On "Impulse" Control And The Demand For Money

Arujuna Sivananthan

Department of Political Economy

University of Glasgow

Submitted in partial fulfilment of the requirements

of the degree of Doctor of Philosophy

3 July 1997

Abstract

The Objective

Money affects every aspect of life and yet its impact on a macro or micro level is not clearly understood. From an individual's point of view an efficient cash management policy could free resources for consumption which otherwise may have been wasted on either holding or transaction costs. But few models analyse a risk-averter's cash management decision and its impact on the money stock. Part of this deficiency can be attributed to the difficulties which arise from the non-linearities inherent in concave utility functions. The sheer complexity of modelling the dynamic evolution of variables which influence the cash management decision, and, the interaction between them has been another factor.

The history of research into the demand for money is vast and has been an important feature in the evolution of macroeconomic theory. Numerous modelling approaches have been utilised to study the many properties of money,

varying from general equilibrium analysis to micro-based models in which the agent behaves like a private optimiser.

Important contributions have been made by eminent economists on how the money stock behaves. Fisher's quantity theory identity $MV = PY$; M is the nominal stock of money, V is the velocity of circulation, P is the price level, and Y is the volume of transactions or real income, which was developed in 1911 still features prominently in economic analysis. The velocity of circulation is assumed to be determined by an exogenous payments mechanism and therefore constant. Hence any change to the money stock yields neutral effects over the long run. Pigou (1917) changes this to include the consumer allocation problem, interest rates and wealth, which subsequently comes to form the basis of the Cambridge equation. These models set the tone for the literature which later followed from the various Classical schools arguing in favour of a passive monetary policy.

Keynes in his *General Theory of Employment, Interest and Money* (1936) radically challenges this view by arguing that velocity was not constant, but varied with the price level and income, which, therefore, required an interventionist monetary authority. He divides the money stock into three components proposing that agents hold money for three very different reasons. The first he concludes is the transactions motive where agents hold money to satisfy planned expenditure. The second is the precautionary motive where money is held as a buffer stock to absorb any unanticipated expenditure shocks. The third is the speculative motive where agents hold money because it is an asset. At the time Keynes wrote his general theory real appreciations in the value of the nominal money stock were not uncommon. Therefore the role of money as a speculative asset was more important then, than it is now.

Baumol (1952) and Tobin (1956) formalise the transactions motive by placing it within a dual asset optimisation framework. Agents in these models optimally determine their money stock by minimising the associated opportunity costs. Miller and Orr (1966) develop this further by introducing uncertainty through a discrete steady state random walk. By limiting the type of agent considered to be risk neutral, they effectively model the problem as a dual asset management exercise in which the agent optimises his utility of his wealth, similar to Tobin (1958). Constantinides and Richard (1978) model the cash management decision as a net present value problem. Increasing the time horizon reduces the frequency of transactions in which agents switch from cash to the interest earning asset or vice versa but increases their magnitude. Smith (1989) expands on this by allowing for interest rate uncertainty. A critical review of the current literature on the transactions money demand for money is presented in Chapter 3.

The original objective of this thesis was to expand on Smith (1989) by developing a model that studied a risk-averter's cash management decision which included genuine aspects of risk and a discretely varying stochastic interest rate. The motive behind this was to study the impact of increased risk sensitivity on an agent's money demand function and also capture the discrete jumps which interest rates exhibit in the real world.

The Thesis

The standard approach to modelling a stochastically varying cash inventory assumes that net disbursements follow a Wiener process. This assumption is also made here ensuring that the new results presented here are not driven by prescribing a different evolution of the state. This reduces the management problem to one of optimal "impulse" control. The standard methodology for obtaining a solution requires

1. constructing the cost function,
2. expanding it in a Taylor series using Ito's lemma to obtain the Hamilton-Jacobi-Bellmen (HJB) equation and
3. determining the optimal targets and thresholds using the "smooth pasting" and "value matching" conditions.

In other areas of economics the "smooth pasting" condition has also been used as an auxiliary condition to satisfy *perceived* economic assumptions.

However, within the stochastic optimal control literature the use this condition has not been observed.

Increased risk sensitivity is introduced through a Von Neumann-Morgenstern utility function. For risk averse individuals these are assumed to be concave and give rise to a non-linear relationship between interest rates and money holdings in the inhomogeneous term of the HJB equation. Thus requiring the problem to be numerically solved. The algorithm involves

1. solving the HJB equation using the natural boundary conditions, and,

2. optimising it with respect to the targets and thresholds.

On the other hand, applications of "smooth pasting" only requires gradient conditions to be imposed with respect to the initial state. This strange feature along with unexpected numerical results led me to explore both the Ito stochastic differential equation and the Chapman-Kolmogorov equation in more detail. This led to the discovery of the natural boundary conditions which are presented in Chapter 1.

Chapter 2 analyses their impact on the simple menu cost model in Dixit (1991a). The results obtained highlight some limitations of the "smooth pasting" condition. Although the economic intuition does not differ from what is suggested in Dixit (1991a), situations could be envisaged where it could.

Chapters 3 and 4 return to the original objective of this thesis. Chapter 3 critically analyses the key contributions on the transactions demand for money. Their strengths and weaknesses are highlighted. Some models which were previously assumed to be robust, under the detailed scrutiny of this chapter, appear to be logically inconsistent. Chapter 4 solves the problem which was initially outlined. The results present a different image of agent behaviour to what existed before. The optimal targets and thresholds do not appear to be as obvious as perviously believed.

The Results

This thesis makes four unique contributions to the current literature. These are dealt with in the four core chapters.

Chapter 1 demonstrates that "smooth pasting" fails to quantify the costs faced by agent in a more general class of problem. Questions are raised about its validity as a first-order optimisation condition. The natural boundary conditions for optimal "impulse" controlled problems are derived and are shown to be the "value matching" conditions. Thus, enabling "impulse" control problems to solve in a way which is consistent with the principles of optimal control. However, it does not seek to detract from its immense value as a heuristic tool. In simple problems like Dixit(1991a) it yields the same answer as the more rigorous approach. Also, from a non-scientific view it provides fundamental insights into how agents determine their optimal exercise targets for American option type models.

Chapter 2 provides a solution to the Dixit menu cost model using the rigorous formulation of an impulse control problem. The richer solutions obtained yield insights into agent behaviour which were previously unobservable. Also various properties which were assumed are now proven. An analytical equation specifying relationship between the discount rate and the zone of inertia is derived. Formerly this could be only deduced by making an empirical link.

A critical review of the current literature on the transactions demand for money is provided in Chapter 3. The strengths and weaknesses of the "seminal" contributions are highlighted. Also a contrast between the results presented in these models and the empirical literature is provided.

Chapter 4 returns to the original objective of this thesis. The similarity between liquidity preference and transaction money demand models is briefly illustrated in section 4.2. The results clearly show that the demand for money is not well behaved as the existing literature predicts. In fact they demonstrate the existence of multiple optima which point to a sequence of utility maximising strategies. Unlike most rational expectations models, the existence all but one optimum cannot be dismissed through partial equilibrium arguments.

A brief summary of the results is offered in the final chapter.

List of Symbols

$\in$	Element of
$[.,.)$	Open Interval
$E[.]$	Expectation Operator
inf	Infimum
max	Maximum
$\int$	Integral
$\sum$	Summation
ln	Natural log
lim	Moving Towards the Limit
A	Second Order Differential Operator
sin	sine
cos	cosine
arcos	Inverse cosine
sinh	Hyperbolic sine
cosh	Hyperbolic cosine
tanh	Hyperbolic tangent
$\mathbb{R}^n$	n-dimensional space
$Lip_\gamma(D)$	Lipschitz continuous
$\|\| \cdot \|\|$	Norm
$C^i(x)$	Differentiable To The Order i in x

Acknowledgements

This undertaking would not have been possible if not for the unwavering support of my parents. For this and providing me with the opportunity to purse this thesis through their generous financial support I shall be eternally grateful.

I wish to thank Stan Hurn and Anton Muscatelli for supervising thorough out my studies. Their advise and patient tutoring, from the moment the first ideas crossed my mind till they finally crystallised has been invaluable.

In April 1995 I had the privilege of meeting Ken Lindsay, the cleverest man that I have met. What is presented here is entirely due to his guidance, encouragement and close supervision that I enjoyed over the past two years. For this and and all the help he has given me I am deeply indebted to him.

I grateful to Terry Moody and Dan Muldoom, my internal and external examiners, who gave me valuable insights into my work which I had overlooked. Their help on presentational matters is also appreciated.

I am thankful to members of the Department of Economics, who in various seminars provided me with helpful comments and Professors Alain Bensoussan, Wendell Fleming, Jean Lepeltier, Ulrich Haussman and George Constantinides for confirming the validity of the boundary and first-order optimisation conditions described in Chapter 1.

All errors and omissions in this thesis are entirely mine and none of the views expressed here reflect those of the University of Glasgow or any person mentioned here.

Contents

Chapter 1

A Re-Evaluation of the "Smooth Pasting" Condition in Problems of "Impulse Control"

1.1 Introduction

Stochastic optimal control has become increasingly popular in economics and finance as a tool for modelling optimising behaviour within an environment of ongoing uncertainty. Its applications have been numerous, ranging from option pricing theory to target zone and menu cost models, e.g. Pindyck (1988), Krugman (1988), Dixit (1991a), and, Dixit and Pindyck (1994). Under costs of adjustment, or any other form of friction these models demonstrate the existence of inertial responses where it is optimal for an agent to wait before acting. The boundaries of the optimal zone of inertia are derived through heuristically motivated boundary and first-order conditions, commonly referred to as the "value matching" and "smooth pasting" conditions.

This chapter considers the stochastic optimal control of a Wiener process in the presence of any cost of adjustment including "impulse" control. Analytical boundary conditions are derived for the problem *explicitly* from the martingale or optimal stopping framework and do not rely on any heuristic motivation. Although the results yield a condition similar to the "value matching" condition, the "smooth pasting" condition, which is also used in many applications of stochastic optimal control, does not feature in any way.

The chapter is structured as follows. Section 1.2 provides a general overview of the "smooth pasting" condition, highlighting some of its perceived strengths and weaknesses and the need for a strict analytical solution to the boundary value problem. Section 1.3 uses a general example to describe the problem of impulse control. The Bellman value function (value function) is formulated in Sections 1.4 and 1.5 and the boundary conditions are derived. The ideas are expounded in one dimension, although the methodology extends naturally to any dimension. In addition, the analysis is restricted to the case with constant coefficients but a variety of problems with non-constant coefficients can be dealt with in a similar way.

1.2 The "Smooth Pasting" Condition

"Smooth pasting" is a useful heuristic first-order optimisation condition for framing many target-threshold type models and helps convey the economic intuition behind numerous situations in a way that is easily understood. Indeed, Chapter 2 confirms that both "smooth pasting" and the rigorously framed optimal stopping strategy yield algebraically equivalent answers. But, the critique offered later in this section and an analysis of the necessary conditions for optima in Section 1.6 suggests why this need not be the case for more complex problems, or, situations in which only the value function needs to be ascertained such as option valuation.

The arguments for and against "smooth pasting" are outlined in detail, including reasons as to why it is absent from a rigorous formulation of the method of impulse control. However, it must be emphasised that, in the absence of any contradictory results to the stochastic optimal control approach to solving a problem, it still remains a valuable first-order condition.

1.2.1 The Optimal Stopping Problem And "Smooth Pasting"

The utility of "smooth pasting" and its applicability to a wide range of problems within an environment of ongoing uncertainty is best illustrated through the simple optimal stopping problem provided in Dixit and Pindyck

(1994). Consider an entry-exit decision in where a firm is faced with a simple binary choice at every instant. It can either wait and accrue a profit or exercise an option at an endogenously determined barrier for a termination payoff. Both the profit earned and the termination payoff will functions of state and time. Assume that the state follows a Wiener process

$$dx_t = \mu(x,t)dt + \sigma(x,t)dz_t. \tag{1.1}$$

In an exit decision a firm chooses to stop production and sell its equipment for scrap value. Entry decisions can be framed in a similar way. The value accrued during the waiting period is zero. Entry implies investment. The associated termination payoff is the expected net present value of future profits less investment costs.

Self evidently there exists a critical value of the state X at each point in time for which $x < X$ would imply continuation will be optimal and $x > X$ for which stopping will be optimal. Therefore, there must exist some condition which helps us determine X. Let $\gamma(x,t)$ denote the flow profits and $\lambda(x,t)$ be the termination payoff. The payoff facing a firm at each instant can be characterised as being

$$\gamma(x,t)dt + \frac{\rho dt}{1+\rho dt}\lambda(x,t),$$

ρ being the discount rate. In an entry decision $\gamma(x,t)-\rho\lambda(x,t)$ has to increase as x increases. If x is large. For an exit decision this expression must decrease in x. To illustrate the link between optimal stopping and "smooth pasting" I shall only consider the former.

It is obvious that X must divide state and time space into two regions, where continuation and termination are optimal. Of course, an a priori knowledge of X is not possible. Instead it must be endogenously determined.

The Bellman value function for this optimal stopping problem takes the form

$$V(x,t) = \max\left[\lambda(x,t), \gamma(x,t) + \frac{E[V(x+dx, t+dt)|x_0 = x]}{1+\rho dt}\right].$$

In the bounded region in which the state moves this can be expanded in a stochastic Taylor series expansion through Ito's lemma to yield

$$\frac{\sigma(x,t)^2}{2} V_{xx}(x,t) + \mu(x,t)V_x(x,t) + V_t(x,t) - \rho V(x,t) + \gamma(x,t) = 0.$$

In the stopping region clearly $V(x,t) = \lambda(x,t)$, therefore

$$V(X,t) = \lambda(X,t) \qquad \forall t.$$

This is referred to as the value matching condition since it equates values of the yet to be solved value function $V(x,t)$ to the already established termination payoff $\lambda(X,t)$. Determination of X requires another auxiliary condition. This is the "smooth pasting" condition and it requires that $\gamma(x,t)$ and $\lambda(x,t)$ to meet tangentially at X. That is

$$\gamma_x(x,t) = \lambda_x(x,t) \qquad \forall t.$$

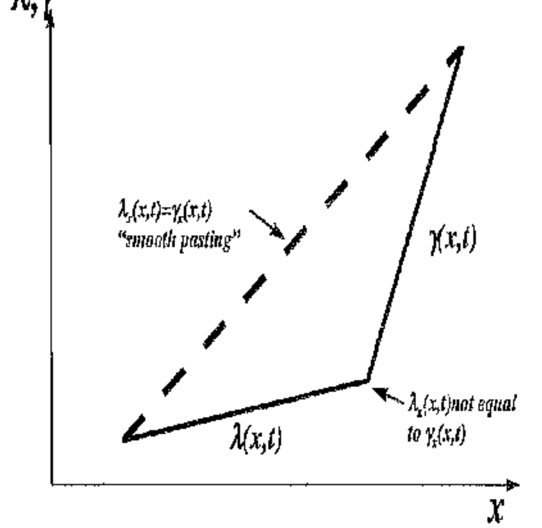

Figure 1.1: "Smooth Pasting" Gradients

Therefore the optimal stopping time or the zone is exactly determined. On the face of it this sounds a perfectly acceptable argument. Indeed, it provides valuable insights into how "smooth pasting" optimally evaluates the stopping times at which control is exercised.

1.2.2 Smooth Pasting and Impulse control

The method of "impulse" control has its genesis in the famous Scarf (1960) inventory control model in which the agent is tasked with optimally managing the stock of a commodity for retail sale in the presence of a random flow of sales and lump sum purchasing costs. If the stock Z_t falls below a critical lower barrier a the quantity $(l - a)$ is ordered, where l is the point of replenishment. The purchasing cost has the effect of reducing the frequency and increasing the size of the orders. Dynamic cash management models increase the dimension of this problem by adding a similar policy at the top end. Stochastic income flows are assumed to add to the inventory Z_t, while planned and unplanned expenditure requirements are assumed to deplete it. If holding costs are continuously incurred at a rate proportional to the money stock, and transaction costs are assumed to be linear, the agent's decision is to choose not only how much cash to withdraw $(l - a)$, but also how much to convert into another asset $(b - u)$, where b is the upper barrier and u is the point to which Z_t is restored (see Constantinides and Richard (1978), and Smith (1989)). The optimal magnitudes of a, b, l and u are determined by applying the so called first-order "smooth pasting" condition which is also widely used in other areas of economics. The following two sections consider its use in the literature on irreversible investment and exchange rate target zones.

Irreversible Investment

Irreversible investment and option pricing models use an "impulse" control framework to demonstrate how the optimal investment decision of a firm could differ from the standard Marshallian investment criterion (see Pindyck (1988), Pindyck (1991), Dixit (1992) and Pindyck and Dixit (1994)). If firms face uncertain demand or costs, new capital can be purchased at a random or fixed price and the cost of investment is linear; it can be shown that firms invest until the marginal revenue product equals its full cost. The latter includes both the cost of purchase and installation and the cost of keeping the option to invest alive. Pindyck (1988) shows that this involves optimally regulating the associated costs and revenues at an upper barrier. Similarly Dixit (1992) demonstrates that disinvestment entails the regulation of operating losses at a lower barrier.

Consider a *competitive market* in which a firm has the capacity to produce one unit of output by incurring a sunk cost I. Assume that variable costs are zero and firms which have incurred a sunk cost will want to produce at its capacity level. If the market in which the firm operates suffers from industry wide demand shocks that follow a continuous stochastic process, the price of a single unit of output can be expressed as

$$P = yD(q).$$

P is the price level, y is the industry wide shock, q is the current level of output and $D(q)$ is the deterministic downward sloping component of the demand curve. Let y follow a geometric Wiener process given by

$$dy = \alpha y ds + \sigma y dz_s,$$

where α is the time gradient and σ is the standard deviation of the Wiener increment. Of course, within an infinitesimally small time interval ds no new entry will take place. Therefore q will be fixed and P will be proportional to y, giving rise to the relationship

$$dP = \alpha P ds + \sigma P dz_s. \tag{1.2}$$

The net present value of a firm's expected profits Π will depend on the current price P and also the expected future price level. If the dynamic evolution of the price level is specified by (1.2), then the expected future price level will only depend on P. Therefore Π will exclusively be a function of P, i.e. $\Pi(P)$.

A firm waiting to enter will observe the price level and use a high price as a trigger to invest. Therefore at some upper barrier $\hat{P}$ a new firm will enter, causing q to increase and P to decrease, making $\hat{P}$ a reflecting upper boundary. If a reflecting boundary did not exist at $\hat{P}$, then the value of the firm will be

$$\Pi(P) = \frac{P}{\delta},$$

where $\delta = r - \alpha$, i.e. the difference between the risk free rate r and the mean rate of growth of the price level. However, the reflecting barrier $\hat{P}$ reduces some of the upside to potential profits and prices. Hence $\Pi(P) < P/\delta$.

If $P < \hat{P}$, then over the infinitesimally small interval ds $\Pi(P)$ can be expanded using a Taylor series through Ito's lemma to yield the second order differential equation

$$\frac{\sigma^2 P^2}{2}\Pi''(P) + (r-\delta)P\Pi'(P) - r\Pi(P) + P = 0. \tag{1.3}$$

This can be solved to obtain

$$\Pi(P) = BP^{\beta} + \frac{P}{\delta}, \tag{1.4}$$

where B is an arbitrary constant and β is the positive root of the characteristic equation of (1.3). The value of B can be determined by eliminating the possibility of sure arbitrage profits. To do this the gradient of $\Pi(.)$ at $\hat{P}$ needs to be zero, i.e.

$$\Pi'(\hat{P}) = \beta B\hat{P}^{\beta-1} + \frac{1}{\delta} = 0.$$

Solving for B and substituting the resulting expression into (1.4) yields

$$\Pi(P) = \frac{P}{\delta} - \frac{1}{\delta\beta}P^{\beta}\hat{P}^{1-\beta}. \tag{1.5}$$

Firms make zero profits in a competitive dynamic equilibrium. At $\hat{P}$ firms will be indifferent between entering the market and staying out. The net present value accrued as a result of entering the market must equal the entry cost I. Using this relationship in (1.5) yields

$$\hat{P} = \frac{\beta}{\beta-1}\delta I.$$

If $f(.)$ is the value of the firm's option to enter, it can be shown that it is a function of P and takes the form

$$f(P) = AP^{\beta}.$$

(see Dixit and Pindyck (1994)). A is an arbitrary constant whose value needs to be determined by the first-order "smooth pasting" condition. If a firm enters at a price level P, it incurs a sunk cost I and receives an income $\Pi(P)$. At the optimal entry trigger P^*, $f(.)$ needs to satisfy the "value matching" condition

$$f(P^*) = \Pi(P^*) - I, \tag{1.6}$$

and the "smooth pasting" condition

$$f'(P^*) = \Pi'(P^*). \tag{1.7}$$

Solving (1.6) and (1.7) simultaneously yields

$$P^* = \frac{\beta}{\beta - 1}\delta I,$$

which is the same as $\hat{P}$. Also $A = 0$, which implies that $f(P) = 0$. Pindyck and Dixit (1994) use this property to argue that a firm contemplating entry into a competitive market faces a zero value of waiting, and conclude that the

> "... value of waiting is negative for most of its price range, and only climbs to zero at the upper end of the range of possible prices",[1]

see Figure 1.2. A model of disinvestment entails a similar argument at the lower boundary.

