta_1 T^{-\rho}]^{-1/\rho} \quad \dots [2]$$

where U denotes utility which is ordinal utility, that is, changes in utility are referred to in *relative* terms rather than absolute terms

(as would be in the case of cardinal utility), M denotes money, T denotes time deposits, β_0 and β_1 are both parameters, and ρ is defined such that $\sigma = 1/(1 + \rho)$ where σ is the elasticity of substitution. The parameter β_0 is set equal to unity as ordinal utility is being used in this analysis; this normalisation will not be at the expense of any generality. The next step is to define a two-period 'budget' constraint in which an individual is assumed to allocate M_0 currency units between money and time deposits. Chetty (1969, p.273) assumes that T is the value of time deposits in the *next* period so that, if the current interest rate is taken to be r , the current discounted value of time deposits would therefore be $T/(1 + r)$ in which case the 'budget' constraint is

$$M_0 = M + T/(1 + r) \quad \dots [3]$$

The main object of the constrained optimisation problem is to maximise utility subject to the above constraint. The marginal conditions are thus derived:

$$\frac{\partial U}{\partial M} = \lambda, \quad \frac{\partial U}{\partial T} = \frac{\lambda}{(1 + r)},$$

and

$$M_0 = M + T/(1 + r)$$

where $\partial U/\partial M = (-1/\rho)[U^{(1+\rho)}] \cdot (-\rho\beta_0 M^{-(1+\rho)})$, and similarly for $\partial U/\partial T$. Division of $\partial U/\partial M$ by $\partial U/\partial T$ gives the following expression

$$\frac{\beta_0}{\beta_1} \left[\frac{M}{T} \right]^{-(\rho + 1)} = (1 + r)$$

which, after taking logarithms and some manipulation, gives a regression model:

$$\ln \left[\frac{M}{T} \right] = -\frac{1}{(1 + \rho)} \ln \left[\frac{\beta_1}{\beta_0} \right] + \frac{1}{(1 + \rho)} \ln \left[\frac{1}{1 + r} \right] + \varepsilon$$

... [4]

where ε is a disturbance term that simply has been added to the regression model. Chetty estimated the above equation using OLS for the period 1945–1966, and came up with the following estimated equation:

$$\ln (M/T) = 1.510 + 34.69 \ln [1/(1 + r)]$$

(1.509)

$$R^2 = 0.981, D.W. = 0.57$$

where the number in brackets denotes the standard error. From this equation, it can be deduced that $\sigma = 34.69$, $\rho = -0.971$, and $\beta_1 = 0.957$.

Before proceeding further with the analysis, one would do well to take a little time to consider the concept of economic monetary aggregates. Chetty (1969, p.274) has defined the adjusted monetary aggregate as

$$M_a = [M^{-\hat{\rho}} + \hat{\beta}_1 T^{-\hat{\rho}}]^{-1/\hat{\rho}} \quad \dots [5]$$

where a 'hat' over a parameter denotes an estimate. If money and time deposits were *perfect* substitutes, their elasticity of substitution would be infinitely large such that $\hat{\rho} = -1$. If one defines $\hat{\beta}_1 = \exp(\hat{\alpha}/\hat{\sigma})$ where α denotes the negative constant term in equation [4], then as $\hat{\sigma} \rightarrow \infty$, $\hat{\beta}_1 \rightarrow 1$ in which case the indifference curve will be perfectly linear. This is shown by the line M_0T_0 in Figure 4.1. which is also coincident with the 'budget' constraint. In such a case, a simple-sum aggregation procedure would be perfectly valid to use. Now suppose that one were to find that money and time deposits became less perfect substitutes so that the elasticity of substitution falls. This could be shown in Figure 4.1. by a more convex indifference curve such as M_2T_2 which will approach a L-shape (not shown) as σ approaches zero. Consider the point of equilibrium at which the 'budget' constraint is tangent to the latter indifference curve. Here, the monetary aggregate will contain M_1 of money and T_1 of time deposits, but it can be noted that the same level of utility can be obtained by simply having M_2 of money. In effect, the addition of T_1 units of time deposits to M_1 is equivalent to the addition of $(M_2 - M_1)$ units of money to M_1 . If one were to start with only M_2 units of money, and were to use a simple-sum aggregation procedure which treats money and time deposits as perfect substitutes, it is easy to see that the inclusion of time deposits could lead to the monetary aggregate

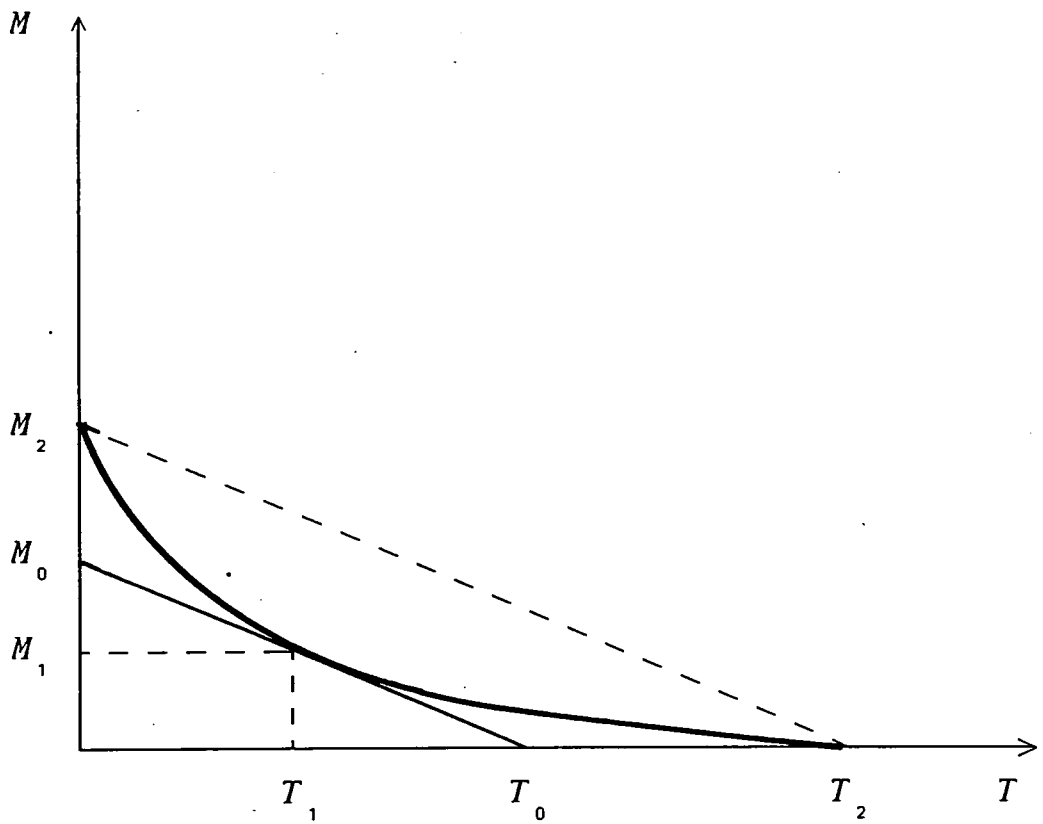


FIGURE 4.1: *The concept of economic aggregation.*

overstating the amount of monetary services available. This is shown in Figure 4.1. by the broken line M_2T_2 which is clearly above the indifference curve. Therefore, the most fundamental concept of an economic monetary aggregate is that a change in its components should leave that aggregate unchanged whereas simple-sum aggregates may give the misleading impression that there has been a change in utility.

Having discussed the basic concept of economic aggregation, the analysis of Chetty (1969) can now be carried further. Given the estimated values of the parameters for the definition of money in equation [5] above, Chetty came up with the following adjusted monetary aggregate:

$$M_a = [M^{-0.971} + 0.957T^{-0.971}]^{-1.03}$$

which Chetty approximates as

$$M_a \approx M + T$$

on the grounds that the substitutability between money and time deposits was so high that it confirmed the contention of Friedman and Meiselman (1963) that money and time deposits can be regarded as 'perfect' substitutes.

The approach is basically the same for the more general case of $(n + 1)$ assets. The second stage of Chetty's analysis was to try and measure the degree of substitutability between money and mutual savings deposits (MS), and between money and savings and loan shares (SL). The main objective of such an analysis was to test the Gurley-Shaw hypothesis that such assets were becoming

even closer substitutes for money so that it would be necessary to include 'near-money' assets in the definition of money. Chetty (1969, pp.276-277) uses what he calls a 'generalised CES utility function' in which the elasticity of substitution between two assets is variable. Boughton (1981, p.377) has pointed out that Chetty's term 'generalised CES utility function' is not quite accurate in that it actually refers to the case in which the elasticity of substitution between any two assets is constant; a condition that Chetty sought to avoid.¹⁸ Boughton therefore suggested the term 'variable elasticity of substitution' (VES) utility function. The VES utility function specified by Chetty is

$$U = \left[\sum_{i=0}^n \beta_i X_i^{-\rho_i} \right]^{-1/\rho_0} \quad \dots [6]$$

where X_i denotes the i th asset ($i = 0, \dots, n$). Asset X_0 is, of course, money with parameter β_0 set equal to unity. The above utility function is maximised subject to the following 'budget' constraint:

$$M_0 = f(Y, r_0, \dots, r_n) = \sum_{i=0}^n \frac{X_i}{(1+r_i)} \quad \dots [7]$$

which lead to the following marginal conditions:

$$\frac{\partial U}{\partial X_i} = (-1/\rho_0) [U^{(\rho_0+1)}] \cdot (-\rho_i \beta_i X_i^{-(\rho_i+1)}) = \frac{\lambda}{(1+r_i)}$$

for $i = 0, 1, \dots, n$,

and

$$M_0 = f(Y, r_0, \dots, r_n) = \sum_{i=0}^n \frac{X_i}{(1+r_i)}$$

The system of equations contained in the first $N + 1$ marginality conditions is then solved to give a system of N equations:¹⁹

$$\begin{aligned} \ln X_i = \frac{-1}{\rho_i + 1} \ln \left[\frac{\beta_0 \rho_0}{\beta_i \rho_i} \right] - \frac{1}{\rho_i + 1} \ln \left[\frac{1}{1 + r_i} \right] + \frac{1 + \rho_0}{1 + \rho_i} \ln X_0 \\ \text{for } i = 1, 2, \dots, n \end{aligned} \quad \dots [8]$$

which is estimated by Chetty using OLS for four assets, viz: money, time deposits, savings and loan shares, and mutual savings deposits. It can be shown that the elasticity of substitution between the i th and j th assets can be derived from the following expression

$$\frac{1}{\sigma_{ij}} = (1 + \rho_j) + \frac{(\rho_i - \rho_j)}{1 + \varphi_{ij}(X_i/X_j)} \quad \dots [9]$$

where φ_{ij} denotes the marginal rate of substitution, and is given by

$$\varphi_{ij} = \frac{\beta_i \rho_i X_i^{-(\rho_i+1)}}{\beta_j \rho_j X_j^{-(\rho_j+1)}}$$

Chetty was able to find the following values of elasticities of substitution between money and alternative assets: 30.864 for time deposits, 35.461 for savings and loan shares, and 23.310 for mutual savings deposits (1969, p.278). This is interpreted by Chetty as

evidence that these assets are all good substitutes for money, and this leads Chetty to derive the adjusted money stock:

$$M_a = [M^{0.954} + 1.020T^{0.975} + 0.880MS^{0.959} + 0.616SL^{0.981}]^{1.026}$$

which is approximated by

$$M_a \approx M + T + 0.880MS + 0.615SL \quad \dots [10]$$

where all the variables have already been defined previously. This is the basis on which a weighted monetary aggregate can be derived.

Feige and Pearce (1977, p.460) have noted that, in the two-asset model, a low interest cross-elasticity would imply a high elasticity of substitution. In particular, they showed that

$$\sigma_{i,j} = \frac{1 + r_j}{r_j} [\eta_{i,j} - \eta_j] \quad \dots [11]$$

where $\eta_{i,j}$ denotes the interest cross-elasticity between the i th and j th assets, η_j denotes the j th asset's interest own-elasticity, and r_j denotes the rate of return for the j th asset. For example, Feige and Pearce chose the values $\eta_{i,j} = -0.4$, $\eta_j = 1.0$, and $r_j = 0.04$ which will give a value of 36.4 for the elasticity of substitution. Thus Feige and Pearce (1977) claim that this value is 'very close' to the actual elasticities reported by Chetty, and go on to conclude that '...a cross-elasticity representing an inelastic response is consistent with a large elasticity of substitution.'(p.460)

However, a number of technical difficulties were discovered with

Chetty's study, and these are discussed in section 4.2.2(c). Meanwhile, the study of Moroney and Wilbratte (1976) will be considered.

(b) *The 'dual' problem that wasn't.* The study by Moroney and Wilbratte (1976) differs from that of Chetty (1969) in three main aspects. Firstly, they assume that households maximise their wealth subject to, what is called, a 'technological transactions' constraint (Moroney and Wilbratte, 1976, p.183). Secondly, they incorporate permanent income and an explicit dynamic adjustment process in the model, and finally, the range of assets included has been extended beyond those deposits of commercial banks and nbfi's to include short-term government bonds and long-term corporate bonds.

The model of Moroney and Wilbratte assumes that households seek to maximise wealth, which is attributable to a mix of money and various interest-bearing assets, subject to a technological transactions constraint. Wealth is defined to be

$$W_t = \sum_{i=0}^n X_{it} (1 + r_{it}) \quad \dots [12]$$

where W_t denotes wealth in period t , X_{it} ($i = 0, \dots, n$) denotes the nominal value of the i th asset, and r_{it} denotes the effective nominal rate of interest. The asset, X_0 , denotes money whose rate of return is set equal to zero, and the other assets X_i ($i = 1, \dots, n$) are interest-bearing assets. The technological transactions constraint is defined as

$$T_t = \left[\sum_{i=0}^n \beta_{it} X_{it}^{-\rho_i} \right]^{-1/\rho_0} \quad \dots [13]$$

Moroney and Wilbratte (1976, p.185) view T as the anticipated volume of transactions that can be accomplished during a given period with the use of money and its various substitutes. The non-negative parameters, β_{it} , are associated with the productivity of various assets in executing transactions, and, as will be seen below, are capable of varying over time. The ρ_i 's are substitution parameters. It has been suggested by Moroney and Wilbratte (1976, p.186) that the coefficients β_{it} need not remain constant over time, and this is attributed to financial innovations which may enhance the convertibility of assets into money or improvements in the effectiveness of money to facilitate exchange. It is proposed by Moroney and Wilbratte, as a working hypothesis, that the β 's are functions of permanent income:

$$\beta_{it} = \beta_i Y_t^{\theta_i} \quad \text{for } i = 0, \dots, n. \quad \dots [14]$$

where Y_t is permanent income, and the θ 's are parameters not constrained to be equal. This is justified on two grounds. Firstly, permanent income is used as a proxy for wealth. Growth in permanent income and wealth may affect the marginal rate of substitution between money and other assets. Secondly, permanent income is expected to reflect the effects of gradual changes in the transactions demand for money.

The first order marginal conditions for the constrained

optimisation problem in equations [12] and [13] are

$$1 + r_{it} = \mu \left\{ \left[\frac{1}{\rho_0} T_t^{(\rho_0+1)} \right] \left[-\rho_i \beta_i Y_t^{\theta_i} X_{it}^{-(\rho_i+1)} \right] \right\}^{\rho_i}$$

for all t , and $i = 0, \dots, n$

and

$$T_t = \left[\sum_{i=0}^n \beta_i Y_t^{\theta_i} X_{it}^{-(\rho_i+1)} \right]$$

for all t .

The first $N + 1$ first-order marginal condition equations (i.e. not including the technological transactions constraint) are then solved by Moroney and Wilbratte to obtain a set of N 'asset-demand' equations:

$$\ln X_{it} = a_{1i} + a_{2i} \ln Y_t + a_{3i} \ln M_t + a_{4i} \ln g_{it} + \varepsilon_{it} \dots [15]$$

where

$$a_{1i} = [-1/(1 + \rho_i)] \ln(\beta_0 \rho_0 / \beta_i \rho_i),$$

$$a_{2i} = (\theta_0 - \theta_i) / [-(\rho_i + 1)],$$

$$a_{3i} = (\rho_0 + 1) / (\rho_i + 1),$$

$$M_t = X_{0t}$$

$$a_{4i} = -1/(\rho_i + 1)$$

$$g_{it} = 1/(1 + r_{it}),$$

and ε_{it} = a disturbance term.

There were five groups of assets that were included in the study, viz; M = currency and demand deposits, X_1 = commercial bank time and savings deposits, X_2 = U.S. government securities with less than one year to maturity, X_3 = bonds issued by private corporations, and X_4 = an aggregate of savings-and-loan liabilities and mutual savings bank deposits. The data period used was from the last quarter in 1956 to the last quarter of 1970. The above equation was estimated by Moroney and Wilbratte using OLS, and after substantial autocorrelation was detected, the Cochrane-Orcutt correction procedure was applied. The estimated elasticities of substitution are shown in the first column of Table 4.2. which is derived from various tables presented in Moroney and Wilbratte (1976).

One feature that can be noted immediately is that short- and long-term bonds should be included in the definition of money because of the relatively high estimated elasticities of substitution for such assets. However, Moroney and Wilbratte (1976, p.190) have said that the 'asset-demand' equation for savings-and-loan liabilities and mutual savings deposits was pathological in that substantial autocorrelation still existed even after the correction procedure, and the coefficient to the interest rate variable was of the wrong sign, and Moroney and Wilbratte say that '[f]or reasons that are not clear, savings-and-loan shares plus mutual savings deposits do not seem compatible with the wealth maximizing framework apparently suitable for analysing the other assets.'(1976, p.190) One possible reason that can be offered here for this result is that the simple-sum aggregation of savings-and-loan shares with mutual savings deposits may not be entirely warranted since this would have

TABLE 4.2: *Estimated elasticities of substitution between money and alternative assets derived from the study by Moroney and Wilbratte (1976).*

<i>Asset</i>	<i>Estimated elasticities of substitution based on equation</i>		
	[15]	[17]	[19]
Time and savings deposits.	8.59	22.48	24.84
Short-term govt. securities with less than 1 year to maturity.	28.57	23.96	34.61
Long-term private corporation bonds.	34.26	27.53	34.86
Savings-and-loan liabilities + mutual savings deposits.	2.79	19.12	22.60

Source: Moroney and Wilbratte (1976), Tables 1, 2, and 3, pp.189, 193, and 194 respectively.

implied that both of these assets had an identical elasticity of substitution. Such a restriction is likely not to be supported by the data, and Moroney and Wilbratte should have included such assets separately.

Another particularly serious problem with the approach is the alleged duality which, Moroney and Wilbratte (1976, p.183) argue, would lead to identical asset-demand equations as those derived in the primal problem (such as that of Chetty (1969)). This would only have been true if the first $N + 1$ first order marginal conditions were solved in both the primal and dual problem, but this would not hold true if *all* (i.e. $N + 2$) first-order marginal conditions were solved to derive the $N + 1$ asset demand equations.²¹ This problem also exists with Chetty (1969). The correct asset demand equations should have included wealth or the anticipated volume of transactions respectively as additional arguments in Chetty (1969) and Moroney and Wilbratte (1976) respectively.

The last two columns of Table 4.2. report alternative estimates of elasticities of substitution. In order to test for homotheticity of the technological transactions constraint, it was necessary to test statistically the null hypothesis that all the ρ 's in equation [12] above were insignificantly different from each other.²² Moroney and Wilbratte were unable to reject this hypothesis, and therefore proceeded to carry out the next stage of their study by using a generalised CES function which is similar to that one defined in footnote 18 of this chapter, and is given below:

$$T = \left[\sum_{i=0}^n \beta_i Y_t^{\theta_i} X_{it}^{-\rho} \right]^{-1/\rho} \quad \dots [16]$$

and the elasticity of substitution would simply be $\sigma = 1/(1 + \rho)$. Wealth is then maximised subject to this constraint to obtain a set of alternative 'asset demand' equations;

$$\ln \left[\frac{M_t}{X_{it}} \right] = b_{1i} + b_{2i} \ln Y_t + b_{3i} \ln g_{it} + v_{it} \quad \dots [17]$$

where

$$b_{1i} = \sigma_i \ln (\beta_0 / \beta_i)$$

$$b_{2i} = \sigma_i (\theta_0 - \theta_i)$$

and

$$b_{3i} = \sigma_i$$

The above system of 'asset-demand' equations were estimated using OLS and the estimates were corrected for autocorrelation. The estimated elasticities of substitution are shown in the second column of Table 4.2. It will be seen from equation [17] that the elasticity of substitution estimate is derivable directly from the coefficient b_{3i} so that they were able to provide some statistical tests regarding the properties of the estimated elasticities of substitution. They find that they are all significantly different from zero, and that they are not able to reject the null hypothesis that the elasticities of substitution are insignificantly different from each other. Moroney and Wilbratte (1976, p.193) concede that this is 'an unusual result' since they contradict a widely held view that short-term assets are more substitutable than long-term assets for money. The results still

confirm that both short- and long-term bonds should be included in the definition of money.

It will have become apparent in the discussions of empirical demand for money studies that it is necessary to incorporate an explicit adjustment process into the demand for money equation when quarterly data is being used. Moroney and Wilbratte (1976, p.191) do the essentially same thing by specifying an adjustment process of the form:

$$\ln\left[\frac{M_t}{X_{it}}\right] - \ln\left[\frac{M_{t-1}}{X_{i,t-1}}\right] = \delta_i \left\{ \ln\left[\frac{M_t^*}{X_{it}}\right] - \ln\left[\frac{M_{t-1}}{X_{i,t-1}}\right] \right\} \dots [18]$$

where δ is the coefficient of adjustment. The adjustment process is then substituted into the system of 'asset demand' equations given in [17]. If the resulting equation were to be estimated by OLS, the existence of a lagged dependent variable may give rise to autocorrelation. An alternative is suggested by Moroney and Wilbratte in which maximum likelihood estimation techniques are used. They lag the system of 'asset demand' equations by one period and multiply it by γ_i . The resulting product is then subtracted from the original system of 'asset demand' equations so that

$$\begin{aligned} \ln m_{it} - \gamma_i \ln m_{i,t-1} &= (1 - \gamma_i)c_{1i} + c_{2i}[\ln Y_t - \gamma_i \ln Y_{t-1}] \\ &+ c_{3i}[\ln g_{it} - \gamma_i \ln g_{i,t-1}] \\ &+ c_{4i}[\ln m_{i,t-1} - \gamma_i \ln m_{i,t-2}] + \epsilon_{it} \\ &\dots [19] \end{aligned}$$

where

$$m_{it} = (M_t / X_{it})$$

$$c_{1i} = \delta_i \sigma_i \ln(\beta_0 / \beta_i)$$

$$c_{2i} = \delta_i \sigma_i (\theta_0 - \theta_i)$$

$$c_{3i} = \delta_i \sigma_i$$

and

$$c_{4i} = (1 - \delta_i)$$

Moroney and Wilbratte (1976, p.194) have reported that the coefficient c_{4i} is insignificantly different from zero for the asset demand equation for savings-and-loan shares plus mutual savings deposits. Thus, they believe that complete adjustment takes place within a quarter for money relative to this class of assets, but takes longer for other assets. Moroney and Wilbratte say that the estimated elasticities of substitution, given in the last column of Table 4.2., show little relative dispersion and are quite similar to those estimates based on equation [17] above. They find it difficult to escape the conclusion that short- and long-term bonds are close substitutes for money, and much closer substitutes than time deposits. But, these results should, of course, be interpreted with a good measure of caution because Moroney and Wilbratte did not follow the correct procedure for deriving the asset demand equations. Another criticism that can be levelled against Moroney and Wilbratte, as well as Chetty (1969), is the specification of wealth and the 'budget' constraint respectively. This is the subject of the following critique by Donovan (1978)

(c) *Donovan's critique.* The main difficulty with the work of Chetty (1969) concerns the nature of the 'budget' constraint given in equation [7] above. The arguments in the utility function fundamentally represent the flows of monetary services yielded by each asset. It is sometimes assumed that the monetary services yielded by each asset are proportional to the average stock of held of that asset. Thus, it is quite correct to regard the demand for an average stock as being equivalent to the demand for the associated service flows. It is important to recognise that, in formulating the 'budget' constraint, the relevant prices *must* be the prices of the service flows yielded by the liquid assets which is referred to by Donovan (1978, p.680) as the rental prices of the assets. The 'budget' constraint should, therefore, express the condition that the sum of the expenditure on the serviced flows yielded by each asset be less than or equal to the total expenditure on monetary services. The concept of rental prices of money has some direct relevance in the construction of Divisia monetary aggregates, and one would do well to consider this concept in sub-section 4.2.2(d) below. It suffices for the time being that the 'budget' constraint specified by Chetty is no more than an accounting identity, namely that the stocks of the assets are constrained to equal wealth. Furthermore, the analysis carried out by Moroney and Wilbratte (1976) would have had more meaning if their wealth function had been re-specified as the expenditure function so that the dual problem noted at the end of footnote 21 to this chapter would make more sense, namely that households seek to *minimise* their expenditure on monetary services with respect to the technological transactions constraint.

Another problem that was also touched upon by Donovan was the possibility of introducing simultaneous equation bias. As noted previously, the studies by Chetty, and Moroney and Wilbratte failed to take into account the last first-order marginal condition when solving the system of equations to derive a system of 'asset-demand' equations. It will have been noticed that the so-called asset demand equations included the demand for money as a function in *every* demand equation (see equations [8], [15], [17], and [19] above). If one were to try to estimate such asset demand equations, it would only be at the peril of inducing simultaneous equation bias which could lead to inconsistent estimates. This problem has apparently been noted by Chetty (1969, p.277) who suggested the two-stage least squares procedure which involves specifying an asset demand equation for money (currency and demand deposits) and then estimating it. The resulting instrumental variable is then used in the second stage of the two-stage least squares procedure. But Chetty refrained from using this procedure, preferring to use OLS on the grounds that income and interest rates may not be strictly exogenous, and claimed that 'the ordinary least squares may be no worse than the two-stage least squares [procedure]...' (1969, p.277) Moroney and Wilbratte (1976) also did not use the two-stage least squares procedure.

(d) *The concept of rental prices of money.* It was previously seen that the correct prices for inclusion in a budget constraint must be able to reflect the rental price of monetary services. This is necessary because the variables of direct interest in a utility function are the service flows yielded by each asset. But, for convenience in collecting empirical data, it is sometimes assumed that the service

flows of an asset is proportional to its average stock so that there sometimes exists a confusion between the two concepts in the literature.

In order to derive the rental prices of money and its substitutes, it is necessary first to make some simplifying assumptions and initial definitions. To keep the notation within manageable proportions, it will be assumed that the factor of proportionality between the stock of an asset and its service flows is equal to unity. Furthermore, define the nominal value of the i th asset in period t as M_{it} , and define the consumer price level in period t as p_t^* . Now, in period t , the real value of the i th asset is given by $m_{it} = (M_{it}/p_t^*)$, that is the real value of the i th asset is derived by deflating the nominal value by the consumer price level.

In order to define the rental price of the i th asset for period t , which is denoted here by π_{it} , it should be noted that the individual must pay p_t^* currency units to be able to hold one unit of a 'basket of real commodities' which is used as the basis for calculating the price index. In period t , the rental price of the i th monetary asset will be equal to p_t^* .

If one were to assume that the price level will not remain constant in the next period, and is expected to rise to p_{t+1}^* in the next period, then the real value of the i th asset held in period t will be expected to decline to M_{it}/p_{t+1}^* . Letting the rate of depreciation be denoted by δ_t , then it can be shown that the rate of depreciation can be derived by solving the following equation for δ_t :

$$\frac{M_{it}}{p_{t+1}^{*e}} = (1 - \delta_t) \frac{M_{it}}{p_t^*} \quad \dots [20]$$

which gives;

$$\delta_t = \frac{p_{t+1}^{*e} - p_t^*}{p_{t+1}^{*e}} \quad \dots [21]$$

Thus, it can be seen that the expected rate of depreciation is analogous to the expected rate of change in the price level. If the nominal rate of interest on a 'benchmark asset' (to be defined in sub-section 4.3.1. below) is denoted by R_t , then the rental price of the i th non-interest bearing asset is defined to be (see Donovan, 1978, p.685, equation 9):

$$\pi_{it} = p_t^* - \frac{(1 - \delta_t)p_{t+1}^{*e}}{(1 + R_t)}$$

If equation [21] is substituted into the preceding equation which then reduces to

$$\pi_{it} = p_t^* - \frac{p_t^*}{(1 + R_t)} \quad \dots [22]$$

Thus, the rental price of the i th non-interest bearing asset is seen to be the price level in period t *minus* the discounted price level. But, when the rate of return on the i th asset, the interpretation of equation [22] will become even clearer. Define the nominal rate of

interest that the i th asset earns in one period as r_{it} . Then equation [20] can be modified such that;

$$\frac{M_{it}(1 + r_{it})}{p_{t+1}^*} = \frac{(1 - \delta_t)M_{it}}{p_t^*} \quad \dots [23]$$

Substituting for δ_t into equation [23] and making all the necessary simplifications, equation [23] will reduce to;

$$\pi_{it} = \frac{p_t^*(R_t - r_{it})}{1 + R_t} \quad \dots [24]$$

which is identical to that equation derived by Donovan (1978, p.686, equation 24) and also to that derived by Barnett (1981, p.197)^{2 3} It will be seen that the differential $(R_t - r_{it})$ may be seen to represent the opportunity cost of holding a liquid asset. In the case of currency, this opportunity cost will be equal to R_t , but will become lower as one obtains a higher rate of return for relatively less liquid assets. Fundamentally, the rental price of the i th asset is the discounted one-period holding cost of the i th asset. This is precisely the prices that should be used in a budget constraint whenever analysing the substitutability between money and alternative assets. One should refrain from specifying budget constraints that are simply accounting identities for the stock of assets. This concept of the rental price of an asset will be used extensively in section 4.3. below, but before departing from this section, an attempt will be made to outline some of the recent literature in the application of economic aggregation theory to the derivation of economic monetary aggregates.

(e) *Recent developments.* The model originally suggested by Chetty (1969) was subjected to further criticism by Boughton (1981). The essential criticism is that Chetty's model is excessively restrictive, but can be generalised by making the utility function homothetic and by relaxing the strong separability restrictions which implied that the determinants of utility were completely independent of each other.

B relaxing the assumption that the marginal rate of substitution between financial assets is independent of income, Boughton proposed a model which modified the strong separability assumption, and restored homotheticity with respect to income.²⁴ He was concerned about the 'temporal constance of relationships' over the period 1953–1975 which witnessed 'frequent and substantial innovations in the market for financial assets.' (1981, p.378) Wishing to take them into account, but without modelling them explicitly, he incorporated some modifications such as an implicit interest rate on demand deposits, filtered data to reduce the effects of common trends, and deflated values of assets. To allow for shifts caused by regulations, a dummy variable was employed.

The substantial finding contradicted Chetty's results. Elasticities of substitution between money (narrowly defined) and other liquid assets were lower than those in Chetty, and 'there has been a significant trend away from money in favor of the substitute assets that is not otherwise explained by the hypothesized relations in the model.' (1981, p.385) The existence of low elasticities of substitution would therefore point to a narrower empirical definition of money.

An experimental work by Ewis and Fisher (1984) used the translog utility function in their study of the demand for money in

the U.S. over the period 1969–1979. The results are of a provisional nature, but are of interest in the context of asset substitutability.

The assets considered were M1 (narrow money), time and savings deposits, short-term Treasury securities, and foreign assets. While, with the exception of foreign assets, no strong substitutability between monetary assets was found, M1 was a weak substitute for domestic assets except time deposits. Evidence of strong substitutability between domestic and foreign assets was also found. Stressing the tentative nature of their results, Ewis and Fisher assert that only the narrow definitions of money would be satisfactory policy targets, and that the low substitutability between domestic financial assets suggests that the simple-summation into broader aggregates would be unwise.

Thus, the results on substitutability of assets are sensitive to model specifications and the use of data. In general, the evidence seems to suggest low substitutability between narrow money and liquid assets. There is unease about high level simple-sum monetary aggregates which implicitly assign to their component assets equal weights.

4.3. Divisia monetary aggregates

4.3.1. The Divisia aggregation procedure

In 1977, the staff of the Federal Reserve Board began to work intensively to identify the definition of money that is most useful to the implementation of monetary policy. One outcome was the construction of a monetary measure using a rigorous application of aggregation theory (discussed in the previous sub-section) and

index-number theory. 'Moneyiness', or a measure of monetary transactions services, is expressed as an index number based on the Divisia quantity index.^{2 5}

(a) *Basic concepts.* The traditional money aggregates, as previously noted, are simple sum totals which implies that their components receive equal weights (of unity) and are thus implicitly considered to be perfect substitutes. The other implicit assumption is that those assets not entering a monetary aggregate have zero substitutability with the assets included. The preceding discussion revealed that there tended to be a low degree of substitutability between money and alternative 'near-money' assets which have been included in very broad monetary aggregates, but it is doubtful if simple-sum aggregation procedures are appropriate.

The basic point is that an aggregation of monetary components, whose object is to try to identify 'money' in terms of the 'flow of services that constitute the output of the economy's monetary-transactions technology'(Barnett and Spindt, 1982, p.4), could be accomplished if one knew the parameters of relevant utility functions. But, such functions are not known and there is an element of arbitrariness in their specification and estimation. Thus, aggregation could be approached on the basis of index-number theory, where there are no unknown parameters, but where prices of component quantities are required in addition to the quantities themselves.

The use of the quantity index dispenses with the use of an aggregator function, but cannot be compiled without both observable prices and quantities. To determine a change in aggregate service flows resulting from changes in component quantities, a quantity

index must have prices as its weights. This reflects the fact that, in equilibrium, prices are proportional to marginal utilities.²⁶

The price of the services of a durable good is its rental price or user cost. Analogously, user costs of financial assets must be derived. According to Barnett *et al* (1981), the user cost is derived which leads to the same expression as given in equation [24] or the tax-adjusted user cost given in footnote 23 to this chapter. But, in computing the monetary quantity index, the user cost formula reduces to

$$\pi_{it} = R_t - r_{it} \quad \dots [25]$$

as $f(R_t, \tau_t, p_t^*) = p_t^*(1 - \tau_t)/(1 + R_t)(1 - \tau_t)$, being independent of the selection of the i th asset, is cancelled out.²⁷

The user cost defined in equation [25] above is the simple differential between the expected maximum available yield, R , on any monetary asset during the holding period and the own rate of return, r , on asset i during that period. Unless services are accrued from a monetary asset i , the asset would not be held. The user costs are thus the equivalent of the price paid for the monetary services. But, it must be stressed that these user costs are *not* the weights used in Divisia monetary aggregates; rather they are the prices on which the weights can be computed.

Before going on to define a suitable quantity index, it is necessary to take a brief look at some of the desirable properties of index numbers. The full discussion is contained in Appendix One to this thesis, but the more salient points can be mentioned here. A quantity index number is said to be *exact* if it exactly equals the

aggregator function whenever the data is consistent with microeconomic maximising behaviour. Since the aggregator function depends only upon quantities, the index number is a quantity index number despite the existence of prices in its formula. Given a quantity index, the corresponding price index then can be computed from Fisher's (1922) weak factor reversal test (see Appendix One).

Although no always-exact index numbers are known, Diewert (1976) has constructed a theory of superlative index numbers. Diewert defines an index number to be *superlative* if it is exact for some aggregator function, f , which can provide a second-order approximation to any linearly homogeneous aggregator function.

Fisher (1922) advocated the following quantity index number, called the Fisher Ideal index:

$$Q_t^F = Q_{t-1}^F \left[\frac{\left[\sum_{i=1}^n \pi_{i,t} m_{i,t} \right] \left[\sum_{i=1}^n \pi_{i,t-1} m_{i,t} \right]}{\left[\sum_{i=1}^n \pi_{i,t-1} m_{i,t} \right] \left[\sum_{i=1}^n \pi_{i,t-1} m_{i,t-1} \right]} \right]^{1/2} \dots [26]$$

Törquinst (1936) advocated the following quantity index number, called the Törquinst-Theil Divisia index:

$$Q_t^D = Q_{t-1}^D \left[\prod_{i=1}^n (m_{i,t}/m_{i,t-1})^{(1/2)(s_{i,t} + s_{i,t-1})} \right] \dots [27]$$

where

$$s_{it} = \pi_{it} m_{it} / \sum_{j=1}^n \pi_{jt} m_{jt}$$

Taking the logarithms of both sides of equation [27], observe that

$$\ln Q_t^D - \ln Q_{t-1}^D = \sum_{i=1}^n s_{it}^* (\ln m_{it} - \ln m_{i,t-1}) \dots [28]$$

where

$$s_{it}^* = (1/2)(s_{it} + s_{i,t-1})$$

Diewert (1976) has proved that both the Fisher Ideal and Törquinst-Theil Divisia indices are 'Diewert-superlative' (see Appendix One). But, Barnett (1980, p.39) has argued that, as a quantity index, the Törquinst-Theil Divisia index is more widely used than the Fisher Ideal index since equation [28] permits an easier interpretation of the index. This equation states that the growth rate of the index is a weighted average of the growth rates of the components. The weights are share contributions of each component to the total value of the services of all components. Because of that 'transparently clear' interpretation, Barnett (1980) advocates the use of the Divisia index to measure the quantity of money at all levels of aggregation.²⁸

Simple-sum aggregates would be the same as Divisia aggregates only if the own rates of return on all component assets

were identical, suggesting that the components would be perfect substitutes. This, however, is not likely to be the case, especially with broad aggregates. If the rate of interest on a component changes, then in the case of the Divisia index, all 'substitution' effects will be internalised (by definition) and the aggregate will not change; it will change only if the rate of interest has an 'income effect' (that is, when the level of utility, or monetary service, changes). A traditional aggregate cannot internalise the effects of substitution following a change in interest rates on a component.²⁹

The weights in the monetary quantity index will change, as interest rates entering the calculations of the weights (and user costs) will indicate. A rise in the general interest rate (indicated by R) will increase weights on liquid or transactions balances and will induce holders to reduce the proportion of such assets in their portfolio. However, if interest rates were to be paid on some hitherto non-interest bearing assets, then this would lower their user costs and increase their holding and they would receive less weight (relatively to remaining non-interest bearing assets, whose holding would decline). Substitution by wealth-holders would terminate only when the marginal return on each asset becomes equal. The weights are likely to be more reliable in the absence of restrictions on interest rates such as ceilings or cartel arrangements.