[1] Dixit and Pindyck (1994) p. 259

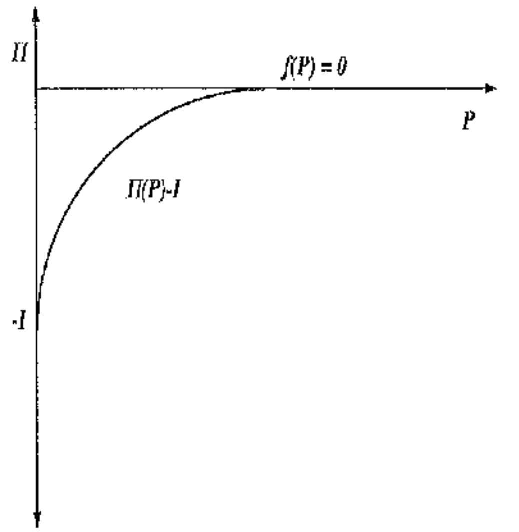

Figure 1.2: The Option of Waiting

Exchange Rate Target Zones

The literature on target zones introduces the method of "impulse" control as a means by which a monetary authority could limit the volatility of its exchange rate. An exchange rate target zone is a hybrid mechanism by which the exchange rate is allowed to freely float, but within a clearly defined region. The monetary authority regulates the exchange rate by selling the currency at an upper barrier and by buying it at a lower barrier, thus keeping the currency within a fixed band. In models of infinitesimal intervention, the exchange rate is restored to a point which is just within the target zone. In the case of discrete intervention an impulse is exercised on the boundary to restore the exchange rate well within the target zone. In Krugman (1991) this gives rise to a S-shaped movement of the exchange rate within the band which is tangential to its boundaries a and b, ruling out expectations of unbounded exchange rate appreciations or depreciations.

Consider a flexible exchange rate which follows the Krugman law of motion, i.e.

$$l(s) = k(s) + \gamma \frac{E[dl(s)]}{ds} \qquad \alpha > 0. \tag{1.8}$$

$l(s)$ is the natural log of the exchange rate, γ is the Cagan interest rate semi-elasticity and the expectation operator $E[.]$ is conditioned on the current information set. Of course, only information on the independent variable k is relevant. k is assumed to reflect the rates of change in the value of foreign currencies, the domestic money supply, real income levels and expectations of money demand shocks. Thus k can be controlled by the monetary authority, specifically to keep the exchange rate within a desired band $l_l < l < l_u$. Let k, absent control, follow a Wiener process of the type

$$dk = \psi ds + \sigma dz_s.$$

The coefficients ψ and σ are assumed to be constants. Using this specification of k, Krugman (1991) explicitly develops a functional form of the exchange rate solution, $l = m(k)$. Flood and Garber (1991) argue that $m(k)$ is a solution to $l(s)$ and expand it using Ito's lemma to obtain

$$\frac{dl}{ds} = \psi m'(k) + \frac{\sigma^2}{2} m''(k). \tag{1.9}$$

This is solved to yield

$$l = m(k) = k + \gamma\psi + Ae^{\lambda_1 k} + Be^{\lambda_2 k}, \tag{1.10}$$

where λ_1 and λ_2 are the roots of the characteristic equation of (1.9). The values of A and B are determined by the "smooth pasting" condition which requires that the exchange rate be tangential to l_u and l_l, i.e.

$$m'(k_u^*) = 0 \qquad \text{and} \qquad m'(k_l^*) = 0$$

see Figure 1.2. Discrete intervention requires the additional conditions

$$m(k_u) = m(Q) \qquad\qquad m(k_l) = q.$$

1.2.3 The Critique

The first-order condition which determines the optimal barriers and thresholds in "impulse" control models is the "smooth pasting" condition. In investment and option pricing models it 'equates the value of waiting with the value of the investment trigger b or the disinvestment trigger a'. The intuition behind this is that the first derivative of the option price must be the same value before and after an option is exercised. In the target zone literature it provides the justification through arbitrage for the tangential relationship between the exchange rate and the upper and lower barriers a

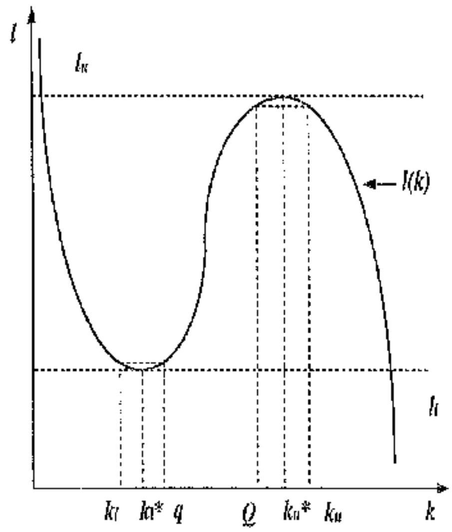

Figure 1.3: A Model of Exchange Rate Target Zones

and b. In inventory theoretic money demand models it is used to pin down the boundaries and thresholds a, l, u and b, equating the marginal cost of being on a boundary with the marginal transaction cost.

These applications of the "smooth pasting" condition would suggest that the controls a, l, u and b are chosen to either optimise the value functions or satisfy some economic argument. But in reality this is not the case. In problems where the "smooth pasting" condition is used as a first-order optimisation condition, it is derived by equating the derivative of the value function, with respect to the initial state and the gradient of the cost of adjustment on the boundary (see Dixit (1994, pp129-130), Dixit (1991b, 667-668), Constantinides and Richard (1977), and Smith (1989)). Although the value function is a function of both the initial state and the set of admissible controls, the principle of Optimal Control requires the value function to be optimised *only* with respect to the set of admissible controls (see Bensoussan and Lions (1975a), (1975b), and Richard (1977)). *It is only by choosing the controls a, l, u and b to optimise the value function that the marginal payoff which flows from controlling a system is set equal to the associated marginal cost.* The initial state is merely an inheritance from a previous unknown history and is *not* a control variable. Its functional relationship with the value function is *fundamentally different* to that of a, l, u, and b. Of course, the initial state may influence the choice of boundaries particularly under high discount rates.

Furthermore, the numerical values of these 'optimal' boundaries and thresholds are implicit in their construction and not variable. For a certain choice of parameters the gradient conditions on the initial state numerically fix the values of a, l, u and b. Another facet of this simplification is manifest through the apparent inability of the current model of impulse control to quantify the extent to which a prescribed strategy deviates from that which is optimal. For instance behaviour of an agent who initially lies outside the 'optimal' boundaries (i.e. outside the interval (a, b)) cannot be compared with one who initially lies within these bounds. Constantinides and Richard (1978), and Smith (1989) provide solutions for the value function outside the "optimal" boundaries, but these do not satisfy the Hamilton-Jacobi-Bellman equation (HJB) they obtain. Here a functional form for the value function is derived from which the costs associated with a sub-optimal choice of a, l, u, and b can be evaluated. Fleming and Rishel (1976, Appendix E), Bensoussan and Lions (1975a), (1975b) and Richard (1977) show that the HJB has convex solutions and therefore has only one control vector, whereas the "smooth pasting" strategy does not seem to suggest any.

In the literature where "smooth pasting" is used to make a model satisfy certain economic arguments, such as the target zone models on exchange rates or the zero profit condition in the model of irreversible investment described earlier, reflecting boundary conditions need to be imposed on the forward state of the Wiener process; not on the initial state as it is currently done. This would ensure that when the exchange rate hit an upper or a lower barrier, it would be instantly restored into the interior of the target zone by either an infinitesimally small amount or a discrete quantity. However, imposing boundary conditions on the forward state is a non-trivial task and would require the dynamic evolution of the exchange rate to be computed through the transition density function. Since the Ito stochastic differential equation has not been adapted to capture boundary conditions on the forward state, it is not suitable for use in problems of this type.

In dynamic programming involving infinite horizons the value function is considered a function of the initial state, despite the fact that it is the forward state which experiences the impulse control. This made possible due to the Markov property which enables the evolution of a stochastic process to be described in terms of its initial state and time. For models in which the state is given by a Wiener process this can be done through either Ito's lemma or the backward Chapman-Kolmogorov equation (see Gihman and Skorohod (1972)). *The initial state plays no other explicit role in an infinite horizon impulse control problem* (see Fleming and Rishel (1975) chapter VI, Bensoussan and Lions (1975a), (1975b) and Richard (1977)).

These arguments, though robust, do not explain the algebraically equivalent results obtained using both the "smooth pasting" and optimal control strategies in Chapter 2. Indeed, this may point to the existence of some undiscovered properties of the HJB equation and the value function. On the other hand, it may be a feature restricted to the Dixit menu cost model due to its unique nature. However, without further evidence which demonstrates that "smooth pasting" and the stochastic optimal control strategy yield differing results, it still remains useful as an approximation of the necessary first-order condition for optima.

1.3 Impulse Control

Let the state follow a Wiener process with the corresponding stochastic differential equation

$$dx_s = \mu ds + \sigma dz_s \qquad s \in [0, \infty) \qquad x_0 = x,$$

over the continuation region (a, b), and, let $\tau_0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \cdots \leqslant \tau_i \leqslant \cdots$ be the series of stopping times; i.e. the series of points in time at which the process being controlled exits the continuation region and a "jump" control is exercised to restore the process to an interior point.

The method of "impulse" control requires the existence of a feedback control law $\mathbf{u}$ which optimises a performance criterion subject to some initial datum and boundary conditions and is described at time s by

$$\mathbf{u}_s = \begin{cases} b - u & \text{if } x_s = b, \\ a - l & \text{if } x_s = a, \end{cases}$$

where u and l are respectively the interior points to which the process is restored to when the upper boundary b or the lower boundary a is encroached. Clearly $a \leqq l$, $u \leqq b$.

Let the instantaneous holding cost be given by the real function $M(x_s)$, $s \in [0, \infty)$, and assume a constant discount rate of ρ. The object, therefore, will be to arrive at a policy

$$\mathbf{p} = \{\tau_1, \mathbf{u}_1; \tau_2, \mathbf{u}_2; \cdots; \tau_i, \mathbf{u}_i; \cdots\},$$

of a series of stopping times and controls which minimise the value function

$$\underset{inf\,\mathbf{u}}{V(x,\mathbf{u})} = E\left[\int_0^\infty e^{-\rho s} M(x_s)\,ds + \sum_{i=0}^{\infty} e^{-\rho\tau_i} R(\mathbf{u})\Big| x, 0\right], \tag{1.11}$$

where $R(\mathbf{u})$ is the cost of adjustment defined by

$$R(\mathbf{u}) = \begin{cases} B(b,u) & \text{if } x_s = b, \\ D(a,l) & \text{if } x_s = a. \end{cases}$$

It is now demonstrated that this value function satisfies the HJB equation.

By using Ito's Lemma and expanding (1.11) in a Taylor series, it follows that

$$AV(x,\mathbf{u}) - \rho V(x,\mathbf{u}) + M(x) = 0, \tag{1.12}$$

where

$$A = \mu\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}. \tag{1.13}$$

Equation (1.12) is referred to as the HJB equation. In particular, the formulation of (1.12) depends on the boundary conditions at a and b. Solving this equation subject to the correct boundary conditions yields the dynamic programming equation, which is also the performance criterion $V(x,\mathbf{u})$. Its infimum with respect to the control law $\mathbf{u}$, is $V(x,\mathbf{u}^*)$. Equally $\mathbf{u}^*$ can be obtained by minimising (1.12). The existence of an optimal feedback control law for this equation has been demonstrated by Bensoussan and Lions (1975a) and Richard (1977).

1.4 The Martingale Formulation of the Bellman Value Function

The martingale formulation of the value function divides into two components. The first measures the net present value of holding costs accrued until the first stopping decision while the second computes the net present value of all costs, i.e. holding costs and costs of adjustment, which follow from this decision. It can be shown easily that this construction of the value function also satisfies Ito's Lemma.

Consider the Wiener process introduced in the last section. When x_s is on the boundary, that is $x_s = a$, or $x_s = b$, the process enters a stopping zone and is absorbed. For convenience let the forward state x_s be denoted by y. Hence the transition density function for this process, $f(y,s|x,0)$, must satisfy the boundary conditions $f(b,s|x,0) = 0$, and $f(a,s|x,0) = 0$, along with the initial condition $f(y,0|x,0) = \delta(y-x)$. It is common knowledge that $f(y,s|x,0)$ satisfies the forward Chapman-Kolmogorov equation (Fokker-Planck equation)

$$f_s(y,s|x,0) = -\mu f_y(y,s|x,0) + \frac{\sigma^2}{2} f_{yy}(y,s|x,0), \tag{1.14}$$

Define a distribution function $F(y,s|x,0)$ with the property that

$$F_y(y,0|x,0) = f(y,s|x,0),$$

$$F(y, s|x, 0) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

Thus, $F(y, s|x, 0)$ must satisfy the backward Chapman-Kolmogorov equation

$$F_s(y, s|x, 0) = \mu F_x(y, s|x, 0) + \frac{\sigma^2}{2} F_{xx}(y, s|x, 0). \tag{1.15}$$

It is also clear that probability mass through the upper and lower boundaries will be respectively

$$F_s(b, s|x, 0), \tag{1.16}$$

and

$$F_s(a, s|x, 0). \tag{1.17}$$

Expressions (1.16) and (1.17) define the probability that the process will enter the zone through the upper and lower boundaries at time s, respectively. Thus, the first costs of adjustment at the upper and lower boundaries, $B(b, u)$ and $D(a, l)$, will be incurred at rates given by (1.16) and (1.17) respectively. Hence the value function, constructed in terms of an optimal stopping problem is

$$\begin{aligned} V(x, \mathbf{u}) &= \int_0^\infty e^{-\rho s} \left[\int_a^b M(y) f(y, s|x, 0)\, dy \right] ds \\ &\quad + \int_0^\infty e^{-\rho s} [B(b, u) + V(u, \mathbf{u})] F_s(b, s|x, 0)\, ds \qquad (1.18) \\ &\quad + \int_0^\infty e^{-\rho s} [D(a, l) + V(l, \mathbf{u})] F_s(a, s|x, 0)\, ds. \end{aligned}$$

The derivation of this is straightforward. The first integral on the right-hand side is the net present value of the holding costs accrued until the first stopping decision. The second and third integrals evaluate the net present value of the sum of the first cost of adjustment, $B(b,u)$ and $D(a,l)$, and all holding costs and costs of adjustment which accrue from this decision, i.e. $V(u,.)$ and $V(l,.)$.

The application of an "impulse" control on the upper boundary at time s, instantly changes the state from b to u. Given that $V(.,.)$ is a functional of a Markov process which contains the net present value of all holding costs and costs of adjustment, this defines a new Markov function with an initial state u over the time horizon $[s,\infty)$ which contains all holding costs and costs of adjustment that accrue from time s onwards, i.e. $V(u,.)$. The multiplier $e^{-\rho s}$ discounts to net present value. The same holds for $V(l,.)$. Since $V(u,.)$ and $V(l,.)$ contain both holding costs and costs of adjustment which accrue as a result of the first stopping decision. It is therefore clear that the above sum yields the expected net present value of a policy of "impulse" control at the upper and lower boundaries.

Integrating (1.18) with respect to time it is clear that

$$\begin{aligned} V(x,\mathbf{u}) &= \int_0^\infty e^{-\rho s}\left[\int_a^b M(y) f(y,s|x,0)\,dy\right] ds \\ &\quad +[B(b,u)+V(u,\mathbf{u})]\left[1-\rho\int_0^\infty e^{-\rho s}F(b,s|x,0)\,ds\right] \\ &\quad +[D(a,l)+V(l,\mathbf{u})]\left[1-\rho\int_0^\infty e^{-\rho s}F(a,s|x,0)\,ds\right]. \quad (1.19)\end{aligned}$$

It is now demonstrated that the constructed $V(x, \mathbf{u})$ satisfies the HJB equation.

Applying the differential operator in (1.13) to (1.19) and use the Chapman-Kolmogorov equations (1.14) and (1.15), it is clear that

$$\begin{aligned} AV(x,\mathbf{u}) = \rho \int_a^b M(y) \left[\int_0^\infty e^{-\rho s} f(y,s|x,0)\, ds \right] dy - M(x) \\ -\rho[B(b,u) - V(u,\mathbf{u})] \left[1 - \rho \int_0^\infty e^{\rho s} F(b,s|x,0)\, ds \right] \\ -\rho[D(a,l) - V(l,\mathbf{u})] \left[1 - \rho \int_0^\infty e^{\rho s} F(a,s|x,0)\, ds \right]. \end{aligned} \quad (1.20)$$

See Appendix A for the derivation of $A(x)V(x, \mathbf{u})$. Elementary calculation now reveals that

$$AV(x,\mathbf{u}) - \rho V(x,\mathbf{u}) + M(x) = 0, \quad (1.21)$$

which is the HJB equation. Minimising this with respect to $\mathbf{u}$ will yield the optimal "impulse" control policy. It is also obvious that we could use Ito's lemma and expand (1.19) in a Taylor series to also obtain (1.21). Therefore, the optimal stopping framework used to set up $V(x, \mathbf{u})$ is also self consistent.

1.5 Computation of the Value Function

The probability density function of the absorption process is first calculated. This function is then used in the constructive definition of the value function $V(x, \mathbf{u})$ in (1.18) for general $B(b, u)$, $D(a, l)$ and $M(y)$.

To ease the computation of the value function in (1.18), it is convenient to introduce the non-dimensional variables

$$\xi = \frac{y-a}{b-a}, \qquad g = \frac{x-a}{b-a}, \qquad \alpha = \frac{\mu(b-a)}{\sigma^2},$$
$$\beta = \frac{2\rho(b-a)^2}{\sigma^2}, \qquad \text{and} \qquad t = \frac{s\sigma^2}{(b-a)^2}. \tag{1.22}$$

With this change of variables the partial differential equation satisfied by $f(y, s|x, 0)$ becomes

$$\frac{\partial f(\xi, t|g, 0)}{\partial t} = -\alpha\frac{\partial f(\xi, t|g, 0)}{\partial \xi} + \frac{1}{2}\frac{\partial^2 f(\xi, s|g, 0)}{\partial \xi^2}, \tag{1.23}$$

where $f(\xi, t|g, 0)$ satisfies the boundary conditions $f(1, t|g, 0) = 0$, $f(0, t|g, 0) = 0$ and the initial condition $f(\xi, 0|g, 0) = \delta(\xi - g)$. The distribution function $F(y, s|x, 0)$ in (1.15) now satisfies

$$\frac{\partial F(\xi, t|g, 0)}{\partial t} = \alpha\frac{\partial F(\xi, t|g, 0)}{\partial g} + \frac{1}{2}\frac{\partial^2 F(\xi, s|g, 0)}{\partial g^2}. \tag{1.24}$$

The value function now becomes

$$\begin{aligned} V(g, \mathbf{u}) \;=\; & \int_0^\infty e^{\frac{-\beta t}{2}}\left[\int_0^1 M(\xi) f(\xi, t|g, 0)(b-a)\, d\xi\right]\frac{(b-a)^2}{\sigma^2}\, dt \\ & + \int_0^\infty e^{\frac{-\beta t}{2}}[B(1, u) + V(u, \mathbf{u})]F_t(1, t|g, 0)\, dt \\ & + \int_0^\infty e^{\frac{-\beta t}{2}}[B(0, l) + V(l, \mathbf{u})]F_t(0, t|g, 0)\, dt. \end{aligned}$$

Hence the expressions for the non-dimensional $V(x, \mathbf{u})$, $B(b, u)$ and $D(a, l)$ become

$$\hat{V} = \frac{\sigma^2 V}{(b-a)^3}, \qquad \hat{B} = \frac{\sigma^2 B}{(b-a)^3}, \qquad \text{and} \qquad \hat{D} = \frac{\sigma^2 D}{(b-a)^3}.$$

1.5.1 Derivation of the transitional PDF

Using the method of separation of variables, it can be verified that

$$\sin(n\pi\xi)e^{\alpha\xi-\frac{1}{2}(\alpha^2+n^2\pi^2)t},$$

is a solution to (1.23) satisfying its boundary conditions for all integers n. Using Fourier methods it can be further proved that

$$f(\xi,t|g,0) = 2\sum_{n=1}^{\infty} e^{\alpha(\xi-g)-\frac{1}{2}(\alpha^2+n^2\pi^2)t}\sin(n\pi\xi)\sin(n\pi g)$$

is the complete solution of (1.23) satisfying the boundary and initial conditions. The details of this calculation appear in Appendix B.