Comparisons of the simple-sum aggregates with the corresponding Divisia aggregates show that narrow aggregates ($M1$) do not differ a great deal, but that the difference between aggregates increases as they are broadened by the inclusion of new components. Thus, one would expect that the simple-sum totals are higher than the Divisia for the higher level aggregates, for

example M3.³⁰ This is because of the low degree of substitutability between narrowly-defined money and other assets, and thus, as noted previously, broad simple-sum aggregates have a tendency to overstate the amount of monetary services available in relation to Divisia monetary aggregates.

(b) *The optimal aggregation procedure.* Barnett (1982) outlined a procedure to identify the 'best' or the optimal level of monetary aggregation.

A three-stage selection process is proposed by Barnett. The first stage is the selection of 'admissible component groupings' to classify assets in the monetary set into separable component subsets. A measure of substitutability between assets is sought through the properties of utility functions with monetary assets as arguments. Stringent conditions of separability must be met in the selection of the admissible component groups (Barnett, 1982, pp.695-696)

The components of each aggregate must include currency (legal tender), but must exclude non-monetary assets. A prior definition of monetary assets is assumed. 'In attaching a name to an aggregate, such as food or money, a prior definition of the components' domain must be selected.' (Barnett, 1982, p.697) It is in the realm of the conditions for grouping the components that research programmes are incomplete (Barnett, 1982, p.707; Goldfeld, 1982, p.717) Problems are encountered most frequently in trying to isolate sets of assets in aggregates intermediate between the 'narrow' and very broad aggregates.

The second stage proceeds after the selection of admissible asset groupings, and refers to the selection of an aggregation formula. As mentioned previously, Barnett's choice is the Divisia

index. When the aggregation is completed, the result is a hierarchy of aggregates nested about money.

Stage three is concerned with the choice of 'best' aggregate out of the available hierarchy. The theoretical choice falls on the highest-level aggregate, which internalises substitution effects between assets. It is also preferred by virtue of the fact that, in choosing a lower-level aggregate, 'we are omitting factors of production from the economy's transactions technology.' (Barnett, 1982, p.699) Thus, the broadest aggregate does not leave out any important information about the economy's flow of transactions services. Methodologically, the identification of the real-life counterpart of the 'money' of economic theory, by using the three-stage approach, is superior to the empirical approaches discussed earlier and to the alternative approach which follows.

There is, however, an alternative third-stage procedure. This refers to the empirical approach to the selection of the optimal monetary aggregate using the criteria discussed in sub-section 4.1.2., namely that the 'best' definition is selected on the basis of whether it works best in meeting macroeconomic policy objectives. It is suggested that, provided that the first two stages in the selection procedure are completed, the choice by 'best results' is acceptable, for it only completes the sole criterion of the definition of money. The empirical results will be discussed briefly in sub-section 4.4.2. below.

4.3.2. A critique

The index number approach to monetary aggregation, suggested by Barnett and his colleagues, is very appealing and is being experimented with in other countries. Its theoretical

arguments are convincing, but there are some practical issues that must be considered. There is the difficulty of trying to account for financial assets that may yield different services to different holders such as households and firms. Barnett admits that this cannot be done because of the relative paucity of sectoral data. The next problem to consider is the specification of the 'benchmark' asset from which the maximum rate of return could be obtained in order to compute the user costs.

At this stage of development of Divisia monetary aggregates, a general critical evaluation is particularly difficult. But it can be said that the use of Divisia monetary aggregates in demand for money studies will not go some way to tackling some of the persistent specification problems involved in estimating an empirical demand for money function. Furthermore, it is on the supply side that the difficulties might be considerably more forbidding. (Goldfeld, 1982, pp.719-720) For instance, what is the supply function of the very broad Divisia monetary aggregate? How does one approach the control of a monetary quantity index which incorporates numerous assets of financial institutions which exhibit different behavioural characteristics? Barnett's answer to the problem concerning the control of money follows the traditional monetarist prescription: use the monetary base control. This is substantiated by evidence that the long-run monetary base multiplier for Divisia L (the very broad monetary aggregate in the U.S.) is stable (Barnett, 1982, pp.692-693). The final problem to consider is on how to present the concept of Divisia monetary aggregates to the general public.

4.4. Empirical evidence and analysis

4.4.1. A demand for money study using Divisia aggregates

The study by Porter and Offenbacher (1984) which may be recalled for their inclusion of the brokerage fee in the demand for money by a controversial method (discussed in the last chapter) went on to specify a demand for money function in which the arguments are Divisia monetary quantity and user cost indices instead of the conventional-sum aggregates and interest rate opportunity costs.

Porter and Offenbacher estimated two demand for money equations in which the Divisia monetary aggregate was postulated to be a function of a scale variable and the Divisia user cost index. They also estimated a conventional specification of the demand for money using conventional-sum aggregates and interest rate opportunity costs. The sample period used in both cases was from 1959:3 to 1982:2 based on quarterly observations. The sample period was sub-divided into two sub-periods, viz: 1959:3-1974:2 and 1974:3-1982:2 for the purpose of carrying out a Chow test of parameter stability. The regressions were run for a number of definitions of money in order to determine the most appropriate definition of money. This approach, by using demand for money stability criteria, is reminiscent of the empirical approach to the definition of money discussed in sub-section 4.1.1. Tables 4.3. and 4.4. report the results based on quarterly and annual mean errors and RMSEs, along with the relevant Chow F-statistics. The full regression results are too numerous to be fully reported here.

Table 4.3. refers to the use of Divisia aggregates and user costs in the demand for money specification, and Table 4.4. refers

TABLE 4.3: Error Summary - Double Logarithmic Money Demand Equations: Divisia Aggregates and

Divisia User Costs

Aggregate	Quarterly Mean Error	Quarterly RMSE	Annual Mean Error	Annual RMSE	Chow Statistic
M1.....	1.5	3.2	1.9	2.1	4.3
M1 + OVERNIGHTS.....	1.2	3.5	1.6	1.9	5.0
M1 + MMMF2.....	0.4	3.0	0.8	1.6	3.6
M1 + OVERNIGHTS + MMMF2.....	0.2	3.3	0.5	1.4	3.8
M1 + MMMF3.....	0.3	3.0	0.6	1.6	4.1
M1 + OVERNIGHTS + MMMF3.....	0.1	3.2	0.3	1.3	4.5
BROAD TRANS.....	-0.6	3.0	-0.3	1.8	1.1
BROAD TRANS + SAV.....	2.9	6.0	3.4	5.2	6.0
M2.....	2.3	5.7	2.7	5.3	5.1
M3.....	2.4	5.0	2.8	4.8	2.9
L.....	2.1	4.7	2.5	4.5	2.9

Source: Porter and Offenbacher (1984), Table 3-10, p.82.

- Notes: 1. Definitions of variables are given in the notes to Table 4.4. below.
 2. Estimation period: 1959:3-1974:2. Simulation period 1974:3-1982:2.
 3. Degrees of freedom: 5,82. The hypothesis that the coefficients are equal over the 1959:3-1974:2 and the 1974:3-1982:2 sample periods is rejected at the 5% significance level if the Chow statistic is greater than the F-statistic 2.33.

TABLE 4.4: Error Summary - Double Logarithmic Money Demand Equations: Conventional Aggregates and Interest Rate Opportunity Costs

Aggregate	Quarterly Mean Error	Quarterly RMSE	Annual Mean Error	Annual RMSE	Chow Statistic
M1.....	2.0	4.0	2.2	2.8	4.6
M1 + OVERNIGHTS.....	1.2	5.1	1.4	2.4	6.6
M1 + MMMF2.....	-2.1	6.4	-2.3	6.1	10.3
M1 + OVERNIGHTS + MMMF2.....	-2.6	6.8	-2.7	5.6	11.0
M1 + MMMF3.....	-2.7	7.3	-3.1	7.1	9.2
M1 + OVERNIGHTS + MMMF3.....	-3.1	7.5	-3.5	6.6	10.3
BROAD TRANS.....	-2.6	5.4	-3.0	5.3	3.2
BROAD TRANS + SAV.....	-0.4	3.2	-0.3	1.5	2.1
M2.....	-0.9	3.2	-0.8	2.5	2.5
M3.....	0.8	3.4	1.0	3.2	2.5
L.....	0.4	2.0	0.5	1.6	1.4

Source: Porter and Offenbacher (1984), Table 3-14, p.88.

Notes: 1. Estimation period: 1959:3-1974:2. Simulation period 1974:3-1982:2.

2. Degrees of freedom: 5,82. The hypothesis that the coefficients are equal over the 1959:3-1974:2 and the 1974:3-1982:2 sample periods is rejected at the 5% significance level if the Chow statistic is greater than the F-statistic 2.33.

Table continued.....

TABLE 4.4 (continued)

3. The variables are defined as follows:

- M1 = demand deposits at all commercial banks other than those due to domestic banks, the US government, and foreign banks and official institutions, less cash items in the process of collection and Federal Reserve float, plus currency outside the Treasury, Federal Reserve Banks, and the vaults of commercial banks, plus travellers' checks of non-bank issuers.
- M1 + OVERNIGHTS = M1 + overnight RPs + overnight Eurodollars.
- M1 + MMMF2 = M1 + non-institutional money market mutual fund shares.
- M1 + MMMF3 = M1 + all money market mutual fund shares.
- BROAD TRANS = M1 + OVERNIGHTS + MMMF3 + large time deposits + bankers' acceptances + commercial paper + short-term Treasury securities + term RPs + term Eurodollars.
- M2 = M1 plus savings and small-denomination time deposits at all depository institutions, overnight RPs at commercial banks, overnight Eurodollars held by US residents other than banks at Caribbean branches of member banks, MMMFs and MMDAs.
- M3 = M2 plus large-denominated time deposits at all depository institutions and term RPs at commercial banks and savings and loan associations.
- L = M3 plus other liquid assets such as term Eurodollars held by US residents other than banks, bankers' acceptances, commercial paper, Treasury bills and other liquid Treasury securities, and US savings bonds.

to the conventional sum aggregates and interest rates.³¹ A comparison across aggregates in each pair of corresponding tables reveals that the broad transactions Divisia aggregate (BROAD TRANS) yields the best overall results. The summary statistics of predictive ability suggest that this Divisia aggregate can be predicted about as well as any other. Its Chow test F-statistic value of 1.1 indicates that the stability hypothesis cannot be rejected both at the 5 and 1 per cent significance levels.

4.4.2. *Other empirical evidence*

The empirical evidence presented by Barnett (1982) indicates that, by the criteria of macroeconomic performance, the broadest aggregate -Divisia L - is superior to the lower-level aggregates (Barnett, 1982, pp.702-706). This was done by counting the number of occasions on which each particular aggregate performed best of all in each statistical test. Table 4.5. confirms the finding of Barnett (1982).

However, this was somewhat at odds with Cagan's (1982) study, which, using the criterion of minimising the variability of the velocity of circulation about a time period isolated the Divisia M1B aggregate as the 'best'. This is certainly shown as the only best performer for this type of criterion by Barnett (1982) in Table 4.5.

Existing evidence in the USA suggests that, on the whole, Divisia aggregates perform better (have better predictive capacity) than the simple-sum aggregates.

The protagonists of the "Divisia money" have little doubt that the Divisia money targets should replace the present targets. Barnett (1982, pp.706-707) could not conceive of any further potential use for any simple-sum aggregates. The replacement of

TABLE 4.5: *The number of times that each aggregate was best performer in statistical tests in Barnett (1982)*

<i>Component Group</i>	<i>Divisia Aggregates</i>	<i>Simple-sum Aggregates</i>
M1B	1	0
M2	6	4
M3	5	1
L	14	0

Source: Barnett (1982), p.705, Table 5.

Note: Definitions of money (except for M1B) are given in Table 4.4. The definition of M1B is given in Table 3.4. in Chapter Three.

M1 by Divisia M1 could be useful, though he considered that the replacement of higher-level aggregates by the Divisia L would offer the best solution for targeting purposes. 'The components of Divisia L should permanently be defined to include all of the money market. Then new money market assets would be incorporated immediately by the definition of the aggregate.' This provision was made to account for financial innovations, but this may not be such a simple matter in practice (see the discussion in Chapter Two, especially sections 2.1. and 2.3.)

4.5. Conclusions

In this chapter, the definition of money was considered. Two approaches were examined regarding the definition of money. The first approach made use of theory to arrive at *a priori* definitions of money. Of particular interest, it was argued that the effect of financial innovation is that the definition of money would be most likely to be modified in favour of broadening the definition to include liabilities of non-bank financial intermediaries in addition to the traditional definition of money. The empirical approach was then considered which is fundamentally defining money by 'best results'. However, there are some methodological weaknesses inherent in such an approach in that an attempt is being made to define money before it is actually conceptualised on *a priori* grounds. The relative stability of the demand for money prior to the 1970s made it difficult to reach any consensus on the 'best' definition of money, and the breakdown in the demand for money made the definition of money by the demand for money stability criteria an infeasible option. An attempt was made to define

money by measuring the 'moneyness' of various assets by reference to the degree of substitutability between narrow money and alternative 'near-money' assets. The evidence surveyed by Feige and Pearce (1977) would not be inconsistent with the view that savings and loan shares and mutual savings deposits were rather poor substitutes. If one were to choose to include them in simple-sum aggregates, it would have to be at the peril of overstating the amount of monetary services available. Chetty (1969) found evidence to support the Gurley-Shaw hypothesis, and Moroney and Wilbratte (1976) found some rather surprising evidence that short-term government debt and corporate debt should be included in the definition of money. Their results are somewhat suspect owing to their failure to use rental prices of money as defined by Donovan (1978). The use of index-number theory is advocated because it enables one to dispense with the problems of specifying and estimating aggregator functions. The Divisia quantity index was particularly favoured by Barnett in his extensive work for its ease of interpretation, namely that the growth rate of the index was equal to the weighted average of the growth rates of its components. The empirical evidence regarding the relative performance of Divisia monetary aggregates to simple-sum aggregates is highly suggestive that Divisia monetary aggregates are to be preferred, and Barnett (1982) favours the Divisia L aggregate as a target. However, at this stage, there still remain formidable problems, mostly of a practical nature, and a plea is made here for better data sources to enable further research to be undertaken in what proves to be a most promising development in monetary economics.

PART TWO

THE FISHER HYPOTHESIS

C H A P T E R F I V E
THE FISHER HYPOTHESIS IN
HISTORICAL PERSPECTIVE

		264
Section	5.1. Basic concepts.	267
	5.1.1. <i>Nature of Fisher's original formulation.</i>	267
	5.1.2. <i>Limiting values of variables.</i>	272
	5.2. Interest rates and prices.	276
	5.2.1. <i>Henry Thornton.</i>	276
	5.2.2. <i>Knut Wicksell.</i>	280
	5.2.3. <i>Inflationary expectations and the Wicksellian model.</i>	283
	5.3. The transition period.	289
	5.3.1. <i>Fisher's theory of the transition period.</i>	290
	5.3.2. <i>Fisher's empirical analysis.</i>	293
	5.3.3. <i>A Wicksellian perspective.</i>	298
	5.4. Conclusions.	305

CHAPTER FIVE

THE FISHER HYPOTHESIS IN HISTORICAL PERSPECTIVE

As the title of this chapter implies, the main emphasis is putting the Fisher hypothesis in historical perspective with a view to clearing up a few misconceptions that surround the Fisher hypothesis. The main justification for including such a chapter in this thesis is that there is a great deal of misinformation concerning the Fisher hypothesis and the concept of real interest rates. That misinformation mainly stems from the failure of the majority of the empirical literature to recognise Irving Fisher's key distinction between situations in which inflation is fully anticipated (full equilibrium), and those situations in which it is not fully anticipated (the transition period). From this failure, two common misconceptions emerge.

The first misconception concerns the 'neutrality' issue in which it is believed that real interest rates are invariant with respect to changes in expected inflation. Such a presumption is only valid in those situations in which the nominal interest rate adjusts *pari passu* to changes in expected inflation, but would certainly not hold true during the transition period. For example, Gibson (1972, p.855) and Fama (1975, p.271) both assume that the real rate of interest remains constant over time, and it would indeed have been difficult to accept their results since their tests were *conditional* on the above assumption.

The second misconception concerns the equivalence of two types of real interest rates. Gebauer (1986, pp.131-135) has drawn

attention to the importance of distinguishing between the real (deflated) rate of interest on financial assets, and the real rate of return on real (i.e. physical) capital; and has argued that a common misconception is that both real rates are always equivalent. This would only be true in a theoretical world of perfect foresight, absence of risk, and perfect competition since then the rates of return on all assets would be equal. However, with the introduction of imperfect foresight and uncertainty, the real (deflated) interest rate on financial assets and the real rate of return on real capital will both diverge from each other so that it becomes extremely important to state precisely which real rate is being referred to.

In order to overcome these two common misconceptions, the discussion in this chapter will be organised as follows. The first section will introduce the basic concepts involved in the relationship between inflation and nominal interest rates as originally formulated by Irving Fisher in his *Appreciation and Interest* (1896). Then, the second section considers the circumstances in which this relationship would arise by making use of the analyses of Henry Thornton (1802), and Knut Wicksell (1898). Although Wicksell recognised the importance of inflationary expectations, his treatment of them in the context of the relationship between interest rates and prices was not adequate in the sense that a further distinction was not made between nominal and real interest rates. Thus the final sub-section of the second section considers explicitly the effect of moving from an assumption of *static* inflationary expectations to one of perfect foresight in the formation of expectations as based on the analysis of Lutz (1974). Following the rejection of the full equilibrium model by Fisher, the third section considers his investigations of the

inflation–interest rate relationship during the transition period, and the analysis is then put into a Wicksellian perspective to allow for imperfect foresight. Finally, the last section (which mainly serves as a prelude to the theoretical discussion in the next chapter) contends that even if perfect foresight were to be assumed, there still would be a good reason why the nominal interest rate may not adjust *pari passu* to changes in expected inflation, and considers the usefulness of the distinction between the two real rates of interest in the context of Tobin's q -theory of investment.

5.1. Basic concepts

Before any formal analysis can be carried out, it is first necessary to consider the basic concepts involved in the inflation–interest rate relationship. Although the correlation between prices and interest rates was recognised well before Irving Fisher, he was certainly one of the first to treat this relationship in a comprehensive and systematic manner in his *Appreciation and Interest* (1896). In order to avoid any confusion, it will be made explicit at the outset that 'appreciation' is defined to be the appreciation in the purchasing power of money brought about by deflation. The converse also holds true. What follows is a discussion of Fisher's original formulation.

5.1.1. Nature of Fisher's original formulation

Fisher stressed that the effect of appreciation on interest rates depended fundamentally on whether or not the appreciation was foreseen and chose to present the theoretical relation in a model assuming full loan market equilibrium, i.e. perfect foresight. Thus Fisher began by saying:

'We must begin by noting the distinction between a

foreseen and unforeseen change in the value of money...
 At present we wish to discuss what will happen *assuming*
 this foresight to exist.' (1896, p.6)

He then went on to derive the formula linking nominal interest rates to inflation rates.¹ His argument ran along the following lines: suppose that loan contracts can be written either in terms of money or in terms of goods. Let R be the nominal or money interest rate and r be the real or commodity interest rate. It is particularly important to distinguish between situations in which price changes are perfectly anticipated (i.e. full equilibrium) and those in which they are imperfectly anticipated. Thus, letting \dot{p}/p be the actual proportionate change in prices, and π be the expected proportionate change in prices, it follows that in full equilibrium,²

$$\dot{p}/p = \pi \quad \dots [1]$$

By such an assumption, the difficult (and often subjective) question of how expectations are formed can be abstracted from for the time being. If prices rise at the expected proportionate rate π over the year so that what costs one currency unit at the beginning of the year will cost $(1 + \pi)$ currency units at the end of the year. Assuming that at the beginning of the year, one currency unit will buy a basket of commodities, an individual has the option of borrowing, say, one currency unit at money rate R for one year or, alternatively, one basket of commodities at real rate r for a year. It is assumed further that there are zero 'storage costs'³ for the goods so that if the individual chooses the former option, he must pay $(1 + R)$ currency units principal and interest when the loan expires. If

the individual chooses the latter option, he must pay back $(1 + r)$ baskets of commodities which he can purchase at a price of $(1 + \pi)$ currency units per basket when the loan becomes due. This price, when multiplied by the number of baskets required to liquidate the loan, results in a total outlay of $(1 + \pi)(1 + r)$ currency units. Therefore, the costs of liquidating the loans expressed in a common unit of account are $(1 + R)$ and $(1 + \pi)(1 + r)$ currency units respectively. If perfect foresight is assumed so that there exists perfect arbitrage, equilibrium requires that these two money sums be equal, that is⁴

$$(1 + R) = (1 + r)(1 + \pi) \quad \dots [2]$$

This expression states that the maturity values of both loans are the same when expressed in terms of a common unit of account.

The main reason for such a result is that if, say, commodity loans were cheaper than money loans, i.e. the inequality

$$(1 + R) > (1 + r)(1 + \pi) \quad \dots [3]$$

holds, then a profit could be made by borrowing commodities, converting them into currency units to be lent out at the money rate, R , and subsequently using the proceeds received from the maturing money loan to purchase commodities with which to retire the commodity debt. The adjustment towards equality involves an increased demand for commodity loans, and an increased supply of money loans which in turn drives up the real (commodity) interest rate and brings down the nominal (money) interest rate. until equality is restored in equation [2].

Now expanding and re-arranging equation [2] gives the following expression relating the nominal interest rate to the real interest rate and expected inflation:

$$R = r + \pi + r\pi \quad \dots [4]$$

Thus, it can be seen that the nominal interest rate is defined precisely to be the sum of the real interest rate, the expected proportionate change in prices, and their product. A comparison of the above equation with that given by Fisher (1896, p.9, equation 3) may reveal, at first glance, that both equations are inconsistent. However, it is only necessary to consider Fisher's definition of equation [4] which is as follows:

'...The rate of interest in the (relatively) depreciating standard is equal to the sum of three terms, viz., the rate of interest in the appreciating standard, the rate of appreciation itself, and the product of these two elements.' (1896, p.9)

Thus by choosing his words carefully, Fisher made the above definition generally applicable, that is, the above definition is applicable to both deflationary and inflationary situations. So, in a deflationary situation (i.e. as in Fisher's original formulation), it is the *commodity* standard that is depreciating relative to the *monetary* standard, and in an inflationary situation, it is the monetary standard that is depreciating relative to the commodity standard, in which case equation [4] holds.

The main implication of the perfect foresight assumption is that nominal interest rates adjust *pari passu* to changes in inflationary

expectations, and therefore it implies that real interest rates remain constant. This was made explicit by Irving Fisher who wrote in 1930 that

'If men had perfect foresight, they would adjust the money interest rate so as exactly to counterbalance or offset the effect of changes in the price level, thus causing the real interest rate to remain unchanged at the normal rate.'
(1930, pp.414-415)

As will be discussed in Section 5.3. below, the assumption of perfect foresight had to be rejected by Fisher on the grounds that the empirical evidence indicated that the *computed* (i.e. deflated) real interest rate was much more variable than the nominal interest rate.

A further point to note is that Gebauer (1986, p.128) has emphasised that the relationship between nominal interest rates and inflationary expectations should not be construed as a theory of interest *per se*, but as being based on Fisher's 'desire to translate interest rates from one standard (unit of account) into another...'. Writing on this point, Fisher said:

'These rates are mutually connected and our task has been merely to state the law of that connection. We have not attempted the bolder task of explaining the rates themselves. Such an explanation constitutes the "theory of interest" in the more usual sense...' (1896, p.92)

With regard to the above quotation, inspection of the preceding equations should confirm that if one interest rate is known in one standard, then the other interest rate in the other standard can be determined, but the equations themselves do not say how one interest rate in one standard gets determined in the first place. That is,

equation [4] above is not a statement of the *theory* of interest, but a statement of how interest rates in two different standards are *related* in theory. Rather than to describe this theoretical relationship as the 'Fisher relation' (as if this relation *already* existed), it does seem more appropriate to refer to it as the Fisher hypothesis for it sums up neatly the very essence of Irving Fisher's work, namely that it was originally *hypothesised* that nominal interest rates adjusted *pari passu* to changes in expected inflation, but that hypothesis had to be rejected by Fisher himself on the basis of the empirical evidence.⁵

5.1.2. Limiting values of the variables

In this sub-section, the limiting values of the variables are derived, and then the behaviour of the theoretical relationship between the nominal interest rate and expected inflation under perfect foresight is considered. In Chapter VI of *Appreciation and Interest*, Fisher considers some theoretically plausible values for the various variables in equation [2] above. If the nominal interest rate equals the rate of inflation, then equation [2] implies that the real interest rate is zero; and by similar reasoning, if the rate of inflation is greater than the nominal interest rate, then the real interest rate will be negative. Here, the assumption of zero 'storage costs' for goods plays an important role in determining whether or not the real (commodity) interest rate can be negative. Given that there are zero 'storage costs' for commodities, it would follow that rather than to lend out commodities at a loss equivalent to the negative rate of commodity interest, individuals could simply hoard their goods. However, by introducing some perishable commodities into the basket of non-perishable commodities, the assumption of zero 'storage costs' will no longer hold, and even if individuals chose to hoard their

goods, they would still earn a negative rate of interest as manifested in nonnegative 'storage costs'. Since the real interest rate can be of any sign in the presence of nonnegative 'storage costs', it follows that the rate of inflation can either be greater than, or less than, or equal to the nominal rate of interest.

In common parlance, the nominal interest rate is often understood to be the rate of interest in the monetary standard. Since money is a durable good, it would follow that the nominal rate of interest can never be less than zero for reasons similar to those outlined above for real interest rates when zero 'storage costs' are assumed. Thus having established the fact that $R \geq 0$, Fisher (1896, p.30) sought to find a constraint for the values of the real interest rate and the rate of inflation such that the nominal interest rate may always be positive. A slight re-arrangement of equation [2] gives:

$$R = (1 + \pi)(1 + r) - 1$$

and since $R \geq 0$, the following inequality is implied

$$(1 + \pi)(1 + r) \geq 1$$

whence

$$(1 + r) \geq \frac{1}{(1 + \pi)}$$

Now, if the rate of inflation, π , is understood as the rate of appreciation in the commodity standard, and letting δ_m be the rate of depreciation in the monetary standard, it has been shown by

Fisher (1896, p.9) that

$$(1 + \pi) = \frac{1}{(1 - \delta_m)}$$

so that the previous inequality becomes

$$r \geq -\delta_m \quad \dots [5]$$

and

$$R \geq 0 \quad \dots [6]$$

that is, the real rate of interest (in the commodity standard) must be greater than or equal to the rate of *appreciation* (i.e. $-\delta_m$) in the monetary standard which is just what Fisher had said. This 'result will not seem mysterious' (1896, p.31) if the behaviour of the market for loanable funds is considered when the rate of appreciation in the monetary standard is greater than the real (commodity) rate of interest. The real rate of return on hoarded money will exceed the real cost of commodity loans so that there would be scope for profitable arbitrage in the market for loanable funds. Here, individuals will wish to secure more commodity loans for conversion into money balances. This behaviour registers itself as a rise in demand for commodity loans so that it bids up the commodity rate of interest until it is equal to the rate of appreciation in the monetary standard. Then equation [2] will hold once more again.

Having discussed the limiting values of the variables, it is now

time to consider how the variables behave. Given the theoretical world of perfect foresight, the real rate of interest will be homogeneous of degree zero in expected inflation, that is, it does not vary with expected inflation. Thus, the nominal rate of interest has to do most of the work in adjusting to expected inflation. Now taking the total differential of equation [4], and multiplying throughout by $1/d\pi$, this gives

$$\frac{dR}{d\pi} = (1 + r) + (1 + \pi) \left[\frac{dr}{d\pi} \right] \quad \dots [7]$$

The assumption of perfect foresight implies that the nominal (or money) interest rate will adjust *pari passu* to changes in expected inflation so that the real (or commodity) rate of interest remains constant. This does not necessarily mean that the real rate of interest is exogenously determined. Rather, the real rate of interest is endogenous because changes in expected inflation do *potentially* have an effect on the real (deflated) rate, but because of the perfect foresight assumption, the real rate is, by implication, assumed constant. Thus, by setting $dr/d\pi$ equal to zero, the rate of change in the nominal interest rate with respect to inflation is

$$\frac{dR}{d\pi} = (1 + r) \quad \dots [8a]$$

but from equation [2], $(1 + r) = (1 + R)/(1 + \pi)$ so that

$$\frac{dR}{d\pi} = \frac{(1 + R)}{(1 + \pi)} \quad \dots [8b]$$

So, Fisher considered the case when the monetary standard is appreciating relative to the commodity standard, and noted in particular that in order '...to offset appreciation, the [nominal] rate of interest must be lowered slightly more than the rate of appreciation.' (1896, p.9) This conclusion is confirmed by equation [8] for the case when $\pi < R$ which is guaranteed by the constraint $R \geq 0$. It can also be noted that from equation [8a] the value of $dR/d\pi$ is constant given that the real rate of interest is a constant.

5.2. Interest rates and prices

Before the relationship between inflation and interest-rates under perfect foresight, sometimes referred to as the original Fisher hypothesis, can be put into its true perspective, it is first necessary to consider theories concerning the relationship between interest-rates *and* inflation. Prominent amongst the various theories is that one put forward by Henry Thornton in 1802, and re-expressed almost a century later by Knut Wicksell.

5.2.1. Henry Thornton

In his work entitled *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain* (1802), Henry Thornton provided an analysis of the relation between interest rates and inflation. He made an important, but fundamental, distinction between the market or loan rate of interest and the expected yield or marginal rate of profit on new capital projects. He stressed that the two rates were separate and distinct phenomena, the former being a money rate determined in the market for loanable funds and the latter, a real rate determined in the commodity market by the supply

of savings and the investment demand for new capital.

Using this two-rate distinction, Thornton developed a theory regarding the connection between interest rates and inflation. In his view, inflation results from a divergence between the two rates of interest, and substantiated his view by his detailed observations. He argued that, with an interest rate of 5 per cent, which the Bank of England was prevented from exceeding by the laws against usury, and supposing that the borrower could expect to obtain a rate of profit higher than 5 per cent, the volume of credit would expand excessively. The note circulation, he said, would grow faster than was desirable, and the price level would rise. He pointed out that this process would persist for as long as the loan rate remained below the rate of profit. In his own words:

'It seems clear...that when the argued quantity of paper shall have...produced its full effect in raising the price of goods, the temptation to borrow at five per cent will be exactly the same as before... [T]he amount of circulating medium alone will have altered, and it will simply have caused the same goods to pass for a larger quantity of paper.' (1802, pp.255-256)

With this framework in mind, Thornton analysed the inflationary consequences of a central bank policy of pegging the rate of interest. If the Bank of England was constrained by usury laws to a ceiling loan rate of 5 per cent which was below the rate of mercantile profit, the Bank would lose control over the volume of its loans and note issue, both of which would expand indefinitely, producing inflation. With the rate pegged, inflation could continue without limit because there existed no automatic corrective mechanism to bring it to

an end. This reasoning constituted the basis of his criticism both of the usury laws and of the Bank's practice of adhering to a fixed discount rate. He contended that the Bank should control the note issue by keeping its discount rate in line with the rate of profit, and if the usury ceiling should threaten to interfere with the operation of discount rate policy, then, the Bank should resort to other forms of credit-rationing to limit its loans. The essential thing was that the Bank should keep a firm rein on the monetary circulation and control the note issue.

Thornton perceived still another connection between the interest rate and inflation. In 1811 he spoke in the House of Commons on the report of the Bullion Committee. According to an account which he gave of this speech, he said that a person who in 1800 had borrowed £1,000 at 5 per cent and paid back the loan in 1810 would, on account of the price rise which in this period had amounted to between 20 and 30 per cent, have paid a real rate (i.e. deflated) rate of two or three per cent.⁶ Hence the borrower made an extra profit which provided an additional incentive to borrow. In Thornton's own words, the situation was perceived as follows: (referring to himself in the third person)

'The temptation to borrow operated on their minds, as he believed, in the following manner:- they balanced their books once a year, and, on estimating the value of those commodities in which they had invested their borrowed money, they found that value to be continually increasing, so that there was an apparent profit over and above the natural and ordinary profit on mercantile transactions. This apparent profit was nominal, as to persons who

traded on their own capital, but not nominal as to those who traded with borrowed money, the borrower, therefore, derived every year from his trade, not only the common mercantile profit, which would itself somewhat exceed the 5 per cent., interest paid by him for the use of his money, but likewise that extra profit which he had spoken of extra profit was exactly so much additional advantage, derived from the circumstance of his being a trader upon a borrowed capital, and was so much additional temptation to borrow.' (1802, p.336)

Here, Thornton was trying to specify the mechanism through which an inflation premium becomes embodied in market rates. To do this, he could not have hypothesised the existence of inflationary expectations. So, according to him, it is profits and profit predictions rather than inflation predictions *per se* that drives up the equilibrium nominal interest rate. By analogy with the discussion in sub-section 5.1.1. above, Thornton's analysis may be viewed in the context of a market for loanable funds. That is, unexpected inflation initially lowers the realised real rate on money loans below the real yield on capital assets. The outcome is a windfall gains for debtors and windfall losses for creditors. Assuming that debtors and creditors predict future profits by extrapolation from past realised profits, and that they adjust demands and supplies of loans accordingly, the rise in demand for loans and the fall in the supply of loans will bid up the nominal interest rate by the full amount of inflation, thereby eliminating the real rate differential between money loans and real capital investment.

5.2.2. Knut Wicksell

According to O'Brien (1984, p.27), Wicksell borrowed from Tooke the observation that prices and the rate of interest were positively correlated both with respect to level and to changes, and also took a formal analysis of the natural rate of interest from Böhm-Bawerk's capital theory. In his work entitled under the German original *Geldzins und Güterpreise* (1898) and later translated in 1936, Wicksell put forward his theory of the cumulative inflationary process resulting from a money rate of interest below the natural rate. This work was 'written in ignorance of Thornton' but his ideas are certainly quite similar to those enunciated by Thornton, and is the subject of the following paragraphs.

The main element of Wicksell's analysis, like Thornton's, is the distinction between two interest rates, viz: the money or market rate and the natural or equilibrium rate. The former is the rate charged on loans in the money market. The latter, as Wicksell pointed out, can be interpreted in several ways. It is the expected marginal yield or internal rate of return on newly created units of physical capital. It is also the rate that would equilibrate desired saving and investment at the economy's full capacity level of output. Alternatively, the natural rate can be defined as the rate that equates aggregate demand for real output with its available supply. It follows, therefore, from this latter definition that the natural rate is also the interest rate that would be neutral with respect to general prices, tending neither to raise or lower them. According to Wicksell, as long as the market rate is equal to the natural rate, desired saving will just equal desired investment, aggregate demand will therefore equal aggregate supply, and price stability will prevail.

Any discrepancy between the two interest rates, however, will cause prices to change.

If, for example, the market rate falls below the natural rate, opportunities now exist for entrepreneurs to make profits. The consequent rise in demand for factor inputs causes their prices to rise so increasing factor incomes. This is further manifested in an increase in aggregate demand which exceeds aggregate supply, and assuming that the excess demand is financed by bank loans resulting in the creation of new money, inflation will occur. Conversely, if the market rate rises above the natural rate, saving will exceed investment, bank loans and the stock of money will contract, there will be a deficiency of aggregate demand, and prices will fall. In order that prices fall in absolute terms, this would require the assumption that prices and wages are flexible downwards as well upwards (c.f. the Keynesian assumption that prices and wages are sticky downwards). Such an assumption will ensure that a deficiency in aggregate demand is manifested in lower prices rather than by a fall in production.⁷

The role of money in Wicksell's analytical framework is that the price level cannot change unless there is a corresponding prior change in the quantity of money. These money stock changes accompany changes in the volume of bank loans used to finance excess aggregate demand. Wicksell specifically states that these changes in the money stock are necessary to permit price level movements to occur, but he insists that such money stock changes are caused by the discrepancy between the two interest rates. By way of illustration, the above example is continued such that banks maintain the market rate of interest below the natural rate. Desired

investment will exceed desired saving. The demand for bank loans will expand, putting upward pressure on the market rate of interest. To prevent the market rate from rising, the banks must be willing to accommodate all borrowers at the fixed market rate. Assuming that the banks are so willing, then the volume of bank lending will rise. And since new loans are granted in the form of increases in the deposits of borrowers, the money stock also expands. The monetary expansion is thus a consequence of the divergence between the market and natural rates of interest, and is the foundation of Wicksell's rejection of the quantity theory of money, namely that the money stock is exogeneously determined, and that prices change in response to a change in the money stock.