The probability flux, or the probability mass exiting through the upper and lower boundaries is respectively

$$\left[\frac{1}{2}f_\xi - \alpha f\right]\Big|_1, \qquad \text{and} \qquad \left[\frac{1}{2}f_\xi + \alpha f\right]\Big|_0.$$

This effectively defines the probability distribution function of the process entering the stopping zone. Here $f(1,s|x,0) = f(0,s|x,0) = 0$, and, therefore the flux on upper and lower boundaries b and a are $(1/2)f_\xi\Big|_1$, and, $(1/2)f_\xi\Big|_0$ respectively, i.e.

$$\begin{aligned}
\frac{1}{2}f_\xi\Big|_1 &= \pi\sum_{n=1}^{\infty} n\sin(n\pi(1-g))e^{\alpha(1-g)-\frac{1}{2}(\alpha^2+n^2\pi^2)t} \\
\frac{1}{2}f_\xi\Big|_0 &= \pi\sum_{n=1}^{\infty} n\sin(n\pi g)e^{-\alpha g-\frac{1}{2}(\alpha^2+n^2\pi^2)t}
\end{aligned} \tag{1.25}$$

It is immediately obvious that

$$F_t(1,t|g,0) = \frac{1}{2}f_\xi\Big|_1 \qquad \text{and} \qquad F_t(0,t|g,0) = -\frac{1}{2}f_\xi\Big|_0. \tag{1.26}$$

It can further be shown that

$$\begin{aligned}
\int_0^\infty e^{\frac{-\beta t}{2}} F_t(1,t|g,0)\,dt &= e^{\alpha(1-g)}\frac{\sinh \chi g}{\sinh \chi}, \\
\int_0^\infty e^{\frac{-\beta t}{2}} F_t(0,t|g,0)\,dt &= e^{-\alpha g}\frac{\sinh \chi(1-g)}{\sinh \chi}.
\end{aligned} \tag{1.27}$$

where $\chi = \sqrt{\beta + \alpha^2}$. See Appendix B for the evaluation of the infinite integrals in (1.27).

1.5.2 Calculation of the value function

The expressions in (1.27) can be used to compute the net present value of all costs which accrue from the first stopping decision at the upper and lower boundary. It remains to evaluate, $\Psi(g)$, the net present value of the holding costs accrued until the first stopping decision which is defined by

$$\begin{aligned}
\Psi(g) &= \int_0^\infty e^{\frac{-\beta t}{2}} \left[\int_0^1 M(\xi) f(\xi,t|g,0)\,d\xi\right] dt \\
&= \int_0^1 M(\xi)\left[\int_0^\infty e^{\frac{-\beta t}{2}} f(\xi,t|g,0)\,dt\right] d\xi.
\end{aligned} \tag{1.28}$$

Using integration by parts it can be shown that

$$\int_0^\infty e^{\frac{-\beta t}{2}} f(\xi,t|g,0)\,dt = 2e^{\alpha(\xi-g)}\frac{\cosh \chi\omega - \cosh \chi\gamma}{\chi \sinh \chi},$$

where $\omega = 1 - |g-\xi|$ and $\gamma = 1 - g - \xi$. See Appendix B for intermediate steps.

Historically linear and quadratic holding cost functions have been used for $M(y)$ (see Dixit (1991a), Constantinides and Richard (1978) and Smith

(1989)). However $\Psi(g)$ can be evaluated for a general holding cost function. Suppose that $M(\xi)$ has a half range cosine Fourier series

$$M(\xi) \;=\; \frac{b_0}{2} + \sum_{j=i}^{\infty} b_j \cos(j\pi\xi),$$

then standard Fourier analysis yields

$$b_j = 2\int_0^1 M(\xi)\cos(j\pi\xi)\, d\xi, \qquad j = 0, 1, 2, \cdots.$$

Using this representation of the holding cost function, the net present value of holding costs accrued until the first stopping decision is

$$\Psi(g) = \frac{2}{\chi \sinh \chi} e^{-\alpha g}\left[\sum_{j=0}^{\infty} b_j \int_0^1 \cos(j\pi\xi) e^{\alpha\xi}[\cosh \chi\omega - \cosh \chi\gamma]\, d\xi\right]. \quad (1.29)$$

This integral can be evaluated using integration by parts resulting in the form

$$\begin{aligned}\Psi(g) \;=\; & 2\Bigg[e^{\alpha(1-g)}\frac{\sinh \chi g}{\sinh \chi}\sum_{j=0}^{\infty} b_j\phi_j(1) \\ & + e^{\alpha g}\frac{\sinh \chi(1-g)}{\sinh \chi}\sum_{j=0}^{\infty} b_j\phi_j(0) - \sum_{j=0}^{\infty} b_j\phi_j(g)\Bigg], \qquad (1.30)\end{aligned}$$

where

$$\phi_j(z) = \frac{(\alpha^2 - \chi^2 - j^2\pi^2)\cos(j\pi z)}{[j^2\pi^2 + (\alpha - \chi)^2][j^2\pi^2 + (\alpha + \chi)^2]}.$$

In view of (1.29) and (1.30), the value function finally simplifies to

$$\begin{aligned}V(g, \mathbf{u}) \;=\; & [B(1, u) + V(u, \mathbf{u})]e^{\alpha(1-g)}\frac{\sinh \chi g}{\sinh \chi} \\ & + [D(0, l) + V(l, \mathbf{u})]e^{-\alpha g}\frac{\sinh \chi(1-g)}{\sinh \chi}\end{aligned}$$

$$+2\left[e^{\alpha(1-g)}\frac{\sinh \chi g}{\sinh \chi}\sum_{j=0}^{\infty} b_j\phi_j(1)\right.$$
$$\left.+e^{\alpha g}\frac{\sinh \chi(1-g)}{\sinh \chi}\sum_{j=0}^{\infty} b_j\phi_j(0) - \sum_{j=0}^{\infty} b_j\phi_j(g)\right].$$

The derivation of the boundary conditions is straight forward. When $x = b$, $g = 1$. Hence

$$V(b,\mathbf{u}) = B(b,u) + V(u,\mathbf{u}). \tag{1.31}$$

On the other hand when $x = a$, then $g = 0$ and

$$V(a,\mathbf{u}) = D(a,l) + V(l,\mathbf{u}). \tag{1.32}$$

It is now obvious that (1.31) and (1.32) define the behaviour of the value function on the boundaries and, therefore, are the boundary conditions to an "impulse" control problem. Solving (1.12) subject to (1.31) and (1.32) will yield $V(x;\mathbf{u})$. The optimal values of a, b, l, and u are obtained through the first order conditions

$$\frac{\partial V(x;\mathbf{u})}{\partial a} = 0, \qquad \frac{\partial V(x;\mathbf{u})}{\partial b} = 0, \qquad \frac{\partial V(x;\mathbf{u})}{\partial l} = 0, \qquad \text{and} \qquad \frac{\partial V(x;\mathbf{u})}{\partial u} = 0. \tag{1.33}$$

1.6 "Smooth Pasting vs. Stochastic Optimal Control

Undoubtedly the most significant conclusion from this analysis is the conspicuous absence of anything resembling the "smooth pasting" condition which suggests the 'optimality' conditions

$$\begin{aligned}\frac{dV(a)}{da} &= \frac{dD(a,l)}{da}, & \frac{dV(b)}{db} &= \frac{dB(b,u)}{db}, \\ \frac{dV(l)}{dl} &= \frac{dD(a,l)}{dl} \quad \text{and} & \frac{dV(u)}{du} &= \frac{dB(b,u)}{du}.\end{aligned} \tag{1.34}$$

It is not exactly clear what these derivative conditions imply. It is stated that they are obtained by differentiating the "value matching" conditions in (1.31) and (1.32), and claim to show that the marginal cost of being on a boundary must equal the marginal cost of adjustment. However, this is clearly not true. Let us consider an agent who initially is on the upper boundary, i.e $x = b$. From (1.31) it can be seen that his marginal costs are given by

$$\begin{aligned}\left.\frac{\partial V(x,\mathbf{u})}{\partial a}\right|_{x=b} &= \left.\frac{\partial V(x,\mathbf{u})}{\partial a}\right|_{x=u} \\ \left.\frac{\partial V(x,\mathbf{u})}{\partial b}\right|_{x=b} &= \frac{\partial B(b,u)}{\partial b} + \left.\frac{\partial V(x,\mathbf{u})}{\partial b}\right|_{x=u} \\ \left.\frac{\partial V(x,\mathbf{u})}{\partial l}\right|_{x=b} &= \left.\frac{\partial V(x,\mathbf{u})}{\partial l}\right|_{x=u} \\ \left.\frac{\partial V(x,\mathbf{u})}{\partial u}\right|_{x=b} &= \frac{\partial B(b,u)}{\partial u} + \left.\frac{\partial V(x,\mathbf{u})}{\partial u}\right|_{x=u}.\end{aligned} \tag{1.35}$$

The left hand side gives the marginal costs of initially being on the boundary b with respect to the choice of controls a, b, l and u. The right hand side gives the marginal costs of transacting down to u with respect to these controls. The choice of a barrier b affects the choice of a, l and u. This is not the case with the "smooth pasting" condition. The first-order conditions described in (1.34) evaluate a, l, u, and b *independently* of each other by equating the gradient of the value function with respect to the initial state to the gradient of transaction costs on boundaries and thresholds. Intuitively this cannot be correct. The choice of one boundary must clearly affect the probability of hitting the other boundary or being restored to a threshold. Hence both boundaries and thresholds must be selected simultaneously to optimise the value function, not independently of each other. Costs are at a minimum only when (1.33) is satisfied concurrently by a, l, u and b.

1.6.1 Optimal Stopping and the Choice of Controls

Since a, l, u and b determine continuation region, it must follow that they also select the stopping zone outside (a, b). In stochastic calculus the expected time at taken for a Weiner process starting at an initial state x to exit (a, b) into the stopping zone can be easily evaluated using the backward Chapman-Kolmogorov equation of the distribution function $F(.)$. Therefore choosing a, l, u and b optimally ensures that the stopping times τ_1, τ_2, ... are also chosen optimally. It is not obvious that choosing stopping times using (1.34) ensures optimality. Indeed, in complex problems it is likely that the process will be stopped prematurely because any costs which flow from moving down to u are ignored. However, in the Dixit menu cost model Chapter 2 clearly demonstrates that both (1.34) and (1.33) yield algebraically equivalent answers.

This analysis shows that the marginal costs of being on a boundary must equal the *marginal costs of transacting* which includes the marginal costs of adjustment on that boundary, not exclusively the marginal costs of adjustment as the "smooth pasting" conditions suggests.

1.7 Concluding Remarks

In this chapter the boundary conditions for an optimal policy of "impulse" control have been derived by constructing a system from first principles using stochastic calculus. It shows that the value of stopping at a state and exercising an "impulse" control must equal the net present value of holding costs accrued up to that state. This also sounds intuitively correct, if the total value of exercising a stopping decision exceeded the net present value of holding costs accrued until this decision was made, it would clearly be sub-optimal to stop. Conversely if the net present value of holding costs exceeded the net present value of the stopping decision, it would imply that the stopping decision should have been taken earlier. By approaching the jump control problem from a different perspective, the natural mathematical boundary conditions for the HJB equation has been motivated in a non-heuristic way. The solution technique exemplified here enables a new class of model to be constructed in economics and finance. These should provide revealing and accurate insights into optimising behaviour within an environment of ongoing uncertainty.

Chapter 2

Control Regimes, Transaction Costs and Business Cycles

2.1 Introduction

Menu cost models have evolved significantly since they were first developed in the mid eighties (see Akerlof and Yellen (1985), Mankiw (1985) and Blanchard and Kiyotaki (1987)). Earlier models analysed the implications of nominal rigidities and sub-optimal welfare outcomes caused by demand shocks within a static environment. Later models expanded on these by ascribing to the firm the net present value of the losses accrued from these shocks within an environment of ongoing uncertainty. The contrast in results between the two approaches is significant. When firms are forced to minimise costs over a much longer time horizon with discounting, the zone of inertia and the strategies adopted change dramatically. In the Dixit model the range of inaction is two orders larger than the Akerlof-Yellen model (see Dixit (1991a)).

The Akerlof-Yellen, Blanchard-Kiyotaki and Dixit models study the behaviour of firms functioning as private optimisers in a monopolistically competitive market. It is obvious from the assumptions which underpin Chamberlinian monopolistic competition that each of these firms will practice horizontal price differentiation. This will create a gap between the price set by each firm and the market price. With a downward sloping demand curve, it can be shown that this gap will give rise to a holding cost, measured in lost profits. These models prove that, if a cost is attached to closing this gap, there exists a zone of inertia in which it will be beneficial for each firm to sustain costs rather than eliminate them through price adjustment. Implicit here is that each firm will face its own unique cost function.

One important feature these models rely on is the pecuniary externality which can be observed in a general equilibrium models involving monopolistic competition. If a firm reduces its price level slightly it increases the demand for its goods. It also increases real money balances, increasing demand for other firms output as well. In monopolistic competition, since output is initially not equal to the social optimum, the increase in real balances has a positive effect on welfare. Of course, the opposite situation could hold as well. This implies that money, at least in the short run, is non-neutral and therefore would require the regular intervention of the monetary authority. If the monetary authority has access to new information on exogenous shocks, after firms have set their prices, systematic feedback rules could stabilise output. If the monetary authority fails to react to these nominal changes, employment levels and output would experience the negative impact forecast by these models.

In labour markets inertial responses in price setting behaviour induces a change in the real wage, the direction of which will depend on both the magnitude and timing of the change in price. The magnitude of the wage change will depend on the elasticity of the labour supply curve. Large fluctuations in employment will result from small menu costs, only with an elastic labour supply curve. In a model such as Blanchard and Kiyotaki (1987) in which price setters do not want to change relative prices to each other and the cost of not adjusting wages is not large, it is not clear why wage setters; whether they be unions, firms or even workers, would settle for large changes in employment for relatively small changes in output.

The appeal of menu cost models is that they predict welfare losses, resulting from inertial responses, which are much larger that the actual cost of adjustment. In the Akerlof-Yellen model not to react instantly to any price change results in a second order loss to the firm. However, the welfare losses are of first order magnitude. In the Dixit model the waiting time between changes in aggregate demand and firms adjusting prices is two orders larger than Akerlof-Yellen. The infinite planning horizon punishes firms more than the preceding static models. This has the effect of reducing the frequency and increasing the size of adjustments. As a result nominal rigidity becomes more entrenched, causing much larger output and welfare distortions. By incorporating uncertainty and a time horizon into the existing menu cost literature Dixit (1991a) reveals a more accurate picture of the effects of nominal friction.

The instantaneous holding cost in the Dixit model is an increasing function of the difference between the price set by firms and the market price. Hence, it is to be expected that the initial cost and the costs accrued in the first few time periods will contribute more towards the value of the cost function than those in later periods, especially if the discount rate is high. However, this is not the case. The results illustrated in Dixit (1991a) show that the zone of inertia is only determined by the exogenous parameters driving the cost function and is independent of the initial cost. Of course, this is implicit in the specification of the heuristically motivated first order "smooth pasting" condition used to evaluate the optimal zone of inaction. The initial cost is excluded from the solution technique. Here it is shown under an optimal stochastic control framework that, even when the initial cost does feature in the evaluation of the optimal zone, it plays no role in determining the optimal zone of inaction. However, this probably has more to do with the type of costs faced by the firm.

The "smooth pasting" condition also excludes the set of admissible controls and the link which exists between them (see Chapter 1) in the computation of the optimal zone of inertia. For a given set of parameters, the gradient conditions with respect to the initial state fix the values of the 'optimal' bounds. This is evident through the Dixit model's apparent inability to quantify the extent to which a prescribed strategy deviates from that which is optimal. Thus the Dixit model fails to capture the price adjustment behaviour of any firm whose initial price gap may lie outside the narrow optimal zone of inertia. Here, the costs faced by such firms along with their prescribed price adjustment strategy is derived.

The assumption that the zone of inertia is symmetrically disposed about the market price is an important feature of models in the current literature. In static models this is self evident because both holding costs and the costs of adjustment are assumed to be symmetric about the market price. However, there exists no *a priori* reason for this to be the case in net present value models within an environment of ongoing uncertainty. Nevertheless, these models make this assumption in deriving the optimal range of inaction. Here it is proved that, within the Dixit model, the optimal zone of inertia will always be symmetric, if firms adjust prices to completely eliminate the price gap between their price and the market price, i.e. follow a zero threshold policy. Analytical expressions which link the behaviour of the boundaries of the optimal zone of inaction to all the exogenous parameters driving the cost function are also derived for this policy. It is difficult to examine the impact of the intertemporal discount rate of each firm on the optimal range of inaction in Dixit (1991a).

However, complete price adjustment, or the zero threshold policy, is only a limited form of the general optimal control policy for such a problem. If firms are allowed the flexibility to choose the magnitude of their price adjustment in an optimal way, it is not obvious that they would opt for a zero threshold policy. The results obtained here confirm that firms always opt for a zero threshold policy even if they are offered this flexibility. Effectively, this chapter confirms the results obtained in Dixit (1991a) through an optimal stochastic control framework and provides a solution technique that allows a more general type of problem to be solved, *quantifying* the extent to which a specified price setting policy deviates from that which is optimal in a *non-heuristic* way.

In Section 2, the Dixit menu cost function is computed using the formulation in Chapter 1 for models of "impulse" control rather than the "smooth pasting" condition. In Section 3, it is proved that the zone of inertia is indeed symmetric about the market price, and analytic expressions are provided linking the exogenous factors driving the net present value cost function and the zone of inertia. In Section 4, the optimal price adjustment policy is derived and results are illustrated in Section 5.

2.2 The zero threshold policy

Let the state variable x be the natural logarithm of the difference between the firm price and the market price and follow a driftless Wiener process

$$dx_t = \sigma dz_t, \qquad x_t \in (a, b),$$

where dz is the Wiener increment, where $a \leqq 0$ and $b \geqq 0$. Let the instantaneous holding cost function be given by kx^2. Furthermore, let the transaction cost g be incurred at each instant the process exits continuation region (a, b) and is restored to zero. Then the net present value of holding and transactions costs will be given by the value function

$$V(x) = \min E\left[\int_0^\infty e^{-\rho t} kx^2\, dt + \sum_{j=1}^{\infty} e^{-\rho\tau_j} g \middle| x(0) = x\right].$$

where τ_j denotes the discrete times at which a "jump" control is exercised to restore the process to zero.

The standard technique of stochastic calculus reveals that $V(x)$ satisfies

$$\frac{\sigma^2}{2} V''(x) - \rho V(x) + kx^2 = 0.$$

with solution

$$V(x) = A\cosh(\alpha x) + B\sinh(\alpha x) + \frac{kx^2}{\rho} + \frac{k\sigma^2}{\rho^2}, \qquad \alpha = \sqrt{\frac{2\rho}{\sigma^2}}. \qquad (2.1)$$

where A and B are arbitrary constants.

For a policy of complete price reconciliation, in which restoration to zero is forced upon the process exiting the continuation region, the value function $V(x)$ must satisfy

$$V(b) - V(0) = g, \qquad V(a) - V(0) = g. \qquad (2.2)$$

on the top and bottom boundaries respectively. Note that the "smooth pasting" condition is not necessary for the evaluation of the the constants A and B (see chapter 2). To ease the treatment of (2.1) let us define the non-dimensional variable w and parameters y, z, and γ by

$$w = \frac{\alpha x}{2} \qquad z = \frac{\alpha a}{2} \qquad y = \frac{\alpha b}{2} \qquad \gamma = \frac{g\rho^2}{2k\sigma^2}. \qquad (2.3)$$

In view of (2.2) and (2.3), equation (2.1) has solution

$$V(w; y, z) = \frac{2w^2}{\gamma} + \frac{1}{\gamma} + \frac{f(y)\cosh(2w - z) - f(z)\cosh(2w - y)}{\sinh(y - z)}, \qquad (2.4)$$

where

$$f(z) = \frac{\gamma - z^2}{\gamma \sinh z}.$$

The intermediate steps in the derivation of (2.3) are provided in Appendix C.