In Wicksell's model, any discrepancy between the two interest rates will set in motion a dynamic sequence of expenditure and price level changes that will continue as long as the gap persists; in Wicksell's own words:

'...a fall in the rate of interest, even though it is causal and temporary, will bring about a perfectly *definite* rise in prices, which, whether it is big or small, will persist as a permanent feature even after the interest rate has returned to its former value. If the rate of interest remains at a low level for a considerable period of time, its influence on prices must necessarily be *cumulative*; that is to say, it goes on repeating itself over equal intervals of time in precisely the same manner.' (1898, trans. 1936, pp. 94-95)

The cumulative process, Wicksell argued, could either be stable or unstable depending upon the type of monetary system a nation possessed. He considered two extreme types of hypothetical monetary

arrangements, viz: a pure cash system embodying the classic characteristics of a gold standard and a pure credit system using no metallic money, any payments being made by means of bookkeeping entries.

In anything other than a pure credit system, Wicksell maintains that the cumulative process plays an equilibrating role. During an inflationary period, for example, the rise in expenditure, prices, and the level of nominal national income results in a drain on the banks' specie reserves so that they are forced to ration credit by raising the market rate of interest until the inflationary process is brought to a halt. Contrariwise, during a period of price deflation, the steady accumulation of excess reserves will eventually induce banks to ease the availability of credit by reducing the rate of interest to stimulate borrowing so that prices will eventually stop falling. Thus, anything other than a pure credit system contains a stabilising adjustment mechanism that brings the cumulative process to a halt.

Having laid out the framework of Wicksell's analysis, it is now necessary to take a look at the role of inflationary expectations in a Wicksellian framework.

5.2.3. *Inflationary expectations and the Wicksellian model*

For the most part, Wicksell's analysis of the inflationary process is conducted on the assumption of the absence of inflationary expectations. No matter how much prices have changed in the past or are changing currently, all individuals expect them to remain unchanged over the indefinite future. This is made quite clear by Wicksell in two places: '...base his calculations on the current price.' (1898, trans. 1936, p.95) and 'If entrepreneurs are not reckoning for the moment on any rise in future prices,...' (p.144). Here, it is

quite explicit that inflationary expectations are *static*, that is, all individuals expect current prices to remain as they are, and with the absence of any other kind of inflationary expectations, the distinction between nominal and real (deflated) interest rates cannot be made. The main distinction, however, is still that between natural and market rates of interest.

Wicksell, of course, did not ignore inflationary expectations altogether. He noted that after the inflationary process has continued for some time (i.e. prices have been rising steadily for some time), the assumption that anticipated future prices are identical to present prices may have to be abandoned. Thus, Wicksell states that

'The upward movement of prices will in some measure "create its own draught". When prices have been rising steadily for some time, entrepreneurs will begin to reckon on the basis not merely of the prices already attained, but of a further rise in prices.' (1898, trans. 1936, p.96)

and

'...once the entrepreneurs begin to rely on this process continuing - as soon, that is to say, as they start reckoning on a future rise in prices - the actual rise will become more and more rapid. In the extreme case in which the expected rise in prices is each time *fully* discounted, the annual rise in prices will be indefinitely great.' (p.148)

In the face of such a development, stabilisation policy would have to be modified somewhat as follows:

'To put an immediate stop to any further rise in prices, it would not be sufficient for the banks to restore the rate of

interest to its original level. This would have the same effect on the business world as would a somewhat lower rate of interest at a time when prices are not expected to alter.' (1898, trans. 1936, p.97)

It is contended that the eradication of inflationary expectations requires that the market rate be raised temporarily above the natural rate associated with zero inflationary expectations. With the market rate established above the natural rate, anticipated price increases will fail to materialise and expectations will be revised downward. Eventually, the market rate will fall back into equality with the natural rate. If, on the other hand, banks persist in trying to peg the market rate below the natural rate, 'two forces will be operating in the direction of higher prices, and the rise will be correspondingly more rapid.' (p.97). These 'two forces' are, of course, the gap between the natural and market rates of interest *and* inflationary expectations.

However, Wicksell did not develop his analysis more fully by, for instance, making a further distinction between nominal and real rates of interest so that inflationary expectations could be incorporated more fully within the model. In view of this inadequate treatment of inflationary expectations, there have been several attempts to combine Wicksell's model with that of Fisher. Two examples of such analyses to be discussed in the following paragraphs are those by Lutz (1974) and Sargent (1969).⁸

The incorporation of inflationary expectations into the Wicksellian model makes it necessary to make a further distinction regarding the rates of interest, namely that between nominal and real (deflated) rates of interest. Their relationship to each other has

already been discussed in Section 5.1 above. It is also necessary to modify an assumption, as Lutz (1974, pp.105-106,109) has already done, regarding the money stock. It was previously argued in the last sub-section that it is the banks themselves who determine the market rate of interest, and they accommodate the money supply to the increased demand for loans at a lower rate so leading to the conclusion that any divergence between the natural and market rates causes the money supply to be endogenously determined. Before introducing the assumption that the money stock is exogeneously determined, it is first necessary to understand the concept of an 'inflationary equilibrium' as described by Lutz (1974, pp.103-104). Here, it is assumed that an economy is growing at a natural exponential rate of n per annum, and it is further assumed that the money supply also grows at the same rate so that prices are said to be stable. If the rate of monetary expansion were to be increased to $n + \pi$ then this excess rate of monetary expansion will produce an inflation rate equal to \dot{p}/p which, assuming perfect foresight, is equal to the anticipated rate of inflation, π . The market rate of interest is also higher by the rate of inflation. Such a state is known as the dynamic steady-state because all second-order rates of growth are zero, and the first-order rates growth in the variables are all constant. More will be said about the dynamic steady-state in the next two chapters.

Now, it has been argued by Lutz (1974) that an inflationary equilibrium in the original Wicksellian model is impossible because inflation is essentially a disequilibrium situation in that the real market rate of interest is below the natural rate, and the endogenous expansion of the money supply 'validates' that disequilibrium. By

allowing the money supply to be exogenously determined, the real market rate of interest becomes endogenously determined, and it becomes possible to achieve inflationary equilibrium by allowing the real market rate to converge on to the natural rate, and with the existence of inflationary expectations, the nominal rate of interest is approximately higher than the real market rate of interest by the rate of inflation.

It is particularly useful to employ an identity used by Sargent (1969, p.130, equation 2.1) to link the various rates of interest, and is given as follows:

$$R \equiv r^* + (r - r^*) + (R - r) \quad \dots [9]$$

where R is for the nominal market rate of interest, r is for the real market rate, and r^* is for the natural (or equilibrium) rate. Here, the nominal market rate of interest is identical to the natural rate of interest plus the differential between the real market and natural rates plus the expected rate of inflation which is approximately equal to $R - r$.

Consider, first, the case of static inflationary expectations. Here, individuals will expect a zero rate of inflation so that the last term of equation [9] will vanish. It, therefore, follows that the fundamental equilibrium condition in the original Wicksellian model is that the real market rate, r , equals the natural rate, that is, $R = r = r^*$. In disequilibrium, the real market rate will diverge from the natural rate so that

$$R = r = r^* + (r - r^*) \quad \dots [10]$$

If the market rate falls below the natural rate, this increases the demand for bank loans which leads to an increase in the money supply and then inflation. By pegging the market rate below the natural rate, the disequilibrium situation is validated.

If the inflationary process continues for sometime, all individuals will come to expect a positive future rate of inflation which is likely to be incorporated as a premium over and above the market rate. It should be noted that perfect foresight will be assumed to exist so that the world be be characterised by the absence of risk and uncertainty. In such a world, the real rates of return on both financial (or monetary) and real (or physical) assets will be equal. If the natural rate is taken to be the marginal efficiency of physical capital goods, then equilibrium would be characterised by the equality of the real (deflated) rate of interest with the natural rate, and the nominal rate of interest would approximately exceed the real (deflated) rate by an inflation premium of π . Thus, the following two equations characterise an inflationary equilibrium in the modified Wicksellian model:

$$r = r^* \quad \dots [11a]$$

and

$$R = r + \pi \quad \dots [11b]$$

Sargent states that

'Irving Fisher...noted that in equilibrium the nominal rate of interest must equal the sum of the marginal rate of return from holding real capital and the expected proportionate rate of change of prices. This condition

follows from the fact that in a riskless world, holding period yields must be equal for all assets.' (1969, p.128)

Gebauer (1986, p.131) claims that '[t]o be sure, Fisher has never "noted" this expressly.' However, it is felt here that the above quotation from Sargent makes it clear that in a world where perfect foresight exists, the real (deflated) market rate of interest must be equivalent to the real rate of return on real assets, and that the nominal market rate of interest exceeds the real market rate by approximately the rate of inflation. Such a conclusion takes into account the contribution made by Wicksell, namely that price stability occurs when the market rate of interest is equal to the natural rate. The incorporation of inflationary expectations gives way to an inflationary or steady-state equilibrium. Similar conclusions to those outlined above were arrived at by Lutz (1974, pp.106-108) using diagrammatic exposition.

In his empirical investigation of the relationship between interest rates and inflation, Fisher (1896, 1930) had to reject the assumption of perfect foresight as the evidence pointed to long adjustment lags implying imperfect foresight. Therefore, the following section is concerned with the transition period.

5.3. The transition period

Empirical evidence presented by Fisher to test the theory of interest adjustment under perfect foresight revealed that market interest rates tended to be high during periods of inflation and low during periods of deflation. However, it was also revealed that interest rates responded slowly and incompletely to inflation. The real rate of interest was three-and-half times more variable than the nominal rate

of interest, and was often negative during periods of rapid inflation.⁹ Since lenders had the option of simply hoarding commodities – and earning a real rate of interest equal to zero – Fisher concluded that the evidence 'must mean that the price movements were inadequately predicted.' (1896, p.67)

Forced to reject the perfect foresight model, Fisher presented an alternative model of the effects of inflation on nominal interest rates based on imperfect foresight. Such a model may be regarded as a disequilibrium model in which transitory changes in real variables (profits, investment, volume of trade) play an important role. Thus, the following sub-section discusses the theory of the transition period, and the second sub-section considers Fisher's empirical work in this field, and the final sub-section places the analysis into a Wicksellian perspective.

5.3.1. *Fisher's theory of the transition period*

The disequilibrium model put forward by Fisher was used for two purposes: firstly, to explain how the nominal rate of interest reaches its equilibrium level consistent with full adjustment to inflation, and secondly, to explain how price changes generate trade cycles. These uses of the model will now be discussed in turn.

To explain how the inflation premium gets embodied in the nominal rate of interest, Fisher assumed that firms were net borrowers, and that borrowers forecast profits extrapolatively. From these assumptions, Fisher argued that unexpected inflation and sluggish nominal interest rates produced falling real interest rates, and hence windfall profits to borrowers. Because borrowers base their profit forecasts on past extrapolation, a windfall profit 'raises an expectation of a similar profit in the future, and this expectation, acting on the

demand for loans, will raise the [nominal] rate of interest.' (1896, p.75) Anticipating the fact that nominal interest rates did not adjust fully, Fisher said 'If the rise is still inadequate, the process is repeated, and thus by continual trial and error the rate approaches the true adjustment.' (1896, pp.75-76) At this point, it would be useful to recall that in sub-section 5.2.1. of this thesis, Thornton could not make his analysis depend on inflationary expectations *per se*. Without being aware of Thornton's contribution, Fisher reiterated Thornton's insight that borrowers formed expectations of higher profits even though they may not be apparently aware that there was an inflation going on.¹⁰

Now regarding the second use of Fisher's disequilibrium model, it is particularly important to note the distinction between *imperfection* and *inequality* of foresight. Firstly, Fisher assumes that all individuals hold identical expectations regarding the future rate of inflation, that is to say that there is a consensus on the expected rate of inflation. In such a case, the nonadjustment of the nominal rate of interest would not affect the volume of loans demanded and supplied. Thus, Fisher (1896, p.77) argues that '[u]nder such circumstances the rate of interest would be below the normal, but as no one knows it, no borrower borrows more and no lender lends less because of it.' Thus, imperfection and equality of foresight produces a transfer of real wealth from creditor to debtor.

Now relaxing the assumption that there is equality of foresight among all individuals so that there now exists inequality of foresight (which is still imperfect), price changes can generate trade cycles. The abnormally low real rate of interest and the resulting over-investment led Fisher to regard it as a major determinant of the

trade cycle.¹¹ Based on his empirical judgement that borrowers are more apt to have superior foresight, a rise in prices will make borrowers more willing to pay a higher rate of interest on their borrowings whilst lenders are still willing to lend at the original interest rate so that the volume of loans and, therefore, investment increases. Fisher concluded that 'This constitutes part of the stimulation to business which bimetalists so much admire.' (1896, p.77). Thus, Fisher recognised that during the adjustment period changes in the rate of inflation will have important effects on *real* variables, real interest rates, the rate of investment, and the volume of trade, rather than a simple adjustment of the nominal rate of interest. In other words, '*inequality* of foresight produces over-investment during rising prices and relative stagnation during falling prices.' (1896, p.78)

Fisher's most comprehensive treatment of the transition period can be found in Chapter IV of his work entitled *The Purchasing Power of Money* (1911). Here, Fisher analyses the dynamics of interest rate adjustment. As prices start to rise, the money profits of businessmen tend to rise faster than prices because of the lag in the adjustment of nominal interest rates, and this increases the businessmen's desire to borrow.¹² Since Fisher was concerned primarily with the relationship between inflation and the trade cycle, he paid particular attention to the effects of interest rate adjustment on real investment and the trade cycle. The sequence of interest adjustments, credit expansion, and inflation is quite similar to Wicksell's cumulative process. The extra borrowing that occurs as a consequence of rising prices and lagged interest rate adjustment takes the form of short-term bank loans which are approximately equal to

the money stock. The increase in the money stock further increases the general level of prices, and leads to a further rise in profits of businessmen. There is now an increase in the demand for loans which may force up the nominal interest rate, but not sufficiently so that it continues to lag behind its normal level. Believing that higher interest rates are being offered, lenders are encouraged to expand the supply of loans so that the money stock increases still further, and leads to a further rise in prices.¹³ Thus, Fisher summarises:

'...a slight initial rise of prices sets in motion a train of events which tends to repeat itself. Rise of prices generates rise of prices, and continues to do so as long as *the interest rate lags behind its normal figure.*' (1911, p.60)

Inevitably, the cumulative process must come to an end. As soon as the nominal rate of interest 'overtakes the rate of rise in prices, the whole situation is changed.' (1911, p.64) Demand for loans will eventually contract, and prices will begin to fall. Interest rates continue to lag behind falling prices so that the process repeats itself over again a few times until the rate of interest has fallen to levels at which borrowers are once more again willing to borrow.¹⁴

5.3.2. Fisher's empirical analysis

Fisher argued that the transition period would be characterised by an increase in the nominal rate of interest, a decrease in the realised real rate of interest, an increase in real business profits, and an increase in the real value of investment. Fisher, in his books on interest rates, usually would discuss the steady-state properties of the inflation-interest rate relationship which already have been discussed in Section 5.1. The usefulness of discussing steady-state properties

rests on the premise that it gives some indication as to how, given an initial exogenous shock, the economy will diverge from its steady-state equilibrium path. If the initial exogenous shock was a once-only occurrence, and other exogenous variables remain constant, the economy would eventually converge by a series of oscillations back on to its steady-state equilibrium path. Unfortunately, the world is simply not kind enough to change slowly enough to allow the steady-state to be observed as there are a variety of random exogenous shocks that continually impinge upon the economy so that the economy is in a perpetual state of flux.¹⁵ Thus, according to Gebauer (1986, p.130), in practice there are a series of transition periods following and overlapping one another. Therefore, Fisher stressed that the *observable* world is in a continual disequilibrium, and that the appropriate framework for analysing real world data is dynamic rather than static.

Chapter XIX of Fisher's work entitled, *The Theory of Interest*, (1930) contains the empirical work dealing with the effects of inflation on interest rates. In Sections 1 to 5 of that chapter, Fisher gave qualitative evidence showing that interest rates expressed in different standards are different when those standards are diverging in value, inflation resulted in high interest rates, but the adjustment was very slow. This was attributed to money illusion:

'...men are unable or unwilling to adjust at all accurately and promptly the money interest rates to changed price levels....The erratic behaviour of real interest is evidently a trick played on the money market by the "money illusion" when contracts are made in unstable money.' (1930, p.415)

Then in Sections 6 and 7, Fisher presented results of correlating

nominal interest rates with lagged inflation rates for both Great Britain and the United States. He argued that a distributed lag of past inflation rates should be employed rather than a discrete lag,¹⁶ and presented simple correlation coefficients which would have been equivalent to carrying out a regression of the form:

$$R_t = \alpha + \beta \sum_{i=0}^T w_i \pi_{t-i} + u_t \quad \dots [12]$$

where α and β are constants, w_i ($i = 1, \dots, T$) are distributed lag weights, T is the order of the estimated lag distribution, and u_t is a stochastic term. For ease in calculation, Fisher constrained the lag weights to decline arithmetically and to sum to unity.

Fisher found that correlations between long-term bond rates and distributed lags of past changes were extremely high. Furthermore, the length of time required for complete adjustment was extremely long so that

'...for Great Britain in 1898-1924, the highest value of r (+0.980) is reached when effects of price changes are assumed to be spread over 28 years or for a weighted average of 9.3 years, while for the United States the highest r (+0.857) is for a distribution of the influence due to price changes over 20 years or a weighted average of 6.7 years.' (1930, p.423)

Using quarterly data on U.S. commercial paper rates, Fisher found that during the period 1915-1927, the highest r was +0.738 for 120 quarters. In most of the empirical literature, there is often some surprise at how long it takes interest rates to adjust fully for

inflation,¹⁷ but Fisher interpreted the lag as largely representing adjustments in real variables such as real interest rates, profits, and the volume of trade, and not as a simple measure of the delay in expectations formation:

'It seems fantastic, at first glance, to ascribe to events which occurred last century any influence affecting the rate of interest today. And yet that is what the correlations with distributed effects of [inflation] show. A little thought should convince the reader that the effects of bumper wheat crops, revolutionary discoveries and inventions,...and similar events project their influence upon prices and interest rates over many future years even after the original casual event has been forgotten...A further probable explanation of the surprising length of time by which the rate of interest lags behind price changes is that between price changes and interest rates a third factor intervenes. This is business, as exemplified or measured by the volume of trade. It is influenced by price change and influences in turn the rate of interest.' (1930, pp.428-429)

Fisher believes that two facts are well established: first, price changes influence the volume of trade, and second, the volume of trade influences the interest rate. He then says that

'The evidence for both relationships is not only empirical but rational. Rising prices increase profits both actual and prospective, and so the profit taker expands his business. His expanding or rising income stream requires financing and increases the demand for loans.' (1930, p.439)

In his summing up of the adjustment of nominal and real interest rates, Fisher said

'The final result, partly due to foresight and partly to the lack of it, is that prices changes do after several years and with the intermediation of changes in profits and business activity affect interest very profoundly. In fact, while the main object of this book is to show how the rate of interest would behave if the purchasing power of money were stable, there has never been any long period of time during which this condition has been even approximately fulfilled. When it is not fulfilled, the money rate of interest, and still more the real rate of interest, is more affected by the instability of money than by those more fundamental and more normal causes connected with income impatience, and opportunity...' (1930, p.451)

Quite clearly, then, Fisher rejected any notion that inflation was 'neutral', that is, real interest rates are certainly not homogeneous of degree zero with respect to prices. Thus, under imperfect foresight, the real (deflated) rate of interest can be expected to vary inversely with inflation. Furthermore, Fisher did not assume a real rate of interest determined independently of past inflation rates. It is precisely the variations in the real factors, according to Fisher's interpretation, which combine to produce the extremely long adjustment period for nominal interest rates, quite apart from the way in which price expectations are formed. However, what is not so clear is that the real (deflated) interest rate can diverge from the natural rate of interest during the transition period. Thus, to dispel any further misconceptions, namely that both real rates are *always*

equal, the preceding analysis is put into a Wicksellian perspective.

5.3.3. A Wicksellian perspective

Before proceeding to analyse the transition period in a Wicksellian perspective, it is first necessary to make it clear that the analysis will still depend on two assumptions made previously for the full equilibrium model (see sub-section 5.2.3 above), namely that the money stock is exogenously determined¹⁸, and that inflationary expectations are formed by all individuals. These assumptions can be contrasted with Wicksell's original assumptions that the money stock is endogenously determined, and that no individuals expect inflation at all. A further assumption, which is closely related to the second assumption, is that identical inflationary expectations are formed by all individuals. Such assumption has been made by Lutz (1974, p.105). However, the way in which this assumption has been presented by Lutz is likely to lead to some confusion regarding Fisher's earlier distinction between imperfection and inequality of foresight. In order to avoid any such confusion, it is important to be able to distinguish carefully between the Wicksellian and Fisherian elements in the adjustment of interest rates, and this aspect will be discussed in more detail as the occasion demands it.

The starting point of the following analysis will be that of an inflationary equilibrium in which the economy is experiencing a constant rate of inflation equal to \dot{p}/p , the real deflated market rate of interest is equal to the natural rate, and the nominal market rate exceeds the real deflated market rate by approximately the rate of inflation. This starting point is generally applicable, that is to say, it is equally applicable to the Wicksellian equilibrium situation as characterised by a zero rate of inflation.

The adjustment of interest rates can be broken up *conceptually* into two stages: the first stage embodying Wicksell's analysis, and the second stage embodies that of Fisher. Consider, then, the effects of an increase in the rate of growth in the money supply so that its rate of growth now exceeds the original inflationary equilibrium rate of growth of $n + \pi$ per unit time. This is manifested in an increase in the supply of credit by the banking system which will lower the real deflated market rate of interest below the natural rate. The demand for credit increases until the real market rate rises back to its original level. The first stage has now been completed.

If it is assumed that all individuals expect a rate of inflation that is less than the new actual rate of inflation, it becomes necessary to distinguish between *ex ante* and *ex post* real interest rates. The former refers to the expected real market rate of interest (i.e. deflated by the expected rate of inflation), and the latter refers to the realised real market rate of interest (i.e. deflated by the actual rate of inflation). Thus, given that all individuals expect a rate of inflation less than the actual rate, the *ex ante* real market rate of interest will adjust so that it is equal to the natural rate, and the nominal market rate will be higher than the *ex ante* real market rate by the expected rate of inflation. Now, when all individuals realise that the actual rate of inflation is higher than expected, the *ex post* real market rate will be less than the *ex ante* real market rate which is equal to the natural rate, so that there is a real transfer of wealth from creditors to debtors as noted by Irving Fisher himself (see sub-section 5.3.1 above). The preceding analysis was based on Figure 2 of Lutz (1974) which has an inelastic supply schedule.

Now, the assumption regarding the exogeneity of the money supply is partly relaxed so that a part of the money supply can be endogenously determined. This is reflected in an upward sloping supply schedule for credit as shown in Figure 5.1 which depicts a market for loanable funds. Referring to this diagram, the initial equilibrium is shown by the intersection of the demand and supply schedules DD and SS respectively at point 1, with a real market rate of interest equal to the natural rate, r^* , and the volume of loans transacted in the market will be equal to C^* . An exogenous increase in the money supply is reflected in a parallel shift of the supply schedule from SS to S_1S_1 , which lowers the real market rate below the natural rate. As a consequence of the lower market rate, the demand for credit increases (by a rightward shift of the demand schedule from DD to DD_1) from C^* to a level associated with point 2 of the diagram. An increase in the inflation rate will lead to a further rise in the demand for credit which is met in two ways: firstly by an endogenous increase in the supply of credit, and secondly, by an increase in the market rate back to its original level. The volume of credit demanded and supplied is higher than previously at C .

Consider how the inflation premium gets incorporated within the interest rate. Attracted to the possibility of profits, borrowers will demand more credit as manifested in an upward shift of the demand schedule from DD_1 to D_1D_1 . Now if the *ex ante* real market rate of interest is to be equal to the natural rate, it requires that all individuals (i.e. debtors and creditors) have identical inflationary expectations. Thus, when creditors expect a rate of inflation similar to that expected by borrowers, the supply schedule shifts from S_1S_1

INTEREST
RATE \uparrow

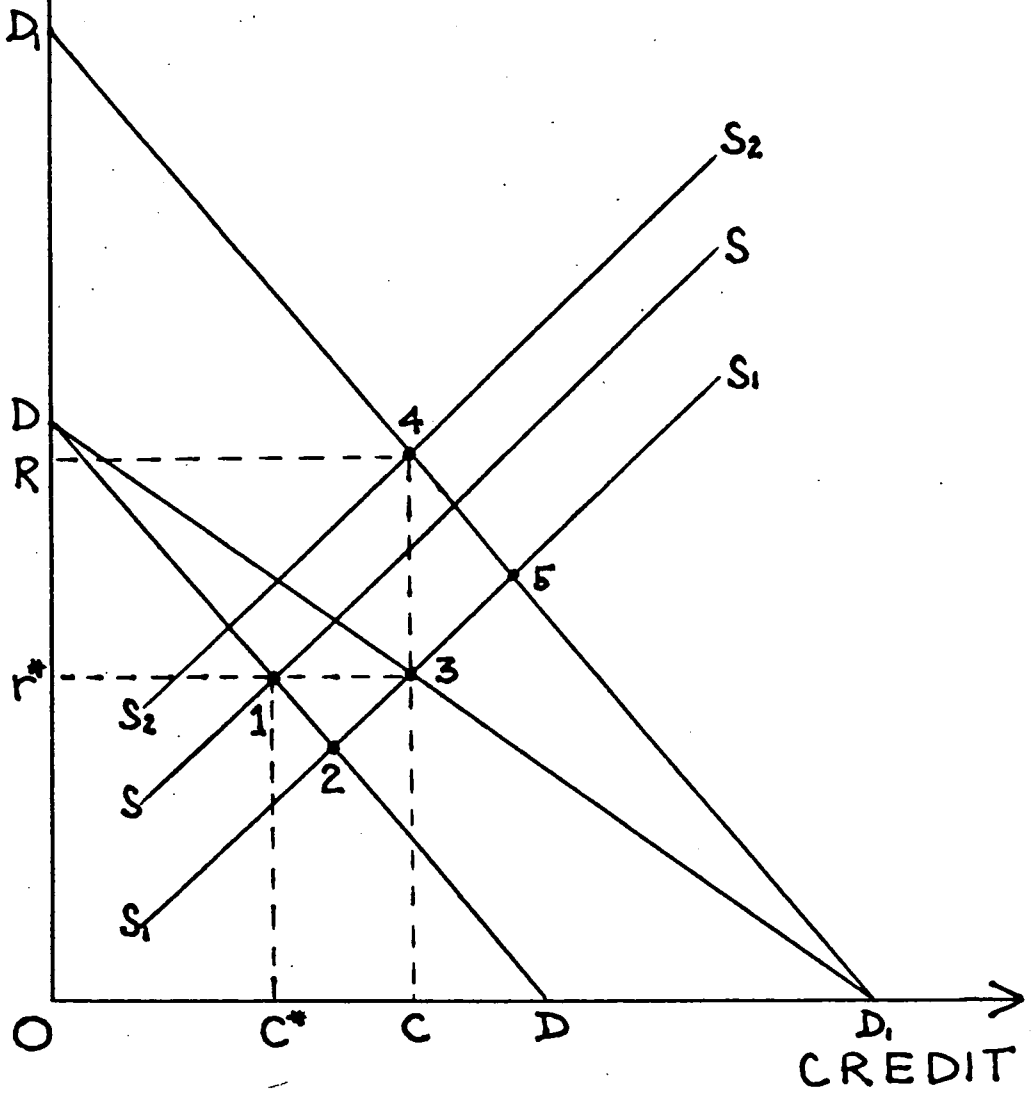


FIGURE 5.1: *Combination of the analyses by Wicksell and Fisher in the context of a market for loanable funds.*

to S_2S_2 . The market rate of interest will now rise to R which will now include the expected rate of inflation as shown by point 4 of the diagram. The volume of credit demanded and supplied remains unchanged at C . Such a conclusion can be rationalised when Fisher's distinction between imperfection and inequality of foresight is taken into account. Recall, from sub-section 5.3.1 that imperfection of foresight produces no change in the volume of credit demanded and supplied, whereas inequality of foresight will produce changes in the volume of credit. In the case of imperfection of foresight, Fisher's conclusion was based on the fact that the volume of credit did not change at all, so implying *equality* of foresight among all individuals.¹⁹ This conclusion could be extended. If the nominal rate of interest rises, but the real volume of credit stays constant, it implies that all individuals have identical inflationary expectations. It is only when inequality of foresight exists that the real volume of credit demanded and supplied is likely to change as argued by Fisher. Thus, for example, if borrowers expect a rate of inflation higher than that expected by lenders, borrowers will demand more credit because the lenders are still willing to lend at a nominal rate of interest that is lower than the nominal rate which borrowers are prepared to pay. Thus the real volume of credit demanded and supplied will increase beyond C , and the nominal rate of interest will rise to somewhere between the level shown for point 5 in the diagram and that level consistent with equality of expectations. So, to drive the point home, Lutz's assumption regarding identical expectations among all individuals is, at best, superfluous when the supply of credit is assumed to be completely inelastic, but can play a crucial role if the supply of credit becomes more elastic.²⁰ Furthermore, confusion could arise out of

the way in which the assumption was presented by Lutz. As previously argued, it is extremely important to distinguish between the Wicksellian and Fisherian components of interest rate adjustment. The Wicksellian component involves an increase in the real volume of credit demanded and supplied, whereas the Fisherian component, assuming identical inflationary expectations, leads to no further change in the volume of credit. But if the assumption of identical inflationary expectations is dropped, then the Fisherian component of interest rate adjustment during the transition period will invariably lead to a further change in the volume of credit. In other words, the assumption of identical inflationary expectations should be taken in the context of interest rate adjustment according to the analysis of Fisher.

As argued by Fisher, the transition period may be characterised by the sluggish adjustment of nominal interest rates to inflation, and it is real interest rates that bear the brunt of the adjustment, that is to say, real deflated interest rates are much more variable than nominal interest rates. According to the above analysis, if the rate of inflation is not fully anticipated, assuming identical expectations for all individuals, then on an *ex ante* basis, interest rates will adjust so that the nominal market rate of interest will exceed the real deflated market rate by approximately the rate of expected inflation, and the *ex ante* real deflated rate will equal the natural rate. On an *ex post* basis, when the actual rate of inflation turns out to be higher than expected, then the *ex post* real deflated rate will turn out to be less than the natural rate. With this in mind, from equation [9] above, the transition period may be characterised by the following set of two expressions:

$$R = r + \pi \quad (\textit{ex ante}) \quad \dots [13a]$$

and

$$r^* \neq r \quad (\textit{ex post}) \quad \dots [13b]$$

Even if inequality of foresight is assumed such that borrowers tend to expect a rate of inflation that is closer to the actual rate than expected by lenders, it is possible that the above conclusion would still hold. As figure 5.1 shows, if lenders do not expect any increase in the rate of inflation, the supply schedule remains as S_1S_1 , so that the market for loanable funds is described at point 5 of the diagram. From the point of view of borrowers, the nominal rate of interest is less than the level consistent with equality of expectations (i.e. point 4 of diagram) so that the *ex ante* real deflated market rate of interest is less than the natural rate. It is possible that another Wicksellian-type interest rate adjustment will take place. Sensing a further opportunity for making profits, borrowers increase their demand for credit which is manifested by a further rightward shift of the demand schedule from D_1D_1 to another demand schedule that will be parallel to DD_1 , so that the nominal market rate of interest is bid up further to R , and volume of credit transactions in the market will be higher than the level consistent with equality of expectations. Since the nominal market rate of interest has risen to R , the *ex ante* real deflated market rate of interest will still be equal to the natural rate, and the nominal market rate will exceed the *ex ante* real deflated market rate by approximately the expected rate of inflation. As before, when the actual rate of inflation turns out to be higher than expected, the *ex post* real deflated market rate of interest will

still be less than the natural rate. In both cases, the *ex post* real deflated market rate of interest is not equal to the natural rate of interest, and thus the discussion above substantiates Gebauer's claim made at the beginning of this chapter that it is incorrect to assume that the real rate of interest on financial assets is always equal to the real rate on real assets.

5.4. Conclusions

Two things have been made clear. Firstly, the notion that inflation is neutral with respect to real deflated interest rates should be firmly rejected for once and all as the empirical evidence presented by Fisher himself showed that real interest rates were much more variable than nominal interest rates. This implies a total rejection of the perfect foresight assumption. In studying the transition period, however, the steady-state relationship between inflation and interest rates with its assumption of perfect foresight can serve as a useful guide as to how far the economy is diverging from its steady-state equilibrium path. Secondly, the notion that real interest rates on both financial and real assets are always equivalent needs to be rejected. By combining the analyses of Fisher and Wicksell, the concept of an inflationary equilibrium was brought out in which both real rates are equivalent. However, according to the analysis in sub-section 5.3.3 above, it was shown that during the transition period, the real rates of interest are certainly not equivalent.

Even if perfect foresight is assumed, there are many good reasons why the hypothetical relationship between inflation and interest rates may not hold. First, it was demonstrated by Mundell (1963) that an inverse relationship between inflation and real interest rates could exist. Second, if taxes are introduced into the economy, one

would expect that the relevant decision variables will be the after-tax rates of return rather than the before-tax rates of return. Both aspects are discussed in the next chapter. A further interesting dimension is provided by Gebauer (1986) who argues that Tobin's q -theory of investment is based on relative changes in both real rates of interest. If, say, the rate of return on financial assets is less than that on real assets, the rate of growth in the capital stock will become greater than the steady-state growth rate so that there exists an actual incentive for investment in additional physical capital. The effects of taxation on Tobin's q -theory of investment will be examined when the disequilibrium properties of the model are discussed in the next chapter.

C H A P T E R S I X
REFINEMENTS OF THE FISHER HYPOTHESIS

Section	6.1. The need for refinement.	310
	6.1.1. <i>Bailey's analysis.</i>	311
	6.1.2. <i>Mundell's analysis.</i>	312
	6.2. A neoclassical monetary growth model.	314
	6.2.1. <i>The production function.</i>	314
	6.2.2. <i>The liquidity function.</i>	315
	6.2.3. <i>The consumption function.</i>	316
	6.2.4. <i>The national income identity.</i>	317
	6.3. All-debt financing model.	318
	6.3.1. <i>The firm's demand for capital.</i>	319
	(a) <i>When taxation is absent.</i>	319
	(b) <i>When taxation is present.</i>	322
	6.3.2. <i>The various after-tax rates of return defined.</i>	324
	6.4. Effects of inflation (1).	325
	6.4.1. <i>Effect on capital intensity.</i>	328
	6.4.2. <i>Effect on gross interest rates.</i>	332
	6.4.3. <i>Effect on net interest rates.</i>	334
	6.5. A critique of Feldstein's model.	336
	6.5.1. <i>A re-consideration of the demand for money function.</i>	336
	6.5.2. <i>Relaxation of more assumptions.</i>	339
	6.5.3. <i>Why are liquidity effects a problem?</i>	339
	6.6. Debt-equity financing model.	344

6.6.1. <i>Initial assumptions.</i>	344
6.6.2. <i>The installment function.</i>	345
6.6.3. <i>The firm's optimal behaviour.</i>	346
6.6.4. <i>Tobin's q-theory of investment under taxation.</i>	351
6.6.5. <i>The firm's demand for capital in a debt-equity model.</i>	356
6.7. Effects of inflation (2).	358
6.7.1. <i>Effect on gross interest rates.</i>	358
6.7.2. <i>Comparison with the all-debt model.</i>	362
6.8. The short-run versus the long-run.	363
6.9. Conclusions.	375

CHAPTER SIX

REFINEMENTS OF THE FISHER HYPOTHESIS

This chapter discusses many of the refinements made to the Fisher hypothesis. Some of the early refinements will be briefly reviewed in Section 6.1. The earliest refinements made use of static analytical techniques to show how and why the nominal interest rate may not rise in tandem with the inflation rate. Owing to the inappropriateness of static analysis for the study of the effects of inflation, a basic neoclassical monetary growth model is set up in Section 6.2. In the sections following Section 6.2., the behaviour of firms is considered using disaggregated models of the financial sector. The main difference between these models lies in the assumptions made regarding how firms finance their marginal investment projects. The simplest, but unrealistic, model assumes that the marginal unit of capital is financed entirely by issues of corporate debt. The inclusion of such a model in this thesis can only be justified on the grounds that if the simplest cases are considered first, one may find it easier to understand the more complex cases which allow for the possibility of debt-equity financing. After these disaggregated models of the financial sector have been set up, it is necessary to specify the firm's demand for capital so as to complete the basic neoclassical monetary growth model. This will permit consideration of the effects of taxes on the inflation-interest rate relationship. Thus, Section 6.3. derives the firm's demand for capital under an all-debt financing model, and Section 6.4. discusses the effects of taxes on the inflation-interest rate

relationship when all-debt financing is assumed. Then Section 6.5. criticises Feldstein's model on the grounds that the predictions generated by the model may not hold if the demand for money is allowed to be a function of income as well as the net nominal interest rate. Section 6.6. then extends the all-debt model to allow for equity financing, and Section 6.7. re-examines the effects of taxation. Finally, Section 6.8. presents a general equilibrium model used by Summers (1983) whose main conclusion is that *any* relationship between inflation and interest rates may exist in the short-run, but could take a more definite form in the long-run.