2.3 Optimal boundary values

The cost function in (2.3) must now be minimised with respect to the choice of boundaries on which firms adjust prices for a given initial position. The "smooth pasting" condition in Dixit (1991a) would require the first-order condition

$$\frac{dV(z)}{dz} = 0. \tag{2.5}$$

Stochastic optimal control requires

$$\begin{aligned}\frac{\partial V(w;y,z)}{\partial y} &= \frac{[f'(y)\sinh(y-z) - f(y)\cosh(y-z) + f(z)]\cosh(2w-z)}{\sinh^2(y-z)} \\ &= 0. \end{aligned} \tag{2.6}$$

and

$$\begin{aligned}\frac{\partial V(w;y,z)}{\partial z} &= \frac{[f(y) - f(z)\cosh(y-z) - f'(z)\sinh(y-z)]\cosh(2w-y)}{\sinh^2(y-z)} \\ &= 0. \end{aligned} \tag{2.7}$$

Equation (2.6) and (2.7) pin the values of a and b so that $V(.)$ is minimised. It is not clear what (2.5) does. The cost function seems to be optimised with respect to the initial state. But in stochastic optimal control the initial state is only a parameter of the problem, an inheritance from an unknown past. *Not a control variable.*

Simplifying (2.6) and (2.7), it is clear that y and z satisfy the conditions

$$f(y) - f(z)\cosh(y-z) - f'(z)\sinh(y-z) = 0, \tag{2.8}$$

$$f'(y)\sinh(y-z) - f(y)\cosh(y-z) + f(z) = 0. \tag{2.9}$$

It is now obvious that the optimal boundaries as provided by the values y and z are independent of the initial state x, and, hence, the optimal solutions of the two boundaries will only depend on the exogenous parameters of the process. These results imply that the zero threshold policy compels all firms whose initial positions lie outside these optimal boundaries to transact instantly and adjust their price level to match the market price. This naturally follows from the optimal stochastic control framework used here due to the unique costs faced by the firm. In applications of the "smooth pasting" condition this is implicit in its construction, irrespective of the initial price gap and the costs faced by the firm.

Equation (2.8) and (2.9) are now solved for y and z. Substituting for $f'(z)$ in (2.8) and multiplying the resulting expression by $\sinh z$ yields

$$f(y)\sinh z + f(z)\sinh(y-2z) + \frac{2z}{\gamma}\sinh(y-z) = 0. \tag{2.10}$$

Similarly substituting for $f'(y)$ in (2.9) and multiplying through by $\sinh y$ gives

$$\frac{-2y}{\gamma}\sinh(y-z) - f(y)\sinh(2y-z) + f(z)\sinh y = 0. \tag{2.11}$$

Subtracting (2.11) from (2.10), dividing resulting expression by $2\cosh(y-z)$ and then substituting for $f(y)$ and $f(z)$ yields

$$[\tanh(y-z) - (y-z)](z+y) = 0 \tag{2.12}$$

Its clear that (2.12) has two possible solutions

$$z + y = 0, \qquad \text{and} \qquad \tanh(y - z) - (y - z) = 0.$$

The latter condition is only satisfied when $z = y$, i.e. $a = b$. However, the boundaries were initially defined as being $a \leqq 0$ and $b \geqq 0$. Hence $z = y$ cannot be a solution. This leaves $z + y = 0$, which is true when $b = -a$. Therefore firms operating a zero threshold policy under a symmetric holding and transaction cost regime will find their optimal zone of inaction symmetrically disposed around the market price. This result is not unexpected and sounds intuitively correct since both costs are symmetric about the market price. However, Dixit (1991a) assumed this property.

Now that $z + y = 0$ it can be shown that z and y must satisfy

$$y - \tanh(y) - \frac{\gamma}{y} = 0. \tag{2.13}$$

Appendix D contains the intermediate steps in the derivation of this equation. Differentiating y with respect to γ, yields

$$\frac{dy}{d\gamma} = \frac{y}{\gamma + y^2 \tanh^2(y)}. \tag{2.14}$$

It is now clear that y is an increasing function of γ. Furthermore

$$\log \gamma = 2 \log \rho + \log g - \log k - 2 \log \delta,$$

and, so

$$\frac{\partial \gamma}{\partial \rho} = \frac{2\gamma}{\rho}, \qquad \frac{\partial \gamma}{\partial g} = \frac{\gamma}{g}, \qquad \frac{\partial \gamma}{\partial k} = \frac{-\gamma}{k} \qquad \text{and} \qquad \frac{\partial \gamma}{\partial \sigma} = \frac{-2\gamma}{\sigma}.$$

Since $y = \alpha b/2$ and $\alpha^2 = 2\rho/\sigma^2$ then

$$b = \sigma y \sqrt{\frac{2}{\rho}}. \tag{2.15}$$

It is clear from this expression for b that for economically realistic values of the parameters, the range of inertia defined by the optimal boundaries could be either small or large depending on how the state is scaled. It is now clear that b is a decreasing function of k and an increasing function of g. This shows that firms will wait longer before adjusting their prices as menu costs increase. But will wait less if the rate at which losses are accrued increases.

The behaviour with respect to σ and ρ is less obvious. From the definition of b it follows that

$$\begin{aligned} \frac{1}{b}\frac{\partial b}{\partial \sigma} &= \frac{1}{b}\sqrt{\frac{2}{\rho}}\left(y + \sigma \frac{\partial y}{\partial \gamma}\frac{\partial \gamma}{\partial \sigma}\right) \\ &= \frac{y\tanh^2(y) - \gamma}{\sigma[\gamma + y\tanh^2(y)]}. \end{aligned}$$

From (2.13) it is obvious that $\gamma = y^2 - y\tanh(y)$ and so

$$\begin{aligned} y^2\tanh^2(y) - \gamma &= y^2\tanh^2(y) - y^2 + y\tanh^2(y) \\ &= y\tanh(y) - y^2\operatorname{sech}^2(y) \\ &= y\operatorname{sech}^2[\cosh(y)\sinh(y) - 2y] \\ &= y\operatorname{sech}^2(y)[\sinh(2y) - 2y)] > 0. \end{aligned}$$

Hence b is an increasing function of σ. That is, the variance of the process and the zone of inertia move in the same direction. This reinforces the intuitively appealing idea that, as uncertainty increases, firms will delay action to avoid the downside of transacting too frequently while trying realise the potential upside of x decreasing. The value of waiting clearly increases with increased uncertainty.

By differentiation of (2.15) with respect to ρ, it can be shown that

$$\frac{\rho}{b}\frac{\partial b}{\partial \rho} = \frac{3\gamma - y^2 \tanh^2(y)}{2(\gamma + y^2 \tanh^2(y))} > 0.$$

See Appendix E. Clearly $2(\gamma + y^2 \tanh^2(y)) > 0$. It is also obvious that when $y = 0$,

$$3\gamma - y^2 \tanh^2(y) = 0. \tag{2.16}$$

If (2.16) is an increasing function of y, then it naturally follows that $\partial b/\partial \rho \geqq 0$. Differentiating (2.16) with respect to y yields

$$y - \tanh(y) > 0.$$

Therefore

$$\frac{\partial b}{\partial \rho} \geqq 0.$$

From this it is apparent that the zone of inertia and the constant discount rate are proportional to each other. It demonstrates that firms will accumulate losses if these losses decrease in value over time. Clearly current losses decrease in value rapidly over time under high discount rates. Therefore as ρ increases, it becomes relatively cheaper for firms to increase their waiting time because in real terms, as time evolves, the value of holding costs being accrued will decrease relative to the cost of adjustment. Although this relationship is intuitive it cannot be deduced from the Dixit model. The discount rate vanishes in simplifying Taylor series expansion used to obtain a simplifying approximation of b. It is only by making an empirical link between ρ and σ^2 that the effect of ρ on b can be analysed.

2.4 The Optimal Price Adjustment Policy

It can also be easily demonstrated that when the zero threshold policy is abandoned in favour of an optimal price adjustment policy, $V(x)$ must satisfy

$$V(b) - V(u) = g \qquad V(a) - V(l) = g, \tag{2.17}$$

where b and a are the upper and lower boundaries respectively, and, u and l are the upper and lower thresholds respectively (see chapter 2). To ease the treatment of (2.1) let us introduce the non-dimensional variables and parameters

$$\gamma = \frac{g\rho^2}{2k\sigma^2} \qquad w = \frac{\alpha(b-u)}{2} \qquad x = \frac{\alpha(b+u)}{2} \qquad y = \frac{\alpha(a-l)}{2} \qquad z = \frac{\alpha(a+l)}{2}. \tag{2.18}$$

Substituting (2.17) into (2.1) and solving the resulting simultaneous equations yields

$$V(v; w, x, y, z) = \frac{2v^2}{\gamma} + \frac{1}{\gamma} + \frac{\Phi(y,z) - \Phi(w,x)}{\tanh(z-2x) - \tanh(x-2v)}, \tag{2.19}$$

where

$$\Phi(w,z) = \frac{\gamma - yz}{\sinh y \cosh(z - 2v)}$$

See Appendix F for intermediate steps in the derivation of (2.19).

For an optima a necessary condition is

$$\frac{\partial V(v;w,x,y,z)}{\partial y} = yz\Big(\frac{\tanh y}{y} - 1\Big) + \gamma = 0. \tag{2.20}$$

It is clear that

$$\frac{\tanh(y)}{y} \leqq 1 \qquad \forall y.$$

Since k, ρ, and g are positive quantises, it must follow from (2.20) that

$$z = \frac{-\gamma}{\tanh y - y} \geqq 0.$$

Hence $a^2 > l^2$, and therefore $|a| > |l|$. Another necessary condition for optima is

$$\frac{\partial V(v;w,x,y,z)}{\partial w} = wx\Big(\frac{\tanh w}{w} - 1\Big) + \gamma = 0. \tag{2.21}$$

Using the same argument, it can be demonstrated that $|b| > |u|$.

In addition to these two conditions, the condition $x \in (a,b)$ such that $a \leqq l, u \leqq b$ must also must be satisfied. This implies that $a \leqq 0$ and $b \geqq 0$.

It is difficult to obtain analytical expressions for the optimal solution in this four dimensional problem. Therefore the problem is reparameterised in terms of four different parameters to obtain the optimal values for boundaries and thresholds using numerical methods.

Let

$$
\begin{aligned}
l &= a + c_1 \qquad c_1 \geqq 0 \\
b &= u + c_2 \qquad c_2 \geqq 0.
\end{aligned}
\tag{2.22}
$$

From this it trivially follows that $b \geqq a + c_1$ and also $b \geqq a + c_2$. The initial price gap x was defined to satisfy the condition $a \geqq x \geqq b$. Therefore a and b can be expressed as

$$
\begin{aligned}
a &= x + c_3 \qquad c_3 \geqq 0 \\
b &= x + c_4 \qquad c_4 \geqq 0.
\end{aligned}
\tag{2.23}
$$

Given the definition of the boundaries and thresholds, (2.22) and (2.23) need to satisfy

$$x + c_4 \geqq x - c_3 + c_1 \qquad \text{and} \qquad c_2 + x - c_3 \leqq x + c_4.$$

Therefore

$$c_4 \geqq c_1 - c_3 \qquad \text{and} \qquad c_4 \geqq c_2 - c_3.$$

This implies that

$$c_3 + c_4 \geqq \max(c_1, c_2),$$

and

$$c_3 + c_4 = \max(c_1, c_2) + c_5 \qquad c_5 \geqq 0.$$

Alternatively we have

$$c_3 + c_4 \geqq c_1 \geqq 0 \qquad \text{and} \qquad c_3 + c_4 \geqq c_2 \geqq 0.$$

Now we can write

$$c_1 = k_1(c_3 + c_4) \qquad 0 \leqq k_1 \leqq 1$$

and

$$c_2 = k_2(c_3 + c_4), \qquad 0 \leqq k_2 \leqq 1.$$

Thus the entire problem can now be reparameterised in terms of four constants k_1, k_2, c_3, and c_4 such that

$$0 \leqq k_1 \leqq 1 \qquad 0 \leqq k_2 \leqq 1,$$

and

$$c_3 \geqq 0 \qquad c_4 \geqq 0.$$

Now we can derive new expressions for a, b, l and u. That is:

$$\begin{aligned}
a &= x - c_3, \\
b &= x + c_4, \\
l &= a + k_1 c_3 + k_1 c_4 \\
&= x + k_1 c_4 + (k_1 - 1)c_3, \\
u &= b - c_2 \\
&= x + c_4 - k_2 c_3 - k_2 c_4 \\
&= x - (1 - k_1)c_4 - k_2 c_3.
\end{aligned}$$

The BFGS method for unconstrained minimisation (a quasi-Newton algorithm) is used to perform the four dimensional optimisation. Details of this are available in Technical Annex 1 and Bulirsch and Stoer (1980).

2.5 Results

Results are provided in two parts. Table 1 provides a comparison of the values obtained from the menu cost function under a zero threshold policy with those presented in Dixit(1991a). The punitive costs associated with following a sub-optimal pricing strategy are quantified by $V(w)$. Firms which initially lie outside the optimal zone of inaction and do not instantly transact down to zero face huge costs. Firms which transact down to zero confront only a fraction of these costs. Thus providing firms with a clear incentive to transact downwards. These observations are in contrast to the Dixit model in which these losses cannot be measured.

Table 2 provides an illustration of the results of the general price adjustment model given. The optimal zone of inertia remains the same, but the values of $V(v)$ are clearly different to the values of $V(w)$ for similar values of x. Although not applicable in this problem, these results indicate that the more general optimising strategy could be relevant in circumstances when the type of costs faced by firms change. A discussion of how the results obtained here could change when some the assumptions dealing with costs are relaxed is also provided. For the source code used to generate these functions see Technical Annexure 4.

2.5.1 Comparison of the zero threshold policy and existing work

The values of the cost function derived here, which does not contain the heuristically motivated "smooth pasting" condition, highlights a pattern of optimising behaviour that previously was not observable. To provide a comparison, the same values are chosen for the exogenous parameters as in Dixit(1991a), i.e. $\sigma = 0.1$, $\rho = 0.05$, $g = 0.1$ and $k = 0.5$. Table 1 compares the optimal values derived for the cost function in section 2 with Dixit (1991a) given by

$$V_{\text{Dixit}}(x) = \frac{-2kh}{\rho\alpha \sinh(\alpha h)} \cosh(\alpha x) + \frac{kx^2}{\rho} + \frac{k\sigma^2}{\rho^2},$$

where h is the so called optimal symmetric boundary value, to the optimal values given in Section 3. The value of h is the first positive root of the equation

$$y - \tanh(y) - \frac{\gamma}{y} = 0. \tag{2.24}$$

$V_{\text{Dixit}}(x)$ increases with respect to x for only a symmetric interval of $x \in (-0.3, 0.3)$. For $x \notin (-0.3, 0.3)$ the "smooth pasting" strategy suggests that firms should instantly incur cost g and eliminate their price gap, reducing (increasing) their total costs to $g + V(0)$, without evaluating the potential costs (or benefits) of following this strategy. In Table 1 $V(w)$ quantifies these costs, providing firms with clear incentive to change x to zero if they initially lie in this outer region.

The findings here are interesting. They confirm what Dixit (1991a) suggests. They show that if firms adopt a zero threshold policy, then the optimal zone of inertia will be constrained by the exogenous parameters driving the cost function. Firms, behaving as private optimisers, initially lying outside this zone will instantly adjust their prices, at time zero, to match the market price and bring themselves inside it, incurring an adjustment cost of g. This effectively implies that no firm will set its initial price so that its initial price gap lies outside this zone. Firms with $x \in (-0.3, 0.3)$ will be faced with a cost of $V(x)$. Firms with $x \notin (-0.3, 0.3)$ will confront a cost of $g + V(0)$.

This analysis clearly demonstrates that the costs associated with letting prices diffuse outside optimal zone of inertia are large. Firms make a significant cost saving by adjusting x down to zero. Therefore firms will only tolerate small deviations in their price from the market price, because, large price gaps are too costly. However, the resulting narrow zone of inaction requires elastic demand curves in output markets and supply curves in labour markets to cause large welfare fluctuations. This is a strong assumption to impose on markets.

2.5.2 The Optimal Price Adjustment Policy

The zero threshold policy compels all firms when adjusting their prices to match the market price. It is silent on the behaviour of firms who may otherwise choose to let their prices to deviate significantly from the market price. This is because the zero threshold policy is only a limited form of the general optimal control policy for dealing with such problems. It allows firms no flexibility in optimally determining the magnitude of price adjustment. In essence, it imposes synthetic constraints on the zone of inaction.

The results for the optimal price adjustment policy are remarkable. It captures the behaviour of firms functioning as complete optimisers. As can be seen from Table 2, firms will always opt for a zero threshold framework, even if their initial price difference is sufficiently large. This is because the potential downside associated with waiting exceeds the upside of matching the market price, albeit by a small amount. But clearly for $x \notin (-0.3, 0.3)$, $V(v) < V(w)$. This is because firms now have the flexibility of partially adjusting their prices. Although, this policy does not yield any further insights here due to the unique costs faced by the firm, these results suggest that a strategy based on firms partially adjusting their prices may yield interesting results in situations where holding costs and discount rates dynamically evolve, rather then being held constant, and the cost of adjustment is changed to include a proportional element.

If the cost attached to adjusting a positive price gap is different to that of adjusting a negative price gap, it is unlikely that the zone of inertia will be symmetrically disposed about the market price. Changing costs of adjustment to include a proportional element is likely to result in some firms adjusting prices to reduce rather than eliminate the price gap. Discount rates and holding costs which dynamically evolve could also have varying effects. In some instances the results may not change from those forecasted by the model derived here. In other cases firms may only partially close the price gap, the optimal zone of inertia could be asymmetric or both. The key feature driving firm behaviour would be the equation of motion governing discount rates and holding costs. Under these circumstances making an *a priori* decision on the type of adjustment policy to follow, such as the zero threshold policy in the Dixit model, would be to abstract too much from the true nature of costs faced by firms. The zone of inaction forecasted by a Dixit type price adjustment policy would clearly be at odds with how firms priced their output in the real world. Only the partial price adjustment policy described here will accurately evaluate how firms truly behave.

2.6 Concluding Remarks

This chapter utilises an optimal stochastic control framework to confirm the results obtained in Dixit (1991a). By adopting a non-heuristic approach it quantifies the potential costs to firms in areas which previously were unobservable due to the "smooth pasting" condition. As a result it becomes clear why firms opt for a zero threshold policy. *Earlier models such as the Mankiw or Dixit models use various heuristic arguments to make this an a priori feature.* Also analytical expressions are derived describing optimal agent behaviour as the underlying parameters change. Obviously, by relaxing some of the assumptions made in the Dixit model on adjustment costs, holding costs, or the discount rate many of the results could be expanded on. However, the aim of this chapter is also to demonstrate some of the *ex ante* restrictions placed on the solution by the heuristically motivated "smooth pasting" condition. In order to do this best, it was considered helpful to confine the analyses to an established model so that all its limitations could be easily observed.

x	$V(w)$	a	b	$V_{\text{Dixit}}(x)$	h
-2.000000	5.815624	-2.010000	0.010000	0.434207	0.340147
-1.900000	5.685384	-1.910000	0.010000	0.434207	0.340147
-1.800000	5.556160	-1.810000	0.010000	0.434207	0.340147
-1.700000	5.427494	-1.710000	0.010000	0.434207	0.340147
-1.600000	5.299132	-1.610000	0.010000	0.434207	0.340147
-1.500000	5.171162	-1.510000	0.010000	0.434207	0.340147
-1.400000	5.044232	-1.410000	0.010000	0.434207	0.340147
-1.300000	4.919831	-1.310000	0.010000	0.434207	0.340147
-1.200000	4.800695	-1.210000	0.010000	0.434207	0.340147
-1.100000	4.691372	-1.110000	0.010000	0.434207	0.340147
-1.000000	4.599035	-1.010000	0.010000	0.434207	0.340147
-0.900000	4.534703	-0.910000	0.421127	0.434207	0.340147
-0.800000	4.515216	-0.810000	0.403579	0.434207	0.340147
-0.700000	4.566686	-0.710000	0.385571	0.434207	0.340147
-0.600000	3.816582	-0.610000	0.368207	0.434207	0.340147
-0.500000	3.466724	-0.510000	0.353139	0.434207	0.340147
-0.400000	0.456546	-0.410000	0.342751	0.434207	0.340147
-0.300000	0.434207	-0.340147	0.340147	0.434207	0.340147
-0.200000	0.395923	-0.340147	0.340147	0.395923	0.340147
-0.100000	0.355556	-0.340147	0.340147	0.355556	0.340147
0.000000	0.339286	-0.340147	0.340147	0.339286	0.340147
0.100000	0.355556	-0.340147	0.340147	0.355556	0.340147
0.200000	0.395923	-0.340147	0.340147	0.395923	0.340147
0.300000	0.434207	-0.340147	0.340147	0.434207	0.340147
0.400000	0.456546	-0.342751	0.410000	0.434207	0.340147
0.500000	3.466724	-0.353139	0.510000	0.434207	0.340147
0.600000	3.816582	-0.368207	0.610000	0.434207	0.340147
0.700000	4.566686	-0.385571	0.710000	0.434207	0.340147
0.800000	4.515216	-0.403579	0.810000	0.434207	0.340147
0.900000	4.534703	-0.421127	0.910000	0.434207	0.340147
1.000000	4.599035	-0.010000	1.010000	0.434207	0.340147
1.100000	4.691372	-0.010000	1.110000	0.434207	0.340147
1.200000	4.800695	-0.010000	1.210000	0.434207	0.340147
1.300000	4.919831	-0.010000	1.310000	0.434207	0.340147
1.400000	5.044232	-0.010000	1.410000	0.434207	0.340147
1.500000	5.171162	-0.010000	1.510000	0.434207	0.340147
1.600000	5.299132	-0.010000	1.610000	0.434207	0.340147
1.700000	5.427494	-0.010000	1.710000	0.434207	0.340147
1.800000	5.556160	-0.010000	1.810000	0.434207	0.340147
1.900000	5.685384	-0.010000	1.910000	0.434207	0.340147
2.000000	5.815624	-0.010000	2.010000	0.434207	0.340147