6.1. The need for refinement

Owing to the fact that the majority of the empirical literature does not show the existence of a relationship between inflation and interest rates in the sense that the latter does not rise by as much as the former, there is a need to refine the Fisher hypothesis to take account of institutional factors that could affect the inflation-interest rate relationship. Mundell (1963, p.280) is correct in pointing out that "...to attribute the discrepancy between theory and reality solely to lack of foresight is to raise doubts about the nature of evidence that would be required to reject the theory [of complete adjustment]." Mundell re-echoes Keynes' main criticism of Fisher in that the discrepancy between reality and theory is not solely due to imperfection of foresight. The point being made here is that, if perfect foresight even existed at all, there are many good reasons why nominal interest rates cannot adjust fully to inflation; and by assuming perfect foresight, the subjective question of how inflation expectations are formed can be abstracted from for the time being.

In this section, early attempts at refining the Fisher hypothesis are considered along with their main shortcomings which have given rise to the need for further refinement. This discussion will serve as a useful prelude to the later sections of this chapter.

6.1.1. *Bailey's analysis*

Although the paper by Bailey (1956) is exclusively concerned with the welfare cost of inflationary finance, there are certainly a few paragraphs outlining the effect of inflation on interest rates (see Bailey (1956). pp.103-4). Bailey's analysis is enshrined in a market for loanable funds in which the banks act as intermediaries by channelling funds from their depositors to their borrowers. The main effect of inflation on interest rates is through money balances deposited with the banks by the public. At the onset of (hyper-) inflation, the banks will find it most profitable to create the maximum amount of non-interest bearing deposits and lend out the funds in the market for loanable funds at high nominal interest rates that reflect the extent of competitive bidding for funds. In times of hyper-inflation, this phenomenon may become so accentuated that, in a competitive banking system, the individual bank may be forced to start paying interest on its non-interest bearing deposits in order to attract deposits away from other banks. The collective action of the banks in competing with one another for more deposits will lead to a situation in which the banks would all offer a nominal interest rate on deposits "approaching, but not equal to, the rate of inflation." (Bailey (1956), p.103)

This conclusion could be explained by considering how banks behave in a banking system in which a reserve ratio requirement is operational (as in the case of inter-war Germany during the years of

hyper-inflation), and this is the explanation offered by Bailey (1956, p.104). During an inflation "that everyone anticipates perfectly", the incentive to substitute interest-bearing bank deposits for currency holdings would be so great that, in a reserve ratio system, all newly-issued currency would immediately flow into bank reserves so that the full reserve requirement of c would be met, where c is the reserve ratio (proportion of total deposits in the form of currency). An individual bank would only be able to expand its loans by the new deposits times $(1 - c)$, and so the maximum rate of interest that could be offered on all deposits would be the rate of inflation times $(1 - c)$. This analysis is not quite in the spirit of Fisher who realised that perfect foresight was not characteristic of the behaviour of individuals, but is certainly unusual in that the effect of inflation on interest rates is through money balances rather than the more orthodox marginal productivity of capital. The main shortcoming of Bailey's analysis is that it is static, and in such an analysis, the rate of inflation cannot, strictly speaking, be correctly included as a variable because it is synonymous to the rate of growth in the price level.

6.1.2. Mundell's analysis

Mundell (1963) uses a variant of the IS-LM apparatus to show that the nominal interest rate will rise (fall) by less than the anticipated rate of inflation (deflation). He states that "...Fisher found verification for a theory of *partial* adjustment of money interest to inflation and deflation but none for his own theory of *complete* adjustment under foresight." (p.280), and in light of Mundell's doubts on the discrepancy between theory and reality as being solely due to lack of foresight, he claims that his theory is

more consistent with the empirical findings of Fisher and also with Keynes' theory of investment. Amongst Mundell's assumptions, it is assumed that wealth is held in the form of money balances and equities, and these are held in a proportion which depends on the nominal interest rate which may be regarded as the true opportunity cost of holding real money balances. It is also assumed that real investment depends inversely on the real rate of interest, and that real savings depends inversely on real money balances. Overall equilibrium occurs when the incentive to invest is matched by the desire to save, and when the demand for equities equals its supply, and when desired real money balances are matched by the existing stock of real money balances. Another feature of this equilibrium is that the inflation rate is zero implying that the nominal and real interest rates are both equivalent.

Suppose now that the government expands the money supply at a rate which exceeds the economy's natural rate of growth so that inflation occurs at a rate which equals the excess rate of growth in the money supply. Real money balances (if the growth rate of the economy is zero) will be depreciating at a rate equal to the rate of inflation, and to this must be added the real return on equities to obtain the total opportunity cost of holding real money balances as reflected in a higher nominal interest rate. This forces individuals to economise on their holdings of real money balances as reflected in an increased desire to save. In order that the extra savings may be absorbed by additional investment, it is necessary for the real interest rate to fall. This leads to Mundell's conclusion that the nominal interest rate will rise by less than the rate of inflation. Whilst Mundell's analysis is instructive in demonstrating the use of a variant

of the IS–LM apparatus in the context of the Fisher hypothesis, it cannot be considered any further due to the static nature of the analysis. In the next section, a neoclassical monetary growth model whose origins can be traced to the inaugural paper of Tobin (1955) is set up and considered in the context of the inflation–interest rate relationship.

6.2. A neoclassical monetary growth model

In this section, a basic neoclassical monetary growth model is set up which will serve as the foundation for the analysis of the inflation–interest rate relation in later sections. The model presented here is a slight variant of the one presented by Feldstein (1976) in that it makes use of a consumption function rather than a savings function, and has its roots in the inaugural paper of Tobin (1955) on dynamic aggregative models.

6.2.1. The production function

Firstly, it is assumed that there are m firms in the economy, all producing a single good which is appropriated into either consumption or investment. It is further assumed that all firms have similar production technologies in that each firm is faced with a linearly homogeneous production function. The i th firm will employ N_i units of labour and K_i units of capital to produce output Y_i as specified in the following production function:

$$Y_i = F_i(K_i, N_i) \quad \dots [1]$$

where the i subscript on the production function emphasises that it is the individual firm's production function. Given certain conditions,

the individual production functions can be aggregated to obtain an aggregate production function that applies to the economy as a whole:¹

$$Y = F(K, N) \quad \dots [2]$$

where it is assumed that both factors of production are both subject to positive, but diminishing, marginal products, and that the marginal product of capital is assumed to vary directly with employment and vice-versa.

6.2.2. The liquidity function

In an all-debt financing model, in which all firms finance their marginal investment projects entirely by issues of corporate debt, there can only be two assets held by households. These two assets are real outside money balances² (M/p) and real corporate debt (B/p) and must satisfy the constraint that total real wealth is matched by the sum of these two assets:

$$W = (M/p) + (B/p) \quad \dots [3]$$

Since outside money balances earn no interest, the ratio of money to corporate debt that households will hold is a decreasing function of the after tax nominal rate of return (R_n). In an all-debt financing model, it is reasonable to assume that the real value of corporate debt is also the real value of the capital stock. Thus, the liquidity-preference relation is expressed as

$$\frac{(M/p)}{K} = L(R_n) \quad \dots [4]$$

with $\partial L / \partial R_n = L_R < 0$. In the long-run steady-state, $(M/p)/K$ must remain constant. That is, M/pK remains constant such that the rate of growth of the money supply M is equal to the rate of growth of the nominal capital stock (pK) .

6.2.3. The consumption function

In addition to their portfolio decisions, households also have to decide on how fast should their wealth accumulate. Their consumption decisions will be influenced by their disposable income (Y^D), and by the net real rate of interest (r_n). In Feldstein's model, it is assumed that consumption is proportional to disposable income.³ Thus, their consumption function is:

$$C = C(r_n) \cdot Y^D \quad \dots [5]$$

where it is assumed that the marginal propensity to consume is inversely related to the net real rate of return (i.e. $\partial C / \partial r_n = C_2 < 0$). It is now necessary to define disposable income which is gross national income less taxes (T) less depreciation in real money balances:

$$Y^D = Y - T - \pi(M/p)$$

where π is the anticipated rate of inflation (which must equal the actual rate of inflation (\dot{p}/p) in the long run steady state). The government uses the proceeds from taxation and the creation of new money (\dot{M}/p) to finance its expenditure (G) so that

$$Y^D = Y - G + \dot{M}/p - \pi(M/p)$$

It is assumed that the rate of inflation is equal to the excess rate of monetary growth over the natural growth rate (i.e. $\pi = \dot{M}/M - n$) and that government expenditure is directly proportional to national income (i.e. γY). Therefore

$$Y^D = Y(1 - \gamma) + n(M/p) \quad \dots [6]$$

6.2.4. The national income identity

Finally, the national income accounting identity is included. National income is equivalent to consumption, plus investment (I), plus government expenditure:

$$Y = C + I + G \quad \dots [7]$$

At this point, it would be useful to take a look at some of the differences and analogies of the current model with that of Feldstein (1976). Firstly, substitution of equation [7] into equation [6] gives

$$Y^D = C + I + n(M/p)$$

and given that $Y^D - C = S$ where S is for saving, and that the capital stock is growing at an exponential rate of n per unit time (in the form of investment), it follows that (not written in per-capita form)

$$S = nK + n(M/p)$$

which corresponds to Feldstein's equilibrium condition (1976, p.812)⁴

Secondly, in common with Feldstein, this model ignores economic depreciation on the capital stock. Thirdly, Feldstein makes the somewhat heroic assumption that saving is proportional to disposable income (1976, p.811) and makes no mention of the demand for money as being a function of income. The effects of relaxing such assumptions will be examined later on.

It now remains to specify the firm's demand for capital and labour in the context of an all-debt financing model which forms the subject matter of the next section.

6.3. All-debt financing model

In this section, the firm's demand for capital is derived, but before doing so, it is necessary to define the nominal interest rate, and consider how a firm might behave in the absence of taxation.

Fisher (1896) deduced that the relationship between nominal and real interest rates took the form (see Section 5.1. of this thesis):

$$(1 + R) = (1 + r)(1 + \pi)$$

where R and r are the nominal and real interest rates respectively. Expanding and re-arranging gives

$$R = r + \pi + r\pi$$

Assuming that there is continuous compounding, the last term of the above expression will become negligible (i.e. $r\pi \approx 0$) so that by definition,

$$R = r + \pi \dots [8]$$

6.3.1. The firm's demand for capital

Considering the behaviour of the firm now, it is extremely important that a distinction be drawn between the short- and long-run. In the short-run, the firm has a putty-clay capital stock which means simply that there is no ready market available if the firm should ever decide to sell some of its capital stock. Only in the long-run, is it considered possible to have a putty-putty capital stock so that the firm is exactly able to equate the marginal product of its capital with its cost. This is the main assumption of this model. More about this will be said in Sections 6.6. and 6.8.

(a) *When taxation is absent.* Feldstein (1976) assumes that the main objective of the firm is to maximise its profits, but it would do no harm here to assume that the main objective of the firm is to maximise its present value instead, and these two approaches are indeed both equivalent. Another point to be made here, is that the all-debt financing model can be regarded as a special case of the more general debt-equity financing model. Therefore, let e be the real rate of return on equity. A more precise definition of e will be given in Section 6.6. rather than here because, as will be shown below, the real rate of return on equity does not enter the marginality conditions for the firm in an all-debt financing model. Then the nominal rate of return on equity can be decomposed into two parts, viz: the capital gain on equity (\dot{V}), and dividends paid out (Div), so that

$$(e + \pi)V(t) = \dot{V}(t) + Div(t)$$

which can be recognised as a first order differential equation. After

imposing the following transversality condition

$$\lim_{s \rightarrow \infty} V(s) \exp^{-(e + \pi)(s - t)} = 0$$

which will guarantee that unstable dynamic behaviour is ruled out, so ensuring a stable and unique solution, the solution to the above differential equation is

$$V(t) = \int_t^{\infty} Div(s) \exp^{-(e + \pi)(s - t)} ds \dots [9]$$

Now, the firm's present value is the discounted value of its future net cash flows.⁵ The firm's cash flow at time s is

$$\begin{aligned} Div(s) = & p(s)[F(K(s), N(s))] - w(s)N(s) \\ & - p(s)K(s)(R - \pi) - pI + pI \end{aligned}$$

The first term represents the firm's gross revenue, and from this, the firm must deduct its wage bill (wN), and the cost of its capital which is the real rate of interest that it must pay on its debt. In addition, the firm also needs to deduct any gross investment expenditure undertaken. But, gross investment is financed entirely by corporate debt in this model so that the last two terms cancel out each other. After substitution for Div in the solution to the differential equation [9], the present value of the firm at time t can now be calculated as

$$V(K, N, t) = \int_t^{\infty} \left\{ p(s)F[K(s), N(s)] - w(s)N(s) - p(s)K(s)(R - \pi) \right\} \cdot \exp^{-(e + \pi)(s - t)} ds \dots [10]$$

where it is to be understood that p , w , K , and N are all functions of time. Among the necessary conditions for this to obtain an extremum are the following Euler equations:⁶

$$\exp^{-(e + \pi)(s - t)} \left[p \frac{\partial F}{\partial N} - w \right] = 0$$

and

$$\exp^{-(e + \pi)(s - t)} \left[p \frac{\partial F}{\partial K} - p(R - \pi) \right] = 0$$

for $t < s < \infty$. These two Euler equations imply that the marginal product of labour must be equated to its real wage, and the marginal product of capital must be equated to the real rate of interest, that is

$$F_N = w/p \quad \text{and} \quad F_K = R - \pi$$

These two conditions describe the firm's demand for labour and capital respectively in the absence of taxation.^{7,8} It can be noted that the optimality condition for capital is equivalent to that of Feldstein (1976, p.810) who expresses it in per-capita form without allowing for depreciation. Note also that the real rate of return on

equity (e) is absent from the above marginality conditions.

(b) *When taxation is present.* Feldstein then goes on to consider how the behaviour of firms might be modified in the presence of corporation taxation. The taxation system introduced can either discriminate between real and inflationary components of income, or remain indifferent between the two components. First, e is redefined as the real after-tax rate of return on equity. Then the nominal after-tax rate of return on equity can be decomposed as follows:

$$(e + \pi)V(t) = (1 - \kappa)\dot{V}(t) + (1 - \epsilon)Div(t)$$

where κ is the capital gains tax rate, and ϵ is the effective rate of tax on dividend income. The solution to the above differential equation, after imposing the transversality condition, is

$$V(t) = \int_t^{\infty} \frac{(1 - \epsilon)}{(1 - \kappa)} Div(s) \exp \left\{ \frac{-(e + \pi)}{(1 - \kappa)} (s - t) \right\} ds$$

In order to discuss the implications of changing the tax treatment of inflation, it will be necessary to specify two different tax rates, each tax rate being on the real and inflationary components of the nominal interest rate. One of the features of the taxation system is that interest paid on corporate debt may be deducted by firms in calculating taxable profits while dividends paid on corporate equity cannot be deducted. Let τ_1 be the corporation tax rate at which the real component of interest payments is deducted, and let τ_2 be the rate at which the inflationary component is deducted. Thus, the

net nominal rate of interest paid by firms is $(1 - \tau_1)r + (1 - \tau_2)\pi$. In an all-debt financing model, the firm deducts from its gross revenue its wage bill, and its real net interest payments on debt. Therefore, at instant s , the firm's net cash flow is now

$$Div(s) =$$

$$[pF(K, N) - wN](1 - \tau_1) - pK[(1 - \tau_1)r + (1 - \tau_2)\pi - \pi]$$

where it is to be understood that p , w , K , and N are all functions of time. Therefore, after substitution for Div in the expression for V ,

$$V(K, N, t) =$$

$$\int_t^\infty \frac{(1 - \epsilon)}{(1 - \kappa)} \left\{ [pF(K, N) - wN](1 - \tau_1) - pK[(1 - \tau_1)r - \tau_2\pi] \right\} \mu_s ds$$

... [11]

where $\mu_s = \exp\{[-(e + \pi)/(1 - \kappa)](s - t)\}$. The Euler equations are now

$$\frac{(1 - \epsilon)}{(1 - \kappa)} \mu_s \left\{ \left[p \frac{\partial F}{\partial N} - w \right] (1 - \tau_1) \right\} = 0$$

and

$$\frac{(1 - \epsilon)}{(1 - \kappa)} \mu_s p \left\{ \left[\frac{\partial F}{\partial K} - r \right] (1 - \tau_1) + \tau_2 \pi \right\} = 0$$

for $t \leq s < \infty$. These two equations together imply that

$$F_N = w/p \quad \dots [12]$$

and

$$(1 - \tau_1) F_K = (1 - \tau_1) r - \tau_2 \pi$$

or

$$F_K = r - \left[\frac{\tau_2}{1 - \tau_1} \right] \pi \quad \dots [13]$$

It can be noted immediately that corporation taxation has no effect on the firm's demand for labour. Equation [13] is now the firm's demand for capital in the presence of taxation. This result can be compared with that of Feldstein (1976, p.811, equation 4).

6.3.2. The various after-tax rates of return defined

It now remains to define the various net rates of return an individual will receive on his holding of corporate debt. The interest payments received by the individual as personal income are taxed at a personal income rate of θ_1 on the real component, and at θ_2 on

the inflationary component, so that the various definitions of net rates of return apply

$$R_n = (1 - \theta_1)r + (1 - \theta_2)\pi \quad \dots [14]$$

and

$$r_n = R_n - \pi = (1 - \theta_1)r - \theta_2\pi \quad \dots [15]$$

Gathering equations [2], [4]-[8], and [12]-[15], the model is now summarised in Table 6.1.

6.4. Effects of inflation (1)

The effect of taxes on the inflation-interest rate relationship can now be examined by taking the total differentials of equations [I]-[X] in Table 6.1., and the differential equations are presented in Table 6.2. It is clear from differential equation [x] that

$$\frac{dR}{d\pi} = \frac{dr}{d\pi} + 1$$

but from differential equation [ii]

$$\frac{dr}{d\pi} = F_{KK} \left[\frac{dK}{d\pi} \right] + F_{KN} \left[\frac{dN}{d\pi} \right] + \frac{\tau_2}{1 - \tau_1}$$

and assuming, as Feldstein (1976, p.810) does, that the labour

TABLE 6.1: *A neoclassical monetary growth model with all-debt financing.*

$$\begin{aligned}
 \text{[I]} \quad & F_N = w/p \\
 \text{[II]} \quad & r = F_K + [\tau_2/(1 - \tau_1)]\pi \\
 \text{[III]} \quad & Y = F(K, N) \\
 \text{[IV]} \quad & M/p = L(R_n) \cdot K \\
 \text{[V]} \quad & C = C(r_n) \cdot Y^D \\
 \text{[VI]} \quad & Y = C + I + G
 \end{aligned}$$

where

$$\begin{aligned}
 \text{[VII]} \quad & Y^D = Y(1 - \gamma) + n(M/p) \\
 \text{[VIII]} \quad & r_n = (1 - \theta_1)r - \theta_2\pi \\
 \text{[IX]} \quad & R_n = (1 - \theta_1)r + (1 - \theta_2)\pi
 \end{aligned}$$

Finally the nominal interest rate is defined as

$$\text{[X]} \quad R = r + \pi$$

TABLE 6.2: *Total differentials of all equations in Table 6.1.*

$$[i] \quad F_{NK}dK + F_{NN}dN = (dw/p) - (w/p)(dp/p)$$

$$[ii] \quad dr = F_{KK}dK + F_{KN}dN + [\tau_2/(1 - \tau_1)]d\pi$$

$$[iii] \quad dY = F_KdK + F_NdN$$

$$[iv] \quad (dM/p) - (M/p)(dp/p) = KL_RdR_n + LdK$$

$$[v] \quad dC = Y^D C_2 dr_n + C dY^D$$

$$[vi] \quad dY = dC + dI + dG$$

where

$$[vii] \quad dY^D = dY(1 - \gamma) + n[(dM/p) - (M/p)(dp/p)]$$

$$[viii] \quad dr_n = (1 - \theta_1)dr - \theta_2 d\pi$$

$$[ix] \quad dR_n = (1 - \theta_1)dr + (1 - \theta_2)d\pi$$

and finally,

$$[x] \quad dR = dr + d\pi$$

supply grows *exogenously* at an exponential rate of n per unit time, this would imply that $(dN/d\pi) = 0$. Then after substituting for $(dr/d\pi)$ in differential equation [x],

$$\frac{dR}{d\pi} = F_{KK} \left[\frac{dK}{d\pi} \right] + \frac{1 - \tau_1 + \tau_2}{1 - \tau_1} \dots [16]$$

6.4.1. Effect on capital intensity

At this stage, it is not possible to determine the sign of $dR/d\pi$ unless the sign of $dK/d\pi$ is determined as well. Therefore, substitute for all terms of the differential equation [vi] to get

$$(1 - \gamma)F_K dK = Y^D C_2 dr_n + C dY^D + n dK$$

after noting that $dI = n dK$ and $dG = \gamma dY$ and that $(dN/d\pi)$ is still equal to zero. After further substitution and manipulation of the preceding expression, an expression for $dK/d\pi$ is obtained as follows

$$\frac{dK}{d\pi} = \frac{C[(1 - \theta_1)(\tau_2/(1 - \tau_1)) + (1 - \theta_2)]nKL_R + Y^D C_2[(1 - \theta_1)(\tau_2/(1 - \tau_1)) - \theta_2]}{(1 - C)(1 - \gamma)F_K - CnL - n - CnKL_R(1 - \theta_1)F_{KK} - Y^D C_2(1 - \theta_1)F_{KK}}$$

After including the term $(nL - nL)$ in the denominator of the preceding expression, it can be shown that the following result is equivalent to that obtained by Feldstein (1976, p.813, equation 17) after letting $C = (1 - \sigma)$, and $-C_2 = \sigma'$ where σ is Feldstein's

notation for savings:

$$\frac{dK}{d\pi} = \frac{(1 - \sigma)[(1 - \theta_1)(\tau_2/(1 - \tau_1)) + (1 - \theta_2)]nKL_R - Y^D_{\sigma'}[(1 - \theta_1)(\tau_2/(1 - \tau_1)) - \theta_2]}{\sigma[(1 - \gamma)F_K + nL] - n(1 + L) - (1 - \sigma)nKL_R(1 - \theta_1)F_{KK} + Y^D_{\sigma'}(1 - \theta_1)F_{KK}} \dots [17]$$

Feldstein (1976, p.812) has shown that the denominator of equation [17] is negative if saving is a nondecreasing function of the net real rate of return, i.e. $\sigma' \leq 0$. If this condition holds, the denominator is unambiguously negative when $\sigma[(1 - \gamma)F_K + nL] - n(1 + L) < 0$.⁹ Thus, the sign of $dK/d\pi$ will be the opposite of the sign of the numerator. Considering the numerator, it can be seen that the first term is negative since the demand for money is inversely related to the nominal rate of interest, i.e. $L_R < 0$. Feldstein has argued that if $\sigma > 0$, the sign of the second term, and therefore the sign of the entire numerator, depends on the nature of the taxation system. Therefore, Feldstein (1976, p.813) points out two special cases in which the second term is zero. The first case arises when there is full indexing of the taxation of interest income, that is, the real component of the nominal interest rate is taxed only ($\theta_2 = 0$), and when corporation tax only allows the deductibility of real interest payments ($\tau_2 = 0$). The second case arises when there is no full indexing, but the corporation and personal income tax rates are equal ($\theta_1 = \theta_2 = \tau_1 = \tau_2$). When these two cases do occur, the second term of the numerator in equation [17] is zero. Therefore, the

numerator is unambiguously negative, and hence $dK/d\pi > 0$.

However, these two cases may not occur in practice. There is often less-than-perfect indexing of taxation of interest income, and the corporate tax rate is certainly not equal to the personal income tax rate. Whilst some form of implicit indexing scheme may operate (for example, in the United Kingdom, tax allowances have to be increased in line with inflation), such indexing is not always perfect. As a simplifying assumption, it is assumed that no full indexing is in operation so that $\theta_1 = \theta_2 = \theta$ and $\tau_1 = \tau_2 = \tau$ approximately. Consider the case in which the corporation tax rate is greater than the personal income tax rate, i.e. $\tau > \theta$. The effect of inflation on equilibrium capital intensity mainly stems from the reinforcing effect of the increase in the savings rate on the reduction in liquidity. This can be quite easily seen if equation [II] is substituted into equation [X] after letting $\tau_1 = \tau_2 = \tau$. Then the nominal rate of interest is

$$R = r + \pi = F_K + \pi/(1 - \tau) \quad \dots [18]$$

implying that, from equation [IX], the net nominal rate of interest is

$$R_n = (1 - \theta)F_K + \left[\frac{1 - \theta}{1 - \tau} \right] \pi \quad \dots [19]$$

Given that F_K is a constant, it is clear from [19] that the net nominal interest rate will rise with inflation, and since liquidity preference is a decreasing function of i_n (see equation [IV]), the demand for money decreases. If, from equation [VIII] the net real rate of interest is

$$r_n = (1 - \theta)F_K + \left[\frac{\tau - \theta}{1 - \tau} \right] \pi \quad \dots [20]$$

then any rise in the inflation rate causes the net real rate of interest to increase (because $\tau > \theta$). The rise in r_n causes an increase in the marginal propensity to save which reinforces the reduction in liquidity. Since savings are composed of money balances and corporate bonds (used to finance capital accumulation), any reduction in liquidity preference implies an increase in capital intensity. Thus, the effect of inflation is to increase $dK/d\pi$ which is positive.

Since coming into office in 1979, the present Government has reduced the rate of corporation tax from 52 to 35 per cent. Owing to this development, it would be useful to consider the possibility that the corporation tax rate may be less than the marginal personal income tax rate, that is $\tau < \theta$. The situation is more complex since an increase in the inflation rate will reduce the net real rate of return received by savers as can be readily ascertained from equation [20] with $\tau < \theta$. It is not possible to give an unambiguous *a priori* answer regarding the effect of inflation on capital intensity because the final outcome is dependent on the relative strength of the savings and liquidity effects. It may be recalled that inflation increases capital intensity if the numerator of equation [17] is negative. With the assumption that there is no indexation in the tax system, this would require the numerator of equation [17] to satisfy the following inequality, as shown by Feldstein (1976, p.814)

$$(1 - \sigma)nk(1 - \theta)L_R + Y^D(\theta - \tau)\sigma' < 0 \quad \dots [21]$$

After some manipulation, [21] is equivalent to

$$\frac{\eta_L}{\eta_S} > \frac{W}{(1 - \sigma)(M/p)} \left[\frac{\theta - \tau}{1 - \theta} \right] \frac{R_n}{r_n} \quad \dots [22]$$

where $\eta_L = -R_n L_R / L =$ the elasticity of real money balances relative to capital with respect to the nominal interest rate, $\eta_S = r_n \sigma' / \sigma =$ the elasticity of the propensity to save with respect to the net real rate of interest. When $\tau \geq \theta$, it is certain that inequality [22] will hold, but may not hold when $\tau < \theta$. It will be more likely that $dK/d\pi$ will be negative if the demand for real balances relative to capital is relatively interest-elastic (η_L is large) and if the savings rate is interest-inelastic (η_S is small). If the inequality is indeed false, then a rise in inflation will lead to a decrease in capital intensity (since the numerator of equation [17] is positive). The effects of inflation on interest rates will now be considered.

6.4.2. Effect on gross interest rates

Fisher's relation between inflation and nominal interest rates is only valid in an economy where there is no taxation. However, it is clear from the preceding equations for the interest rate variables that taxation does play an important role in determining the net real rate of interest. In an economy without taxation, the nominal interest rate, the real interest rate, and the net real interest rate are all equivalent, but not so when there is taxation. In the most general case, substitution of equation [II] from Table 6.1. into [X] and [VIII] shows that the nominal and net real interest rates are respectively

$$R = F_K + \left[\frac{1 - \tau_1 + \tau_2}{1 - \tau_1} \right] \pi \quad \dots [23]$$

and

$$r_n = (1 - \theta_1)F_K + \left[\frac{(1 - \theta_1)\tau_2 - (1 - \tau_1)\theta_2}{1 - \tau_1} \right] \pi \quad \dots [24]$$

Now consider the effect of inflation on the nominal interest rate. Differentiation of [23] with respect to π (or substitution of differential equation [ii] from Table 6.2. into [x]) gives

$$\frac{dR}{d\pi} = \frac{1 - \tau_2 + \tau_1}{1 - \tau_1} + F_{KK} \left[\frac{dK}{d\pi} \right] \quad \dots [16 \text{ repeated}]$$

Fisher's conclusion that $dR/d\pi = 1$ corresponds to the case when there are no taxes, and an interest inelastic demand for money.¹⁰ In the more general case in which taxes are recognised, the nominal interest rate may rise substantially more than the rate of inflation. Without full tax indexing, i.e. $\tau_1 = \tau_2 = \tau$, [16] becomes

$$\frac{dR}{d\pi} = \frac{1}{1 - \tau} + F_{KK} \left[\frac{dK}{d\pi} \right] \quad \dots [25]$$

With no change in capital intensity, $dR/d\pi = 1/(1 - \tau)$. Thus

with a corporation tax of 35 percent, $dR/d\pi = 1.54$. That is, the nominal interest rate would have to rise by almost one-and-half times as much as the rate of inflation. If $\tau > \theta$, $dK/d\pi$ will be positive, implying that the second term in [25] will be negative. So the nominal rate of interest may rise less than one-and-half times the rate of inflation given that $\tau = 0.35$. Similarly, when $\tau < \theta$, $dK/d\pi$ might be negative. In that case, the nominal rate of interest may rise by more than one-and-half times the rate of inflation.

With full tax indexing, i.e. $\tau_2 = 0$, equation [16] becomes

$$\frac{dR}{d\pi} = 1 + F_{KK} \left[\frac{dK}{d\pi} \right] \dots [26]$$

Here, with no change in capital intensity, the original conclusion by Fisher that $dR/d\pi = 1$ holds. It was shown previously that with full tax indexing, i.e. $\tau_2 = \theta_2 = 0$, $dK/d\pi$ is positive. Since $F_{KK}(dK/d\pi) < 0$, it follows that the nominal interest rate may rise by slightly less than the rate of inflation. So, either in the absence of taxation or full tax indexation, this conclusion may be loosely compared with those reached by both Bailey (1956) and Mundell (1963) which were discussed in Section 6.1.

6.4.3. Effect on net interest rates

It would also be useful to discuss the effect of inflation on the net real rate of interest. Differentiation of equation [24] with respect to π (or substitution of differential equation [ii] from Table 6.2. into [viii]) gives

$$\frac{dr_n}{d\pi} = \frac{(1 - \theta_1)\tau_2 - (1 - \tau_1)\theta_2}{1 - \tau_1} + (1 - \theta_1)F_{KK} \left[\frac{dK}{d\pi} \right] \dots [27]$$

If there are no taxes, and the demand for money is interest-inelastic, equation [27] implies that $dr_n/d\pi = 0$ which is in line with Fisher's original conclusion that (in theory) the real interest rate is unaffected by inflation. When the effect of inflation on capital intensity was discussed previously, two special cases were considered, viz: full tax indexing of interest ($\theta_2 = \tau_2 = 0$), and equality of corporation and personal income tax rates ($\tau_1 = \tau_2 = \theta_1 = \theta_2$). Considering these cases, it can be shown that the first term of equation [27] vanishes so that $dr_n/d\pi$ only depends on the second term $(1 - \theta_1)(dK/d\pi)F_{KK}$ which is negative.¹¹ Therefore, the net real rate of interest may decline slightly as the inflation rate increases (c.f. Mundell (1963))

However, in general, with taxes, inflation can have a very substantial (and in some cases, detrimental) effect on the net real rate of interest. If there is no full tax indexing of interest, equation [27] becomes

$$\frac{dr_n}{d\pi} = \frac{\tau - \theta}{1 - \tau} + (1 - \theta)F_{KK} \left[\frac{dK}{d\pi} \right] \dots [28]$$

If the corporation tax rate exceeds the personal tax rate, the first term of equation [28] is positive. Recall that when $\tau > \theta$,

$dk/d\pi > 0$. Therefore, the second term is negative. An unambiguous answer concerning the effect of inflation on the net real rate of interest cannot be given. The likelihood of $dr_n/d\pi$ being negative seems to increase if the difference between the corporation tax rate and personal income tax rate becomes small. If $\tau < \theta$, the first term will be negative. Assuming an interest-inelastic demand for real money balances, it is possible that the second term will become positive. Again it is not possible to give an unambiguous answer regarding the sign of $dr_n/d\pi$, but as the negative difference between the corporation tax and personal income tax rates gets larger, the likelihood that $dr_n/d\pi$ will be negative will increase.

6.5. A critique of Feldstein's model

The main criticism to be made here of Feldstein's model is that it does not postulate that the demand for money is a function of the net nominal interest rate only, and not of income as well. It will now be shown that the income elasticity of the demand for money plays an important role in deciding whether the sign of $dK/d\pi$ can be determined at all. So, it will be useful to re-consider the demand for money in some detail.

6.5.1. A re-consideration of the demand for money function

All households desire to allocate their wealth between money and corporate debt whose demand schedules are as follows:

$$\frac{D}{M} / p = L(R_n, Y, W) \quad \dots [29]$$

and

$$B^D/p = B(R_n, Y, W) \quad \dots [30]$$

where the superscript D denotes desired quantities, and R_n is the net nominal interest rate (to be defined later on). The above two demand schedules are constructed such that for all R_n , Y , and W , they will satisfy the following equation:

$$(M^D/p) + (B^D/p) = W \quad \dots [31]$$

As pointed out by Tobin (1969, pp.18 - 20), when the sum of real assets are constrained to satisfy balance sheet identities, the partial derivatives of the above two demand schedules are related in a certain way. Thus, taking the total differential of equation [29] and adding this to the total differential of equation [30] yields the following expression:

$$d((M^D + B^D)/p) = (L_R + B_R)dR_n + (L_Y + B_Y)dY + (L_W + B_W)dW$$

Subtraction of the total differential of equation [31] from the preceding expression yields the following expression

$$0 = (L_R + B_R)dR_n + (L_Y + B_Y)dY + (L_W + B_W - 1)dW$$

For this expression to hold at all times, it is necessary that the following conditions are satisfied:

$$(L_R + B_R) = 0, \quad (L_Y + B_Y) = 0, \quad \text{and} \quad (L_W + B_W) = 1$$

It is assumed then that the above conditions characterise the asset demand functions of equations [29] and [30]. Notice that Feldstein (1976, p.810) argues that in the long-run steady state, the ratio of real money balances to the real capital stock must remain constant, and such a condition can be quite easily met by assuming that $L_R = -B_R$ and $L_Y = B_Y$. Portfolio equilibrium requires that households be satisfied with their allocation of their wealth between money and corporate debt, that is

$$M^D/p = M/p \quad \text{and} \quad B^D/p = B/p$$

But notice that [3] (in Section 6.2.) and [31] together imply that either one of the above equations is sufficient to describe portfolio equilibrium; so it is only necessary to require that $M^D/p = M/p$ as Walras' Law will ensure that demands for both assets are exactly matched by their supplies. Therefore, portfolio equilibrium is characterised by:

$$M/p = M^D/p = L(R_n, Y, W)$$

where it is assumed that $L_R < 0$, $L_Y > 0$, and $L_W = 0$ so that

$$M/p = L(R_n, Y) \quad \dots [32]$$

which replaces the liquidity preference function in Table 6.1.

6.5.2. Relaxation of more assumptions

Two further assumptions of Feldstein's model will now be relaxed. Firstly, it is assumed that government expenditure is exogenously determined so that

$$G = \bar{G} \quad \dots [33]$$

where a 'bar' over a variable denotes exogeneity. It is still assumed that the supply of labour is exogenously determined, in which case

$$N = \bar{N} \quad \dots [34]$$

Secondly, the consumption function is assumed not to be proportional to disposable income, in which case

$$C = C(Y^D, r_n) \quad \dots [35]$$

where the marginal propensity to consume out of disposable income is positive but less than unity (i.e. $0 < C_1 < 1$, where $C_1 = \partial C / \partial Y^D$), and $C_2 < 0$. This replaces the consumption function in Table 6.1. The revised model is now summarised in Table 6.3.