Table 2.1: The Zero Threshold Policy vs. The Dixit Optimal Policy

x	$V(v)$	a	l	u	b
-2.000000	2.059944	-2.010000	0.191338	0.191338	0.461514
-1.900000	1.944792	-1.910000	0.188907	0.188907	0.459699
-1.800000	1.832973	-1.810000	0.185915	0.185915	0.457472
-1.700000	1.723971	-1.710000	0.182246	0.182246	0.454752
-1.600000	1.617223	-1.610000	0.177767	0.177767	0.451447
-1.500000	1.512131	-1.510000	0.172325	0.172324	0.447459
-1.400000	1.408090	-1.410000	0.165743	0.165743	0.442661
-1.300000	1.304525	-1.310000	0.157823	0.157823	0.436946
-1.200000	1.200951	-1.210000	0.148344	0.148344	0.430187
-1.100000	1.097061	-1.110000	0.137069	0.137069	0.422267
-1.000000	0.992916	-1.010000	0.121611	0.119480	0.416121
-0.900000	0.888780	-0.910000	0.107210	0.105323	0.405815
-0.800000	0.785938	-0.810000	0.088047	0.085194	0.393846
-0.700000	0.686441	-0.710000	0.068057	0.063766	0.382879
-0.600000	0.593698	-0.610000	0.046432	0.043729	0.367149
-0.500000	0.513858	-0.510000	0.021980	0.015662	0.355389
-0.400000	0.456470	-0.410000	0.004171	0.000762	0.344190
-0.300000	0.434207	-0.340147	0.000000	0.000000	0.340147
-0.200000	0.395923	-0.340147	0.000000	0.000000	0.340147
-0.100000	0.355556	-0.340147	0.000000	0.000000	0.340147
0.000000	0.339286	-0.340147	0.000000	0.000000	0.340147
0.100000	0.355556	-0.340147	0.000000	0.000000	0.340147
0.200000	0.395923	-0.340147	0.000000	0.000000	0.340147
0.300000	0.434207	-0.340147	0.000000	0.000000	0.340147
0.400000	0.456470	-0.344190	-0.000762	-0.004171	0.410000
0.500000	0.513858	-0.355389	-0.015662	-0.021980	0.510000
0.600000	0.593698	-0.367149	-0.043729	-0.046432	0.610000
0.700000	0.686441	-0.382879	-0.063766	-0.068057	0.710000
0.800000	0.785938	-0.393847	-0.085193	-0.088046	0.810000
0.900000	0.888756	-0.404789	-0.106189	-0.107472	0.910000
1.000000	0.992916	-0.416121	-0.119480	-0.121611	1.010000
1.100000	1.097061	-0.422267	-0.137069	-0.137069	1.110000
1.200000	1.200951	-0.430187	-0.148344	-0.148344	1.210000
1.300000	1.304525	-0.436946	-0.157823	-0.157823	1.310000
1.400000	1.408090	-0.442661	-0.165743	-0.165743	1.410000
1.500000	1.512131	-0.447459	-0.172324	-0.172325	1.510000
1.600000	1.617223	-0.451447	-0.177767	-0.177767	1.610000
1.700000	1.723971	-0.454752	-0.182246	-0.182246	1.710000
1.800000	1.832973	-0.457474	-0.185914	-0.185915	1.810000
1.900000	1.944792	-0.459699	-0.188907	-0.188907	1.910000
2.000000	2.059944	-0.461514	-0.191338	-0.191338	2.010000

Table 2.2: Results From The Optimal Price Adjustment Policy

Chapter 3

Transactions Demand for Money: A Critical Review

3.1 Introduction

Keynes (1936) identifies three reasons for holding money. The first is the transactions motive, where agents select an optimal cash balance by minimising the costs associated with managing a portfolio of cash and an interest earning illiquid asset. The second is the speculative motive, where money is held as a component of a portfolio of assets optimally selected to maximise return while minimising risk. The third is the precautionary motive, where money is held as a buffer stock to absorb any unplanned expenditure shocks. These ideas were subsequently formalised in models which incorporated either one or two of these motives. This chapter looks at how the literature on the transactions demand for money has evolved, highlighting some of its perceived strengths a weaknesses.

The early contributions are discussed in Section 2. Subsequent developments are analysed in Section 3 highlighting the attempt to synthesise both the transactions and the precautionary motives. The most recent literature is reviewed in Section 4, looking at how valuable information could have been discarded by authors when obtaining analytic solutions. Finally the benefits of combining the inventory theoretic approach with the various portfolio models, which examine both the speculative and precautionary motive, are discussed. *When referring to money or cash here, it is not strictly in the sense of $M0$, but assets with cash like attributes*

3.2 Early Transaction Money Demand Models

Modern transactions money demand models have their genesis in Baumol (1952) and Tobin (1956) (Baumol-Tobin) which look at money holdings from a micro basis. Transactions are assumed to occur at a constant rate and are perfectly foreseen, and agents seek to optimise a portfolio consisting of cash and an illiquid interest earning asset.

Take an agent who spends $\pounds C$ at a uniform rate. To do this he must either borrow or draw on his savings and incur an opportunity cost in the form of forgone interest income, say γ per period. All withdrawals are made in fixed quantities of $\pounds M$ and with each withdrawal a lump sum transaction cost of β is incurred. Thus any $M \leqq C$ will permit the agent to meet expenses if a sufficient number of withdrawals are made. Hence a minimum of C/M withdrawals will be required costing $\pounds\beta C/M$. If $\pounds M$ is expended at a constant rate, then average money holdings will be $\pounds M/2$. If interest is earned every C/M periods then the interest opportunity cost will be $\gamma M/2$. The total cost of holding money will be

$$\frac{\beta C}{M} + \frac{\gamma M}{2}.$$

Minimising this costs with respect to M and making M the subject yields

$$M = \sqrt{\frac{2\beta C}{\gamma}},$$

which is the expression for the optimum level of money holdings. It is obvious from this that, if $C = PY$ where P is the price level and Y is the level of real income, the income elasticity of money will be $1/2$. However, numerous empirical studies have shown this not to be the case. Narrow money has been demonstrated to have a short run income elasticity of close to zero and a long run income elasticity of close to unity.

Many explanations can be forwarded as to why these models fail to match empirical findings. Firstly they are deterministic and therefore do not account for precautionary balances. The interest rate is held constant. Hence any change in the level of expected money holdings will not reflect any inertial responses which occur due to interest rate uncertainty. Because the model is static it also fails to capture the true intertemporal opportunity cost of holding money. Also, it takes a limited view of agent optimising behaviour. Unlike most portfolio based models, where agents seek to maximise wealth by holding a portfolio of assets, of which money may be one, in an environment of ongoing uncertainty, these models assume a world of perfect foresight. Given these limitations it is not surprising that the Baumol-Tobin approach does not reflect empirical findings. However, the model also yields some useful insights. It captures the long run directional sensitivity of the money stock to changes in key macro variables, serving as an approximation of the equilibrium behaviour of transactions money balances.

3.3 The Discrete Stochastic Cash Flow Models

The criticism levelled at Baumol-Tobin in the literature that immediately followed it was that it failed to reflect the money holdings of firms both from positive and normative points of view. Miller and Orr (1966), (Miller-Orr) argue that the typical pattern of money holdings is not as simple as the deterministic view in Baumol-Tobin, but that they typically follow a random walk. This assumption instantly changes the dimension of the problem. Firms not only have to decide on how much to withdraw when cash holdings hit a minimum level, but also how much surplus cash to switch into the interest earning asset. Transfer costs like Baumol-Tobin are assumed to be lump sum, say $£\beta$. All transfers between the two accounts occur instantaneously. Also the cash balance is not allowed to fall below a minimum level. Firms seek to minimise the long run average cost of managing their money stock by using a two parameter control policy. The two parameters being the upper limit of cash holdings h and a threshold of z.

Given these assumptions, the cost of managing a firm's cash inventory over a finite horizon of say T days will be

$$E[C] = \beta\frac{E[N]}{T} + \varphi E[M],$$

where $E[N]$ is the expected number of cash transfers over time horizon T, β is the marginal cost of each transfer, $E[M]$ is the expected daily cash balance, and φ is the interest earnings per day. Obviously the firm's objective will be to minimise $E[C]$ with respect to h and z.

To solve this problem Miller and Orr express $E[N]/T$ in terms of the control variables h, and z. If the time span between transfers into and out of the cash account are given by $x_1, x_2, \cdots$, which are independent random variables from a population with a well defined probability distribution with a mean D and a finite variance, then

$$E[x_1 + x_2 + \cdots, +x_n] \leqq T < E[x_1 + x_2 + \cdots, +x_{n+1}],$$

or

$$D.E(N) \leqq T < D.E[N] + D,$$

since $E[x_1 + x_2 + \cdots, +x_n] = D.E[N]$. From the above equation it can be implied that

$$\frac{1}{D} - \frac{1}{T} < \frac{E[N]}{T} \leqq \frac{1}{D}.$$

It is obvious that if T grows unboundedly, then $E[N]/T$ will tend towards $1/D$.

For a symmetric random walk the mean first passage time out of the continuation region (0,h), $D(z,h)$ is given by Feller (1957) to be

$$D(z,h) = (z)(h-z).$$

Miller-Orr convert this value into the expected duration between cash transfers per day letting $z' = z.m$ and $h' = h.m$ to arrive at

$$D(z',h') = \frac{(z')(h'-z')}{m^2 t}.$$

Next they express $E[M]$ in terms of z and h. The probability that the cash balance will contain precisely x units is

$$f(x) = pf(x-1) + qf(x+1) \qquad x \neq z,$$

which must satisfy the boundary conditions

$$f(z) = p[f(z-1) + f(h-1)] + q[f(z+1) + f(1)], \tag{3.1}$$

and,

$$f(0) = 0, \qquad f(h) = 0,$$

and the distribution condition

$$\sum_{x=0}^{n} f(x) = 1.$$

Solving these equations yield a solution of the form

$$f(x) = A_1 + B_1 x \qquad 0 < x < z, \tag{3.2}$$

and

$$f(x) = A_2 + B_2(h-x) \qquad z < x < h. \tag{3.3}$$

The linearity of (3.2) and (3.3) gives rise to a mean of the distribution they form of $(h+z)/3$. Further letting $Z = h - z$ the cost function to be minimsed will now be

$$\min_{Z,z} E[C] = \beta \frac{m^2 t}{zZ} + \gamma \frac{(Z+2z)}{3}.$$

This is nothing more than the gambler's ruin problem which is described in Feller (1957). Minimising this expression it is clear that the optimal threshold z is

$$z = \left(\frac{3\beta m^2 t}{4\gamma}\right)^{\frac{1}{3}}$$

and the optimal size of the upper boundary h will be

$$h = 3z.$$

By substituting the optimal level of z into the above equation it can be easily seen that a firms steady state average cash balance M^* is given by

$$M^* = \frac{4}{3}\left(\frac{3\beta m^2 t}{4\gamma}\right)^{\frac{1}{3}}.$$

If $m = PY$ it is clear that the income elasticity of average steady state money holdings will be 2/3.

Here again like Baumol-Tobin the model deviates from what has been empirically observed. Although Miller-Orr introduce uncertainty to expand on the deterministic nature of Baumol-Tobin, the model still remains discrete, failing to take account of the continuous process by which cash flows occur. Interest rates are assumed to be constant. By leaping to the steady state, the model is constrained to being essentially static. Also, in reality it could be argued that costs faced by agents are not in fact lump sum but linear and asymmetric. By assuming that agents do not borrow, the effect of debt and the asymmetric responses which this gives rise to are also excluded.

However, Miller-Orr breaks new ground on many fronts. It marries the transactions and precautionary motives for holding money. Cash balances are a risky asset whose value is stochastic and given by a random walk. If interest revenue from the illiquid asset is normalised to zero, then the risky asset cash will generate an income of return $-\varphi$. The agent effectively maximises the *expected value* of the payoff from holding the risky asset by optimally deciding whether to go long or short on cash in the form of the interior threshold, in the presence of fixed hedging costs. When cash balances hit zero, the agent buys an optimal quantity of cash at a fixed cost to restore himself to the interior point. If balances hit some upper boundary, the agent sells cash due to its punitive return at a fixed cost. This effectively, makes Miller-Orr an *expected utility maximisation problem.* With all the risky and risk free return transferred to cash, it is clear from the mathematical properties of linear utility functions, that utility maximisation by a risk neutral firm will yield the same optimal targets and thresholds as a cost minimisation exercise. Effectively, agents maximise the payoff of playing a fair game. The appealing idea of this approach is that money balances are only adjusted when they hit only an upper or lower boundary. Temporary and short-term changes in the in the money stock are voluntarily held.

3.4 The Continuous Stochastic Cash Flow Models

3.4.1 Steady State Models

Milbourne, Buckholtz and Wasan (1983) (MBW) try expand on Miller-Orr by introducing continuous cash flows in the form of a Wiener process. The results obtained are identical to Miller-Orr. Frenkel and Jovanovic (1980) allow for continuous cash flows as well, but opt for a single parameter control policy like Baumol-Tobin, which yields a short run income elasticity of money holdings of less than a half.

The robustness of these models lie in the modelling techniques utilised. Frenkel and Jovanovic (1980) impose an upper boundary which is unbounded.

This forces the agent to incur holding costs which are punitive, with a probability that is non-trivial, when cash balances diffuse upwards. No action is available to mitigate these losses. Therefore, it is not surprising that despite introducing uncertainty they obtain an income elasticity smaller than the purely deterministic perfect foresight model described by Baumol-Tobin. A finite upper boundary which is chosen optimally like Miller-Orr will serve to make this model realistic and also yield a smaller income elasticity.

MBW should yield similar results to Miller-Orr, but not identical since they use a continuous time framework. Agents restore balances to an interior point z if cash balances hit an upper limit h or a lower limit of zero. MBW obtain a steady state differential equation for the distribution of cash holdings of the type

$$0 = \frac{\sigma^2}{2}\frac{\partial^2 \pi(x)}{\partial x^2} + \frac{\sigma^2}{2}(C_1 - C_2)\delta(x - z), \tag{3.4}$$

where x denotes the stock of money and σ^2 its variance. They solve (3.4) to obtain a cash distribution

$$\pi(x) = \begin{cases} \frac{2x}{hz} & 0 \leqq x \leqq z \\ \frac{2(h-x)}{h(h-z)} & z \leqq x \leqq h. \end{cases}$$

C_1 and C_2 are defined to be

$$\lim_{t\to\infty} \frac{\partial f(x,t)}{\partial x}\Big|_{x=0} = C_1, \qquad \text{and} \qquad \lim_{t\to\infty} \frac{\partial f(x,t)}{\partial x}\Big|_{x=h} = C_2.$$

Firstly

$$\frac{\sigma^2}{2}\frac{\partial^2 \pi(x)}{\partial x^2}$$

does not exist, since the second derivative of $\pi(x)$ does not exist. Thus the distribution of net cash holdings is that of a purely deterministic process, something like Baumol-Tobin. Secondly $\pi(x)$ does not satisfy (3.4). Also for (3.4) to hold C_1 must equal C_2. In other words

$$\lim_{t\to\infty} \frac{\partial f(x,t)}{\partial x}\Big|_{x=0} = \lim_{t\to\infty} \frac{\partial f(x,t)}{\partial x}\Big|_{x=h}$$

or

$$\frac{2}{hz} = \frac{2}{h(z-h)}.$$

This obviously cannot be true unless $h = 0$. But h was initially defined to be not equal to zero. Thus the solution that MBW provide for the distribution of cash holdings does not satisfy the differential equation from which it is obtained.

Clearly the logical foundations of both models are suspect. From an economic standpoint they do not significantly add to the insights provided by Baumol-Tobin and Miller-Orr. The cost minimisation criterion used to determine the optimal boundaries and thresholds allow for only the optimal cash management decision of risk neutral agents to be analysed. And, like the preceding models the so called solutions are only valid for a steady state view of the world.

3.4.2 Net Present Value Cost Minimisation Models

Constantinides and Richard (1978) ascribe to the agent the net present value of a cost minimisation problem, where net cash disbursements dx_t follow a Wiener process, i.e

$$dx_t = \mu dt + \sigma_x dz_x.$$

μ is the mean flow of net cash disbursements, σ_x is the standard deviation of net cash flows and dz_x is a Wiener increment. The net present value of holding costs are given by

$$\underset{inf\,\mathbf{u}}{V(m,r;\mathbf{u})} = E\left[\int_0^\infty e^{-\rho t}C(x_t)\,dt + \sum_{i=1}^{\infty} e^{-\rho\tau_i}B(\phi_i)\middle| x_0 = x\right]. \tag{3.5}$$

Holding costs $C(x)$ are assumed to be

$$C(x) = \begin{cases} hx & x \geqq 0 \\ -px & 0 \geqq x. \end{cases}$$

ρ is the discount rate, and the transfer costs $B(\phi_i)$ are given by

$$B(\phi_i) = \begin{cases} K^+ + (u-U)k^+ & x \geqq u \\ K^+ + (D-d)k^+ & d \geqq x. \end{cases}$$

Expanding (3.5) using a stochastic Taylor series expansion yields

$$\rho V(x) + \mu V'(x) - \frac{\sigma^2}{2}V''(x) - C(x) = 0. \tag{3.6}$$

Solving (3.6) with respect to the "smooth pasting" and "value matching" conditions given by

$$V'(D) + k^+ = 0 \qquad V'(U) - k^- = 0,$$

and

$$V(d) = V(D) + K^+ + k^+(u-U) \qquad V(u) = V(U) + K^- + k^-(D-d),$$

respectively, yields a solution of the form

$$V(x) = \begin{cases} V(u) + (x-u)k^-, & u \leqq x \\ hx/\rho + h\mu/\rho^2 + c_1e^{\lambda_1 x} + c_2e^{\lambda_2 x}, & 0 \leqq x \leqq u \\ -px/\rho - p\mu/\rho + c_3e^{\lambda_1 x} + c_4e^{\lambda_2 x}, & d \leqq x \leqq 0 \\ V(d,r) + (d-x)k^+, & x \geqq d. \end{cases}$$

c_1, c_2, c_3 and c_4 are constants. Note that the overdraft rate only affects $V(x)$, *iff* initial cash holdings x are less than zero. Otherwise it has no effect, despite the fact that it may still apply when cash holdings drop below zero.

The results generated differ significantly compared to preceding models. Constantinides and Richard (1978), by looking at the net present value of costs associated with managing a cash inventory increase the time horizon over which costs are generated. Discounting ensures that transfer costs are punished more than in the steady state literature. Clearly the zone of inertia increases. This is important, because it points to a longer adjustment lag and a higher long run income elasticity than predicted by steady state models. Proportional costs are continuously incurred at a rate proportional to the storage level of money, and costs accrue at a much faster rate than the finite horizon considered in steady state models. This has the effect of reducing the frequency and increasing the size of withdrawals and deposits. Of course some of this increased inertia could be partly due to assuming that transfer costs are linear in the size of the transaction. The modification of these costs to reflect the asymmetric costs encountered when depositing and withdrawing cash yield the asymmetric targets and thresholds encountered in reality. The overdraft rate of $-p$ could also contribute towards this.

The weakness in this model lies in its use of the "smooth pasting" condition as an optimisation tool. This is a heuristically motivated condition which cannot be reconciled with the optimal stochastic control theoretic framework used. Effectively, gradient conditions are imposed on the cost function with respect to the initial level of money holdings. A critique of this condition is provided in Chapter 1, where the natural conditions for this problem are derived. It is also clear that the solution to the cost function in the regions $u \leqq x$ and $x \leqq d$ do not satisfy (3.6). Due to these inconsistencies the forecasted behaviour of this model could detract from the underlying assumptions of the model.

Smith (1989) tries to develop Constantinides and Richard (1978) further by introducing a mean reverting stochastic process to model interest rates. This yields a significant breakthrough allowing the optimal targets and thresholds to capture not just cash flow uncertainty, but also interest rate uncertainty. From an economic standpoint, the assumption of mean reversion is questionable. It is only a long run observation. To confirm this in most economies would require data sets spanning a long time period. Given that structural and institutional regimes influencing the prices and rates of return on alternative assets to money have experienced numerous changes within each decade, it is unlikely that matching interest rates over a lengthy time span would reflect the same opportunity cost of holding money. In fact it would be more realistic to assume that interest rates follow a Poissonian type process in which the interest rate experiences discrete jumps at discrete time intervals.