6.5.3. Why are liquidity effects a problem?

The revised model can now be examined by taking the total differentials of all equations in Table 6.3., and these differential equations are presented in Table 6.4. As before, the sign of $dR/d\pi$ (as expressed in equation [16]) still depends on the sign of $dK/d\pi$. Therefore, substitution of all terms in the differential equation [vi] yields the following expression

TABLE 6.3: *Revised neoclassical monetary growth model with all-debt financing.*

$$\begin{aligned}
 \text{[I]} \quad & F_N = w/p \\
 \text{[II]} \quad & r = F_K + [\tau_2/(1 - \tau_1)]\pi \\
 \text{[III]} \quad & Y = F(K, N) \\
 \text{[IV]} \quad & M/p = L(R_n, Y) \\
 \text{[V]} \quad & C = C(Y^D, r_n) \\
 \text{[VI]} \quad & Y = C + I + G
 \end{aligned}$$

where

$$\begin{aligned}
 \text{[VII]} \quad & Y^D = Y - G + n(M/p) \\
 \text{[VIII]} \quad & N = \bar{N} \\
 \text{[IX]} \quad & G = \bar{G} \\
 \text{[X]} \quad & r_n = (1 - \theta_1)r - \theta_2\pi \\
 \text{[XI]} \quad & R_n = (1 - \theta_1)r + (1 - \theta_2)\pi
 \end{aligned}$$

Finally the nominal interest rate is defined as

$$\text{[XII]} \quad R = r + \pi$$

TABLE 6.4: *Total differentials of all equations in Table 6.3.*

$$[i] \quad F_{NK}dK + F_{NN}dN = (dw/p) - (w/p)(dp/p)$$

$$[ii] \quad dr = F_{KK}dK + F_{KN}dN + [\tau_2/(1 - \tau_1)]d\pi$$

$$[iii] \quad dY = F_KdK + F_NdN$$

$$[iv] \quad (dM/p) - (M/p)(dp/p) = L_RdR_n + L_YdY$$

$$[v] \quad dC = C_1dY^D + C_2dr_n$$

$$[vi] \quad dY = dC + dI + dG$$

where

$$[vii] \quad dY^D = dY - dG + n[(dM/p) - (M/p)(dp/p)]$$

$$[viii] \quad dN = d\bar{N}$$

$$[ix] \quad dG = d\bar{G}$$

$$[x] \quad dr_n = (1 - \theta_1)dr - \theta_2d\pi$$

$$[xi] \quad dR_n = (1 - \theta_1)dr + (1 - \theta_2)d\pi$$

and finally,

$$[xii] \quad dR = dr + d\pi$$

$$F_K dK = C_1 dY^D + C_2 dr_n + n dK + dG$$

After further substitution and manipulation, an expression for $dK/d\pi$ is obtained as follows:

$$\frac{dK}{d\pi} = \frac{C_1 n L_R [(1 - \theta_1)(\tau_2 / (1 - \tau_1)) + (1 - \theta_2)] + C_2 [(1 - \theta_1)(\tau_2 / (1 - \tau_1)) - \theta_2] + (1 - C_1)(dG/d\pi)}{F_K - C_1(1 + nL_Y)F_K - (1 - \theta_1)[C_1 n L_R + C_2]F_{KK} - n}$$

but, by assumption, $dG/d\pi = 0$, so that

$$\frac{dK}{d\pi} = \frac{C_1 n L_R [(1 - \theta_1)(\tau_2 / (1 - \tau_1)) + (1 - \theta_2)] + C_2 [(1 - \theta_1)(\tau_2 / (1 - \tau_1)) - \theta_2]}{F_K - C_1(1 + nL_Y)F_K - (1 - \theta_1)[C_1 n L_R + C_2]F_{KK} - n} \dots [36]$$

Considering the sign of the numerator, it is clear that the first term is negative since $C_1 > 0$, and $L_R < 0$. The sign of the second term in the numerator depends on the tax parameters as before. With regard to the denominator of [36], the last two terms are certainly negative. If the predictions of Feldstein's model are still to hold true, it is therefore necessary that the first two terms of the denominator satisfy the following inequality:

$$\frac{1}{C_1} < nL_Y + 1 \quad \dots [37]$$

The possible magnitude of L_Y cannot be properly ascertained without reference to the income elasticity of the demand for money, which is defined as $\varepsilon = L_Y(Y/L)$. If $\varepsilon = 1$, as suggested by the empirical evidence^{1 2}, then $L_Y = L/Y$ which is the proportion of real money balances to income, in which case inequality [37] becomes

$$C_1 > \frac{Y}{nL + Y}$$

This inequality states that if the sign of $dK/d\pi$ is to be determinate, it is necessary that the marginal propensity to consume out of disposable income be greater than or equal to $Y/(nL + Y)$. Now, the proportion of real money balances to income is less than unity, and most probably nearer to zero than to unity, that is, L is small relative to Y . Since the natural growth rate is small, it would follow that the ratio $Y/(nL + Y)$ is probably very close to unity. By implication, this would require a marginal propensity to consume that is close to unity implying that savings were negligible. Such an observation would not be supported by empirical evidence. It therefore follows that inequality [37] may not be satisfied so that the sign of $dK/d\pi$ remains indeterminate.

One possible way of overcoming the problem of determining the sign of $dK/d\pi$ would be to neglect real balance effects, as Summers (1983, p.205) has already done, so that all terms containing L vanish from equation [36]:

$$\frac{dK}{d\pi} = \frac{C_2[(1 - \theta_1)(\tau_2/(1 - \tau_1)) - \theta_2]}{(1 - C_1)F_K - (1 - \theta_1)C_2F_{KK} - n}$$

Summers (1983, p.208) then considers the special case in which consumption is interest inelastic ($C_2 = 0$). Clearly $dK/d\pi = 0$ so that from equation [25], $dR/d\pi = 1/(1 - \tau)$.

One further shortcoming of Feldstein's model is that it fails to take account of the possibility of equity financing, and this is now the subject of the following section.

6.6. Debt-equity financing model

6.6.1. Initial assumptions

The all-debt financing model was considered as a special case of a more general model involving both debt and equity finance. It was previously assumed that the firm financed its marginal unit of capital entirely by issuing corporate debt. That assumption is now relaxed so that the marginal unit of capital can also be financed by equity. In contrast with the all-debt model, the real after-tax rate of return on equity will enter the marginality conditions which makes it necessary to define e precisely. On the one hand, creditors of the firm are guaranteed a fixed nominal rate of return on their holdings of the firm's debt. On the other hand, equity holders are the last to make a claim on the firm's profits, and this makes the rate of return on equity variable.¹³ Therefore, in order to compensate equity holders for undertaking extra risks associated with uncertain rates of return on their investments, the firm must pay a risk premium (ρ) over and above the real after-tax rate of return on its

corporate debt so that investors may be induced to hold the firm's equity. The real after-tax rate of return on equity is now defined as $e = (1 - \theta)R - \pi + \rho$.

In order to keep the present analysis within manageable proportions, it will be assumed that the tax system is indifferent between real and inflationary components of income so that there is only one tax rate applied to both components (e.g. $\theta_1 = \theta_2 = \theta$ and $\tau_1 = \tau_2 = \tau$). Furthermore, the present model will include depreciation on capital, and the taxation system is so designed that full allowance is made for depreciation.

As before, the nominal after-tax rate of return on equity can be decomposed into two parts as follows:

$$(e + \pi)V(t) = (1 - \kappa)\dot{V}(t) + (1 - \epsilon)Div(t) \dots [38]$$

where κ is the capital-gains tax rate, and ϵ is the effective tax rate on dividends.

Now, equation [38] is recognised as a first-order differential equation whose solution is

$$V(t) = \int_t^{\infty} \frac{(1 - \epsilon)}{(1 - \kappa)} Div(s) \exp \left\{ \frac{-(e + \pi)}{(1 - \kappa)} (s - t) \right\} ds \dots [39]$$

after imposing the transversality condition.

6.6.2. The installment function

In defining dividends, it will be assumed that there are some adjustment costs for the firm carrying out its marginal investment

project. When the firm adjusts its capital stock, either by installing new capital, or by not replacing existing capital as it wears out, it will incur adjustment costs which can be characterised by, for example, managerial time and effort, and disruptions to production whilst the capital stock is being adjusted. Hayashi (1982, p.215) has shown that such adjustment costs may be introduced into the production process by an 'installment function'. This would be a function of K as well as I because the cost of installing I units of investment goods is likely to depend on the size of I relative to K . Letting $H(I,K)$ be the installment function, H will be an increasing and convex function of I so that $H_I > 0$, and $H_{II} > 0$. This would reflect the presumption that the cost of installment per unit of investment will be greater, the greater the rate of investment for any given K . A corollary of the above proposition is that if the capital stock increases relative to the rate of investment, then the cost of installment would fall at a declining rate (i.e. $H_K < 0$, and $H_{KK} > 0$).

6.6.3. The firm's optimal behaviour

The firm arrives at its taxable profits by deducting from gross revenue, its installment costs (pH), its wage bill (wN), and its interest payments on debt which is assumed to form a proportion b of the firm's total capital stock. Thus, the firm's interest payments amount to $pbKR$. The profits are then taxed at the corporate rate of τ to arrive at after-tax profits.¹⁴

In each period, the firm carries out its investment project which involves a gross investment expenditure of pI . Gross investment is defined as net investment (\dot{K}) plus depreciation (δK) where δ is the economic rate of depreciation so that

$I = \dot{K} + \delta K$. With the existence of debt finance, it is possible for the firm to finance a proportion b of its gross investment from corporate debt issues. Thus, the net cost of gross investment would be $(1 - b)pI$.

Since the value of corporate debt is fixed in nominal terms, equity holders stand to gain from inflation because creditors of the firm see the real value of their debt falling. The gain that accrues to equity holders is $pbK\pi$. Furthermore, in the preceding paragraph on investment, it was implicit that the firm only allowed for depreciation on a proportion $(1 - b)$ of its total capital stock. It is therefore necessary to make further allowances for depreciation on the remainder of the capital stock which amounts to $pb\delta K$. Then the total depreciation on the firm's capital stock is equivalent to $p\delta K$. If D is the term for depreciation allowances under a taxation system that makes full allowance for depreciation, then D is equal to $p\delta K$. Following the practice of Feldstein *et al* (1978, p.559) and Summers (1983, p.206), it is assumed that the firm uses historic-cost and first-in first-out (FIFO) inventory accounting conventions so that inflation causes depreciation requirements to be underestimated and taxable profits to be overstated respectively. Feldstein *et al* and Summers both use a parameter to denote the effect of inflation on the tax system which is denoted here by ι so that the firm suffers a capital depreciation equivalent to $pK(\iota\pi)$. The firm's dividends at instant s can now be defined as

$$\begin{aligned}
 Div(s) = & [pF(K, N) - pH(I, K) - wN - pbKR](1 - \tau) \\
 & - (1 - b)pI + D \\
 & + pbK(\pi - \delta) - pK(\iota\pi) \quad \dots [40]
 \end{aligned}$$

This definition of dividends can be compared with those supplied by Summers (1981, p.121, equation A-4) and Hayashi (1982, p.215, equation 2). In the former case, the main difference is that this thesis does not assume that installment costs are proportional to the rate of investment. Also, the depreciation allowances system here is much simpler than those formulated by Summers and Hayashi. Furthermore, the definition of dividends here excludes investment tax credits, but includes the effects of inflation on historic cost and FIFO inventory accounting conventions. Note also that when all-debt financing exists, and when there is no economic depreciation so that there is no need for depreciation allowances in the taxation system, and when there are zero adjustment costs (i.e. $b = 1$, $\delta = D = H = 0$), the definition of dividends becomes the one supplied in the all-debt model when $\tau_1 = \tau_2 = \tau$.

Equation [40] can then be substituted into equation [39] to give the firm's present value

$$\begin{aligned}
 V(t) = & \int_t^{\infty} \left[\frac{1 - \epsilon}{1 - \kappa} \right] \{ [pF(K, N) - pH(I, K) - wN - pbKR] (1 - \tau) \\
 & - (1 - b)pI + p\delta K \\
 & + pbK(\pi - \delta) - pK(\iota \pi) \} \mu_s ds \quad \dots [41]
 \end{aligned}$$

where it is to be understood that p , w , K , N , and I are all functions of time s , and where

$$\mu_s = \exp \left\{ \frac{-(e + \pi)}{(1 - \kappa)} (s - t) \right\}$$

Before proceeding to derive the marginality conditions, it is necessary to set a constraint on capital accumulation such that it will equal net investment, that is

$\dot{K} = I - \delta K$. Thus the constrained dynamic optimisation problem can be set up by introducing a shadow price, λ , for the constraint. Then equation [41] becomes:

$$\begin{aligned} L(t) = & \int_t^\infty \left\{ \frac{(1 - \epsilon)}{(1 - \kappa)} \left\{ [pF(K, N) - pH(I, K) - wN - pbKR](1 - \tau) \right. \right. \\ & - (1 - b)pI + p\delta K + pbK(\pi - \delta) - pK(\iota \pi) \left. \right\} \\ & \left. - \lambda(K - I + \delta K) \right\} \mu_s ds \quad \dots [42] \end{aligned}$$

Among the necessary conditions for an extremum are the following Euler equations:¹⁵

$$\frac{\partial L}{\partial N} = \frac{(1 - \epsilon)}{(1 - \kappa)} \mu_s \left\{ \left[p \frac{\partial F}{\partial N} - w \right] (1 - \tau) \right\} = 0,$$

$$\frac{\partial L}{\partial I} = \mu_S \left\{ \frac{(1 - \epsilon)}{(1 - \kappa)} \left[-pH_I(1 - \tau) - p(1 - b) \right] + \lambda \right\} = 0,$$

and

$$\frac{\partial L}{\partial K} - \frac{d}{ds} \left[\frac{\partial L}{\partial \dot{K}} \right] =$$

$$\mu_S \left\{ \frac{(1 - \epsilon)}{(1 - \kappa)} \left[[pF_K - pH_K - pbR](1 - \tau) + p\delta + pb(\pi - \delta) - p(\iota\pi) \right] - \lambda\delta - \left[\lambda[(e + \pi)/(1 - \kappa)] - \dot{\lambda} \right] \right\} = 0$$

The above equations all imply that

$$F_N = w/p \quad \dots [43],$$

$$\frac{(1 - \epsilon)}{(1 - \kappa)} \left[H_I(1 - \tau) + (1 - b) \right] = \frac{\lambda}{p} \quad \dots [44]$$

and

$$\dot{\lambda} = \lambda \left[\frac{(e + \pi)}{(1 - \kappa)} + \delta \right] - \left\{ \frac{(1 - \epsilon)}{(1 - \kappa)} p \left[[F_K - H_K - bR](1 - \tau) + \delta + b(\pi - \delta) - \iota\pi \right] \right\} \quad \dots [45]$$

Equation [43] is the familiar marginal productivity condition for labour which is clearly unaffected by the presence of taxes. Summers (1981, p.123) interprets equation [44] as characterising an implicitly defined investment function linking investment to the real shadow price of capital (λ/p), to the tax parameters, and to the cost of adjustment. Equation [45] can be recognised as a first order differential equation in λ . When this equation is solved for λ , λ can be regarded as the present discounted value of additional future after-tax profits that are due to one additional unit of investment. This can be quite easily seen if investment leads to an increase in the capital stock of the firm which would lead to an increase in F_K , and a decrease in H_K (i.e. increase in H_K in *absolute* terms) implying increased profitability. This has already been noted by Hayashi (1982, p.217).

6.6.4. Tobin's q -theory of investment under taxation

It is not too difficult to relate equations [44] and [45] with Tobin's q -theory of investment. Tobin (1969, p.19) suggests that the rate of investment is an increasing function of the ratio of the market value of new *additional* investment goods to their replacement cost. By defining marginal- q as the real shadow price of capital, that is $q = \lambda/p$, Hayashi (1982, p.217) has made his interpretation of q consistent with Tobin's q -theory of investment because of the reasonable presumption that the market value of an asset is determined by its net discounted present value of future profits as defined by λ . It is possible to solve equation [44] for I in terms of q and K so that

$$I = I(Q) \cdot K \quad \dots [46a]$$

where it is assumed that $I'(Q) > 0$, and that the investment function is linearly homogeneous in K , and Q is called the 'modified- q ' which is defined as follows

$$Q = \frac{\left[\frac{(1 - \kappa)}{(1 - \epsilon)} q - (1 - b) \right]}{(1 - \tau)} \quad \dots [47]$$

Now, in an all-equity financing model without taxation (i.e.

$b = \kappa = \epsilon = 0$), equation [46a] becomes

$$I = I(q - 1) \cdot K$$

Tobin's q -theory of investment postulates that when the market value of new additional investment goods is greater than its replacement cost (i.e. $\lambda > p$, implying $q > 1$), it will be profitable for the firm to invest until $\lambda = p$. The converse also holds true when $\lambda < p$. It is only when $q = 1$ that the firm will not undertake any investment at all. Another way of putting Tobin's q -theory of investment is to regard investment as a function of the difference between the real shadow price of the marginal unit of capital and its replacement cost. When the firm finances its marginal unit of capital entirely by issues of equity in the absence of taxation, $Q = q - 1$ because the cost of the marginal unit of capital to the firm is simply the cost of obtaining funds by equity issues. The firm will not undertake any investment when the real shadow price of capital equals the cost of capital, that is when $Q = 0$.

Equation [47] shows the modified- q when the firm finances its marginal unit of capital both by debt and equity issues in the

presence of taxation. In the discussion leading up to the definition of the firm's dividends in equation [40], it was argued that the existence of debt finance reduced the cost of capital to the firm. In contrast with the case of all-equity finance without taxes, the firm would have to increase its rate of investment until the real shadow price of capital equals its cost, that is

$$q = \frac{(1 - \epsilon)}{(1 - \kappa)} (1 - b) \quad \dots [48]$$

However, there is a small difficulty here. If, as Summers (1981, p.84, equation 11) suggests, the firm equates the real shadow price of capital to its cost (i.e. $Q = 0$), then the rate of investment would be equal to zero.¹⁶ But do consider the shadow price constraint, $\dot{K} = I - \delta K$. This would imply that there would be a decumulation of the capital stock equal to δK , and if, in the long-run steady-state, the standard assumption of a constant capital stock is made, then it would require a gross rate of investment of δK per unit time to keep the capital stock intact. In order to make this model consistent with that of Feldstein (1976) which assumes that the labour force grows exogenously at an exponential rate of n per unit time, it is necessary to make some modifications. Near the beginning of Section 6.3., an important distinction was made between the short- and long-run with regard to the firm's capital stock. Due to putty-clay capital in the short-run, the real shadow price of capital will deviate from the cost of capital. Consider what happens if the labour force grows exogenously at an exponential rate of n per unit time whilst the firm does not carry

out any gross investment at all. The capital stock will decline relative to the labour force so that the capital-labour ratio falls. Since it is being assumed that the marginal product of capital varies directly with employment, and inversely with the capital stock (see equation [2] in Section 6.2.), the marginal product of capital will rise whilst installment costs decline. This leads to an increase in λ so that $Q > 0$. The firm now has no option but to increase its rate of investment until it reaches the point when the rate of investment equals the natural rate of growth in the capital stock plus the rate of depreciation. Therefore, the long-run steady state equilibrium condition for investment is

$$I = (n + \delta)K = I(0) \cdot K \quad \dots [49]$$

when $Q = 0$, and equation [46] is the firm's investment function in the short-run. Such an equilibrium condition will now guarantee that the capital-labour ratio remains constant in the steady-state. A corollary of this observation is that in the long-run steady state, Q must remain equal to zero throughout time. Making use of the fact that $\dot{Q} = 0$, it is possible to derive the firm's demand for capital which takes account of both debt and equity finance, and of taxation. Now differentiation of equation [47] with respect to time gives

$$\dot{Q} = \frac{\left[\frac{(1 - \kappa)}{(1 - \epsilon)} \dot{q} \right]}{(1 - \tau)}$$

implying that, when $\dot{Q} = 0$, $\dot{q} = 0$. But since $q = \lambda/p$, it can be noted that

$$\dot{q} = \frac{\dot{\lambda}}{p} - \frac{\lambda}{p} \left[\frac{\dot{p}}{p} \right]$$

Substitution for $\dot{\lambda}/p$ (from equation [45]) into the preceding expression gives

$$\begin{aligned} \dot{q} = & q \left[\frac{(e + \pi)}{(1 - \kappa)} + \delta - \pi \right] \\ & - \left\{ \frac{(1 - \epsilon)}{(1 - \kappa)} \left[[F_K - H_K - bR](1 - \tau) + \delta + b(\pi - \delta) - \iota\pi \right] \right\} \end{aligned}$$

after noting that $\dot{p}/p = \pi$ in the long-run steady-state. The preceding expression can then be solved to obtain another investment equation so that there is a set of two simultaneous investment equations in [46a] and [46b].

$$\begin{aligned} I = I \left\{ \left[\frac{(1 - \kappa)}{(1 - \epsilon)} \dot{q} - \frac{1}{(1 - \epsilon)} q \left[(e + \pi) - (1 - \kappa)(\pi - \delta) \right] \right. \right. \\ \left. \left. + \left[(F_K - bR)(1 - \tau) + \delta + b(\pi - \delta) + \iota\pi \right] \right] \frac{1}{(1 - \tau)} \right\} \cdot K \\ \dots [46b] \end{aligned}$$

where it is assumed that equation [46b] is also linearly homogeneous in K . In the long run steady state, $\dot{q} = 0$, and the expression

within the outer brackets should equal zero. This implies that, after some re-arrangement, the expression within the outer brackets becomes

$$q = \frac{(1 - \epsilon)[(F_K - bR)(1 - \tau) + \delta + b(\pi - \delta) - \iota \pi]}{(e + \pi) - (1 - \kappa)(\pi - \delta)} \dots [50]$$

The two conditions for q set out in equations [48] and [50] will both guarantee that investment in equations [46a] and [46b] will equal the steady-state growth in the capital stock of $(n + \delta)K$.

6.6.5. The firm's demand for capital in a debt-equity model

The final step is to derive an expression linking the marginal product of capital with the real interest rate and the rate of inflation. As explained previously, the firm will be carrying out the optimal rate of investment when the real shadow price of capital equals its cost which requires that equation [48] must hold. Therefore substitution for q from equation [48] into equation [50], and re-arrangement gives the firm's demand for capital when the marginal unit of capital can be financed both by debt and equity, and when there is economic depreciation of which full provision is made for in the taxation system: (continued on next page)

$$\begin{aligned}
F_K &= \left\{ \frac{(1-b)(1-\theta)}{(1-\kappa)(1-\tau)} + b \right\} r \\
&+ \left\{ \frac{(1-b)(1-\theta)}{(1-\kappa)(1-\tau)} + b - \frac{1}{(1-\tau)} (1-\iota) \right\} \pi \\
&+ \left\{ \frac{(1-b)}{(1-\kappa)(1-\tau)} \right\} \rho \quad \dots [51]
\end{aligned}$$

It is clear that when all-debt financing is assumed ($b = 1$), and when there are no effects of inflation on historic cost accounting conventions ($\iota = 0$), equation [51] reduces to equation [13] showing the firm's demand for capital when $\tau_1 = \tau_2 = \tau$. The reason for the existence of the first term within the curly brackets in the firm's demand for capital is that the firm finances its marginal unit of capital not only by debt, but by equity issues as well. In doing so, the firm has to take account of the tax rates on capital gains and dividends. For simplicity, it will be assumed that the debt-equity ratio, b , is exogenously determined.¹⁷ The effect of having introduced equity finance into this model is to reduce the firm's demand for capital because the firm now has to strive for a higher marginal product of capital before equation [51] can hold. For convenience, equation [51] is re-arranged so that

$$r = \alpha F_K - \nu \pi - P \quad \dots [52]$$

where

$$\begin{aligned}
\alpha &= [(1-\kappa)(1-\tau)] / [(1-b)(1-\theta) + b(1-\kappa)(1-\tau)], \\
\nu &= 1 - [(1-\kappa)(1-\iota)] / [(1-b)(1-\theta) + b(1-\kappa)(1-\tau)],
\end{aligned}$$

and

$P = [(1 - b)/(1 - b)(1 - \theta) + b(1 - \kappa)(1 - \tau)]^{\rho}$, and it is assumed that $dP = 0$.

Before the revised monetary growth model assuming debt-equity finance can be set out, it is necessary to make a few minor modifications to reflect changes in the assumptions made regarding taxation and depreciation. Since depreciation charges on the capital stock reduce disposable income by δK , equation [VII] in Table 6.3. is modified accordingly. The definitions of the after-tax rates of interest are also modified to reflect the assumption that the taxation system is indifferent between real and inflationary components of income. The revised model is summarised in Table 6.5 with the addition of an investment equation.

6.7. Effects of inflation (2)

6.7.1. Effect on gross interest rates

The model set out in Table 6.5. can now be examined by taking the total differentials of all equations which are then presented in Table 6.6. As before, equation [xiii] of Table 6.6. shows that

$$\frac{dR}{d\pi} = \frac{dr}{d\pi} + 1$$

But from equation [ii],

$$\frac{dr}{d\pi} = \alpha F_{KK} \left[\frac{dK}{d\pi} \right] - \nu$$

TABLE 6.5: *Revised neoclassical monetary growth model with debt-equity financing.*

$$\begin{aligned}
 \text{[I]} & \quad F_N = w/p \\
 \text{[II]} & \quad r = \alpha F_K - \nu \pi - P \\
 \text{[III]} & \quad Y = F(K, N) \\
 \text{[IV]} & \quad M/p = L(R_n, Y) \\
 \text{[V]} & \quad C = C(Y^D, r_n) \\
 \text{[VI]} & \quad I = (n + \delta)K \\
 \text{[VII]} & \quad Y = C + I + G
 \end{aligned}$$

where

$$\begin{aligned}
 \text{[VIII]} & \quad Y^D = Y - G + n(M/p) - \delta K \\
 \text{[IX]} & \quad N = \bar{N} \\
 \text{[X]} & \quad G = \bar{G} \\
 \text{[XI]} & \quad r_n = (1 - \theta)r - \theta \pi \\
 \text{[XII]} & \quad R_n = (1 - \theta)(r + \pi)
 \end{aligned}$$

Finally the nominal interest rate is defined as

$$\text{[XIII]} \quad R = r + \pi$$

TABLE 6.6: *Total differentials of all equations in Table 6.5.*

- [i] $F_{NK}dK + F_{NN}dN = (dw/p) - (w/p)(dp/p)$
- [ii] $dr = \alpha(F_{KK}dK + F_{KN}dN) - \nu d\pi$
- [iii] $dY = F_KdK + F_NdN$
- [iv] $(dM/p) - (M/p)(dp/p) = L_RdR_n + L_YdY$
- [v] $dC = C_1dY^D + C_2dr_n$
- [vi] $dI = (n + \delta)dK$
- [vii] $dY = dC + dI + dG$

where

- [viii] $dY^D = dY - dG + n[(dM/p) - (M/p)(dp/p)] - \delta dK$
- [ix] $dN = d\bar{N}$
- [x] $dG = d\bar{G}$
- [xi] $dr_n = (1 - \theta)dr - \theta d\pi$
- [xii] $dR_n = (1 - \theta)(dr + d\pi)$

and finally,

- [xiii] $dR = dr + d\pi$

after assuming that $dN/d\pi = 0$. Thus,

$$\frac{dR}{d\pi} = \alpha F_{KK} \left[\frac{dK}{d\pi} \right] + (1 - \nu) \dots [53]$$

It can be seen that the term $1 - \nu$ is unambiguously positive, and that the sign of $dR/d\pi$ depends on the sign of $dK/d\pi$ as before. By neglecting real balance effects, it can be shown that

$$\frac{dK}{d\pi} = \frac{-C_2 [(1 - \theta)\nu + \theta]}{(1 - C_1)F_K - C_2(1 - \theta)\alpha F_{KK} - (n + \delta) + C_1\delta}$$

By assuming that consumption is interest inelastic, $dK/d\pi$ equals zero so that $dR/d\pi = 1 - \nu$ or

$$\frac{dR}{d\pi} = \frac{(1 - \kappa)(1 - \iota)}{(1 - b)(1 - \theta) + b(1 - \kappa)(1 - \tau)} \dots [54]$$

This equation is different from that derived by Summers (1983, p.208, equation 4) because it makes explicit allowance for the debt-equity ratio, and capital gains taxes whereas Summers assumes that these parameters are subsumed within his definition of tax parameters.¹⁸ It is therefore possible (and more enlightening) to analyse the effects of a change in the debt-equity ratio on $dR/d\pi$. Consider the case in which the firm finances its marginal unit of capital by debt only. Then $dR/d\pi = (1 - \iota)/(1 - \tau)$. This shows that when inflation starts to have a detrimental effect on historic cost and FIFO inventory accounting conventions, the nominal

interest rate will rise by less than one-and-half times the rate of inflation given that the current corporation tax rate is 35 per cent.

6.7.2. A comparison with the all-debt model

In order to examine the effect of introducing equity finance on the inflation-interest rate relationship, it is necessary to differentiate equation [54] with respect to b . This gives

$$\frac{d^2R}{d\pi db} = \frac{-(1 - \kappa)(1 - \iota)[(1 - \kappa)(1 - \tau) - (1 - \theta)]}{\{(1 - \theta) + b[(1 - \kappa)(1 - \tau) - (1 - \theta)]\}^2}$$

The denominator of the above expression is unambiguously positive, and the sign of $d^2R/d\pi db$ depends on the term in square brackets in the numerator. Given that the current British tax parameters are $\theta = 0.29$, $\tau = 0.35$, and $\kappa = 0.3$, it follows that the numerator is positive so that $d^2R/d\pi db > 0$. Thus, under the current tax laws, the introduction of equity finance into the model decreases $dR/d\pi$. The nominal interest rate will now have to rise by less for each point rise in the inflation rate than in the case of all-debt finance. Such a prediction is consistent with those comments made by Patterson and Ryding (1984, p.302) regarding the introduction of equity finance and personal capital gains taxes.

So far, this thesis has shown that in the long-run steady-state, a definite relationship between the inflation rate and the nominal interest rate can exist. But, now an important question needs to be asked: can the inflation-interest rate relationship hold in the short-run? This will have important implications for empirical research because if it is shown that no definite relationship exists at

all, then any empirical research on the relationship between short-term interest rates and inflation rates is unlikely to be fruitful. This question is the subject of the following section.

6.8. The short-run versus the long-run

The main essential difference between the long-run steady-state and the short-run is that the economy is allowed to deviate from its long-run equilibrium path in the short-run. So, the long-run equilibrium conditions stated in this model may not hold, and it is therefore necessary to introduce some modifications.

First, consider the labour market. It will be assumed that wages follow a time path described by the following equation

$$w(s) = w \exp^{\pi(s - t)}$$

so that the rate of growth in wages is equal to the actual rate of inflation (\dot{p}/p) which equals the anticipated rate of inflation (π) in the long-run, that is

$$\frac{\dot{w}}{w} = \pi \quad \dots [55]$$

The firms' demand for labour is a function of the real wage rate, and it is assumed that the labour supply is exogenously determined so that the labour market is described by the following two equations:

$$N = F_N^{-1}(w/p) \quad \text{and} \quad N^S = \bar{N}$$

A necessary condition for long-run equilibrium in the labour market is that the firms must be able to meet all their labour requirements at the going real wage rate, and that households supply all the labour they wish. When the labour market is in equilibrium, there is neither excess demand or supply so that

$$N = N^S$$

By introducing the Keynesian assumption that prices and wages are inflexible in the short run, it can be postulated that the labour market will not be in equilibrium during the short-run so that $N \neq N^S$. Thus, the rate of growth in money wages will be determined by the expected rate of inflation and also by the ratio of excess demand to labour supply in the labour market

$$\frac{\dot{w}}{w} = \nu \left[\frac{N}{N^S} - 1 \right] + \pi \quad \dots [56]$$

which can be recognised as the short-run Phillips relation as included in Summers' model (1983, p.204, equation 1i). Clearly, when the labour market is in equilibrium, equation [56] reduces to equation [55].

Secondly, a further distinction is made between the short-run and the long-run in the context of inflationary expectations. As before, it will be assumed that in the long-run steady-state, the expected rate of inflation equals the actual rate of inflation so that

$$\pi = \frac{\dot{p}}{p}$$

Summers (1983, p.205) postulates that inflationary expectations are formed adaptively. In this monetary growth model, it is assumed that expectations are formed of the rate of change in the inflation rate so that

$$\frac{\dot{\pi}}{\pi} = h \left[\frac{\dot{p}}{p} - \pi \right] \quad \dots [57]$$

where $h > 0$. In contrast with Summers' own formulation, the preceding expression states that the expected rate of change in the inflation rate is an increasing function of the difference between the actual rate and expected rate of inflation. Such a formulation will permit the existence of a steady rate of inflation to be expected in the long-run which is assumed to equal the excess rate of growth in the money supply over the natural rate of growth.

Thirdly, the investment equation set out in equation (VI) of Table 6.5. will not hold since Q will not be equal to zero. Therefore, the short-run investment equation, given in equation [46a] is introduced into the model. In order to analyse the model in the short-run, it is necessary to know the signs of the partial derivatives of Q . The signs of the partial derivatives of Q can be determined by taking the total differential of Q , and setting the relevant differentials equal to zero. From equations [47] and [50], it can be shown that

$$\begin{aligned}
& [(1 - \tau)dQ] \cdot [(1 - \theta)R + \rho - (1 - \kappa)(\pi - \delta)] \\
& \quad + [(1 - \tau)Q + (1 - b)] \cdot [(1 - \theta)dR - (1 - \kappa)d\pi] \\
& = (1 - \kappa) \left\{ (F_{KN}dN + F_{KK}dK - bdR)(1 - \tau) + (b - \iota)d\pi \right\}
\end{aligned}$$

where it is being assumed that $d\delta = d\rho = 0$, and that the debt-equity ratio is invariant during the short-run. Setting $dK = dR = d\pi = 0$, the partial derivative of Q with respect to N is

$$\frac{\partial Q}{\partial N} = \frac{(1 - \kappa)F_{KN}}{(1 - \theta)R + \rho - (1 - \kappa)(\pi - \delta)}$$

If the sign of Q_N is to be determinate, it is necessary to make some assumptions regarding the parameters in the above partial differential. Consider the denominator first: it can be seen that the sign depends on the last term of the denominator. For the denominator to be *unambiguously* positive, it is required that the rate of inflation be less than the economic rate of depreciation. With such an assumption, it is clear that $Q_N > 0$. By similar reasoning, it can also be shown that $Q_K < 0$. These conclusions can be justified when it was explained previously that if the capital stock declined relative to the labour supply, Q would increase. This is the reason why the long-run equilibrium condition for investment was formulated in equation [49]. The signs of the partial derivatives of Q with respect to the nominal rate of interest and the rate of inflation are much more difficult to determine because it all depends on the sign of q . Considering Q_R first, it can be shown that

$$\frac{\partial Q}{\partial R} = \frac{-\{(1 - \theta)[(1 - \tau)Q + (1 - b)] + b(1 - \tau)(1 - \kappa)\}}{(1 - \tau)[(1 - \theta)R + \rho - (1 - \kappa)(\pi - \delta)]}$$

But from equation [47], $(1 - \tau)Q + (1 - b) = [(1 - \kappa)/(1 - \epsilon)]q$, and it follows that Q_R depends on the sign of q . If the capital stock declines relative to the labour force, the term within the curly brackets in the numerator will be unambiguously positive because q is positive. Thus $Q_R < 0$. However, if the capital stock increases relative to the labour supply, Q will become negative since the real shadow price of capital will now be less than the cost of capital. It is still possible that Q_R will remain negative if the firm carrying out its investment plans is rational in the sense that it will *not* undertake any more investment if the nominal shadow price becomes negative. This can be quite easily seen if equation [45] is considered again: a firm with a disproportionately large capital stock can expect a negative marginal product of capital so that it becomes possible that the solution to the differential equation will become negative (i.e. $\lambda < 0$). Thus, as long as the shadow price of capital remains positive, it will be certain that $Q_R < 0$. So, if the nominal interest rate rises, Q will decline and thus reduce investment. Such observations are consistent with the Keynesian theory of investment.

It can be shown that the partial derivative of Q with respect to the rate of inflation is

$$\frac{\partial Q}{\partial \pi} = \frac{(1 - \kappa)\{[(1 - \tau)Q + (1 - b)] + (b - \iota)\}}{(1 - \tau)[(1 - \theta)R + \rho - (1 - \kappa)(\pi - \delta)]}$$

It can be seen that the sign of the numerator, and therefore the sign of Q_π , depends on the last term in the numerator. Recall that the term ι stands for the effect of inflation on historic cost accounting and FIFO inventory accounting conventions. Feldstein *et al* (1978, p.S68-S69) have given a 'conservative' estimate of the value of ι as being around 0.2. On that basis, with a relatively high debt-equity ratio, the numerator will be positive so that $Q_\pi > 0$. However, the conclusions are less clear-cut for the case when a relatively low debt-equity ratio is assumed so that the numerator is of ambiguous sign. To overcome this difficulty, it will be assumed that $b > \iota$. Thus, if the inflation rate rises, Q will rise. This is not too difficult to see if, *ceteris paribus*, the inflation rate reduces the real interest rate so that investment becomes more worthwhile.

It will be convenient to write the modified- q in a more general form so that

$$] \quad Q = Q(K, N, R, \pi) \quad \dots [58$$

where $Q_N > 0$, $Q_K < 0$, $Q_R < 0$, and $Q_\pi > 0$.