Smith (1989) models the mean reverting interest rate as an Ornstien-Uhlenbeck process

$$dr = \alpha(\gamma - r)dt + \sigma_r dz_r \qquad \alpha \geqq 0$$

where r is the interest rate. The coefficient $\alpha(\gamma - r)$ captures the mean reverting effect, and dz_r is a standard Wiener increment. Applying a stochastic Taylor series expansion on the resulting cost function with respect to the initial state yields a partial differential equation of the type

$$\begin{aligned} \rho V(x,r) - \mu V_x(x,r) - \alpha(\gamma - r)V_r(x,r) \\ \frac{\sigma^2}{2}V_{xx}(x,r) - \sigma_{x,r}V_{xr}(x,r) - \frac{\sigma_{rr}}{2}V_{rr}(x,r) = C(x,r). \end{aligned} \tag{3.7}$$

This is solved using the "smooth pasting" and "value matching" conditions to obtain the cost function

$$V(x,r) = \begin{cases} V(u,r) + (x-u)k^{-}, & u \leqq x \\ (\rho+\alpha)^{-1}[xr + \alpha\gamma x/\rho + (\rho+\alpha)^{-1}\mu r & \\ +\mu\gamma\alpha(1+\rho(\rho+\alpha)^{-1})/\rho^2 + \sigma_{xr}/\rho] + c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x}, & 0 \leqq x \leqq u \\ -px/\rho - p\mu/\rho + c_3 e^{\lambda_1 x} + c_4 e^{\lambda_2 x}, & d \leqq x \leqq 0 \\ V(d,r) + (d-x)k^{+}, & x \geqq d. \end{cases}$$

The above solutions to (3.7) do not necessarily correspond to the assumptions made in Smith (1989). Consider the solution for the region $0 \leqq x \leqq u$. This is only valid if $\sigma_r = 0$, i.e. if the stochastic component of interest rates is removed. But this would imply that interest rates are deterministic. In fact a meaningful solution to (3.7) can only be obtained if $\alpha(\gamma - r) < 0$. If $\alpha(\gamma - r) > 0$, then the anti diffusion effect of this coefficient would cause extreme instabilities. This effectively implies as $t \to \infty$ interest rates will be unbounded and negative. Consider the solutions for $V(x,r)$ in the regions $x \geqq d$, and $u \leqq x$. These like Constantinides and Richard (1978) do not satisfy the differential equation (3.7). Hence $V(u,r) + (x-u)k^{-}$ and $V(d,r) + (d-x)k^{+}$ cannot be solutions to (3.7).

These inconsistencies along with the heuristically motivated "smooth pasting" condition may explain why Smith (1989) does not yield any significantly different results to Constantinides and Richard (1978). If interest rates over time become negative and grow unboundedly, agents after a finite period of time will hold all their wealth in cash. This effectively reduces the time horizon under consideration to be a finite one. Assuming that $\sigma_r = 0$ eliminates any inertial responses that may occur due to interest rate uncertainty from the model, effectively reducing the model to a type similar to Constantinides and Richard (1978). Therefore it is not surprising that Smith (1989), like previous models, concludes that the dynamic process governing aggregate money demand is a product of the "chattering" of cash balances between the targets and thresholds.

3.5 Empirical Evidence

Inventory theoretic money demand models suggest a stable money demand function and a lagged adjustment in the money stock to exogenous expenditure shocks. Here the empirical evidence is discussed. The literature in this area is vast. Therefore rather than provide an exhaustive account of all the research done, certain illustrative examples of the most recent work shall be discussed. The literature is discussed without any recourse to problems which arise from aggregation bias. However, this issue shall be dealt with in the subsection which follows the empirical survey.

3.5.1 An Empirical Survey

Artis and Lewis (1976) estimate a first order partial adjustment model, and look at the stability of the estimated coefficients for the period 1963(2) to

1973(1). The rate of return on broad money is measured as the difference between the own rate of money and gilts. Risk in all equations is incorporated through introducing the variance of bond prices in all equations. They find that all equations fail the Chow test for the period 1971(1)-1973(1). For broad money this instability extends over a much longer period. Artis and Lewis (1976) argue that these instabilities are caused by disequilibrium in the money market. Hendry (1979,1985) studies the demand for transactions balances in the non-bank private sector. It is assumed that the long run demand for M1 is determined by real income and an opportunity cost which is considered to be the 3 month local authority rate. The findings suggest a short run income elasticity of less than 0.5. The error-correction component has a one to one relationship between the money stock, prices and income indicating a long run income elasticity of money of unity. An interesting finding is that velocity is negatively correlated to real money balances, suggesting a smaller a zone of inertia with increased velocity. Evidence of a lagged adjustment in the money stock is also found. Milbourne (1983), and Cuthbertson (1986) effectively confirm these findings. Cuthbertson and Taylor (1991) suggest that the the demand for M1 seems to experience structural changes in the late eighties, and also find some instability in the period 1968(4)-1983(4). Using a longer data set Artis and Lewis (1981) show that M2 has a long run elasticity of unity. Hendry and Erricson (1988), examining data for the period 1867-1975 for broad money, using the Engle-Granger two step cointegrating technique, obtain a long run income elasticity of unity. Muscatelli (1989) demonstrates that whilst M3 has a unitary long run elasticity M1 does not. The cointegrating vector for M1 only appears in the demand for money equation, implying short run divergences from the equilibrium demand equation. It demonstrated that the M3 equilibrium error exists in the price, income and interest rate equations highlighting a more complex adjustment pattern. Hurn and Muscatelli (1991a,b) also find evidence of a small short run income elasticity for M0 and M4 and a long run elasticity of close to unity.

In the USA most models demonstrate instabilities in the post 1973 period. Goldfeld (1976) finds a stable money demand function with a low short run income elasticity for the periods 1952(2)-1973(4), but in dynamic simulations the model over predicts money balances for the period 1979-1982 (see Cuthbertson (1985)). Laidler (1980) finds that M2 is much less stable than M1, which is interesting because in this period the targeted aggregate was M1. Gordon (1984) uses an ADL-ECM approach to model the demand for narrow money, but finds considerable instability in the estimated equations. A comprehensive account for narrow money is provided by Baba et al. (1988) for the periods 1960(2)-1984(2). They find a short run income elasticity which is 0.34 and a long run elasticity of 0.5, the interesting aspect about this study are the various measures of opportunity costs used.

The empirical evidence is clearly mixed. The perceived stability or instability of the money stock clearly depends on the specification of the money demand equation and the kind of statistical techniques used, for example GLS, ADL-ECM or the Engle-Granger two step technique. There exists clear evidence of lagged adjustment, however, a significant minority of the surveyed literature seems to find an unstable money demand function. Whilst the former is consistent with agents using a target threshold inventory management technique, the latter does not support this view. Theoretical models need to account for these two differing views on how the money stock behaves in order to be consistent with the empirical evidence.

3.5.2 Aggregation Bias

The previous section draws conclusions on the validity of micro money demand models based on the results derived from macro aggregate empirical models. However, there exist a number of problems associated with reconciling both these approaches. A large body of the theoretical econometric literature addresses the issue of aggregation bias, which is defined as being the deviation of macro parameters from the average of the corresponding micro parameters, in detail highlighting some of the important issues which need to be dealt with. See Pesaran, Pierse and Kumar (1989), and, Lee, Pesaran and Pierse (1990).

Consider the following disaggregated model

$$H_a: \quad \mathbf{y_i} = \mathbf{X_i}\beta_i + \mathbf{u_i} \qquad i = 1, 2, \cdots, m. \tag{3.8}$$

$\mathbf{y_i}$ is a $n \times 1$ vector of the dependent variable, $\mathbf{X_i}$ is a $n \times k$ matrix of observations on the regressors in (3.8) and β_i is a $k \times 1$ vector of coefficients and $\mathbf{u_i}$ is the associated disturbance term. The aggregate equation, which satisfies the Klien-Nataf consistency condition is

$$H_b: \quad \mathbf{y_a} = \mathbf{X_a}\mathbf{b_a} + \mathbf{v_a}. \tag{3.9}$$

and

$$\mathbf{y_a} = \sum_{i=1}^{m} \mathbf{y_i} \qquad \mathbf{X_a} = \sum_{i=1}^{m} \mathbf{X_i}.$$

$\mathbf{b_a}$ is a $k \times 1$ vector of aggregate coefficients. The basic test of aggregation concerns itself with the problem of

$$H_\psi: \quad \psi = \sum_{i=1}^{m} \mathbf{X_i}\beta_\mathrm{i} - \mathbf{X_a b_a} = 0.$$

The test statistic ψ is assumed to be a Gaussian variable. There exist many reasons for the null hypothesis H_ψ being rejected most of which result from the micro dynamics of the variables concerned. Structural breaks in the micro data for specific dependent variables $\mathbf{y_i}$ may lead to the null hypothesis being rejected. Misspecification of either the aggregate or disaggregate models could be another important factor. Furthermore, there exists a large array of micro based estimation issues which could lead to either a significant upward or downward bias in the computed values of both the long and short run elasticities of the real money stock with respect to the interest rate or income. Aggregation bias could also enter when one moves from a narrow measure of money to a broader measure. The inferences drawn from the empirical literature must consider this problem before accepting or rejecting hypothesis based on the results derived in the current micro based theoretical literature.

3.6 Concluding Remarks

The inventory theoretic approach generates targets and thresholds using a dual asset management framework for only risk-neutral agents. The effects of changes in macroeconomic variables only affect the money stock through the associated opportunity costs of holding money. Although this approach yields significant insights into how the transactions motive affects the aggregate money stock and captures the precautionary motive for holding money in cases of risk neutrality; it does not analyse the effects of risk. This may explain why some of the results described above may not necessarily be compelling as alternative approaches, or be consistent with the empirical evidence. Clearly the contentious issue of instability in the empirical literature needs to be dealt with in models of the target threshold type described in this chapter if these models are to cover the gamut of accumulated evidence.

One direction in which the theoretical literature could develop is to consider a different type of agent such as a risk averter. The non-linearities caused by a concave utility function could significantly alter the 'neat' results obtained in existing models. Liquidity preference models developed by Markowitz (1952), (1959), Tobin (1958) Feldstein (1969) and Courakis (1988), and also expected utility theory would be the natural starting point here. The Lucas critique offers some promise in this direction. A sound mathematical approach to solving these problems may also yield differing results to the current heuristically motivated techniques used.

Chapter 4

Optimal Money Holdings

4.1 Introduction

Transaction money demand models explain the sensitivity of money balances to interest rates, and, the lagged adjustment of monetary aggregates to exogenous changes in macroeconomic variables through a dual asset optimisation approach. The agent, usually assumed to be a risk neutral firm, optimally selects a portfolio consisting of an interest earning illiquid asset and cash in the presence of transaction costs and an exogenously specified stream of cash flows. On the other hand, risk aversion models represent money holdings as a component of a portfolio of assets, optimally selected by an agent to maximise his utility of wealth, trading off risk and return. This chapter links these two ideas to determine the optimal portfolio choice of a risk averter in the presence of stochastic shocks to asset prices and an equilibrium net income stream. Numerical solutions are obtained for the optimal zone of inaction using a utility maximisation framework. The results do not yield the well behaved inertial responses derived in conventional cost minimisation problems. The multiple optima observed demonstrates that the optimal value of waiting for going long or short on cash discretely varies despite holding the underlying parameters and risk preferences constant. This clearly indicates that a one off change in a key macroeconomic variable will result in several discrete adjustments being made to the money stock over a long period of time. This is also confirmed by the intuition behind the popular General To Specific (GTS) empirical modelling technique. The use of flexible lags in formulating the money demand equation in the presence of a multiplicity of accessible long run equilibrium relationships allows for the possibility of an exogenous shock forcing the economy onto a new equilibrium. Of course, the GTS method *a priori* allows for only a unique long run equilibrium relationship, thus discounting the existence of multiple optimising strategies. However, the use of some kind of spectral estimation technique will overcome this limitation. Based on the varying lags of the error correction mechanism found in the current empirical literature, there exists strong evidence to believe in the presence of more than a single long run equilibrium relationship.

Section 2 provides a brief illustration of risk aversion models and highlights their similarity to transactions money demand models of the Miller-Orr type. A simple example is also provided. Section 3 sets out the model, and the underlying assumptions. The initial value problem is solved in Section 4. Although this deals with a very special case, its values are necessary to obtain a numerical solution to the general problem. Finally, numerical solutions obtained by solving the model are presented in Section 5. All the technical detail is relegated to the Appendices and Technical Annexures.

4.2 Risk Aversion Models vs. Stochastic Transaction Money Demand Models

Risk aversion liquidity preference models analyse the optimal portfolio choice of a risk averter. Given a risk preference, an initial level of wealth and a wealth constraint, the agent is faced with the problem of allocating his wealth among a portfolio of assets which maximises his utility (of wealth) over a given time horizon. This allocation is made among both risky and risk free assets. Holding assets with a higher risk may increase his return, but at the same time increase the possibility of a capital loss. Money on the other hand does not yield a return but is also risk free and thus may prove to be attractive as a component of the portfolio (see Markowitz (1952), (1959), Tobin (1958), Feldstein (1969), Dalal (1983)). Herein lies the justification for agents holding money as a component of a portfolio of assets. It is explicitly demonstrated here that risk aversion and the initial level of wealth plays an important role in determining the optimal choice of cash and the interest earning asset.

In many ways risk aversion and transaction money demand models overlap. An optimal portfolio selection exercise equates the loss in marginal utility to an agent as a result holding a portfolio of a risk free and a risky interest earning asset to the gain in marginal utility arising from both these assets. A risk averter attaches a diminishing marginal utility to each unit of additional wealth. A change in utility resulting from a change in wealth specifically depends on the level of wealth itself. *Hence wealth needs to be explicitly included in an optimal portfolio selection exercise for a risk averter* *like Tobin (1958) does.* On the other hand, risk neutral agents attach a constant marginal utility to increasing levels of wealth. Thus normalising the wealth level and the interest income earned with it to zero will not change the optimal portfolio decision of the agent. This is what Miller-Orr and subsequent stochastic transaction money demand models do. The introduction of a stochastic interest rate in Smith (1989) allows for a price varying asset. In fact Miller-Orr explicitly conclude that the closed form solutions obtained are analogous to a dual asset portfolio selection exercise for firms with risk neutral preferences.

Consider a simple example in which an agent who is initially endowed with an interest earning asset of amount w and a stock of cash m_0. Assume that w earns a rate of return r per time period, payable on its outstanding balance at the end of each period after all portfolio adjustments have been made. Also assume that at the end of the current period cash flows can cause the stock of cash to be in one of three states, i.e. m_1 with probability p, m_2 with probability q and m_3 with probability $1-p-q$. The states have the ordering $m_3 < m_2 < m_0 < m_1$. Then the expected wealth of the agent at the end of period can be characterised as

$$p\Big[w(1+r)+m_1\Big]+q\Big[w(1+r)+m_2\Big]+(1-p-q)\Big[w(1+r)+m_3\Big].$$

The distribution of probabilities in this example implies that only one state above and two below the current state are accessible. Suppose that the agent can only hold a quantity of money that is either less than m_1, or greater than m_3. If the agent has a cash level of m_1 at the end of the period he will need to convert money into the interest earning asset. On the other hand if he has m_3 he will need to convert some of the illiquid asset into cash. If the agent has m_2 he may or may not wish to draw on his illiquid asset. But clearly he faces an increased chance of hitting m_3 in the following time period.

Let the agent convert a quantity of money m_a into the illiquid asset in state 1, and in states 2 and 3 withdraw amounts m_b and m_c respectively. The expected cost of these transfers are given by the function $C(m_a, m_b, m_c)$ where $C(.)$ is an increasing function of its arguments. The agent's expected payoff will be

$$\begin{aligned} p\Big[(w + m_a)(1 + r) + (m_1 - m_a) - C(m_a, m_b, m_c)\Big] \\ +q\Big[(w - m_b)(1 + r) + (m_2 + m_b) - C(m_a, m_b, m_c)\Big] \\ +(1 - p - q)\Big[(w - m_c)(1 + r) + (m_3 + m_c) - C(m_a, m_b, m_c)\Big]. \end{aligned}$$

If the agent wishes to maximise his expected utility, he faces the problem

$$\begin{aligned} \max_{m_a, m_b, m_c} pU\Big(\Big[(w + m_a)(1 + r) + (m_1 - m_a) - C(m_a, m_b, m_c)\Big]\Big) \\ +qU\Big(\Big[(w - m_b)(1 + r) + (m_2 + m_b) - C(m_a, m_b, m_c)\Big]\Big) \\ +(1 - p - q)U\Big(\Big[(w - m_c)(1 + r) + (m_3 + m_c) - C(m_a, m_b, m_c)\Big]\Big). \end{aligned}$$

If the return on the illiquid asset r is normalised to zero, the problem becomes

$$\begin{aligned} \max_{m_a, m_b, m_c} pU\Big(\Big[(w + m_a) + (m_1 - m_a)(1 - r) - C(m_a, m_b, m_c)\Big]\Big) \\ +qU\Big(\Big[(w - m_b) + (m_2 + m_b)(1 - r) - C(m_a, m_b, m_c)\Big]\Big) \\ +(1 - p - q)U\Big(\Big[(w - m_c) + (m_3 + m_c)(1 - r) - C(m_a, m_b, m_c)\Big]\Big). \end{aligned}$$

If the agent is risk neutral, it can be easily be seen that the first order conditions for maximising the above equation with respect to m_a, m_b and m_c will be same as those obtained from

$$\max_{m_a, m_b, m_c} p(rm_a) - q(rm_b) - (1 - p - q)(rm_c) - C(m_a, m_b, m_c),$$

or

$$\min_{m_a,m_b,m_c} -p(rm_a) + q(rm_b) + (1-p-q)(rm_c) + C(m_a, m_b, m_c).$$

This is a straight forward cost minimisation exercise, similar to the Martingale framework utilised by Miller-Orr. However, if the agent is a risk-averter, then it is easily seen that w explicitly enters into the first-order optimising condition similar to liquidity preference models. In fact the only difference between this example and the liquidity preference models discussed earlier, other than the obvious simplification, is that the risky asset here is cash. In the liquidity preference models the interest earning illiquid asset is the risky asset.

Undoubtedly this analysis demonstrates that cost minimisation stochastic transaction money demand models not only analyse the effects of cash shocks, but also implicitly derive the optimal portfolio choice of risk neutral agents. They also contain a wealth constraint, which enters through the specification of the interest rate and the shocks to cash. Effectively transaction money demand models of the Miller-Orr type and the liquidity preference models pioneered by Tobin (1958) are two sides of the same coin. The first difference lies in the transfer of risk from the illiquid interest earning asset in liquidity preference models to cash in Miller-Orr. The second is that the latter looks at risk neutral agents, whereas the former considers risk aversion.

4.3 The Money Demand Model

Each agent is assumed to be endowed with an initial level of wealth, which consists of an illiquid asset w and cash m. w earns a return in the form of interest, which could be negative, whereas m does not. The objective of the agent is to maximise

$$\int_0^{\infty} e^{-\rho t} U(w_t + m_t[1 - (g + r_t)H(-m_t)])\, dt,$$

where $U(.)$ is an increasing and strictly concave Von Neumann-Morgenstern utility function satisfying the conditions $U(0) = 0$, and $U'(\infty) = 0$. r_t is the interest rate, g is the constant overdraft premium, and $H(-m_t)$ is a Heavyside step function which helps capture the overdraft charges.

Interest rates are assumed to follow a Poissonian Birth-Death type process whose probabilistic evolution is given by

$$\frac{dP_{r_t}(t)}{dt} = \lambda(r_t - 1)P_{r-1}(t) - \{\lambda(r_t) + \gamma(r_t)\}P_{r_t}(t) + \gamma(r_t + 1)P_{r_t+1}(t). \quad (4.1)$$

This implies that at time t if the interest rate is r_t ($r_t = ..., 1, 2, ...$), then the probability of transition $r_t \to r_t + 1$ in the infinitesimally small time interval $(t + dt)$ is given by $\lambda(r_t)dt + o(dt)$, where $o(dt)$ contains higher order terms of dt. Similarly, if at time t the interest rate is r_t ($r_t = ..., 1, 2, ...$), the probability of the transition $r_t \to r_t - 1$ in the interval $(t + dt)$ is $\gamma(r_t)dt + o(dt)$. The probability of a transition to any other state other than a neighbouring state is $o(dt)$. The probability of interest rates remaining constant is $1 - \{\lambda(r_t) + \gamma(r_t)\}dt + o(dt)$. It follows that if the evolution of the transition density function for interest rates is given by (4.1), it can be shown that expected interest rates evolve according to the law

$$\frac{dr_t}{dt} = \lambda(r_t) - \gamma(r_t) \qquad r_0 = r. \quad (4.2)$$

See Gardiner (1985. pp. 237-238) for a derivation of (4.2).