Finally, some stochastic terms are included in the aggregate production function, the consumption function, and the liquidity preference function (or LM locus). The purpose of these stochastic terms is to denote random exogenous shocks to the economy that might cause it to deviate from its long-run equilibrium path. Thus, letting ϵ denote the stochastic terms, the relevant equations are modified such that

$$Y = F(K, N)[1 + \epsilon_1] \quad \dots [59]$$

$$C = C(Y^D, r_n) + \varepsilon_2 \quad \dots [60]$$

$$M/p + \varepsilon_3 = L(R_n, Y) \quad \dots [61]$$

Gathering all equations, the short-run model which is almost identical to that of Summers (1983) is summarised in Table 6.7.

It has been argued by Summers (1983, pp.206-207) that in the short-run, there can be no definite relationship between the rate of inflation and the nominal interest rate. His argument rests on two examples of exogenous shocks to the economy that might occur in the short-run. The first example given is an exogenous demand shock brought about by an exogenous change in government expenditure, or by the stochastic term ε_3 which could be regarded as an exogenous change in consumption habits. The immediate effects of such shocks can be analysed by substituting the differential equations [vi] and [vii] into equation [viii] of Table 6.8.¹⁹, after noting that from equation [ii]

$$dN = \frac{dY - F_K dK(1 + \varepsilon_1) - F d\varepsilon_1}{F_N(1 + \varepsilon_1)}$$

(continued...)

TABLE 6.7: *Short-run monetary growth model.*

- [I] $F_N = w/p$
- [II] $Y = F(K, N)[1 + \varepsilon_1]$
- [III] $\frac{\dot{w}}{w} = \nu \left[\frac{N}{N^S} - I \right] + \pi$
- [IV] $M/p + \varepsilon_2 = L(R_n, Y)$
- [V] $\frac{\dot{\pi}}{\pi} = h \left[\frac{\dot{p}}{p} - \pi \right]$
- [VI] $C = C(Y^D, r_n) + \varepsilon_3$
- [VII] $I = I(Q) \cdot K$
- [VIII] $Y = C + I + G$
- where
- [IX] $Y^D = Y - G + n(M/p) - \delta K$
- [X] $N^S = \bar{N}$
- [XI] $G = \bar{G}$
- [XII] $Q = Q(N, K, R, \pi)$
- [XIII] $r_n = (1 - \theta)R - \pi$
- [XIV] $R_n = (1 - \theta)R$

TABLE 6.8: Total differentials of equations in Table 6.7.

$$[i] \quad F_{NK}dK + F_{NN}dN = (dw/p) - (w/p)(dp/p)$$

$$[ii] \quad dY = [F_KdK + F_NdN] \cdot [1 + \varepsilon_1] + Fd\varepsilon_1$$

$$[iii] \quad d(\dot{w}/w) = \nu'(dN/N^S) + d\pi$$

$$[iv] \quad (dM/p) - (M/p)(dp/p) + d\varepsilon_2 = L_RdR_n + L_YdY$$

$$[v] \quad d(\dot{\pi}/\pi) = h'd(\dot{p}/p) - h'd\pi$$

$$[vi] \quad dC = C_1dY^D + C_2dr_n + d\varepsilon_3$$

$$[vii] \quad dI = I_KdQ + IdK$$

$$[viii] \quad dY = dC + dI + dG$$

where

$$[ix] \quad dY^D = dY - dG + nd(M/p) - \delta dK$$

$$[x] \quad dN^S = d\bar{N}$$

$$[xi] \quad dG = d\bar{G}$$

$$[xii] \quad dQ = Q_NdN + Q_KdK + Q_RdR + Q_\pi d\pi$$

$$[xiii] \quad dr_n = (1 - \theta)dR - d\pi$$

$$[xiv] \quad dR_n = (1 - \theta)dR$$

so that the total differential of the IS locus is as follows

$$\left[1 - C_1 - nC_1L_Y - \frac{I^KQ_N}{F_N(1 + \varepsilon_1)} \right] dY =$$

$$(1 - C_1)dG + [I^KQ_K - I^KQ_N(F_K/F_N) - C_1\delta + I]dK$$

$$+ [(1 - \theta)(nC_1L_R + C_2) + I^KQ_R]dR$$

$$+ (I^KQ_\pi - C_T)d\pi - \frac{I^KQ_N F d\varepsilon_1}{F_N(1 + \varepsilon_1)} - nC_1 d\varepsilon_2 + d\varepsilon_3$$

... [62]

and, from equation [iv] of Table 6.8., the total differential of the LM locus is

$$d(M/p) + d\varepsilon_2 = L_R(1 - \theta)dR + L_Y dY \quad \dots [63]$$

Taking the partial derivatives of output with respect to G and ε_3 , the following expressions show that

$$\frac{\partial Y}{\partial G} = \frac{(1 - C_1)}{y} \quad \text{and} \quad \frac{\partial Y}{\partial \varepsilon_3} = \frac{1}{y}$$

where y is the coefficient to dY in equation [62]. It was argued towards the end of Section 6.5., when criticising the model of Feldstein (1976), that the expression comprising the first three terms in y was unlikely to be negative, and that if the income elasticity of the demand for money was equal to unity, the term nC_1L_Y was likely to become negligible so that the first three terms in y will

closely approximate the marginal propensity to save out of disposable income. It can be shown that the last term in y is equivalent to the marginal propensity to invest out of national income:

$$\frac{\partial I}{\partial Y} = \frac{\partial I}{\partial Q} \cdot \frac{\partial Q}{\partial N} \cdot \frac{\partial N}{\partial Y} = \frac{I'KQ_N}{F_N(1 + \epsilon_1)}$$

So, y will be positive as long as the marginal propensity to save out of disposable income (approximately) exceeds the marginal propensity to invest out of national income. Thus, on the basis of this assumption, either an increase in government expenditure or an exogenous shock will raise output as reflected in a rightward shift of the IS locus. Considering the effect on nominal interest rates, it can be shown that the partial derivatives of R with respect to G and ϵ_3 are both positive, so that an increase in G or ϵ_3 will cause nominal interest rates to rise.

The effect of an increase in output on prices can be determined by substituting dN into equation [i] of Table 6.8. Thus, the partial derivative of the price level with respect to output is

$$\frac{\partial p}{\partial Y} = \frac{-p^2 F_{NN}}{wF_N(1 + \epsilon_1)}$$

which is of positive sign. Thus, a rise in output will lead to rising prices. The rise in prices reduces real money balances which pushes up nominal interest rates even further. This continues until output is restored to its former rate. However, rising prices cause individuals

to expect that the inflation rate will begin to rise so that money wages will tend to rise faster than the actual rise in prices. This leads to an increase in the real wage rate which reduces the demand for labour by firms. The decline in employment causes output to decline. As prices decline, real money balances will rise, leading to a fall in interest rates. On the whole, an exogenous demand shock leads to an eventual decline in output, prices and nominal interest rates. Notice especially that there is positive correlation between the inflation rate and nominal interest rates in this example.

Summers then considers a second example in which there is a liquidity shock as reflected by an exogenous increase in real money balances or in the stochastic variable ε_2 . As can be verified from the total differential of the LM locus shown in equation [63], the initial effect of an increase in real money balances is to reduce nominal interest rates as prices start to rise. Rising prices reduce real money balances so that nominal interest rates start to rise as the inflation rate begins to fall. The economy will ultimately converge on to a new equilibrium at the former level of output and nominal interest rates, with higher prices. In such an example, it is interesting to note that a *negative* Fisher effect could be observed.

Thus, Summers concludes his discussion of the above two examples in the following words:

'As a first approximation, demand shocks will tend to lead to a positive relation between [nominal] interest rates and inflation while liquidity shocks lead to negative covariation. This suggests that there is little reason to expect any stable relation between short-term movements in interest rates and inflation.' (1983, p.207)

The main implication of such a conclusion is that it will be particularly difficult to analyse the short-run relationship between nominal interest rates and inflation owing to the variety of shocks that impinge on the economy. One possible strategy suggested by Summers (1983, p.215) is to analyse long-run relationships by 'filtering out' high frequency short-run variations in the variables by using band spectral regression techniques.

6.9. Conclusions

This chapter has been largely devoted to a theoretical examination of the Fisher hypothesis which, when originally formulated, predicted that nominal interest rates would rise (fall) in tandem with the inflation rate. However, it was shown that such a relationship would not hold in the presence of taxes. In a model, in which the firm was assumed to finance its marginal unit of capital entirely by debt, it was shown that in a taxation system which was indifferent between real and inflationary components of income, the nominal interest rate would have to rise by about one-and-half times the rate of inflation, given current tax laws. By introducing equity finance, it was shown that, for a one point rise in the inflation rate, the nominal interest rate would have to rise by less than in the case of all-debt financing. Such relationships between interest rates and inflation may exist in the long-run steady state, but not in the short-run as shown by Summers (1983). The implication for future empirical research on the Fisher hypothesis is that any attempt to analyse the short-run movements between nominal interest rates and inflation may not prove to be fruitful. Furthermore, the strategy suggested by Summers can be carried over in order to formulate a

dynamic steady-state demand for money which took explicit account long-run information otherwise ignored in the short-run. Once having derived the dynamic steady-state demand for money, it is of interest to examine the effect of the Fisher hypothesis on the demand for money by including first order growth variables in prices and income as well as the standard variables. The purpose of the next chapter is to show that the steady-state demand for money may turn out to be inherently unstable because of frequent changes in tax regimes which are parameterised in the refined Fisher hypothesis, and to consider the claim made by Hendry and Mizon (1978) that such a dynamic model 'appears to be fully consistent with standard economic theory statements of the demand for money function.' (p.562)

C H A P T E R S E V E N
THE REFINED FISHER HYPOTHESIS AND THE
STEADY-STATE DEMAND FOR MONEY

Section	7.1.	The dynamic steady-state demand for money.	380
	7.2.	Incorporation of the Fisher hypothesis.	384
	7.3.	Empirical evidence and analysis.	389
	7.4.	Consistency with theory.	402
	7.5.	Conclusions.	409

CHAPTER SEVEN
THE REFINED FISHER HYPOTHESIS
AND THE STEADY-STATE DEMAND FOR MONEY

In a paper on long-run features of dynamic time series models, Currie (1981) drew attention to the importance of assessing the dynamic long-run properties of estimated equations. Using a number of empirical examples, which included the demand for money, he showed that the equilibrium value of the dependent variable was sensitive to the rates of growth in the explanatory variables, and the magnitude and speed with which these effects were transmitted may be such to be of concern when using the equations for short- and medium term forecasting. With this in mind, Patterson and Ryding (1982) developed a statistical framework in which to test the null hypothesis that k th order growth coefficients were not significantly different from zero. In applying it to the reduced form of the dynamic steady-state demand for money, Patterson and Ryding (1984) gave some indication as to possible values of the rate of change in the nominal interest rate with respect to the inflation rate in which the hypothesis that the reduced form dynamic multiplier on prices is not significantly different from zero may be rejected. On a theoretical note, Currie (1981) also expressed some doubts about whether or not the parameters of the dynamic steady-state demand for money have economically sensible values.

This chapter seeks to address two issues. The first one concerns the absence of the rate of change in the nominal interest rate from the structural form of the dynamic steady-state demand for

money. Whilst Hendry and Mizon (1978) and Currie (1981) *assume* that such a variable is equal to zero, Patterson and Ryding (1984) offer an explicit rationale for its absence from the structural form of the dynamic steady-state demand for money. They believe that the Fisher hypothesis has been implicitly used in making the rate of change in the nominal interest rate term redundant. Thus Section 7.1 considers the derivation of the dynamic steady-state demand for money, and then Section 7.2 demonstrates two uses of the Fisher hypothesis. Firstly, the second-order growth rate in prices is related to the rate of change in the nominal interest rate, and by explicitly incorporating the Fisher hypothesis in its second order steady state form, the structural form of the steady-state demand for money is obtained in which the rate of change in the nominal interest rate is absent. Secondly, making use of the fact that the original Fisher hypothesis is a first-order steady-state relationship between the rate of change in the price level and the nominal rate of interest, the Fisher hypothesis is again explicitly incorporated to obtain the reduced form of the steady-state demand for money in which the nominal rate of interest itself is absent. Then Section 7.3 considers the properties of the reduced form dynamic multiplier in prices, and then goes on to consider the empirical evidence presented in the above-cited studies. Currie (1981) has suggested that the *structural* form of the dynamic multipliers are not significantly different from zero, and therefore argues that these can be constrained to zero in the estimation process. However, Patterson and Ryding (1982) show that the imposition of such constraints can lead to pervasive results. Thus, in both papers, Patterson and Ryding (1982, 1984) show that the dynamic multipliers are significantly different from zero. In the latter

paper, whilst testing the hypothesis that the reduced form dynamic multiplier in prices is not significantly different from zero, Patterson and Ryding give some indication as to possible values of the rate of change in the nominal interest rate with respect to inflation. The third section concludes with the proposition that, on the basis of the evidence, the reduced-form dynamic steady-state demand for money function may be inherently unstable due to frequent changes in tax regimes.

In Section 7.4., the second issue of this chapter is addressed. All the above-cited studies do not seem to offer any satisfactory *a priori* explanations for the possible magnitudes in the rate of growth variables, and their effect on the dynamic steady-state demand for money. Thus, the empirical evidence is also considered in its second perspective, that is, with a view to developing a dynamic version of the theory of the demand for money. It is noted here with particular concern that attempts are being made to construct a theory of the dynamic steady-state demand for money on an *ad hoc* basis, whereas the correct approach would be to develop a theory on an *a priori* basis, and then to test such a theory empirically. It is concluded that the lack of consensus in the results regarding the magnitudes of the rate of growth explanatory variables should be investigated further, and the *a priori* determination of such magnitudes is worthy of future theoretical research.

7.1. The dynamic steady-state demand for money

In deriving the dynamic steady-state properties of the demand for money, Currie (1981, p.705) bases his analysis on a conventional autoregressive distributed lag of the form:¹

$$M_t = \alpha_0 + \sum_{j=0}^J \left[\alpha_{1j} p_{t-j} + \alpha_{2j} y_{t-j} + \alpha_{3j} R_{t-j} + \gamma_{(j+1)} M_{t-j-1} \right]$$

where all variables, except the nominal interest rate, are entered [1] as logarithms, j is the number of periods lagged, and the α 's and γ 's are all constants. The nominal interest rate, R , may either be entered without any transformation, or as the logarithm of the level plus one. M stands for a definition of money, p stands for the price level, and y stands either for income or expenditure. Implicit in the model above, there is an assumption of uniform lag lengths in all variables, and such an assumption can be relaxed, as Patterson and Ryding (1984, p.21) have already done, by rewriting equation [1] so as to allow for different lag lengths in each variable:

$$M_t = \alpha_0 + \sum_{j=0}^J \alpha_{1j} p_{t-j} + \sum_{n=0}^N \alpha_{2n} y_{t-n} + \sum_{k=0}^K \alpha_{3k} R_{t-k} + \sum_{\ell=1}^L \gamma_{\ell} M_{t-\ell} \dots [2]$$

Consider now the *static* steady-state demand for money function; in such a state, it is assumed that all exogenous variables remain constant over time, that is, for example, $p_t = p_{t-j}$ for all t and j . Thus, equation [2] reduces to a conventional specification of the demand for money:

$$M_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 R_t \dots [3]$$

where, for example, $\beta_1 = \sum \alpha_{1j} / (1 - \sum \gamma_\ell)$.

In order to examine the *dynamic* steady-state properties of the demand for money, it is necessary to crystallise the concept of what is meant by *k*th order growth rates. When a steady-state is said to be of order *k*, it means that an exogenous variable, say x_t , is growing in such a way that the *k*th order rate of growth of x_t is non-zero and constant over time, while all growth rates of order greater than *k* are zero. That is, $\Delta^j x_t = 0$ for all $j \geq k + 1$ in which case the steady-state is said to be of order *k*. Therefore, the static steady-state refers to the case when $k = 0$, and the dynamic steady-state to the case when $k = 1$. The static steady-state model of the demand for money can be generalised to a dynamic steady-state model by assuming that the first order rates of growth in all exogenous variables remain constant throughout time. Currie (1981, p.705, equation (3)) has shown that the *first order* rate of growth in the dependent variable is related to the first order rates of growth in the explanatory variables as follows:

$$\left[1 - \sum_{\ell} \gamma_{\ell} \right] \pi_M = \sum_j \alpha_{1j} \pi_p + \sum_n \alpha_{2n} \pi_y + \sum_k \alpha_{3k} \pi_R \dots [4]$$

where π_p , π_y , π_R , and π_M refer to the first order rates of growth in prices, real income, nominal interest rates, and the money supply respectively. Now, it can be shown that $x_{t-j} = x_t - j\pi_x$ so that, for example, $p_{t-j} = p_t - j\pi_p$, and $M_{t-\ell} = M_t - \ell\pi_M$. The first step is to substitute these expressions into equation [2] which then reduces to

$$\begin{aligned}
 M_t &= \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 R_t \\
 &- \left[\frac{\sum j \alpha_{1j}}{(1 - \sum \gamma_\ell)} \right] \pi p^- \left[\frac{\sum n \alpha_{2n}}{(1 - \sum \gamma_\ell)} \right] \pi y^- \left[\frac{\sum k \alpha_{3k}}{(1 - \sum \gamma_\ell)} \right] \pi R \\
 &- \left[\frac{\sum \ell \gamma_\ell}{(1 - \sum \gamma_\ell)^2} \right] \pi M
 \end{aligned}$$

It should be noted that it is not strictly correct to assume that πM is independent of the rates of growth in the explanatory variables owing to the existence of a relationship between πM and the rates of growth in the explanatory variables as shown in equation [4]. Thus, in the second step, Currie (1981, p.709, footnote 1) has argued that the term πM becomes redundant when equation [4] is substituted into the preceding expression to obtain the dynamic steady-state demand for money function:

$$\begin{aligned}
 M_t &= \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 R_t \\
 &+ \psi_1 \pi p + \psi_2 \pi y + \psi_3 \pi R \quad \dots [5]
 \end{aligned}$$

where the β 's are the coefficients of the level of the explanatory variables which have already been defined in equation [3], and where the ψ 's are the coefficients of the rate of growth in the explanatory variables, where, for example,²

$$\psi_1 = - \left\{ \left[\frac{\sum j \alpha_{1j}}{(1 - \sum \gamma_\ell)} \right] + \left[\frac{\sum \alpha_{1j} \sum \ell \gamma_\ell}{(1 - \sum \gamma_\ell)^2} \right] \right\}$$

Such coefficients are termed as the structural form of the dynamic multiplier in the explanatory variable.

It is particularly interesting to note that the term describing the rate of change in nominal interest rates is included in the above formulation of the dynamic steady-state demand for money whereas Hendry and Mizon (1978, p.561, equation (22)) and Currie (1981, pp.709–712, equations (9), (11), and (12)) assume that such a term is equal to zero, and there apparently seems to be no explanation offered by the above-mentioned researchers for its absence. It does seem surprising that Currie (1981, p.709, footnote 1) offered an explanation on why the term representing the rate of growth in the money supply was absent from his formulation of the dynamic steady-state demand for money, and yet did not make it clear why the term π_R was absent. Patterson and Ryding (1984) have given an explanation for the absence of the rate of change in the interest rate, and the reason that they give is *similar* to that given by Currie for the absence of the π_M term; namely that the rates of change in the explanatory variables are *still not* independent because there exists a second-order steady-state relationship which is the 'Fisher relation'. Thus, the next section explicitly considers the effect of such a relationship on the dynamic steady-state demand for money.

7.2. Incorporation of the Fisher hypothesis

Patterson and Ryding (1984, p.301) have made a useful distinction between the reduced form of a model in its conventional sense, and the *steady-state* reduced form. Consider a model consisting of a system of equations as follows:

$$By = \Gamma x \quad \dots [6]$$

where y is the column vector consisting of jointly dependent variables, x is the column vector consisting of predetermined variables, and the matrices B and Γ are the coefficient matrices. [6] refers to the structural form of the model, and the reduced form may be obtained by pre-multiplying both sides of [6] by the inverse of B to obtain

$$y = B^{-1}\Gamma x \quad \dots [7]$$

The reduced form of the model expresses the jointly dependent variables in terms of predetermined variables only. Patterson and Ryding (1984) suggest that it may be necessary to treat the nominal interest rate as *predetermined* because of the difficulty of modelling short-run relationships. In the context of the Fisher hypothesis, it was shown in Section 6.8 that a definite relationship between inflation and interest rates may not exist in the short-run because of the variety of shocks impinging on the economy. In particular, it was shown by Summers (1983) that, as a first approximation, demand shocks led to a positive relationship between inflation and interest rates whereas a liquidity shock led to negative covariance. However, in the long-run steady state, as Chapter 6 of this thesis has already shown, a definite relationship between inflation and interest rates can exist if real balance effects are neglected. Therefore Patterson and Ryding argue that, in the long-run steady-state, it is no longer legitimate to treat the nominal interest rate as predetermined, that is, it becomes necessary to treat it as a *jointly dependent* variable. Hence, Patterson and Ryding make the distinction between the reduced form in its conventional sense and the steady-state reduced

form.

Consider, then, the relationship between inflation and the nominal interest rate as exemplified by the Fisher hypothesis. When it was originally stated by Irving Fisher in 1896, the hypothesis postulated a relationship between the inflation rate and the nominal rate of interest in which the latter would adjust *pari passu* with changes in the inflation rate. As Chapter 6 of this thesis has already shown, this relationship between inflation and the nominal interest rate is likely to be modified in the presence of taxes. Given current tax law and all-debt financing, the nominal interest rate would have to rise by about one-and-half times as much as the inflation rate whereas the introduction of equity finance would bring this rate of change somewhere between unity and one-and-half. For convenience, equations [25] from Section 6.4 (after neglecting real balance effects and assuming that consumption is interest-inelastic) and [54] from Section 6.7 are reproduced below:

$$\frac{dR}{d\pi} = \frac{1}{1 - \tau} \quad \dots [8]$$

and

$$\frac{dR}{d\pi} = 1 - \nu \quad \dots [9]$$

where equation [8] refers to the rate of change in the nominal interest rate with respect to the inflation rate in an all-debt financing model, and equation [9] refers to the case of a debt-equity model, and where ν is defined in equation [52] of section 6.6. Now define

μ as the parameter incorporating all the relevant tax parameters so that equations [8] and [9] can be written as $dR/d\pi = \mu$. Integrating this expression with respect to π gives the following relationship between the nominal interest rate and the rate of inflation:

$$R_t = c + \mu\pi_{p,t} \quad \dots [10]$$

where $\mu \approx 1.5$ in an all-debt financing model, and $1 < \mu < 1.5$ in a debt-equity financing model, and the constant, c , refers to the marginal productivity of capital (see, for example, equation [18] of Section 6.4) It should also be remembered that, in the steady-state, it is assumed that the anticipated rate of inflation (π_p) is equal to the actual rate (\dot{p}/p): such an assumption serves as a useful way of abstracting from the subjective question of how inflationary expectations are formed, and therefore any empirical tests will not be conditioned by the way inflationary expectations are formed.

Regarding the first use of the Fisher hypothesis, it has been suggested by Patterson and Ryding (1984, p.302) that it could be assumed that there is zero *second order* rate of growth in the price level so that the rate of change in the nominal interest rate is equal to zero: this can be shown by taking the first difference of equation [10] so that

$$\pi_R = \Delta^2 p$$

Clearly, when $\Delta^2 p = 0$, $\pi_R = 0$. They believe that such an assumption is justified because it is unlikely that the United Kingdom

will experience second order rates of growth. Therefore the term πR in equation [5] is now redundant because of the existence of a second-order steady-state relationship between the rate of change in the inflation rate and the rate of change in the nominal interest rate.

There is now a system of two equations consisting of the dynamic steady-state demand for money and the relationship between the nominal interest rate and the inflation rate. Regarding the second use of the Fisher hypothesis, the system of two equations can be written in matrix form to give the structural form of the model:

$$\begin{bmatrix} 1 & -\beta_3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M \\ R \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \psi_1 & \psi_2 \\ c & 0 & 0 & \mu & 0 \end{bmatrix} \begin{bmatrix} d \\ p \\ y \\ \pi p \\ \pi y \end{bmatrix}$$

where d is a term set equal to unity. Pre-multiplying both sides of the preceding expression by the inverse of the left-hand side coefficient matrix gives:

$$\begin{bmatrix} M \\ R \end{bmatrix} = \begin{bmatrix} 1 & -\beta_3 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \psi_1 & \psi_2 \\ c & 0 & 0 & \mu & 0 \end{bmatrix} \begin{bmatrix} d \\ p \\ y \\ \pi p \\ \pi y \end{bmatrix}$$

whence

$$\begin{bmatrix} M \\ R \end{bmatrix} = \begin{bmatrix} (\beta_0 + \beta_3 c) & \beta_1 & \beta_2 & (\psi_1 + \beta_3 \mu) & \psi_2 \\ c & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} d \\ p \\ y \\ \pi p \\ \pi y \end{bmatrix}$$

and it follows that the dynamic steady-state *reduced form* of the demand for money, taking explicit account of the Fisher hypothesis is now:

$$M_t = (\beta_0 + \beta_3 c) + \beta_1 p_t + \beta_2 y_t + (\psi_1 + \beta_3 \mu) \pi p + \psi_2 \pi y \quad \dots [11]$$

The coefficient to the rate of change in the price level is termed the reduced form dynamic multiplier in prices which is the sum of two effects, the first one being termed as the direct effect, ψ_1 , of a rise in the rate of inflation on the steady-state demand for money, and the second term as the indirect effect, $\beta_3 \mu$, of a change in the rate of inflation on the nominal interest rate.

7.3. Empirical evidence and analysis

This section will now consider the empirical evidence presented by Patterson and Ryding regarding possible values of the rate of change in the nominal interest rate with respect to inflation whilst testing the null hypothesis that the reduced form dynamic multiplier in prices is not significantly different from zero.

In deriving their dynamic steady-state demand for money function, Hendry and Mizon (1978, p.560) firstly estimated equation [1] above in an unrestricted form for both $J = 2$ and $J = 4$, and after testing various common root restrictions, the final restricted form which they arrived at was: (all variables in logarithmic form)

$$\begin{aligned} \Delta(M/p)_t &= 1.61 + 0.21\Delta y_t + 0.81\Delta R_t + 0.26\Delta(M/p)_{t-1} \\ &\quad (0.65) \quad (0.09) \quad (0.31) \quad (0.12) \\ &- 0.40\Delta p_t - 0.23(M/py)_{t-1} - 0.61R_{t-4} + 0.14y_{t-4} \\ &\quad (0.15) \quad (0.05) \quad (0.21) \quad (0.04) \\ \hat{\sigma} &= 0.0091, \quad R^2 = 0.69 \quad \dots [12] \end{aligned}$$

where the figures in paratheses denote standard errors. Then, Hendry and Mizon (1978, p.561) set the rates of change in the price level and real disposable income equal to constants, and, of course, set the rate of change in the nominal interest rate equal to zero so that the preceding equation became

$$\begin{aligned} (\pi_M - \pi_p) &= 1.61 + 0.21\pi_y + 0.26(\pi_M - \pi_p) - 0.40\pi_p \\ &- 0.23(M/py) - 0.61R + 0.14y \end{aligned}$$

where M stands for nominal sterling M3 balances, y stands for real disposable income at 1970 prices, p is the implicit deflator for y , and R stands for the yield on consols. Their specification of the dynamic steady-state demand for money is, however, not quite complete as Currie (1981) has already pointed out. As previously discussed in Section 7.1, it is not quite correct to assume that π_M is independent of π_p and π_y . Therefore, the structural form of the

dynamic steady-state demand for money is obtained by Currie as described by equations [2], [4], and [5]. The calculations are presented in Table 7.1, and the resulting structural form of the dynamic steady-state demand for money is presented as equation 1 of Table 7.2 which summarises some of the results of studies purporting to examine the dynamic steady-state properties of the demand for money. No test statistics were given by either Hendry and Mizon or Currie for this particular equation.

In advancing the hypothesis that the coefficients to the rate of growth explanatory variable were all not significantly different from zero, Currie (1981) estimated a long-run solution whose form is given as equation 3 in Table 7.2, and where the variables have the same definition as that of Hendry and Mizon. Currie was able to reject the joint hypothesis that all coefficients to the level variables were not significantly different from zero, and accepted the other joint hypothesis that all the coefficients to the rate of growth variables (including the rate of change in the nominal interest rate) were not significantly different from zero, and concluded that these variables 'may be constrained to zero without loss of explanatory power.' (p.712). Such a conclusion is viewed with some scepticism by Patterson and Ryding (1982, p.22) who estimate a dynamic steady-state demand for money function based on an unrestricted autoregressive model with uniform lag lengths of two. This is reproduced as equation 4 of Table 7.2, where M stands for M1 balances, y stands for total final expenditure at 1975 prices, p is the implicit deflator for total final expenditure, and R is the local authority three-month rate. The

TABLE 7.1: Solved coefficients from equation [12], and summary of calculations leading up to coefficients of equation 1 in Table 7.2.

Lag	M	p	y	R
0	-	0.60	0.21	0.81
1	1.03	-0.63	0.02	-0.81
2	-0.26	0.26	0	0
3	0	0	0	0
4	0	0	0.14	-0.61
T O T A L S	$\sum \gamma =$ 0.77	$\sum \alpha =$ 1 0.23	$\sum \alpha =$ 2 0.37	$\sum \alpha =$ 3 -0.61
	$\sum \ell \gamma =$ 0.51	$\sum j \alpha_1 =$ -0.11	$\sum n \alpha_2 =$ 0.58	$\sum k \alpha_3 =$ -3.25
	$\beta_0 = 7.0$	$\beta_1 = 1.0$ $\psi_1 = -1.74$	$\beta_2 = 1.61$ $\psi_2 = -6.09$	$\beta_3 = -2.65$ $\psi_3 = -20.01$

TABLE 7.2: Summary of results of studies leading up to the derivation of the dynamic steady-state demand for money whose structural form is

$$M_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + \psi_1 \pi_p,t + \psi_2 \pi_y,t$$

Equation no.	Study	Coefficients						Value of reduced form of dynamic multiplier in prices when $\mu = 1\%$	Remarks: (e)
		CNST	ρ	γ	R	π_p	π_y		
1.	Hendry & Mizon (1978)	7.0	1.0 (a)	1.6	-2.6	-1.7	-6.1	-5.6	(i) Test $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ (ii) Test $H_0: \psi_1 = \psi_2 = \psi_3 = 0$ (iii) Test $H_0: \psi_1 + \mu\beta_3 = 0$
2.	Coghlan (1979)	5.08	1.0	0.424	(b)	2.488	1.014	**	
3.	Currie (1981)	(c)	0.32 (0.62)	2.87 (1.4)	-1.17 (2.9)	2.5 (7.2)	-17.8 (19.2)	0.745	(i) $\chi^2(3) = 224.5$ (ii) $\chi^2(3) = 1.8$
4.	Patterson & Ryding (1982)	-11.27	0.935 [18.40]	2.183 [2.12]	-10.41 [1.48]	-11.41 [0.909]	-36.22 [0.43]	-27.025	(ii) $M = 1.205$ where M is distributed as $\chi^2(3)$ under H_0 .
5.	Patterson & Ryding (1984)	0.38	1.0 (d)	1.0	-4.66 [10.31]	-10.69 [17.36]	-9.87 [26.14]	-17.68	(iii) $M_\mu = 80.75$ when $\mu = 1\%$, and where M_μ is distributed as $\chi^2(1)$ under H_0 .

Notes: (a) Figures in parentheses denote asymptotic standard errors. If these are omitted, this was because they were not reported in the study.
 (b) Not included, and therefore the reduced form dynamic multiplier in prices is not calculable - see text.
 (c) The value of the constant term was not reported.
 (d) Figures in square brackets denote Wald test statistics which are distributed as $\chi^2(1)$ under the null hypothesis that the coefficient is not significantly different from zero.
 (e) The relevant statistics against which the null hypothesis may be tested are as follows:

$$\chi^2(1) = 3.84, \quad \chi^2(1) = 6.63, \quad \chi^2(3) = 7.81, \quad \text{and} \quad \chi^2(3) = 11.34.$$

0.05 0.01 0.05 0.01

the M1 rather than the sterling M3, definition of money was chosen 'to avoid, as far as possible, complications arising out of the joint endogeneity of some part of the money supply and the nominal interest rate;...' (Patterson and Ryding (1984, p.306). Patterson and Ryding (1982) state that they are unable to reject the joint hypothesis that all coefficients to the rate of growth variables are not significantly different from zero. Then, they imposed a zero constraint on the rate of growth in expenditure in order to examine the effects on the dynamic steady-state demand for money of imposing such a constraint. They note that such a constraint can lead to pervasive results since the mean lag on prices and nominal interest rates increased from about 3 or 4 years to an implausible $13\frac{1}{2}$ and $16\frac{1}{2}$ years respectively (1982, pp.23-24), and Patterson and Ryding state that '[the] imposition of the constraint, that a dynamic multiplier be set equal to zero, has far-reaching effects which may well outweigh the problems associated with non-zero, but insignificant, dynamic multipliers.' (p.24)

Rather than to impose a constraint on dynamic multipliers, Patterson and Ryding (1982, pp.24-25) chose to reduce the lag length of the model based on equation [2] in Section 7.1, and estimated another dynamic steady-state demand for money function which is reproduced in their later paper (1984, p.306) and as equation 5 in Table 7.2, where the variables also have the same definition as the variables in equation 4 of Table 7.2. Note that the price level and expenditure have been constrained such that the *steady-state* (i.e. *not* the short-run) demand for money is linearly homogeneous in prices and expenditure. They note that the static multiplier in the nominal interest rate, and the *structural* dynamic multipliers in prices

and expenditure are all now significant. However, Patterson and Ryding (1984, p.303) argue that testing the null hypothesis that the reduced form of the dynamic multiplier in prices is insignificant is more appropriate because it also takes into account the effect of a change in the rate of inflation on the nominal interest rate. They also contend that there is a quite substantial difference in the values of the structural form of the dynamic multiplier and its reduced form counterpart. As the basis for their argument, Patterson and Ryding used Currie's formulation of the dynamic steady-state demand for money in its structural form (see Currie (1981, p.712, equation (12)) which is reproduced as equation 3 in Table 7.2. In the previous section it was shown that the reduced form of the dynamic multiplier in prices was equal to $\psi_1 + \beta_3\mu$. Dependent on the value of μ , the reduced form of the dynamic multiplier for Currie's formulation is therefore $(2.5 - 1.17\mu)$. Given the theoretical values of μ which were derived in the previous chapter, the dynamic multiplier in prices will lie somewhere between 0.745 (when $\mu = 1\frac{1}{2}$) and 1.33 (when $\mu = 1$). Other values of the reduced form dynamic multiplier in prices, when $\mu = 1\frac{1}{2}$, are presented in the ninth column of Table 7.2. The reduced form dynamic multiplier in prices is not calculated for equation 2 of Table 7.2, and the reasons for this will become clear later on. It appears, therefore, that the only case in which the structural form of the dynamic multiplier in prices is unlikely to be different from its reduced form counterpart occurs when $\mu = 0$ which lies well beyond the range of theoretically plausible values. Moreover, it may be noted that in the special case when the reduced form dynamic multiplier in prices is equal to zero, the direct effect of inflation on the steady-state demand for money is

exactly counterbalanced by the indirect effects of inflation on nominal interest rates which affect the steady-state demand for money as well. Patterson and Ryding, having argued that there is a difference in the values of the structural and reduced forms of the dynamic multiplier in prices, suggest that the null hypothesis of direct interest should be

$$H_0: \psi_1 + \beta_3 \mu = 0 \quad \dots [13]$$

Patterson and Ryding (1982) have shown that testing the hypothesis that the k th order growth coefficients are zero, either singly or jointly, can be carried out using the Wald principle which only requires estimation from the unrestricted model. In particular, they show that the Wald test statistic is given by (see Patterson and Ryding (1982), p.18):

$$W = h(\hat{\theta})' (\hat{H}\hat{\sigma}H')^{-1} h(\hat{\theta})' \quad \dots [14]$$

where an 'hat' over a variable denotes an unrestricted estimator, $h(\theta)$ is the vector of constraints on the coefficient vector expressed in the form $h(\theta) = 0$, H is the matrix (a vector if one constraint) of derivatives of $h(\theta)$ with respect to θ evaluated at $\hat{\theta}$, and $\hat{\sigma}$ is a consistent estimator of the asymptotic covariance matrix of $\hat{\theta}$. Under the null hypothesis $h(\theta) = 0$, W is distributed asymptotically as $\chi^2(r)$ where r is the number of restrictions, that is, the dimension of $h(\theta)$.