Net cash flows follow a Wiener process with non-constant coefficients, i.e.

$$dm_t = \mu(m_t + w_t)dt + \sigma(m_t + w_t)dz_t \qquad m_0 = m \qquad m_t \in (a, b). \quad (4.3)$$

The assumption of non-constant coefficient is driven by the feature that net cash flows at each point in time are dependent on the level of wealth $(m_t + w_t)$ agents hold. The inconvenience losses which arise from cash asset transfers are

$$B(\mathbf{u}) = \begin{cases} K^+ + k^+\mathbf{u} & \text{if } \mathbf{u} \geqq 0 \\ K^- + k^-\mathbf{u} & \text{if } \mathbf{u} \leqq 0, \end{cases}$$

where

$$\mathbf{u} = \begin{cases} b - u & \text{if } x_t = b, \\ a - l & \text{if } x_t = a. \end{cases}$$

The vectors $(b - u)$ and $(a - l)$ are the size of cash transfers at the upper and lower boundaries respectively, and K^+, K^-, k^+, k^- are constants. The loss in utility arising from these transfers is $U(B(\mathbf{u}))$ since $B(\mathbf{u})$ is an unrecoverable outflow of wealth. Obviously if $E[r_t] \leqq 0$ agents will hold all their wealth as m to avoid incurring unrecoverable transfer costs. It can be seen easily seen that in this case (4.3) reduces to a geometric Wiener process. It is obvious that w will implicitly depend on the control vector and the state variable. However, it only enters the HJB equation through its inhomogeneous term, that is the instantaneous utility function U. Equation (4.3) has been formulated to eliminate the possibility that the total wealth of agents will be negative since a negative wealth level is precluded by some utility functions. In others it yields economically unacceptable solutions.

The objective of the agent is to arrive at a policy

$$\mathbf{p} = \{\tau_1, \mathbf{u}_1; \tau_2, \mathbf{u}_2; \cdots; \tau_i, \mathbf{u}_i; \cdots\},$$

of a series of stopping times and transfers which maximise his infinite horizon utility subject to cash transfers $\mathbf{u}$, i.e.

$$\begin{aligned} \underset{sup\mathbf{u}}{V(m, r; \mathbf{u})} \quad &= \quad E\left[\int_0^\infty e^{-\rho t} U(w_t + m_t[1 - (g + r_t)H(m_t)])\, dt \right. \\ &\left. + \sum_{i=0}^{\infty} e^{-\rho \tau_i} U(B(\mathbf{u}_i)) \middle| m, r, 0\right], \end{aligned} \tag{4.4}$$

where ρ is the rate of time preference, or the subjective discount rate which is assumed to be strictly positive.

It can be shown using Ito's Lemma that (4.4) must satisfy the Hamilton-Bellman-Jacobi equation

$$\begin{aligned} \rho V(m, r; \mathbf{u}) - \frac{\sigma^2 (w + m)^2}{2} V_{mm}(m, r; \mathbf{u}) - \mu(w + m) V_m(m, r; \mathbf{u}) \quad \quad (4.5) \\ -(\lambda(r) - \gamma(r)) V_r(m, r; \mathbf{u}) - U(w + m[1 - (g + r)H(-m)]) = 0. \end{aligned}$$

In obtaining (4.5) it is assumed, as with any other stochastic control problem, that the probability transacting in the first infinitesimally small time interval $[0, dt]$ is zero. Hence the absence of any transactions costs in the inhomogeneous term $U(w + m[1 - (g + r)H(-m)])$.

4.4 The Initial Value Problem

Equation (4.5) is a parabolic partial differential equation. To solve it, the behaviour of $V(.)$ needs to be explicitly specified when $r = 0$. Obtaining a profile of $V(.)$ when $r = 0$ is referred to as the initial value problem. Of course, this profile will be independent of the dynamics of the problem and will only be given by economic arguments.

As stated earlier, if the rate of return on w is either zero or less, agents will store all wealth in m to avoid unrecoverable wealth outflows in the form of transaction costs, and negative rates of return on w. Since $w = 0$, r will cease to influence $V(.)$. Obtaining this $V(.)$ will provide the solution to the initial value problem.

The Wiener process governing net cash flows now becomes

$$dM_t = \mu M_t dt + \sigma M_t dz_t \qquad M_0 = M \qquad M_t \in [0, \infty), \tag{4.6}$$

where $M = w + m$. The infinite horizon utility function will be

$$V(M) = E\left[\int_0^\infty e^{-\rho t} U(M_t)\, dt \middle| M, 0 \right]. \tag{4.7}$$

Equation (4.7) gives the expected utility over an infinite horizon when cash (wealth) flows are specified by (4.6). Naturally (4.7) ceases to be a control problem because both opportunity and transfer costs vanish.

Expanding (4.7) in a Taylor series using Ito's Lemma yields

$$\rho V(M) - \frac{\sigma^2}{2} M^2 V_{MM}(M) - \mu M V(M) - U(M) = 0. \tag{4.8}$$

Solving (4.8) will give the profile of $V(M)$ when $r = 0$.

Equation (4.8) is a differential equation of the Cauchy-Euler type and can be shown to have a solution of the form

$$\begin{aligned} -V(M) \quad = \quad & AM^{\alpha_2} + BM^{\alpha_1} \\ & + \frac{2M^{\alpha_2}}{\sigma^2(\alpha_2 - \alpha_1)} \int_0^M U(y) y^{-(1+\alpha_2)}\, dy \\ & + \frac{2M^{\alpha_1}}{\sigma^2(\alpha_2 - \alpha_1)} \int_M^\infty U(x) x^{-(1+\alpha_1)}\, dx, \end{aligned} \tag{4.9}$$

See Appendix G for the derivation of (4.9).

Critical to this problem is the behaviour at $V(M)\Big|_{M=0}$. It is obvious that $U(M)\Big|_{M=0} = 0$ by the defining property of a utility function. Since the evolution of M is given by a geometric Weiner process, if initially $M = 0$, then $M_t = 0$. As a result it is clear that $V(M)\Big|_{M=0} = 0$. Effectively if an agent is endowed with a wealth level of zero, the geometric Wiener process constrains both his wealth and utility over the infinite horizon to be zero. Equation (4.9) needs to have this property to be economically consistent.

It can be demonstrated that

$$\lim_{M\to 0} \frac{2M^{\alpha_2}}{\sigma^2(\alpha_2-\alpha_1)} \int_0^M U(y)y^{-(1+\alpha_2)}\,dy = 0,$$

and

$$\lim_{M\to 0} \frac{2M^{\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)} \int_M^\infty U(x)x^{-(1+\alpha_1)}\,dx = 0.$$

Since $U(M)\Big|_{M=0} = 0$, then provided that $A \equiv 0$ it is clear that

$$\lim_{M\to 0^+} V(M) = 0.$$

See Appendix H for the proof. As a result (4.9) now becomes

$$\begin{aligned} -V(M) \quad &= \quad BM^{\alpha_1} + \frac{2M^{\alpha_2}}{\sigma^2(\alpha_2-\alpha_1)} \int_0^M U(y)y^{-(1+\alpha_2)}\,dy \qquad (4.10) \\ &\quad + \frac{2M^{\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)} \int_M^\infty U(x)x^{-(1+\alpha_1)}\,dx, \end{aligned}$$

In order for $V(M)$ to be congruous with the assumed attributes of $U(M)$, it must also satisfy the condition $\lim_{M\to\infty} V'(M) = 0$, i.e. if an agent is initially endowed with infinite wealth, then his marginal infinite horizon utility with respect to his initial endowment must be zero. This intuitively follows from the property that $\lim_{M\to\infty} U'(M) = 0$.

Differentiating (4.10) with respect to M it is clear that

$$\begin{aligned} -V'(M) &= B\alpha_1 M^{\alpha_1 - 1} + \frac{2\alpha_2 M^{\alpha_2 - 1}}{\sigma^2(\alpha_2 - \alpha_1)} \int_0^M U(y) y^{-(1+\alpha_2)}\, dy \\ &\quad + \frac{2\alpha_1 M^{\alpha_1 - 1}}{\sigma^2(\alpha_2 - \alpha_1} \int_M^\infty U(y) y^{-(1+\alpha_1)}\, dy. \end{aligned}$$

See Appendix I for intermediate steps. It can be demonstrated that

$$\lim_{M\to\infty} M^{\alpha_2 - 1} \int_0^M U(y) y^{-(1+\alpha_2)}\, dy = 0,$$

and

$$\lim_{M\to\infty} M^{\alpha_1 - 1} \int_M^\infty U(y) y^{-(1+\alpha_1)}\, dy = 0.$$

See Appendix J for intermediate steps. Hence

$$\lim_{M\to\infty} V'(M) = B\alpha_1 \lim_{M\to\infty} M^{\alpha_1 - 1}.$$

Setting $B = 0$ will yield

$$\lim_{M\to\infty} V'(M) = 0.$$

This satisfies the marginal utility condition imposed on the utility function. Thus

$$-V(M) = \frac{2}{\sigma^2(\alpha_2 - \alpha_1)} \left[M^{\alpha_2} \int_0^M U(y) y^{-(1+\alpha_2)}\, dy + M^{\alpha_1} \int_M^\infty U(y) y^{-(1+\alpha_1)}\, dy \right].$$

Now let $\alpha_2 - \alpha_1 = -2C$. Then the solution to (4.8) satisfying the properties $V(M)\Big|_{M=0} = 0$ and $\lim_{M\to\infty} V'(M) = 0$ will be

$$\begin{aligned} V(M) &= \frac{1}{\sigma^2 C} \left[M^{\alpha_2} \int_0^M U(y) y^{-(1+\alpha_2)}\, dy \right. \\ &\quad \left. + M^{\alpha_1} \int_M^\infty U(y) y^{-(1+\alpha_1)}\, dy \right]. \end{aligned} \tag{4.11}$$

For (4.11) to be consistent with the conditions imposed on $U(M)$, it also needs to have the property $\lim_{M\to\infty} V(M) = \infty$. This follows from the choice theoretic assumption $\lim_{M\to\infty} U(M) = \infty$. It requires that, if an agent is endowed with an infinite quantity of wealth his instantaneous utility at time zero will be unbounded. Since an infinite wealth level cannot ever be exhausted to a finite quantity, even over an infinite time horizon, it follows that the infinite horizon utility function must also be unbounded. It can be shown that (4.11) satisfies this condition, and is therefore the complete solution (4.8). See Appendix K for the proof of $\lim_{M\to\infty} V(M) = \infty$.

Equation (4.11) can be further simplified into a form that will make it amenable to numerical evaluation. Let $y = Mx$ in both integrals of (4.11). Then $V(M)$ simplifies to

$$\begin{aligned} V(M) &= \frac{1}{\sigma^2 C}\Big[M^{\alpha_2}\int_0^1 U(Mx)M^{-(1+\alpha_1)}x^{-(1+\alpha_2)}M\,dx \\ &\quad +M^{\alpha_2}\int_1^\infty U(Mx)M^{-(1+\alpha_1)}x^{-(1+\alpha_1}M\,dx\Big] \\ &= \frac{1}{\sigma^2 C}\left[\int_0^1 U(Mx)x^{-(1+\alpha_2)}\,dx + \int_1^\infty U(Mx)x^{-(1+\alpha_1)}\,dx\right] \end{aligned}$$

Clearly

$$\begin{aligned} \int_0^1 U(Mx)x^{-(1+\alpha_2)}\,dx &= \left[\frac{-U(Mx)x^{-\alpha^2}}{\alpha_2}\right]_0^1 + \int_0^1 \frac{MU'(Mx)x^{\alpha^2}}{\alpha^2}\,dx \\ &= \frac{-U(M)}{\alpha_2} + \frac{M}{\alpha_2}\int_0^1 U'(Mx)x^{-\alpha_2}\,dx. \end{aligned}$$

It is also obvious that

$$\begin{aligned} \int_1^\infty U(Mx)x^{-(1+\alpha_1)}\,dx &= \left[\frac{-U(Mx)x^{-\alpha^1}}{\alpha_1}\right]_1^\infty + \int_1^\infty \frac{MU'(Mx)x^{\alpha^1}}{\alpha^1}\,dx \\ &= \frac{U(M)}{\alpha_1} + \frac{M}{\alpha_1}\int_1^\infty U'(Mx)x^{-\alpha_1}\,dx. \end{aligned}$$

Hence, $V(M)$ can be further reduced to

$$V(M) = \frac{1}{\sigma^2 C}\Big[\frac{-U(M)}{\alpha_2} + \frac{M}{\alpha_2}\int_0^1 U'(Mx)x^{-\alpha_2}\,dx$$

$$+\frac{U(M)}{\alpha_1}+\frac{M}{\alpha_1}\int_1^{\infty}U'(Mx)x^{-\alpha_1}\,dx\Bigg].$$

In the latter integral on the the R.H.S., setting $x = 1/z$ yields

$$\begin{aligned}\int_1^{\infty}U'(Mx)x^{-\alpha_1}\,dx &= \int_1^0 U\left(\frac{M}{z}\right)z^{\alpha_1}\left(\frac{-dz}{z^2}\right)\\ &= \int_0^1 z^{\alpha_1-2}U'\left(\frac{M}{z}\right)dz.\end{aligned}$$

It is also clear that

$$\begin{aligned}\frac{1}{\alpha_1}-\frac{1}{\alpha_2} &= \frac{\alpha_2-\alpha_1}{\alpha_1\alpha_2}\\ &= \frac{C\sigma^2}{\rho}.\end{aligned}$$

Therefore $V(M)$ can expressed as

$$V(M)=\frac{1}{\sigma^2 C}\left[\frac{U(M)C\sigma^2}{\rho}+\frac{M}{\alpha_2}\int_0^1 U'(Mx)x^{-\alpha_2}\,dx+\frac{M}{\alpha_1}\int_0^1 z^{\alpha_1-2}U'\left(\frac{M}{z}\right)dz\right].$$

Letting $v = M/z$ it can be shown that

$$M\int_0^1 z^{\alpha_1-2}U'\left(\frac{M}{z}\right)dz = \frac{1}{\alpha_1}MU'(M)+\frac{M}{\alpha_1}\int_0^1 z^{\alpha_1-2}\Big(vU'(v)\Big)'\,dz.$$

See Appendix L for intermediate steps. Therefore the reduced form of the solution to the initial value problem becomes

$$\begin{aligned}V(M) &= \frac{U(M)}{\rho}+\frac{1}{\sigma^2 C}\Bigg[\frac{M}{\alpha_2}\int_0^1 U'(Mx)x^{-\alpha_2}\,dx\\ &\quad +\frac{MU'(M)}{\alpha_1^2}+\frac{M}{\alpha_1^2}\int_0^1 z^{\alpha_1-2}\Big(vU'(v)\Big)'\,dz\Bigg]. \qquad (4.12)\end{aligned}$$

Equation (4.12) is the exact expression to compute the profile for the infinite horizon utility function for specific utility functions for given levels of M when $r = 0$.

4.5 Numerical Solutions

Many functional forms could be substituted for $U(M)$. Here it assumed that

$$U(M) = \ln(1 + M), \tag{4.13}$$

where $\ln(1 + M)$ is a constant relative risk aversion utility function. Now, from (4.13) it is clear that

$$\begin{aligned} vU' &= \frac{v}{1+v} \\ &= 1 - \frac{1}{1+v}. \end{aligned}$$

Hence

$$\Big(vU'(v)\Big)' = \frac{1}{(1+v)^2}.$$

Therefore

$$\begin{aligned} &\frac{1}{\alpha_2}\int_0^1 U'(Mx)x^{-\alpha_2}\,dx + \frac{1}{\alpha_1^2}\int_0^1 z^{\alpha_1-2}\Big(vU'(v)\Big)'\,dz \\ &\quad= \frac{1}{\alpha_2}\int_0^1 \frac{x^{-\alpha_2}}{1+Mx}\,dx + \frac{1}{\alpha_1^2}\int_0^1 \frac{z^{\alpha_1}}{z^2(1+\frac{M}{z})^2}\,dz \\ &\quad= \frac{1}{\alpha_2}\int_0^1 \frac{x^{-\alpha_2}}{1+Mx}\,dx + \frac{1}{\alpha_1^2}\int_0^1 \frac{z^{\alpha_1}}{(M+z)^2}\,dz. \end{aligned}$$

Substituting this into (4.12) will yield the solution to the initial value problem for a logarithmic utility function.

Analytical solutions of the type described in the literature would necessitate the discarding of valuable information in both (4.5) and (4.12). Usually such solutions also require the use of strong assumptions which could detract from reality. To overcome these limitations (4.5) and (4.12) shall be solved numerically. The numerical algorithm used here requires a detailed knowledge of spectral methods, integration rules, and optimisation techniques. Therefore all technical detail will be relegated to the Technical Annexures.

A description of how (4.5) is solved is as follows. An agent is endowed with some level of illiquid assets w, and cash $m \in (a, b)$, and follows a control policy of the type

$$\mathbf{u} = \begin{cases} b & \text{if } x_t = b, \\ a & \text{if } x_t = a. \end{cases}$$

The analysis is restricted to a zero threshold policy because later results confirm that a more general four parameter control policy described in the current literature is unlikely to yield any further insights. Also, the four parameter control policy forms the envelope of the zero threshold policy. Therefore a detailed analysis of $V(.)$ under a zero threshold policy will also reveal the important properties of $V(.)$ under the more general four parameter control policy.

The initial endowments enable the computation of a solution to (4.12) over the interval (M_a, M_b) where $M_a = w + a$, and $M_b = w + b$. Since (4.12) is in the form of an integral, its numerical solution is evaluated through a Simpson's rule adaptive integrator. See Technical Annex 2 and the source code in Technical Annex 6 for details of this.

The integration technique used to obtain a solution to (4.5) requires the computation of the infinite horizon utility function $V(m, r, \mathrm{u})$ and its partial derivatives $V_m(m, r, \mathrm{u})$ and $V_{mm}(m, r, \mathrm{u})$ as Chebyshev polynomials. This requires the interval $m \in (a, b)$ to be mapped into $\theta \in (-1, 1)$, and is done through the transformation

$$\theta = \frac{2(m-a)}{(b-a)} - 1, \qquad \theta = \arccos(x).$$

Next, the derivatives $V_M(M)$ and $V_{MM}(M)$ are computed from the numerical solution to (4.12) through collocation differentiation. Once this is done, and the probabilistic evolution of interest rates has been specified, i.e.

$$\lambda(r) - \gamma(r) = \begin{cases} > 0 \\ = 0 \\ < 0. \end{cases}$$

the solution to $V(m, r, \mathbf{u})$ can be computed through a fourth order Runge-Kutta integration scheme for $r > 0$ utilising the boundary conditions derived in Chapter 2. For details on the Runge-Kutta scheme and Chebyshev's polynomials see Technical Annexures 3,4 and 6. Here a Chebyshev polynomial of order twenty is used to estimate $V(m, r, \mathbf{u})$. If the computed solution of $V(m, r, \mathrm{u})$ proves to be a hill with a single maxima with respect to the boundary values (or controls) a and b, then this maxima and the corresponding values of a and b can be deduced by an optimisation routine such as BFGS.

4.6 Results

The results highlight some very interesting properties. Risk averse agents with identical risk preferences do not have homogenous money demand functions. The optimal *exercise prices* at which cash is bought and sold vary by discrete amounts, even if the underlying parameters driving asset prices are the same. As a result any one off change in a macroeconomic variable will result in several discrete adjustments being made to the money stock over a long period of time.

Let $\lambda(r) - \gamma(r)$, be given dynamically for all r and assume that r is reflected upwards when $r = 0$, effectively ensuring that $r \geqq 0$. This requires that

$$\lambda(0) = 1, \qquad \text{and} \qquad \gamma(0) = 0.$$

Furthermore, let r be reflected downwards when $r = R$, where R is the upper bound of r. Thus

$$\lambda(R) = 0, \qquad \text{and} \qquad \gamma(R) = 1.$$

If

$$\lambda(r) = \cos(5\pi r), \qquad \text{and} \qquad \gamma(r) = \sin(5\pi r),$$

then the reflecting properties described by the two conditions above ensure an equilibrium interest rate of 5 percent per year, i.e. when $\cos(5\pi r) = \sin(5\pi r)$. This also corresponds to the long term equilibrium interest rate observed in empirical studies. The properties of the coefficients are such that the further away r_t is from the equilibrium, the more rapidly it converges to the equilibrium, which is also consistent with reality. The overdraft premium is set at 2 percent annually. To ensure that distortions are not caused by differing marginal rates of substitution and transformation (exclusive of transaction costs), ρ is also set to five percent. The mean cash flow in each time period is assumed to be zero, i.e. $\mu = 0.0$. This is analogous to assuming that the agent's income stream and consumption path are in a long term equilibrium situation, and, any changes to the level of cash is caused by exogenous shocks. The standard deviation σ of the process governing net cash flows is assumed to be normalised to 0.05.