The constraint of interest here is the zero constraint on the reduced form dynamic multiplier in prices. Now substituting for ψ_1 and β_3 into the null hypothesis gives

$$H_0: h(\theta)_\mu = - \left\{ \left[\frac{\sum \alpha_{1j} \sum \ell \gamma_\ell}{(1 - \sum \gamma_\ell)^2} \right] + \left[\frac{\sum j \alpha_{1j}}{(1 - \sum \gamma_\ell)} \right] \right\} \\ + \mu \left[\frac{\sum \alpha_{3j}}{(1 - \sum \gamma_\ell)} \right] \dots [15]$$

where the subscript μ on $h(\theta)$ indicates that the hypothesis is evaluated conditional on a given value for μ . Since there is only one constraint to be tested, the Wald statistic may be simplified to:

$$W(\theta)_\mu = \frac{h^2(\theta)_\mu}{(\widehat{H\Phi H'})_\mu} \dots [16]$$

which is distributed as $\chi^2(1)$ under H_0 , and where $(\widehat{H\Phi H'})_\mu$ is a consistent estimator of the asymptotic variance, denoted *avar*, of the linear combination $\hat{\psi}_1 + \mu \hat{\beta}_3$ conditional on μ , that is:

$$^2 (\widehat{H\Phi H'})_\mu = \widehat{avar}(\hat{\psi}_1) + 2 \widehat{acov}(\hat{\psi}_1, \hat{\beta}_3)_\mu + \widehat{avar}(\hat{\beta}_3)_\mu$$

Having laid out the statistical framework for testing the null hypothesis that, for a known or assumed value of μ , say μ^+ , the reduced form dynamic multiplier in prices is not significantly different from zero, Patterson and Ryding (1984, p.307) calculate the Wald test statistic as

$$W(\hat{\theta})_{\mu^+} = \frac{114.33 + 99.66\mu^+ + 21.72(\mu^+)^2}{6.58 - 4.97\mu^+ + 2.11(\mu^+)^2} \dots [17]$$

which is distributed as $\chi^2(1)$ under H_0 . The Wald test statistic given in equation 5 of Table 7.2 is for an assumed value of $1\frac{1}{2}$ for μ . Conditional on this value of μ , the null hypothesis that the reduced form dynamic multiplier in prices is rejected at conventional significance levels. It is of some interest to calculate the value of μ for the special case in which the reduced form dynamic multiplier in prices is equal to zero. The value of μ is given by $-\hat{\psi}_1/\hat{\beta}_3$ which, given the value of the estimates in Patterson and Ryding, is about -2.29 which may be compared with a value of -2.23 given by Patterson and Ryding (1984, p.308). Thus, given that both $\hat{\psi}_1$ and $\hat{\beta}_3$ are of the same sign, it follows that the special case in which the direct effects of inflation are exactly offset by the indirect effects of inflation on nominal interest rates can only occur for a value of μ which is not economically feasible.

Table 7.3 shows some values for μ given different corporate tax rates, and the Wald test statistics are also given. On the basis of the figures, it is not possible to reject the null hypothesis that the reduced form dynamic multiplier in prices is zero within a range of economically feasible values of μ . The corporate tax rates were chosen so that the effects of a change in tax regimes can be analysed. Prior to 1973, United Kingdom operated a classical corporation tax system in which the corporate tax rate was set at 40 per cent, and in 1973, as a part of the United Kingdom's entry into the European Economic Community, she moved into an imputation system in which the effects of economic double taxation on corporate income are mitigated by imputing a tax credit on dividend income. The corporate tax rose to 52 per cent.³ Since coming into office, the present Government has reduced the corporate tax rate from 52

per cent to 35 per cent, and at the same time, bringing inflation down to low levels, in which case, two inflation rates are given in Table 7.3, one reflecting a 'low' rate of inflation, and the other one reflecting a 'high' rate of inflation. The last two columns of Table 7.3 show the estimated elasticity, denoted here by $e_{v,\pi}$ of the reduced form dynamic steady-state inverse velocity of money (i.e. M/py) with respect to the rate of inflation. It is clear from the table that increasing the corporate tax rate will, *ceteris paribus*, increase the elasticity (in absolute terms) of the inverse velocity of money, and a decrease in the inflation rate will lead to a lower elasticity. Thus, on the basis of Patterson and Ryding's estimated reduced form steady-state demand for money function, and in accordance with Feldstein's all-debt financing model, it is inferred that with a lower corporate tax rate and a lower rate of inflation, the elasticity of the inverse velocity of money with respect to the inflation rate is now lower than it was in the 1970s. Such evidence would be highly suggestive that the dynamic steady-state demand for money would be inherently unstable, and if the short-run is considered, the lack of a definite relationship between nominal interest rates and inflation may tend to exacerbate the instability of the demand for money. Such a proposition would need to be carefully investigated, of course.

The foregoing analysis was on the basis that the value of μ was either known or assumed, and it has been shown by Patterson and Ryding (1984, pp.309-310) that it is possible to allow for uncertainty in the value of μ . One approach, that was ruled out by Patterson and Ryding, would have been to estimate a dynamic relationship between nominal interest rates and expected inflation.

TABLE 7.3: Values of μ , Wald test statistics, and elasticities of inverse income velocity with respect to inflation for key corporate tax rates.

Corporate tax rate	Value of μ	Wald test statistic $W(\hat{\theta})_{\mu}$	$e_{v, \pi}$ when	
			$\pi = 0.04$	$\pi = 0.10$
0	1	63.36	-0.614	-1.535
0.35	1.54	81.25	-0.715	-1.787
0.4	1.67	81.96	-0.739	-1.847
0.52	2.08	77.38	-0.815	-2.038
0.65	2.85	60.16	-0.959	-2.397

Source: Wald test statistics calculated from equation (13) in Patterson and Ryding (1984, p.307).

There are, of course, two main difficulties inherent in such an approach. The first difficulty stems from choosing an appropriate proxy for the expected rate of inflation, and any tests carried out would be conditional on how (no matter how subjective) inflationary expectations are formed. The second difficulty lies in the estimation of the parameter μ . It is difficult to deny that tax regimes in the United Kingdom have been constantly changing. If the value of μ is interpreted in accordance with Feldstein's all-debt financing model, it is not too difficult to be convinced that the value of μ has not been constant in view of changes in corporate tax rates. When a debt-equity model is considered, the situation becomes even more difficult in that there are even more tax parameters, not to mention the debt-equity ratio, that will also determine the value of μ . Instead, Patterson and Ryding carried out a sensitivity analysis in which uncertainty was introduced by finding the ratio of an assumed (or 'estimated') value of μ to its standard error which is in effect a t -statistic whose limits are zero and infinity, the lower limit corresponding to the case of perfect uncertainty, and the upper limit to the case of perfect certainty. By evaluating the Wald test statistic for different values of t , Patterson and Ryding (1984, p.308, figure 2) show that the Wald test statistic declines as the value of the t -statistic gets smaller. By comparing the values of the Wald test statistics with those at conventional significance levels, Patterson and Ryding conclude that the results obtained in the case of known or assumed values of μ 'are not materially altered' by the introduction of uncertainty in the value of μ .(p.310)

7.4. Consistency with theory

Having now discussed the empirical evidence regarding possible values of μ , it is necessary to consider the claim made by Hendry and Mizon (1978) that the dynamic steady-state demand for money is consistent with economic theory statements. The basis of the argument in this section will rest on the fact that it is possible to derive similar dynamic steady-state demand for money functions from different models so that it becomes apparent that some future theoretical research needs to be undertaken in order to determine the *a priori* effects of a change in the rate of inflation, and of a change in the rate of growth of income or expenditure on the steady-state demand for money.

First of all, it can be recalled that the reduced form dynamic steady-state demand for money was derived by treating the nominal interest rate as a jointly dependent variable which is a function of the inflation rate as in equation [10] above, and then substituting it into the structural form of the steady-state demand for money so arriving at equation [11]. The nominal interest rate is absent from equation [11] because the Fisher hypothesis was explicitly included. Consider equation 2 of Table 7.2 which was derived directly from a small monetary model developed at the Bank of England by Coghlan (1979). The long-run steady-state properties of the model are analysed at some length in Currie (1982) from which this discussion is derived. Inspection of the equation reveals that the nominal interest rate term is absent, and the equation has a structure which is quite similar to that of equation [11] above. In particular, Currie (1982, pp.68 - 70) derived the dynamic steady-state demand for money by using a price-equation reported in Coghlan (1979, p.34,

equation B(ii)) which has the following form:⁴

$$\begin{aligned}
 \dot{p}_t = & -1.162 + 0.0703\dot{p}_{z,t} + 0.2283\dot{p}_{z,t-1} + 0.2276\dot{p}_{z,t-2} \\
 & + 0.0988\dot{M}_t - 0.1941\dot{M}_{t-2} \\
 & - 0.6768\dot{y}_t + 0.3349\dot{y}_{t-1} + 0.1003\dot{y}_{t-2} \\
 & + 0.1318\dot{y}_{t-1} - 0.0561\dot{R}_{t-1} \\
] & + 0.2287 (M/py)_{t-1} \dots [18
 \end{aligned}$$

where all the variables, except for the nominal interest rate, have been entered in logarithmic form, and where, for example, $\dot{p}_t = \ln p_t - \ln p_{t-1}$. M stands for sterling M3 balances, y stands for private sector total final expenditure at 1970 market prices, p stands for the implicit deflator for total final expenditure, p_z stands for the price deflator for the sterling value of imports, and R stands for the interest rate on bank deposits. By setting the rate of growth variables in equation [18] equal to zero, Coghlan (1979, p.20) derives the implicit long-run *static* demand for money as:

$$(M/p) = 5.081 + 0.4237y$$

The above static steady-state demand for money function can be generalised to a *dynamic* steady-state demand for money by re-arranging equation [18] above so that M_t is placed on the left hand side of the equation, and following the same procedure as in Table 7.1 above, the following dynamic steady-state demand for money function is obtained:

$$M_t = 5.08 + p_t + 0.424y_t + 4.789\pi_{p,t} - 2.301\pi_{pz,t} \\ + 1.014\pi_{y,t} + 0.245\pi_{R,t}$$

and by assuming, as Currie (1982, p.69) does, that $\pi_p = \pi_{pz}$ and, of course, $\pi_R = 0$, the preceding equation then becomes the dynamic steady-state demand for money function as presented in equation 2 of Table 7.2.

The reason for the absence of a nominal interest rate term from equation 2 of Table 7.2 can be found by examining equation [18] which reveals that there is no term in the *present* nominal interest rate whereas the other steady-state demand functions in Table 7.2 were chiefly derived from a model, such as equation [2] above which included the *present* nominal interest rate. Now *if* equation 2 of Table 7.2 was to be interpreted as the reduced form dynamic steady-state demand for money function, it is possible to determine the signs of the estimated values of the parameters ψ_1 and β_3 by considering the cases in which the reduced form dynamic multiplier in prices is likely to be positive. One possible case can occur when the direct effect of inflation on the steady-state demand for money outweighs the indirect effect of inflation on nominal interest rates so that the structural form dynamic multiplier in prices for equation 2 of Table 7.2, ψ_1 , would have positive sign and β_3 would have negative sign. Therefore, on that basis, of all dynamic steady-state demand for money functions reported in Table 7.2, two out of five (i.e. equations 2 and 3) functions have positive (structural form) dynamic multipliers in prices, and only one out of five (i.e. equation 2) functions has a positive dynamic multiplier in expenditure/income.

In discussing possible *a priori* magnitudes for the rate of growth in prices, one needs to look no further than Cagan's pioneering study of hyperinflations. According to Cagan (1956, p.35), the demand for real balances would be inversely related to the expected rate of inflation, which in the steady-state is assumed to equal the actual rate. An increase in the rate of inflation increases the opportunity cost of holding money balances so that there is a substitution away from monetary assets into real assets. Such a justification has already been provided by Currie (1981, p.709). Such a relationship may be reinforced by the indirect effect of inflation on nominal interest rate: an increase in the rate of inflation will lead to a rise in the nominal interest rate which then increases the opportunity cost of holding money balances. This *a priori* presumption is confirmed by the results presented as equations 1, 4, and 5 of Table 7.2, but not by equations 2 and 3. According to the results presented in Table 7.3 for equation 5, it is apparent that an increase in the rate of inflation leads to an increase in the reduced form elasticity of inverse velocity with respect to inflation. This would be consistent with the *a priori* presumption that at higher rates of inflation, the inverse velocity of money would increase, and become more responsive to changes in the inflation rate.

With respect to the rate of growth in income/expenditure, all above-cited studies are unable to offer any definite explanation for its negative sign in equations 1, and 3 to 5 of Table 7.2. Consider, for example, the theory of the demand for money as presented by Baumol (1952) which was discussed in Chapter Three. One implication of the 'square root' formula is that there are economies

of scale in holding money balances. If the volume of transactions (as proxied by income) rises, Baumol's theory predicts that money holdings will rise less than proportionately so leading to economies of scale in money balances. However, Currie (1981, p.709) suggests that economies of scale may be absent due the fact that as the volume of transactions rises, the demand for money rises more than proportionately as can be seen from equation 1 of Table 7.2. This observation is also reflected in equations 3 and 4. Furthermore, given a *level* of transactions, all equations (except equation 2) of Table 7.2 suggest that the demand for money balances is inversely related to rate of growth in transactions which is difficult to rationalise given the current state of knowledge on the theory of the dynamic steady-state demand for money. Apparently, the *only* justification for its inclusion is, as Patterson and Ryding (1982) have already noted, that the exclusion of the growth variable in income/expenditure may lead to pervasive results.

7.5. Conclusions

This chapter has looked at the properties of the dynamic steady-state demand for money, and examined the effects of incorporating the Fisher hypothesis explicitly within it. It was shown that the absence of a growth term in nominal interest rates from the dynamic steady-state demand for money function can be explained by the assumption that there are zero second-order rates of growth in prices so that there would, in effect, be a zero first-order rate of growth in nominal interest rates according to the Fisher hypothesis. It was particularly noted that the nominal interest rate is best left in the short-run demand for money function owing to difficulties in

modelling the behaviour of nominal interest rates in the short run as demonstrated by Summers (1983) and section 6.8 of this thesis. However, in the long-run, it is believed that a steady-state relationship exists between inflation and interest rates so that such a relationship has to be explicitly incorporated. Thus the long-run dynamic steady-state demand for money exhibits the property that it is a function of the inflation rate which works through two effects, viz: directly, and indirectly via nominal interest rates as the reduced-form dynamic multiplier shows. Owing to frequent changes in tax regimes, the reduced-form steady-state demand for money may turn out to be inherently unstable.

Regarding the consistency of the dynamic steady-state demand for money with theoretical considerations, it was possible to show, with the aid of the pioneering work of Cagan (1956), that the dynamic steady-state demand for money is inversely related to the rate of inflation. However, a major difficulty still remains in explaining why the steady-state demand for money is, on the basis of the empirical evidence, directly related to the rate of growth in incomes. This would be a subject worthy of further research.

A P P E N D I C E S

- | | | |
|-------------|--|-----|
| Appendix 1. | A Note on Index Numbers and their
Desirable Properties. | 409 |
| 2. | A Note on Aggregate Production
Functions. | 415 |

APPENDIX ONE
A NOTE ON INDEX NUMBERS
AND THEIR DESIRABLE PROPERTIES

This appendix discusses briefly a few aspects of index numbers by defining the concept of exact and superlative index numbers. A proof that the Törquinst–Theil Divisia quantity index is superlative will be offered.

Firstly, define the quantity index as $Q(\mathbf{p}_0, \mathbf{p}_1; \mathbf{x}_0, \mathbf{x}_1)$ which is a function of a N -dimensional vector of prices in periods 0 and 1 where $\mathbf{p}_0 > \mathbf{0}_N$ and $\mathbf{p}_1 > \mathbf{0}_N$ ($\mathbf{0}_N$ is a N -dimensional null vector), and of the corresponding quantity vectors $\mathbf{x}_0 > \mathbf{0}_N$ and $\mathbf{x}_1 > \mathbf{0}_N$.[†] Similarly define a price index as $P(\mathbf{p}_0, \mathbf{p}_1; \mathbf{x}_0, \mathbf{x}_1)$. One important property of index numbers is that, given either a price or quantity index, the other function can be defined implicitly by the following equation which is the Fisher (1922) weak factor reversal test:

$$P(\mathbf{p}_0, \mathbf{p}_1; \mathbf{x}_0, \mathbf{x}_1) \cdot Q(\mathbf{p}_0, \mathbf{p}_1; \mathbf{x}_0, \mathbf{x}_1) = \mathbf{p}_1' \mathbf{x}_1 / \mathbf{p}_0' \mathbf{x}_0 \quad \dots [A1]$$

that is, the product of the price index times the quantity index should yield the expenditure ratio between two periods. This is known as the 'adding up' property. This is useful for empirical work for calculating divisia user cost indices from the divisia

[†] Emboldened lower case letters are used to denote vectors, and emboldened upper case letters are for matrices.

quantity indices and expenditure ratios.

An important theorem in index number theory makes it possible to use the ideal quantity index in order to calculate the exact aggregate. Let $\mathbf{p}_r \gg \mathbf{0}_N$ for periods $r = 1, \dots, R$ and suppose that $\mathbf{x}_r \gg \mathbf{0}_N$ is a solution to the following problem:

$$\max f(\mathbf{x}) = (\mathbf{x}'\mathbf{A}\mathbf{x})^{\frac{1}{2}} = \left(\sum_{i=1}^N \sum_{j=1}^N x_i a_{ij} x_j \right)^{\frac{1}{2}}$$

$$\text{subject to } \mathbf{p}'_r \mathbf{x} \leq \mathbf{p}'_r \mathbf{x}_r ; \quad \mathbf{x} \gg \mathbf{0}_N \quad \dots [A2]$$

where $a_{ij} = a_{ji}$ (all i, j). Provided that maximisation takes place in the region where $f(\mathbf{x})$ is concave and positive, then

$$\frac{f(\mathbf{x}_r)}{f(\mathbf{x}_0)} = Q_F(\mathbf{p}_0, \mathbf{p}_r; \mathbf{x}_0, \mathbf{x}_r) \quad r = 1, \dots, R \quad \dots [A3]$$

The implication of this theorem is that, given the normalisation $f(\mathbf{x}_0) = 1$, the ideal quantity index may be used to calculate the aggregate $f(\mathbf{x}_r) = (\mathbf{x}'_r \mathbf{A} \mathbf{x}_r)^{\frac{1}{2}}$ for $r = 1, \dots, R$. Thus, it is not necessary to estimate the unknown parameters of the \mathbf{A} matrix. If a quantity index $Q(\mathbf{p}_0, \mathbf{p}_r; \mathbf{x}_0, \mathbf{x}_r)$ and a functional form for the aggregator function satisfies [A3], then Q is said to be *exact* for f .

Superlative index numbers are only exact for a function that can provide a second order approximation to a linear homogeneous function. Diewert (1974, p.125) has shown that the aggregator function defined in [A2] is capable of providing such an approximation, and it is in this context that the ideal quantity index may also be regarded as being *superlative*. It will be shown that

the Divisia quantity index is superlative, but the quadratic approximation will be derived since it will prove to be useful in this proof.

First, define a quadratic function of the form

$$\begin{aligned} f(z) &= a_0 + a'z + \frac{1}{2} z'Az \\ &= a_0 + \sum_{i=1}^N a_i z_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N z_i a_{ij} z_j \quad \dots [A4] \end{aligned}$$

where a_i , a_{ij} are constants, and $a_{ij} = a_{ji}$ (all i, j). Furthermore, z is a N -dimensional vector. Consider the derivation of a quadratic approximation to [A4] in which $f(z)$ is evaluated at z_1 and z_0 so that

$$\begin{aligned} f(z_1) - f(z_0) &= a'z_1 + \frac{1}{2} z_1'Az_1 - a'z_0 - \frac{1}{2} z_0'Az_0 \\ &= a'(z_1 - z_0) + \frac{1}{2} z_1'A(z_1 - z_0) + \frac{1}{2} z_0'A(z_1 - z_0) \\ &= \frac{1}{2} [a + Az_1 + a + Az_0]'(z_1 - z_0) \end{aligned}$$

since $A' = A$. But it can be seen from [A4] that

$$\frac{\partial f}{\partial z_i} = a_i + \frac{1}{2} \sum_{j=1}^N a_{ij} z_j \quad i = 1, \dots, N$$

or in matrix notation,

$$\nabla f(z) = a + Az$$

Thus,

$$f(z_1) - f(z_0) = \frac{1}{2} [\nabla f(z_1) + \nabla f(z_0)]'(z_1 - z_0) \dots [A5]$$

which is the quadratic approximation, and $\nabla f(z_r)$ is the gradient vector evaluated at z_r .

It will now be shown that the Torquinst-Theil discrete time approximation to the Divisia quantity index is superlative. Suppose that a homogeneous translog aggregator function is given:

$$\ln f(x) = \alpha_0 + \sum_{i=1}^N \alpha_i \ln x_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \ln x_i \ln x_j \dots [A6]$$

Diewert (1976) has shown that the above function is capable of providing a second order approximation to an arbitrary twice-continuously-differentiable linear homogeneous function. Using the parameters given in the translog functional form, define a quadratic function such that

$$f^*(z) = \alpha_0 + \sum_{i=1}^N \alpha_i z_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} z_i z_j \dots [A7]$$

Applying the quadratic approximation [A5], the following expression is obtained:

$$f^*(z_1) - f^*(z_0) = \frac{1}{2} [\nabla f^*(z_1) + \nabla f^*(z_0)]'(z_1 - z_0) \dots [A8]$$

It is now necessary to relate the function defined in [A7] with the translog function as follows

$$\frac{\partial f^*(z_r)}{\partial z_j} = \frac{\partial \ln f(x_r)}{\partial \ln x_j} = \left[\frac{\partial f(x_r)}{\partial x_j} \right] \left[\frac{x_{rj}}{f(x_r)} \right] \dots [A9]$$

(and $f(z_r) = \ln f(x_r)$, $z_{rj} = \ln x_{rj}$ for $r = 0,1$ and $j = 1, \dots, N$). Substituting [A9] into [A8], the following expression is obtained

$$\ln f(x_1) - \ln f(x_0) = \frac{1}{2} \left[\hat{x}_1' \frac{\nabla f(x_1)}{f(x_1)} + \hat{x}_0' \frac{\nabla f(x_0)}{f(x_0)} \right]' [\ln x_1 - \ln x_0] \dots [A10]$$

where $\ln x_r = [\ln x_{1r}, \dots, \ln x_{Nr}]$ and \hat{x}_r is the vector x_r diagonalised into a matrix ($r = 0,1$).

Assume that x_r is a solution to the following problem

$$\max f(x) \quad \text{s.t.} \quad p_r' x = m \quad (m = p_r' x_r) \dots [A11]$$

where $f(x)$ is the translog function. After the elimination of the Lagrangian multiplier in the first order conditions, the relations $p_r/p_r' x_r = \nabla f(x_r)/x_r' \nabla f(x_r)$ ($r = 0,1$) are obtained. Since f is linear homogeneous, $x_r' \nabla f(x_r)$ may be replaced by $f(x_r)$ in the preceding relations. Substitution of these relations into [A10] leads to the following expression

$$\ln f(x_1) - \ln f(x_0) = \frac{1}{2} \left[\frac{\hat{x}_1' p_1}{p_1' x_1} + \frac{\hat{x}_0' p_0}{p_0' x_0} \right] [\ln x_1 - \ln x_0]$$

$$= \sum_{n=1}^N \frac{1}{2} [w_{1n} + w_{0n}] \ln (x_{1n} / x_{0n})$$

Finally,

$$\begin{aligned} f(x_1) / f(x_0) &= \prod_{n=1}^N (x_{1n} / x_{0n})^{\frac{1}{2}(w_{1n} + w_{0n})} \\ &\equiv Q_D(p_0, p_1; x_0, x_1) \end{aligned}$$

where $w_{rn} = p_{rn}x_{rn} / p_1x_r$. It has now been shown that the Divisia quantity index is superlative for a translog aggregator function because of that function's second order approximation property.

APPENDIX TWO
A NOTE ON AGGREGATE PRODUCTION FUNCTIONS

This note discusses how, given certain conditions, individual production functions can be aggregated to form the aggregate production function. It is assumed that there are m perfectly competitive firms who all use similar production techniques to produce quantities of a single good for consumption and investment at any instant of time. The production function for the i th firm is

$$Y_i = F_i(K_i, N_i) \quad \dots [A2-1]$$

where Y_i is the output of the single good by the i th firm, K_i is the capital stock of the i th firm, and N_i is employment by the i th firm per unit time. The firm's production function is characterised by positive though diminishing marginal products of capital and labour:

$$F_N, F_K > 0, \text{ and } F_{NN}, F_{KK} < 0.$$

It is assumed that the production function is linearly homogeneous such that a proportionate increase in both capital and employment will lead to an equi-proportionate increase in output:

$$\lambda F(K_i, N_i) = F(\lambda K_i, \lambda N_i)$$

A corollary of the linear homogeneity property of the production function is that all partial derivatives will be homogeneous of degree zero; thus for example

$$\frac{\partial F(K_i, N_i)}{\partial K_i} = \frac{\partial F(\lambda K_i, \lambda N_i)}{\partial \lambda K_i}$$

Letting $\lambda = 1/N_i$, the preceding expression is re-written as

$$\frac{\partial F(K_i, N_i)}{\partial K_i} = \frac{\partial F(K_i/N_i, 1)}{\partial (K_i/N_i)}$$

In words, the expression states that the marginal product of capital depends only on the ratio of capital to employment. This result also holds true for the marginal product of labour.

Owing to putty-clay technology, it is not possible for firms to dispose of their capital stock as there is no ready market available. The capital stock represents the accumulation of the single good which is appropriated to assist in the production of further output. Regarding the labour market, it can be assumed that prices and wages are sufficiently flexible such that firms are able to hire all the labour they need at the going rate.

Now, the profits made by the *i*th firm is the difference between its gross revenue, and its costs which comprise the wage bill and the cost of capital:

$$\Pi_i = pF_i(K_i, N_i) - wN_i - (R - \pi)K_i$$

where $(R - \pi)$ is the real rate of interest (return) on capital, and π is the anticipated inflation rate.

In order to maximise its profits, the firm must employ sufficient labour so that the marginal product of labour equals the real wage rate:

$$\frac{\partial \Pi_i}{\partial N_i} = pF_N(K_i, N_i) - w = 0$$

that is,

$$F_N(K_i, N_i) = w/p \quad \dots [A2-2]$$

Aggregate output is obtained by aggregating over the individual outputs of the firms in the economy:

$$Y = \sum_{i=1}^m Y_i = \sum_{i=1}^m F_i(K_i, N_i)$$

By Euler's Theorem,

$$\sum_{i=1}^m Y_i = \sum_{i=1}^m \left[F_K(K_i, N_i)K_i + F_N(K_i, N_i)N_i \right]$$

It was previously shown that the marginal products of capital and labour depend only on the capital-labour ratio. Using this fact, and the fact that, because of similar production technologies amongst firms, the capital-labour ratio is the same for all firms, it follows

that the marginal products are the same for all firms. Therefore,

$$\sum_{i=1}^m Y_i = F_K(K_i/N_i, 1) \sum_{i=1}^m K_i + F_N(K_i/N_i, 1) \sum_{i=1}^m N_i$$

Since the capital-labour ratio is the same for all firms, it follows that the economy's capital-employment ratio, $\sum_{i=1}^m K_i / \sum_{i=1}^m N_i$, is the same. Thus,

$$Y = F_K(K/N, 1)K + F_N(K/N, 1)N$$

where $K = \sum_{i=1}^m K_i$ and $N = \sum_{i=1}^m N_i$. Finally, by Euler's Theorem,

$$Y = F(K, N) \quad \dots [A2-3]$$

There is just one final important point to be made about the aggregation procedure. The aggregate production function linking Y with K and N is only valid for a certain distribution of the N_i across the firms. That distribution is given by the marginal productivity condition for labour in equation [A2-2]. If this condition is not fulfilled, then the aggregation procedure is violated.

NOTES

Chapter One

1. An exposition of the *IS-LM* model can be found in most macroeconomics textbooks.
2. R^2 statistics, used on their own, do not necessarily provide the best overall measure regarding the performance of an estimated relation. At best, such statistics only measure the 'goodness-of-fit' over the sample period. This is further elaborated in Chapter Four during the discussion of the definition of money by 'best' results.
3. The other conclusions regarding the possibility of a disequilibrium situation can be found in Artis and Lewis (1976) and Gafga (1985b).
4. A full discussion can be found in Judge (1983).
5. An aggregator function is a neutral term used by economic aggregation theory to refer to either production or utility functions.

Chapter Two

1. See, for example, Podolski (1986), chapter 2, and Silber (1975), pp. 63-64.
2. See especially Schumpeter (1934), chapter 2, and Schumpeter (1939), chapter 3.
3. See Schumpeter (1934), pp.65-66.
4. For example, see Kamien and Schwartz (1982), p.2, and Scherer (1980), pp.409-410.
5. For a concise theoretical treatment of process innovation, see Koutsoyiannis (1979), pp.85-86.
6. Some care has been taken here to avoid excessive oversimplification into a two-input type of analysis by collectively referring to various inputs as capital- and labour-related inputs as the case may demand it. It must be recalled that the marginal rate of technical substitution (MRTS) is negative so any increase in MRTS represents a decrease in *absolute* terms. The converse of capital-deepening is 'labour-deepening' in which the marginal productivities of labour-related inputs increase relative to those of capital-related inputs.
7. It is interesting to note that the existence of automated dealing systems on the stock exchange is not entirely an innovation because some dealers prefer the security of direct communication with their counterparts when dealing in large lots because they feel more secure doing so. For smaller lots, they can be transacted via the automated dealing system. Therefore, such dealing systems are an innovation to some extent in changing the way small transactions are handled, but not quite an innovation for larger lots. For a discussion on

technological developments in the financial sector, see Hamilton (1986), Chapter Two, especially p.45).

8. See Hester (1982), p.43, and Podolski (1986), p.107.
9. See Podolski (1986), p.108.
10. See Hester (1981), pp.149-151.
11. *ibid*, pp.155-161.
12. Regulation Q was one of the regulations set by the Federal Reserve which forbade payment of interest on bank deposits in excess of the interest rate ceiling. This was a measure designed to counteract the banks' tendency to bid for funds by bidding up interest rates.
13. See Hester (1981), pp.151-155.
14. Further examples for this category are listed in Table 1 in Kane (1981), p.360.
15. See Silber (1975), pp.68-69, and Scherer (1980), p.424.
16. See Scherer (1980), pp.425-426.
17. The following discussion is based on Scherer (1980), pp.426-428.
18. For example, see Schumpeter (1934), p.65, and Schumpeter (1939), p.73.
19. See Schumpeter (1939), p.111.
20. For a survey on the objectives of firms, see Koutsoyiannis (1981), section E. The objectives of financial firms may differ from those of firms operating in the real sector of the economy, and would indeed be worthy of further investigation.
21. The discussion need not be confined to regulated firms. It is indeed possible for unregulated firms to get involved in the sense that they can exploit more economically feasible options

created by regulated firms' innovations brought about by a desire to avoid regulations.

22. See Midland Bank (1966), p.3.
23. For a detailed account of the development of the parallel markets, see Grady and Weale (1986), chapter 6.
24. See Podolski (1986), p.152.
25. The special deposits scheme is not to be confused with the supplementary special deposits scheme introduced in 1973.
26. See Grady and Weale (1986), pp.48 and 49.
27. See for example Goodhart (1984), pp.103 and 166.
28. It is not too difficult to understand why there was a bias towards delay in adjusting the Bank Rate. One does not have to look no further than to consider the unfavourable political consequences of rising interest rates.
29. The formula was to add $\frac{1}{2}$ per cent on the Treasury Bill rate, and to round up the resulting rate to the nearest $\frac{1}{4}$ per cent giving the MLR.
30. Useful statistics on calls for SSDs may be found in the Appendix to the article on the SSD scheme, Bank of England (1982), p.85.
31. For instance, see Wilson Committee *Appendices* (1980, p.507, paragraph 3.362).
32. Other correlation coefficients presented by Goodhart (1984, p.156) are as follows: for annual rates of growth in M1 and M3 during the 1960s, the correlation coefficient was +0.97, and for the 1970s, it was -0.45; for quarterly rates of growth in M1 and sterling M3 during the 1960s, it was +0.91, and during the 1970s, it was +0.55; for quarterly rates of growth in M1

and M3 during the 1960s, it was +0.92, and during the 1970s, it was +0.46.

33. This would be a typical example of regressive expectations: in such a model, expectations of a variable tend to 'regress' towards the 'normal' value of a variable. See Gafga (1985a), Chapter 4, section 2 for more details.

Chapter Three

1. It should be noted that it is being implicitly assumed that buying and selling transactions costs are indistinguishable from each other. Contrast this with the assumption made by Baumol (1952), p.548.
2. The only essential difference lies in the assumption regarding transactions costs - see note 1.
3. This may be shown if the partial derivative of $E(\pi)$ in equation [27] is taken with respect to σ_π such that

$$\frac{\partial E(\pi)}{\partial \sigma_\pi} = \frac{E(r)}{\sigma_r} - \frac{2k}{\sigma_r} \left[1 - \frac{\sigma_\pi}{\sigma_r} \right]^{-\frac{1}{2}}$$

When $E(\pi)$ is at a maximum, $\partial E(\pi)/\partial \sigma_\pi = 0$. Thus, solving for σ_π gives

$$\sigma_\pi = \sigma_r \left[1 - \left[\frac{2k}{E(r)} \right]^2 \right]$$

The expression within the square brackets must be positive since it is necessary that $2k < E(r)$. It therefore follows that any increase in uncertainty (σ_r) will lead to an increase in σ_π . Substitution of σ_π into $E(\pi)$ gives the maximum expected revenue which is independent of any change in uncertainty. Thus the opportunity locus is seen to be shifting from OP_0 to OP_1 in Figure 3.5. without any change in $E(\pi)_{\max}$.

4. From the expression for σ_π given in the previous note, it can be seen that as k approaches zero, σ_π approaches σ_r .
5. Attention is being concentrated on the effects of the inclusion

of a time-trend variable on the conventional specification for the demand for money which includes GNP as a proxy for the volume of transactions, and so those specifications containing debits as a proxy for transactions are not reported here. However, results can be found in Lieberman (1977).

6. A sketch of the $g(\lambda)$ function can be found in Porter and Simpson (1980), p.194.
7. The full regression results that include previous-peak variables are not reported in Porter and Simpson (1980), Table B-6, pp.224-225, and so are not included in Table 3.2.
8. The results reported in Table 3.4. are based on the best performing specification that includes a linear-times-log ratchet variable, which Porter and Offenbacher (1984, p.92, n.5) say has been 'used exclusively'. Other results can be found in Porter and Simposn, Table B-4, pp.219-220. Note that the results reported in Table 3.5. can be compared with those given in Table 3.3.
9. See Porter and Offenbacher (1984), p.92, n.8.
10. *Ibid*, p.93, n.8. Note that the value of the constant has been included explicitly in equation [43].

Chapter Four

1. Note that if one of the commodities is nominated as a unit of account, there is only the need for $N - 1$, and not N , exchange ratios since the accounting price of money is set equal to unity.
2. Only a highly eccentric individual would be capable of hoarding large amounts of currency such that the main function of currency in this example would be that as a store of value. There are obvious disadvantages to the hoarding of large amounts of currency because one incurs an opportunity cost of holding currency which bears a zero nominal rate of interest. This opportunity cost is likely to be exacerbated in times of inflation. Thus, it may be safely assumed that the proportion of currency having the attribute of store of value is so negligible that it can be completely disregarded in the weighted monetary aggregate which is to be discussed in section 4.2.2.
3. Debit cards are distinct from credit cards in that the former may be regarded as both means of payment and mediums of exchange whereas the latter is a medium of exchange only in that the credit card bill has to be settled at some future date. Some discussions of EFTPOS technology and other aspects of the bank-customer relationship affected by technology can be found in Goodhart (1984), Chapter V, pp.169-180, and Hamilton (1986), Chapter 2, especially pp.30-40.
4. Friedman and Meiselman (1963) chose to regress the level of consumption, rather than income, on the money stock and autonomous expenditure in order to avoid as far as possible any problems posed by the definition of income which was

$Y \equiv C + A$. If income was used as the dependent variable, one would be faced with the situation in which part of income was correlated with itself.

5. There have also been other contributions to the causality debate, and each contribution has tended to use different definitions of money in order to determine which particular definition of money gives the 'best' results. For example, see Crockett (1970) who used quasi-M1 and quasi-M2 definitions of money which were based on the London clearing banks only. Other contributions stemming from the work of Granger (1969) who devised statistical methods for measuring causality, have used differing definitions of money. For example, see Sims (1972), and Williams *et al* (1976). The latter study used the same definitions of money as used by Crockett (1970) but the results were very poor in that income seemed to 'lead' money rather than the other way around.
6. See Judd and Scadding (1982), p.993.
7. In some empirical studies of the demand for money, the income elasticity is sometimes constrained to unity reflecting the view that there are no economies of scale in money balances. Hence the dependent variable would be (M/Y) . Of course, one can choose to impose the assumption of no money illusion in which case the dependent variable would be (M/PY) .
8. Examples can be found in Gafga (1985b), Table 3, and Artis and Lewis (1976), Table 2, p.154.
9. See Chow (1966).
10. See Laidler (1981), p.4.
11. Interest rate differentials have been used by, for example, Lee

(1967), and Artis and Lewis (1976).

12. Since the cross-elasticities estimated by Lee (1967) were based on interest rate differentials, it was necessary for Feige and Pearce (1977) to adjust the reported values so that they would be more comparable with the other studies reported which did not use interest rate differentials. The values of the cross-elasticities are based on the mean-values reported in Table 1(b) of Feige and Pearce (1977), pp.453-455.
13. See Lee (1967), p.1175.
14. See Hamburger (1969), pp.407 and 411.
15. See Lee (1969), p.417.
16. An aggregator function is a neutral term used by economic aggregation theory to refer to either a production or a utility function.
17. A discussion of the properties of the CES aggregator function can be found in several mathematical economics texts; see for example Silberberg (1978), pp.313-322, Kogiku (1971), p.67.
18. A 'generalised CES' utility function has the property that the elasticity of substitution between any two assets is always constant. Thus the function takes on the form;

$$U = \left[\sum_{i=0}^n \beta_i X_i^{-\rho} \right]^{-1/\rho}$$

which can be compared with equation [6] in the main text.

19. In the strictest sense, this regression model is not really a system of asset demand equations. This point is discussed in sub-section 4.2.2(c) below.

20. Indeed, this is the main reason why the utility function given in equation [6] is more appropriately labelled as a VES utility function.
21. This may be seen if the different versions of the constrained optimisation problem posed by Moroney and Wilbratte (1976) can be stated in the most general terms, namely that;

$$\left. \begin{array}{l} \text{maximise } T = f(X_0, X_1, \dots, X_n) \\ \text{subject to } W = \sum_{i=0}^n X_i(1 + r_i) \end{array} \right\} \text{version I}$$

and

$$\left. \begin{array}{l} \text{maximise } W = \sum_{i=0}^n X_{it}(1 + r_{it}) \\ \text{subject to } T = f(X_0, X_1, \dots, X_n) \end{array} \right\} \text{version II}$$

Version *I* is the *correct* primal problem, but version *II* is *not* the correct dual problem. A theorem in duality theory states that if the utility function is maximised subject to a budget constraint in the primal problem, and if, in the dual problem, the, say, expenditure function, is *minimised* subject to the utility function which is now a constraint, then one can obtain asset demand functions that are identical to each other. For both versions of the problem above, the first-order marginality conditions would be

$$\left. \begin{array}{l} (\partial f / \partial X_i) - \lambda(1 + r_i) = 0 \\ \text{and } W = \sum_{i=0}^n X_i(1 + r_i) \end{array} \right\} I$$

and for the second version

$$\left. \begin{aligned} (1 + r_i) - \mu(\partial f / \partial X_i) &= 0 \\ \text{and } T = f(X_0, X_1, \dots, X_n) \end{aligned} \right\} II$$

for all $i = 0, \dots, n$, and where λ and μ are the Lagrangian multipliers for each version respectively. It can be seen that if the first $N + 1$ first-order marginal conditions are solved in each version, one would be able to replicate the misinterpretation of the duality theorem made by Moroney and Wilbratte (1976). It should be clear, therefore, that if the last first-order marginal condition is also used to solve for $N + 1$ instead of N asset demand equations such that they were also a function of wealth and the anticipated volume of transactions, one would obtain a *very different* set of asset demand equations for each version. The correct dual problem would be stated as follows:

$$\text{minimise } W = \sum_{i=0}^n X_i(1 + r_i)$$

$$\text{subject to } T = f(X_0, X_1, \dots, X_n)$$

which would be a rather strange thing for households to do! There are several discussions of duality theory in many mathematical texts; among the best are, for example, Madden (1986), chapter 12, especially pp.180-4.

22. A function is said to be homogeneous of degree r if $f(tx) = t^r f(x)$ where x is a vector of variables. But if such

a function is not homogenous of degree r such that $f(t\mathbf{x}) \neq t^r f(\mathbf{x})$ but if $f(\mathbf{x})$ can be shown to have been a monotone increasing transformation of a homogeneous function, say $g(\mathbf{x})$ such that $f(\mathbf{x}) = h[g(\mathbf{x})]$, then $f(\mathbf{x})$ is said to be homothetic. Thus, it can be shown that the VES aggregator function is not homothetic since $\rho_i \neq \rho_j$ for all i and j , but if $\rho_i = \rho_j$ for all i and j , it can be shown that this generalised CES function will be homothetic since a monotone increasing transformation of that function will be homogeneous. See Madden (1986), pp.240-243.

23. If so desired, one can adjust the rental price to account for taxation such that the adjusted rental price would be;

$$\pi_{it} = \frac{p_t^*(R_t - r_{it})(1 - \tau_t)}{(1 + R_t)(1 - \tau_t)}$$

as suggested by Barnett (1981, p.197, equation 7.5).

24. The GES (generalised elasticity of substitution) utility function as defined by Boughton (1981, p.378) possesses the property that it is homothetic with respect to income.
25. For a simple overview of the Divisia aggregation procedure, see Barnett and Spindt (1982) and Barnett *et al* (1981). More technical and systematic expositions can be found in Barnett (1980) and Barnett (1981), Chapter 7. The empirical evidence regarding the relative performance of Divisia to simple-sum aggregates is discussed in section 4.4 below, but the reader may like to refer to Barnett (1982), Cagan (1982), and Porter and Offenbacher (1984).

26. See Barnett and Spindt (1982), pp.2-3.
27. See Barnett *et al* (1981), p.498.
28. Numerical examples of the computation of Divisia index numbers can be found in Barnett (1980), pp.39-43, and another example is provided in Barnett and Spindt (1982).
29. See Barnett (1982), pp.690-691 and Barnett and Spindt (1982), p.7. If one refers back to figure 4.1. of this chapter, it will be seen that a change in relative interest rates could be broken up conceptually into a 'substitution effect' and an 'income effect'. The former will involve the 'budget' constraint sliding along the original indifference curve, and then the latter will involve the new 'budget' budget constraint shifting parallel to itself to a higher indifference curve which represents higher utility. This is one thing that simple-sum aggregates cannot do: internalise substitution effects.
30. The current analysis is based mainly on Barnett (1982), pp.690-691.
31. Note that there is an error in the title of Table 3-14 in Porter and Offenbacher (1984), p.88. From the sequence of the tables given in that study, the table should refer to conventional aggregates and not to Divisia aggregates.

Chapter Five

1. See Fisher (1896), pp.8-9.
2. A superscript dot over a variable denotes that it is a time derivative, i.e. rate of change with respect to time. Thus, for example, $\dot{p} = dp/dt$ where t stands for time.
3. The effect of relaxing the assumption of zero 'storage costs' is discussed below in sub-section 2.1.2.
4. This expression is also used to discuss exchange rate arbitrage. This was recognised by Fisher (1930), pp.403-407.
5. This is discussed further in Section 5.3.
6. See Thornton (1802, pp.335-336)
7. See Wicksell (1898, trans. 1936, p.149)
8. See especially Lutz (1974), pp.102-111 and Sargent (1969), pp.130-132.
9. See Fisher (1896, p.67) and also Fisher (1930, pp.415-416).
10. See also Fisher (1930, pp.399-400).
11. See Fisher (1896, pp.76-77).
12. See Fisher (1911, pp.58-59).
13. *ibid*, pp.59-60.
14. *ibid*, pp.64-70.
15. *ibid*, pp.70-71.
16. See Fisher (1930, pp.418-420)
17. For a discussion on the criticism regarding Fisher's "implausibly long" lags, see Sargent (1973), Sections I and II.
18. See Lutz (1974, p.109).
19. See the quotation from Fisher (1896, p.77) given in sub-section 5.3.1.
20. Compare Figure 5.1. with those of Lutz (1974).

Chapter Six

1. See Appendix 3 for a fuller derivation of the aggregate production function.
2. A distinction was made by Gurley and Shaw (1960) between 'outside' and 'inside' money. Outside money is money backed by assets that do not represent a claim on members inside the economy, for example, fiat currency backed by government securities, gold, or foreign exchange reserves. The holding of fiat currency by members of the economy does not impose any offsetting obligations upon them, so money is a net asset. Conversely, inside money is money backed by assets that do represent a claim on members of the economy: for example commercial bank deposits backed by investments and lending to the private sector. In this case, money is not a net asset.
3. This is easily seen if one considers that the definition of savings is disposable income minus consumption. Thus the statements made in this thesis and by Feldstein (1976) are both equivalent.
4. This result comes about if investment is considered as the net addition to the capital stock per unit time. If Feldstein assumes that the labour force grows at an exponential rate of n per unit time such that $N = N_0 e^{nt}$, then, in order to keep the capital-labour ratio constant, it would imply that $K = K_0 e^{nt}$. Differentiating this with respect to time gives $I = \dot{K} = nK$.
5. The notation '*Div*' is used here to emphasise the point that the all-debt financing model is indeed a special case of a more general debt-equity financing model. In an all-debt financing

model, 'dividends' may be equal to zero, or might even represent 'retained earnings' in which Feldstein (1976, p.810, footnote 3) describes equity profits as 'intramarginal'.

6. Given an expression of the form:

$$g(y, X, \dot{X}, t) = \int_0^{\infty} f(y, X, \dot{X}, t) dt,$$

the necessary conditions for this to obtain an extremum are the following Euler equations:

$$\frac{\partial f}{\partial y} = 0 \quad \text{and} \quad \frac{\partial f}{\partial X} - \frac{d}{dt} \left[\frac{\partial f}{\partial \dot{X}} \right] = 0$$

for $0 < t < \infty$. But if the above expression is not explicitly a function of \dot{X} such that

$$g(y, X, t) = \int_0^{\infty} f(y, X, t) dt,$$

then, of course, the Euler equations become

$$\frac{\partial f}{\partial y} = 0 \quad \text{and} \quad \frac{\partial f}{\partial X} = 0$$

For a discussion on the classical 'calculus of variations' problem, see Intriligator (1971), Chapter 12.

7. These two conditions can also be derived if the firm chooses to maximise its own profits rather than its present value in

which case profits are defined by gross revenue, less the wage bill, and less the cost of capital:

$$\Pi = pF(K, N) - wN - p(R - \pi)K$$

Differentiation of the preceding expression with respect to N and K leads to the same marginality conditions.

8. The equivalence of the marginality condition for capital with Feldstein's per-capita marginality condition can be demonstrated as follows. Utilising the linear homogeneity of the production function, it follows that

$$\lambda Y = F(\lambda K, \lambda N)$$

Setting $\lambda = 1/N$,

$$Y = NF(K/N)$$

Substitution for Y in the profit function, and then differentiating with respect to K gives

$$\Pi_K = pNF'(K/N) \cdot (1/N) - p(R - \pi) = 0$$

which implies that $F'(K/N) = (R - \pi)$.

9. The requirement for the denominator of equation [17] to be negative when $\sigma' < 0$ is that

$$\sigma[(1 - \gamma)F_K + nL] - n(1 + L) < 0$$

To show that this inequality holds true, multiply it by K to obtain

$$\sigma[(1 - \gamma)KF_K + nKL] - nK(1 + L) < 0$$

Letting $(M/p) = KL$, the above inequality becomes

$$\sigma[(1 - \gamma)KF_K + n(M/p)] < nK + n(M/p)$$

But Feldstein's equilibrium condition states that (1976, p.812 - not in per-capita form)

$$\sigma[(1 - \gamma)F(K, N) + n(M/p)] = nK + n(M/p)$$

Hence, the inequality becomes

$$\sigma[(1 - \gamma)KF_K + n(M/p)] < \sigma[(1 - \gamma)F(K, N) + n(M/p)]$$

or equivalently, $KF_K < F(K, N)$ which holds.

10. By considering the numerator of equation [17], it is clear that $dK/d\pi = 0$ when $\theta_1 = \theta_2 = \tau_1 = \tau_2 = 0$, and when $L_R = 0$.
11. From equation [17], it can be seen that $dK/d\pi > 0$ when $\tau_2 = \theta_2 = 0$ or when $\theta_1 = \theta_2 = \tau_1 = \tau_2$. Since $F_{KK} < 0$, it follows that the second term of equation [25] is negative.
12. For brief surveys of the empirical evidence, see, for example, Artis and Lewis (1981, Chapter 2), and Coghlan (1980, Chapter 5).
13. With the exception of cumulative preference shareholders. In

this model, they are ruled out for simplicity.

14. The formulation here is based on the first formulation of Hayashi (1982, p.215, equation 4a).
15. Further to note 6, consider the following constrained dynamic optimisation problem:

$$\max \int_0^{\infty} g(\mathbf{x}, \dot{\mathbf{x}}, t) dt = \int_0^{\infty} f(\mathbf{x}, \dot{\mathbf{x}}, t) dt$$

subject to

$$\mathbf{z} = \mathbf{h}(\mathbf{x}, \dot{\mathbf{x}}, t)$$

where $\mathbf{x} = \{x_1, \dots, x_n\}$, $\dot{\mathbf{x}} = \{\dot{x}_1, \dots, \dot{x}_n\}$, and $\mathbf{z} = \{z_1, \dots, z_n\}$.

The Langrangian function is now defined as

$$L(\mathbf{x}, \dot{\mathbf{x}}, t, \lambda) = f(\mathbf{x}, \dot{\mathbf{x}}, t) + \lambda [\mathbf{z} - \mathbf{h}(\mathbf{x}, \dot{\mathbf{x}}, t)]$$

Amongst the necessary conditions for an extremum are the following Euler equations

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left[\frac{\partial L}{\partial \dot{\mathbf{x}}} \right] = 0$$

for $0 \leq t < \infty$. See Intriligator (1971), pp.317 - 320 for further details.

16. In deriving equation 11, Summers (1981) draws on the work of Hayashi (1982) who demonstrated that in a perfectly competitive world,

$$q = \frac{V - A}{pK}$$

(see p.218, equation 14) where $V - A$ is the market valuation of the firm net of the present value of tax deductions. In comparing Summers' equation 11 with equation [46] of this thesis, $(V - A)/pK$ should be replaced by marginal- q . The above relationship is useful because it allows the q -theory of investment to be tested empirically - hence the theme of Summers' paper.

17. For a model in which the debt-equity ratio is determined endogenously, see Feldstein *et al* (1978).
18. Note that in Summers' notation, r stands for the nominal interest rate, and not the real interest rate as defined in this thesis.
19. It is still being assumed that the labour supply is exogenously determined so that $dN^S = 0$ in equation [iii].

Chapter Seven

1. The notation used here is different from that of Currie (1981, pp.704–705) since it relates specifically to the demand for money.
2. When comparing the coefficients of the first-order rate of growth variables given by Currie (1981, p.705) and those given by Patterson and Ryding (1982, Table A, p.10), let $\ell = j + 1$.
3. See James and Nobes (1983), Chapter 12, especially p.267.
4. Equation (3) of Currie (1982, p.67) has been erroneously reported since some coefficients have been mixed up with those of equation B(i) in Coghlan (1979); however, the results derived are correct on the basis of equation B(ii).

REFERENCES

- ARTIS, M.J. & M.K. LEWIS (1976), 'The Demand for Money in the United Kingdom, 1963-1973.', *Manchester School*, v.44, no.2, pp.147-181.
- ARTIS, M.J. & M.K. LEWIS (1981), *Monetary Control in the United Kingdom*, Oxford, Philip Allan.
- BAILEY, M.J. (1956), 'Welfare Cost of Inflationary Finance.', *Journal of Political Economy*, v.64, no.2, pp.93-110.
- BANK OF ENGLAND (1971a), 'Key Issues in Monetary and Credit Control.', *Bank of England Quarterly Bulletin*, v.11, no.2, pp.195-198.
- BANK OF ENGLAND (1971b), 'Reserve Ratios: Further Definitions.', *Bank of England Quarterly Bulletin*, v.11, no.4, pp.482-489.
- BANK OF ENGLAND (1974), 'Credit Control: a Supplementary Scheme.', *Bank of England Quarterly Bulletin*, v.14, no.1, pp.37-39.
- BANK OF ENGLAND (1982a), 'The Supplementary Special Deposits Scheme.', *Bank of England Quarterly Bulletin*, v.22, no.1, pp.74-85.
- BANK OF ENGLAND (1982b), 'Transactions Balances - a New Monetary Aggregate.', *Bank of England Quarterly Bulletin*, v.22, no.2, pp.224-225.
- BANK OF ENGLAND (1983), 'Competition, Innovation, and Regulation in British Banking.', *Bank of England*

- Quarterly Bulletin*, v.23, no.3, pp.363-376.
- BARNETT, W.A. (1980), 'Economic Monetary Aggregates: an Application of Index Numbers and Aggregation Theory.', *Journal of Econometrics*, v.14, no.1, pp.11-48.
- BARNETT, W.A. (1981), *Consumer Demand and Labour Supply: Goods, Monetary Assets, and Time*, Amsterdam, North-Holland.
- BARNETT, W.A. (1982), 'The Optimal Level of Monetary Aggregation.', *Journal of Money, Credit, and Banking*, v.14, no.4, pp.687-710.
- BARNETT, W.A., E.K. OFFENBACHER, & P.A. SPINDT (1981), 'New Concepts of Aggregated Money.', *Journal of Finance*, v.36, no.2, pp.497-505.
- BARNETT, W.A. & P.A. SPINDT (1982), *Divisia Monetary Aggregates: Their Compilation, Data, and Historic Behaviour*, Federal Reserve Board Staff Study No.116, Washington D.C., Publications Services, Federal Reserve Board.
- BAUMOL, W.J. (1952), 'The Transactions Demand for Cash: an Inventory-Theoretic Approach.', *Quarterly Journal of Economics*, v.66, no.4, pp.545-556.
- BEN-HORIM, M. & W.L. SILBER (1977), 'Financial Innovation: a Linear Programming Approach.', *Journal of Banking and Finance*, v.1, pp.277-296.
- BOORMAN, J.T. (1985), 'A Survey of the Demand for Money: Theoretical Formulations and Pre-1973 Empirical Results.', in *Modern Concepts in Macroeconomics*, edited by T.M. Havrilesky, Arlington Heights, IL, Harlan

- Davidson. Formerly printed in *Money Supply, Money Demand, and Macroeconomic Models* (2nd edition), edited by T.M. Havrilesky and J.T. Boorman, Arlington Heights, IL, Harlan Davidson, 1982.
- BOUGHTON, J.M. (1981), 'Money and Its Substitutes.', *Journal of Monetary Economics*, v.8, no.3, pp.375-386.
- BOYD, J.H. (1976), 'Household Demand for Checking Account Money.', *Journal of Monetary Economics*, v.2, no.1, pp.81-98.
- BRUNNER, K. & A.H. MELTZER (1967), 'Economies of Scale in Cash Balances Reconsidered.', *Quarterly Journal of Economics*, v.81, no.3, pp.422-436.
- BRUNNER, K. & A.H. MELTZER (1971), 'The Uses of Money: Money in the Theory of an Exchange Economy.', *American Economic Review*, v.61, no.5, pp.784-805.
- CAGAN, P. (1956), 'The Monetary Dynamics of Hyper-Inflation.', in *Studies in the Quantity Theory of Money*, edited by M. Friedman, Chicago, University of Chicago Press.
- CAGAN, P. (1982), 'The Choice among Monetary Aggregates as Targets and Guides for Monetary Policy.', *Journal of Money, Credit, and Banking*, v.14, no.4, pp.661-686.
- CAGAN, P. & A.J. SCHWARTZ (1975), 'Has the Growth of Money Substitutes Hindered Monetary Policy?', *Journal of Money, Credit, and Banking*, v.7, no.2, pp.137-159.
- CHETTY, V.K. (1969), 'Measuring the Nearness of Near-Money.', *American Economic Review*, v.59, no.3, pp.270-281.
- CHOW, G. (1966), 'On the Short-Run and Long-Run Demand for Money.', *Journal of Political Economy*, v.44, no.2,

p.111-131.

- CLOWER, R.L. (1971), 'Theoretical Foundations of Monetary Policy.', in *Monetary Theory and Monetary Policy in the 1970s: Proceedings of the Sheffield Money Seminar*, edited by G. Clayton, J.C. Gilbert, and R. Sedgewick, Oxford, Oxford University Press.
- COGHLAN, R.T. (1979), 'A Small Monetary Model of the UK Economy.', *Bank of England Discussion Paper*, no.3 (May).
- COGHLAN, R.T. (1980), *The Theory of Money and Finance*, London, Macmillan.
- CROCKETT, A.D. (1970), 'Timing Relationships between Movements of Monetary and National Income Variables.', *Bank of England Quarterly Bulletin*, v.10, no.4, pp.459-472.
- CURRIE, D. (1981), 'Some Long-Run Features of Dynamic Time-Series Models.', *Economic Journal*, v.91, no.3, pp.704-715.
- CURRIE, D. (1982), 'The Long-Run Properties of the Bank of England Small Monetary Model of the UK Economy.', *Applied Economics*, v.14, no.2, p.63-72.
- DePRANO, M. & T. MAYER (1965), 'Tests of the relative importance of autonomous expenditures and money.', *American Economic Review*, v.55, no.4, pp.729-752.
- DIEWERT, W.E. (1974), 'Functional Forms for Revenue and Factor Requirement Functions.', *International Economic Review*, v.15, no.1, pp.119-130.
- DIEWERT, W.E. (1976), 'Exact and Superlative Index Numbers.',

- Journal of Econometrics*, v.4, no.2, pp.115-145.
- DONOVAN, D.J. (1978), 'Modelling the Demand for Liquid Assets: an Application to Canada.', *IMF Staff Papers*, v.25, no.4, pp.676-704.
- EWIS, N.A. & D. FISHER (1984), 'The Translog Utility Function and the Demand for Money in the US.', *Journal of Money, Credit, and Banking*, v.16, no.1, pp.34-51.
- FAMA, E. (1975), 'Short Term Interest Rates as Predictors of Inflation.', *American Economic Review*, v.65, no.3, pp.269-282.
- FEIGE, E.L. & D.K. PEARCE (1977), 'The Substitutability of Money and Near-Monies: a Survey of the Time-Series Evidence.', *Journal of Economic Literature*, v.15, no.2, pp.439-469.
- FELDSTEIN, M. (1976), 'Inflation, Income Taxes, and the Rate of Interest.', *American Economic Review*, v.66, no.5, pp.809-820.
- FELDSTEIN, M., J. GREEN, & E. SHESHINSKI (1978), 'Inflation and Taxes in a Growing Economy with Equity and Debt Finance.', *Journal of Political Economy*, v.86, no.2, pt.2, pp.S53-S70.
- FISHER, I. (1896), *Appreciation and Interest*, New York, Macmillan. Reprinted in *Mathematical Investigations in the Theory of Value and Price (1892)*, and *Appreciation and Interest (1896)*, Reprints of Economics Classics, New York, Augustus Kelley, 1965.
- FISHER, I. (1911), *The Purchasing Power of Money*, New York, Macmillan.

- FISHER, I. (1922), *The Making of Index Numbers*, Boston, Houghton-Mifflin.
- FISHER, I. (1930), *The Theory of Interest*, New York, Macmillan.
- FRIEDMAN, M. (1964), 'Post-War Trends in Monetary Theory and Policy.', *National Banking review*, v.2, September, pp.1-9.
- FRIEDMAN, M. & D. MEISELMAN (1963), 'The Relative Stability of Monetary Velocity and the Investment Multipliers in the U.S., 1897-1958.', in Commission on Money and Credit, *Stabilisation Policies*, Eaglewood Cliffs, NJ, Prentice-Hall, pp.165-223.
- FRIEDMAN, M. & A.J. SCHWARTZ (1969), 'The Definition of Money: Net Wealth and Neutrality Criteria.', *Journal of Money, Credit, and Banking*, v.1, no.1, pp.1-14.
- GAFGA, P.H. (1985a), 'To What Extent can Inflationary Expectations be Regarded as Rational?', *B.A. Dissertation*, mimeo., Dept. of Economics, University of Durham.
- GAFGA, P.H. (1985b), 'The Instability of the Demand for Money in the UK, 1960-1983.', *Applied Econometrics Project*, mimeo., Dept. of Economics, University of Durham.
- GEBAUER, W. (1986), 'Real Rates and Capital Investment: A Fisher-Wicksell Perspective.', in *Shifting Frontiers in Financial Markets*, edited by D.E. Fair, Dordrecht, Martinus-Nijhoff.
- GIBSON, W.E. (1972), 'Interest Rates and Inflationary Expectations: New Evidence.', *American Economic Review*, v.62, no.5, pp.854-865.

- GOLDFELD, S.M. (1973), 'The Demand for Money Revisited.', *Brookings Papers on Economic Activity*, Part 3, pp.577-638.
- GOLDFELD, S.M. (1976), 'The Case of the Missing Money.', *Brookings Papers on Economic Activity*, Part 3, pp.683-730.
- GOLDFELD, S.M. (1982), 'Comment on the Optimal Level of Monetary Aggregation.', *Journal of Money, Credit, and Banking*, v.14, no.4, pp.716-720.
- GOODHART, C.A.E. (1984), *Monetary Theory and Practice: the UK Experience*, London, Macmillan.
- GRADY, J. & M. WEALE (1986), *British Banking, 1960-85*, London, Macmillan.
- GRANGER, C.W.J. (1969), 'Investigating Causal Relations by Econometric and Cross-Spectral Methods.', *Econometrica*, v.37, no.3, pp.424-438.
- GURLEY, J. & E.S. SHAW (1955), 'Financial Aspects of Economic Development.', *American Economic Review*, v.45, no.4, pp.515-538.
- GURLEY, J. & E.S. SHAW (1960), *Money in a Theory of Finance*, Washington D.C., Brookings Institution.
- HAMBURGER, M.J. (1969), 'Alternative Interest Rates and the Demand for Money: Comment.', *American Economic Review*, v.59, no.3, pp.407-412.
- HAMBURGER, M.J. (1977), 'Behaviour of the Money Stock: Is There a Puzzle?', *Journal of Monetary Economics*, v.3, no.3, pp.265-288.
- HAMBURGER, M.J. (1984), 'Financial Innovation and Monetary

- Targeting.', in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets*, Federal Reserve Bank of St. Louis - Conference Papers, Amsterdam, Kluwer-Nijhoff.
- HAMILTON, A. (1986), *The Financial Revolution*, Harmondsworth, Middlesex, Penguin.
- HAYASHI, F. (1982), 'Tobin's Marginal-q and Average-q: A Neoclassical Interpretation.', *Econometrica*, v.50, no.1, pp.213-224.
- HEIN, S.E. (1980), 'Dynamic Forecasting and the Demand for Money.', *Federal Reserve Bank of St. Louis Review*, v.62, no.6, pp.13-23.
- HEIN, S.E. (1984), Discussion on Porter and Offenbacher (1984), in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets*, Federal Reserve Bank of St. Louis - Conference Papers, Amsterdam, Kluwer-Nijhoff.
- HENDRY, D.F. & G.E. MIZON (1978), 'Serial Correlation as a Convenient Simplification, not a Nuisance.', *Economic Journal*, v.88, no.3, pp.549-583.
- HESTER, D.D. (1981), 'Innovations and Monetary Control.', *Brookings Papers on Economic Activity*, Part 1, pp.141-189.
- HESTER, D.D. (1982), 'On the Adequacy of Policy Instruments and Information when the Meaning of Money is Changing.', *American Economic Review*, v.72, Papers and Proceedings, pp.40-44.
- INTRILIGATOR, M.D. (1971), *Mathematical Optimisation and*

- Economic Theory*, Eaglewood Cliffs, NJ, Prentice-Hall.
- JAMES, S. & C. NOBAY (1983), *The Economics of Taxation*, 2nd edition, Oxford, Philip Allan.
- JOHNSTON, R.B. (1984), *The Demand for Mon-Interest Bearing Money in the United Kingdom*, Government Economic Service Working Paper no.66, February, London, HMSO.
- JUDD, J.P. & J.L. SCADDING (1982), 'The Search for a Stable Money Demand Function: a Survey of the Post-1973 Literature.', *Journal of Economic Literature*, v.20, no.3, pp.993-1023.
- JUDGE, G. (1983), 'The Stability of Behavioural Relationships and their Empirical Estimates.', *British Journal of Economic Issues*, v.5, no.12, pp.1-14.
- KAMIEN, M.J. & N.I. SCHWARTZ (1982), *Market Structure and Innovation*, Cambridge, Cambridge University Press.
- KANE, E.J. (1977), 'Good Intentions and Unintended Evil. The Case Against Selective Credit Allocation.', *Journal of Money, Credit, and Banking*, v.7, no.1, pp.55-69.
- KANE, E.J. (1981), 'Accelerating Inflation, Technological Innovation, and the Decreasing Effectiveness of Banking Regulation.', *Journal of Finance*, v.36, no.2, pp.355-367.
- KANE, E.J. (1983), 'Policy Implications of Structural Changes in Financial Markets.', *American Economic Review*, v.73, no.2, pp.96-100.
- KANE, E.J. (1984), 'Microeconomic and Macroeconomic Origins of Financial Innovation.', in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets*, Federal Reserve Bank of St. Louis - Conference Papers,

Amsterdam, Kluwer-Nijhoff.

- KOGIKU, K.C. (1971), *Microeconomic Models*, , .
- KOUTSOYIANNIS, A. (1979), *Modern Microeconomics*, 2nd edition, London, Macmillan.
- LAILLER, D.E.W. (1969), 'The Definition of Money: Theoretical and Empirical Problems.', *Journal of Money, Credit, and Banking*, v.1, no.3, pp.509-525.
- LAILLER, D.E.W. (1981), 'Monetarism: An Interpretation and an Assessment.', *Economic Journal*, v.91, no.1, pp.1-28.
- LEE, T.H. (1967), 'Alternative Interest Rates and the Demand for Money: The Empirical Evidence.', *American Economic Review*, v.57, no.5, pp.1168-1181.
- LEE, T.H. (1969), 'Alternative Interest Rates and the Demand for Money: Reply.', *American Economic Review*, v.59, no.3, pp.412-418.
- LIEBERMAN, C. (1977), 'The Transactions Demand for Money and Technological Change.', *Review of Economics and Statistics*, v.59, no.3, pp.307-317.
- LUCAS, R.E. (1981), *Studies in Business Cycle Theory*, Oxford, Basil Blackwell.
- LUTZ, F.A. (1974), 'Inflation and the Rate of Interest.', *Banca Nazionale del Lavoro Quarterly Review*, no.109, pp.99-117.
- MADDEN, P. (1986), *Concavity and Optimization in Microeconomics*, Oxford, Basil Blackwell.
- MELTZER, A.H. (1963), 'Demand for Money: the Evidence from Time Series.', *Journal of Political Economy*, v.71, no.2, pp.219-246.

- MEYER, L.H. (1984), Discussion on Hamburger (1984), in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets*, Federal Reserve Bank of St. Louis - Conference Papers, Amsterdam, Kluwer-Nijhoff.
- MIDLAND BANK (1966), 'London's New Markets for Money.', *Midland Bank Review*, no.2, pp.3-10.
- MILLER, M.H. & D. ORR (1966), 'A Model of the Demand for Money by Firms.', *Quarterly Journal of Economics*, v.80, no.3, pp.413-435.
- MORGAN, G.V. (1969), 'The Essential Qualities of Money.', *Manchester School*, v.37, no.3, pp.237-248.
- MORONEY, J.R. & B.J. WILBRATTE (1976), 'Money and Money Substitutes.', *Journal of Money, Credit, and Banking*, v.8, no.2, pp.181-198.
- MUNDELL, R.A. (1963), 'Inflation and Real Interest.', *Journal of Political Economy*, v.71, no.3, pp.280-283.
- NEWLYN, W.T. (1964), 'The Supply of Money and its Control.', *Economic Journal*, v.74, no.2, pp.327-346.
- NEWLYN, W.T. & R. BOOTLE (1978), *The Theory of Money*, Oxford, The Clarendon Press.
- NIEHANS, J. (1978), *The Theory of Money*, Baltimore, The John Hopkins University Press.
- O'BRIEN, D.P. (1983), 'Monetary Theory in Historical Perspective.', in *Economic Analysis in Historical Perspective*, edited by D.P. O'Brien and J. Creedy, Sevenoaks, Butterworths.
- ORR, D. (1970), *Cash Management and the Demand for Money*, New York, Praeger.

- PATTERSON, K.D. & J. RYDING (1982), 'Deriving and Testing Rate of Growth and Higher Order Growth Effects in Dynamic Economic Models.', *Bank of England Discussion Paper*, no.21, August.
- PATTERSON, K.D. & J. RYDING (1984), 'The Modified Fisher Hypothesis and the Steady-State Demand for Money.', *Manchester School*, v.52, no.3, pp.300-313.
- PESEK, B. & T.R. SAVING (1967), *Money, Wealth, and Economic Theory*, New York, Macmillan.
- PESEK, B. & T.R. SAVING (1968), *Foundations of Money and Banking*, New York, Macmillan.
- PODOLSKI, T.M. (1986), *Financial Innovation and the Money Supply*, Oxford, Basil Blackwell.
- PORTER, R.D. & E.K. OFFENBACHER (1984), 'Financial Innovations and Measurement of Monetary Aggregates.', in *Financial Innovations: Their Impact on Monetary Policy and Financial Markets*, Federal Reserve Bank of St. Louis - Conference Papers, Amsterdam, Kluwer-Nijhoff.
- PORTER, R.D. & T.D. SIMPSON (1980), 'Some Issues Involving the Definition and Interpretation of the Monetary Aggregates.', in *Controlling Monetary Aggregates III*, Federal Reserve Bank of Boston, Conference Series no.23.
- RADCLIFFE COMMITTEE (1959), 'Committee on the Working of the Monetary System.', *Report*, Cmnd. 827, London, HMSO.
- SARGENT, T.J. (1969), 'Commodity Price Expectations and the

- Interest Rate.', *Quarterly Journal of Economics*, v.83, no.1, pp.127-140.
- SARGENT, T.J. (1973), 'Interest Rates and Prices in the Long Run.', *Journal of Money, Credit, and Banking*, v.5, no.1, pp.385-449.
- SCHERER, F.M. (1980), *Industrial Market Structure and Economic Performance*, Boston, Houghton-Mifflin.
- SCHUMPETER, J.A. (1934), *The Theory of Economic Development*, translated by R. Opie, New York, Oxford University Press, 1961.
- SCHUMPETER, J.A. (1939), *Business Cycles* (2 volumes), New York, McGraw-Hill.
- SHACKLE, G.L.S. (1971), Discussion on Clower (1971), in *Monetary Theory and Monetary Policy in the 1970s: Proceedings of the Sheffield Money Seminar*, edited by G. Clayton, J.C. Gilbert, and R. Sedgewick, Oxford, Oxford University Press.
- SILBER, W.L. (1975), 'Towards a Theory of Financial Innovation.', in *Financial Innovation*, edited by W.L. Silber, Farnborough, Hants., Lexington Books/D.C. Heath, pp.53-85.
- SILBER, W.L. (1983), 'The Process of Financial Innovation.', *American Economic Review*, v.73, no.2, pp.85-89.
- SILBERBERG, E. (1978), *The Structure of Economics: A Mathematical Analysis*, Internattional Student Edition, New York, McGraw-Hill, 1981.
- SIMS, C.A. (1972), 'Money, Income, and Causality.', *American Economic Review*, v.62, no.3, pp.540-552.

- SUMMERS, L.H. (1981), 'Taxation and Corporate Investment: A q -Theoretic Approach.', *Brookings Papers on Economic Activity*, Part 1, pp.67-127.
- SUMMERS, L.H. (1983), 'The Nonadjustment of Nominal Interest Rates: a Study of the Fisher Effect.', in *Macroeconomics, Prices, and Quantities*, edited by J. Tobin, Oxford, Basil Blackwell.
- SYLLA, R. (1982), 'Monetary Innovation and Crises in American Economic History.', in *Crises in the Economic and Financial Structure*, edited by P. Wachtel, New York, Lexington Books/D.C. Heath.
- THORNTON, H. (1939), 'Two Speeches of Henry Thornton on the Bullion Report, May 1811.' Reprinted in Appendix III in *An Enquiry Into the Nature and Effect of the Paper Credit of Great Britain (1802)*, edited with an introduction by F.A. von Hayek, London, George Allen and Unwin.
- TOBIN, J. (1955), 'A Dynamic Aggregative Model.', *Journal of Political Economy*, v.23, no.1, pp.103-115.
- TOBIN, J. (1956), 'The Interest Elasticity of Transactions Demand for Cash.', *Review of Economics and Statistics*, v.38, no.3, pp.241-247.
- TOBIN, J. (1958), 'Liquidity Preference as Behaviour Towards Risk.', *Review of Economic Studies*, v.25, no.1, pp.65-86.
- TOBIN, J. (1965), 'Money and Economic Growth.', *Econometrica*, v.33, no.4, pp.671-684.
- TOBIN, J. (1969), 'A General Equilibrium Approach to Monetary

- Theory.', *Journal of Money, Credit, and Banking*, v.1, no.1, pp.15-29.
- TÖRQUINST, L. (1936), 'The Bank of Finland's Consumption Index.', *Bank of Finland Monthly Bulletin*, v.10, pp.1-8.
- WHALEN, E.L. (1966), 'A Rationalisation of the Precautionary Demand for Cash.', *Quarterly Journal of Economics*, v.80, no.2, pp.314-324.
- WICKSELL, K. (1898), *Interest and Prices*, translated by B. Ohlin, London, Macmillan, 1936.
- WILLIAMS, D., GOODHART, C.A.E., & GOWLAND, D.H. (1976), 'Money, Income, and Causality: The U.K. Experience.', *American Economic Review*, v.66, no.3, pp.417-423.
- WILSON COMMITTEE (1978), 'Committee to Review the Function of Financial Institutions.', *Evidence on the Financing of Industry and Trade* (8 volumes), London, HMSO.
- WILSON COMMITTEE (1980), 'Committee to Review the Function of Financial Institutions.', *Report*, Cmnd.7937, London, HMSO. Published in two parts, *Report* and *Appendices*.
- YEAGER, B. (1968), 'Essential Properties of the Medium of Exchange.', *Kyklos*, v.21, pp.45-69.