Transaction costs are assumed to be linear and asymmetric. The cost of selling cash and buying the illiquid asset is assumed to be less than the cost

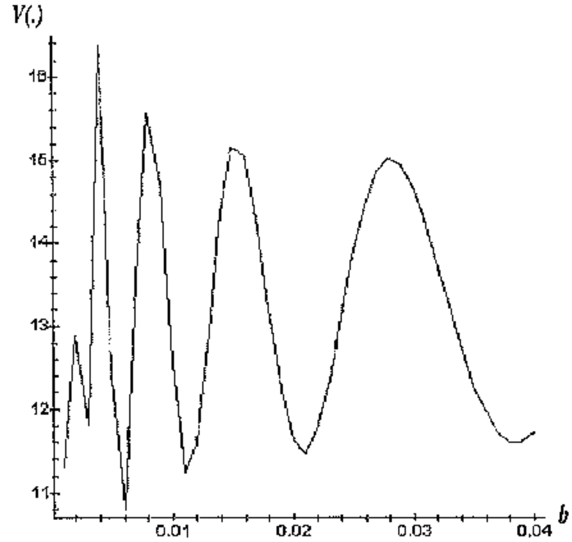

Figure 4.1: $\lambda(r) - \gamma(r)$ Dynamically Given

of selling the illiquid asset and buying cash, which is in line with what is observed in the real world where the cost of buying cash is much higher than the costs of selling cash. The proportional cost of selling m_t for w_t, and the cost of selling w_t into m_t are assumed to be 1 and 3 percent of the size of the transaction respectively. The fixed cost component of transaction costs is assumed to be symmetric and is set at 0.001 percent of the value of the initial portfolio. This is a relative quantity based on the fact that w has been normalised to be 1.0. It also corresponds to reality where the fixed costs of adjusting a portfolio of assets are very small indeed. Initial cash holdings m are assumed to be zero. These endowments are given exogenously and therefore can be arbitrarily specified. In Figure 4.1 a cross section of the value function $V(m, r, \mathrm{u})$ is taken at $a = -0.02$ and is plotted against values for $b \in [0.001, 0.04]$. Clearly $V(m, r, \mathrm{u})$ is undulating, and an optimising routine could converge to any of the local optima. Significantly these optima are not lumped together in a localised region, but are spread out. The important feature driving this is the constant nature of cash deposits and withdrawals. Many parameters such as the size of the withdrawals, ρ and r also contribute towards this. If b is only slightly larger than the initial cash endowment of zero, then the agent is likely to hit the upper boundary more frequently and will be forced to endure frequent and irreversible outflows of wealth in the form of transaction costs. As b increases these outflows are likely to diminish and thus conserve the pool of wealth. As b increases even more the opportunity cost associated with tolerating a large zone of inaction will take effect penalising the agent.

Why then the second, and third hills? The answer to this lies in the difference between the rate of time preference and the actual rate at which wealth grows. The key is to understand the intertemporal dynamics of $V(m, r, \mathbf{u})$.

If the agent is risk neutral, it is easily observed that the trade-off between the MRS and the expected MRT of utility will be of a linear nature. This follows naturally from the linearity of a risk neutral utility function. In Section 4.2 it was demonstrated that utility maximisation by a risk neutral agent is equivalent to cost minimisation. This implies that the MRT of a risk neutral agent will be independent of his level of wealth, and, only depend on the rate at which costs evolve. In this problem holding costs accrue at a rate proportional to the expected interest rate. Transaction costs accumulate at the rate at which the Wiener process exits the continuation region through the boundaries a and b. Therefore, it intuitively follows that a linear transformation of the net present value of the sum of these costs will be minimised for a choice of a and b at the turning point of a convex hull.

However, in situations where the agent is risk averse, the concavity of the utility function imposes a non-linear relationship between the MRS and MRT. Furthermore, it is easily seen that the MRT will explicitly depend on wealth holdings. As wealth varies the MRT will be either larger, equal to, or, smaller than the MRS. It is this feature which causes the aperiodic fluctuations in $V(.)$. Clearly, if wealth is large, $U(.)$ will also be large. Also, $V(.)$ will be larger when $U(.)$ is large in earlier time periods rather than in later time periods. This is an obvious effect of the rate of time preference ρ. If the agent fixes his upper boundary beyond the first optimum value of b, then his marginal rate of substitution (MRS) will be greater than his marginal rate of transformation (MRT), i.e. $MRS > MRT$ (inclusive of transaction costs). This is because the gain in $V(.)$ due holding a certain level of illiquid assets and excess cash in earlier time periods is less than offset by the gain in utility in later periods resulting from the growth in wealth caused by switching some of the excess cash into illiquid assets. Effectively, the effect of ρ dominates that of r_t causing $V(.)$ to increase again. The transfer of excess cash into the illiquid asset results in unrecoverable losses in the form of transaction costs. To offset this loss the MRT must exceed the MRS. This may require a relationship to be specified between the rate of time preference ρ, the equilibrium rate of interest, and also the nature of transaction costs. But, making this link is difficult. Even if such a relationship was specified it may only serve to dampen the amplitude and change the period of the infinite horizon utility function. For a link to be made one would need to model the complicated feedback relationship between transaction costs, r_t and w_t which would re-

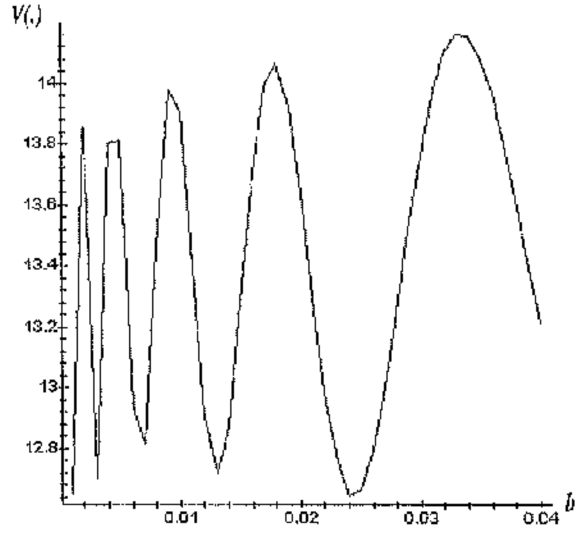

Figure 4.2: $\lambda(r) - \gamma(r) = -1.0$

sult in a unique optima, and, then try and establish some kind of link with ρ. The concavity of $U(.)$ would almost certainly require such a problem to be numerically solved. To simplify the problem it is much easier to make an empirical link between the four parameters as has been done here. The concave nature of the utility function is clearly the major factor driving the oscillatory nature of the solution. If state varying controls were used where a and b were allowed to dynamically evolve rather than being held constant, it could be the case that these hills vanish. However, again, there is no *a priori* reason to believe this.

In the previous example it was assumed that interest rates converged to an equilibrium rate of 5 percent per year. Here it is assumed that $\lambda(r) - \gamma(r) = -1.0$. As a result expected interest rates follow a downward course, and in the long run push the agent towards holding all his wealth in cash with a profile of $V(.)$ similar to the initial value problem. However, it can be still seen that $V(.)$ is undulating due to the various points at which the substitution and transformation dominate each other. The peaks observed in Figure 4.2 are also of a smaller amplitude and a larger period than in the previous example. This is because, the variance associated with the rate of return on illiquid assets is much less here, and, therefore does not expose the agent to the same degree of wealth volatility as before. It can be clearly seen that the optimal values of waiting increase as the thresholds increase. This follows from the property that, as time increases, expected interest rates will become negative and unbounded. Agents will choose not to opt for small targets because they will not be able to recover the frequent transfer costs incurred through any interest income they may earn. Although this situation is not likely, it confirms the validity of the previous set of results and the underlying intuition of the modelling approach used here.

Intuitively it sounds plausible that increasing the discount rate will front load the problem. This would imply that, for the same utility function, both the value of of $V(.)$ and the amplitude of the aperiodic fluctuations observed in the two previous examples must decrease. Indeed, this is exactly what is observed. To make this feature obvious an annual discount rate of 100 percent was chosen. The dampening effect of this can be clearly seen in the Figure 4.3.

Diminished risk sensitivity should yield an optimal region that exhibits less volatility. Figure 4.4 plots the profile of $V(.)$ against b for another utility function with constant relative risk aversion. $U(.)$ is an exponential of the

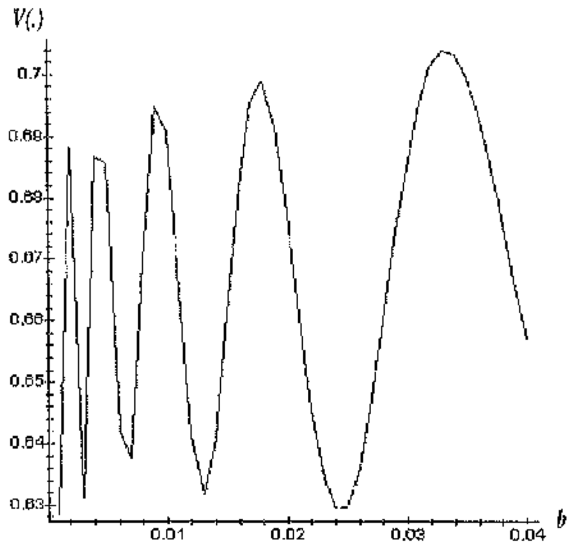

Figure 4.3: $\rho = 1.0$

type $U(x) = x^{\psi}$, where $\psi = 0.9$. Risk aversion requires $0 < \psi < 1$. The agent exhibits less risk sensitivity as $\psi \to 1$. The results clearly confirm that stability increases with decreasing risk aversion. Thus reinforcing the validity of the modelling approach used here.

These findings shed a new light on how agents behave. The existence of multiple optima clearly demonstrates that the optimal value of waiting for going long or short on cash discretely varies, even with homogenous risk preferences and constant parameters. It also gives rise to a series discretely varying lags between a one off change in a macroeconomic variable and the money stock being adjusted. Significantly, the aggregate money demand function will exhibit discreet jumps over the different triggers at which agents choose to exercise their option to go long or short on cash.

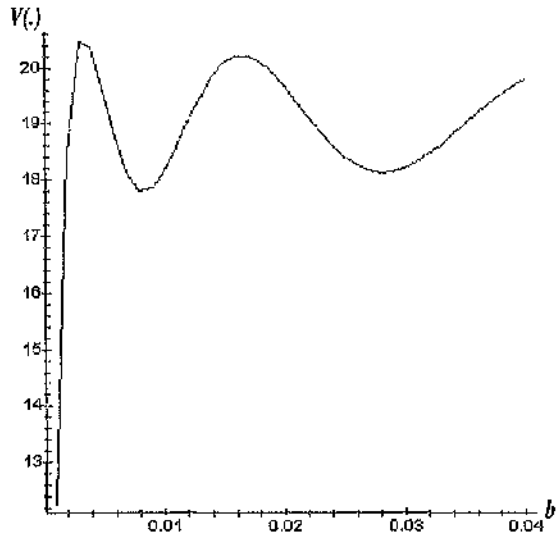

Figure 4.4: $U(x) = x^{\psi}$

In most rational expectations models the existence of multiple equilibria is associated with either unstable or saddle point solutions (in higher dimension models). Most of these solutions, for example the divergent dynamic path specified by the complex roots of an ordinary differential in a standard two dimensional macroeconomic model can be dismissed through a partial equilibrium argument. Of course these require extremely strong assumptions that rely on a degree of foresight and rationality that is unobservable in practice. In contrast, none of the observed optima here are unstable, and therefore cannot be dismissed through a partial equilibrium argument. What is exactly the *fundamental solution* here is not clear since all of them share the unique feature that the marginal utility of the infinite horizon utility function with respect to a boundary is stationary.

If agents learn by doing, does this ensure that agents converge to the global optimum? Rational expectations models have studied the possibility of agents converging to an unstable solution under simple linear learning models. De Canio (1979) and Evans (1985) argue that agents use observations over a finite period say T to estimate the parameters of a system. They then use this for another period of duration T after which they recompute the parameters again. Bray (1982) assumes that agents recursively estimate the parameters driving the system each period through a least squares method. The outcome of both these techniques is that agents converge to the fundamental solution as they continue to refine their estimates of the parameters. Of course in this model there is no learning to be done because the exact values of the exogenous parameters driving the system are assumed to be known with perfect foresight.

If the multiple equilibria observed here cannot be dismissed through learning or by using stability arguments, how does an agent converge to a global optima? If all optima share the same property, that is the marginal utility of the infinite horizon utility function with respect to a boundary is stationary, even the most advanced optimisation routines such as the Quasi-Newtonian BFGS technique will not be capable of distinguishing a global optimum from local optima. The only way in which one could arrive at the global optimum would be to evaluate the infinite horizon utility function for all possible values of boundaries, which is clearly unbounded, and then use some kind grid search technique. Of course a grid search technique is an ad-hoc method by which the value of one optima is compared with the value of another. If the utility function is not evaluated for all possible values of boundaries and thresholds, it could be the case that a grid search method only returns a local optima. A multiplicity of optima in a utility function could make this a very costly technique to use. *In fact its ad-hoc non-scientific nature, its high cost, and the near impossibility of pinning down a global optima reinforces the notion that agents are most likely to converge to the local optima that is most accessible to them.*

4.7 Concluding Remarks

In this model it has been demonstrated that, under time invariant controls, the MRS and MRT effect alternatively dominate each other over certain ranges of the control vector yielding solutions with multiple optima. This shows that the optimal value of waiting for buying and selling cash discretely varies, despite all other parameters and risk preferences being held constant. The aggregate demand function for a population of homogenous agents will not converge to the well behaved functional forms hypothesised in the preceding literature. If agents converge to different optima, the full effect of an exogenous shock may not be felt all at once as the current literature suggests, due to all agents adjusting their targets simultaneously, but take effect slowly at staggered time intervals. Its full effect taking time to work through the whole economy. This also sounds intuitively correct. In earlier models this effect could be only be explained by assuming that initial endowments were heterogonously distributed among agents. This model provides an explanation of the slow adjustment of targets by showing that agents, within a homogenous framework, may *optimally* vary the extent to which they hedge their risk.

Naturally, the assumptions made here could be further expanded by utilising state varying controls, but in the absence of a theoretical basis providing the natural boundary conditions for such a tool, the results obtained here could prove to be the most accurate approximation of the syntheses between the inventory theoretic approach to modelling the transactions demand for money, and the risk sensitive wealth maximisation approach to modelling.

Chapter 5

Conclusions

This thesis presents three important results, and, a critique of the existing inventory theoretic money demand models which sheds a completely different light on them. Chapter 1 provides an analysis of the heuristically motivated "smooth pasting" condition which is used in stochastic "impluse" control models as an optimisation tool and derives the natural boundary conditions for solving such problems. The necessary first-order optimisation conditions are also discussed. Chapter 2 deals with a simple application of the "smooth pasting" condition, highlighting some of its shortcomings. Unique insights into the Dixit menu cost model are obtained. Also previously assumed properties are proven. Chapter 3 analyses the strengths and weaknesses of the existing body of literature on the transactions demand for money. Key models are dissembled and critically analysed. Some of them which were previously not subject to the same degree of scrutiny, now do not hold up. In Chapter 4 a more robust and logically sound alternative to the existing approach is presented. It tries to reconcile the two different findings of the empirical literature; that is the lagged adjustment of the money stock to changes in other variables, and, the perceived instability of the money demand equation. Concepts from the existing transaction money demand models are gelled together with those from liquidity preference models to obtain a pattern of optimising behaviour which goes against the 'neat' results obtained in the current literature.

5.1 Chapter 1

The results in this chapter can be summarised as follows. "Smooth pasting" condition is a heuristically motivated condition which is absent from the stochastic optimal control framework for dealing with "impulse" control problems. This chapter provides a more rigorous approach to solving such problems. However, later results in Chapter 2 confirm that both techniques yield the same strategy. The results also confirm the "value matching" condition as being the natural boundary condition for "impulse" controlled problems. It demonstrates that the value of stopping at a state and exercising an "impulse" control must equal the net present value of holding costs accrued up to that state. This also sounds intuitively correct. If the total value of exercising a stopping decision exceeded the net present value of holding costs accrued until this decision was made, it would clearly be sub-optimal to stop. Conversely if the net present value of holding costs exceeded the net present value of the stopping decision, it would imply that the stopping decision should have been taken earlier. Optimisation with respect to the set of admissible controls occurs as with any other stochastic optimal control problem.

5.2 Chapter 2

A solution to the Dixit menu cost model is obtained using the "value matching" and natural first-order optimisation conditions. *Previously assumed properties such as symmetricity and the zero threshold policy are now proven to be optimal.* A link between the intertemporal discount rate and the zone of inertia is derived. This could not be deduced from Dixit (1991a) unless an empirical link between variance of the Wiener process and the discount rate is specified. Also costs faced by a firm initially lying outside optimal zone are quantified, providing firms with a clear incentive to transact down to zero. The benefits of following a partial price adjustment policy in cases where the costs faced by firms are different is also provided. From a technical point of view, a more accurate estimate of the zone of inertia is obtained. The simplifying expansion used in Dixit (1991a) is less accurate.

5.3 Chapter 3

The key contributions in existing body of literature on the transactions demand for money is surveyed in this chapter. Each model is scrutinised in detail highlighting its strengths and weaknesses. The logical foundations of some models are shown to be not robust as perceived before. The mathematical analysis used is clearly questionable. Those models which appear to be robust, analyse agent behaviour under restrictive conditions, e.g. the deterministic and steady state model of Baumol-Tobin, or the steady state model of Miller-Orr.

The 'neat' results which characterise most models is also compared with the empirical literature. The general consensus is that although a majority of empirical models find evidence which supports the current target threshold modelling approach, a significant minority finds that the money demand function is unstable. This goes against the stable behaviour forecasted by current models. A promising direction in which the the current body of literature on the transactions demand for money could evolve is discussed in the final section.

5.4 Chapter 4

Chapter 4 returns to the initial objective of this thesis. The money demand problem for a risk averter is solved. The mean reverting diffusion process used to capture interest rate variations is replaced with a more realistic Poissonnian jump stochastic process. Spectral methods and numerical integrations schemes such as Simpson's rule and Runge Kutta 4 are introduced for solving the HJB equation for the first time. The findings here significantly differ from preceding models. The key conclusion is the existence of multiple optima, which has interesting implications for the money demand function and implicitly for the demand illiquid assets. This is in contrast to the static liquidity preference models in which intertemporal effects are not considered. However, unlike other rational expectations models, *these optima cannot be dismissed as being bubble solutions.* Learning by agents or the use of ad-hoc 'global' optimising routines also do not discount this possibility.

The multiple optima results from the MRS and MRT effects alternatively dominating each other over alternating ranges of the control vector. This demonstrates that *ceteris paribus* the optimal value of waiting for buying and selling cash discretely varies. If agents converge to different optima, the full effect of an exogenous shock may not be felt all at once, but take effect slowly, at discreetly staggered time intervals, as different agents discretely adjust their at targets varying points in time.

Appendices

Appendix A

By differentiating (1.19) with respect to x it is clear that

$$\begin{aligned} V_x(x,\mathbf{u}) &= \int_0^\infty e^{-\rho s}\left[\int_a^b M(x(s)) f_x(x(s),s|x,0)\,dx(s)\right] ds \\ &\quad -\rho[B(b,u)-V(u,\mathbf{u})]\int_0^\infty e^{-\rho s}F_x(b,s|x,0)\,ds \qquad (\text{A.1}) \\ &\quad -\rho[D(a,l)-V(l,\mathbf{u})]\int_0^\infty e^{-\rho s}F_x(a,s|x,0)\,ds. \end{aligned}$$

Differentiating (A.1) with respect to x yields

$$\begin{aligned} V_{xx}(x,\mathbf{u}) &= \int_0^\infty e^{-\rho s}\left[\int_a^b M(x(s)) f_{xx}(x(s),s|x,0)\,dx(s)\right] ds \\ &\quad -\rho[B(b,u)-V(u,\mathbf{u})]\int_0^\infty e^{-\rho s}F_{xx}(b,s|x,0)\,ds \qquad (\text{A.2}) \\ &\quad -\rho[D(a,l)-V(l,\mathbf{u})]\int_0^\infty e^{-\rho s}F_{xx}(a,s|x,0)\,ds. \end{aligned}$$

Multiplying (A.1) by μ and (A.2) by $\sigma^2/2$ and adding both gives

$$\begin{aligned} A(x)V(x,\mathbf{u}) &= \int_0^\infty e^{-\rho s}\left[\int_a^b M(x(s)) Af(x(s),s|x,0)\,dx(s)\right] ds \\ &\quad -\rho[B(b,u)-V(u,\mathbf{u})]\int_0^\infty e^{-\rho s}AF(b,s|x,0)\,ds \\ &\quad -\rho[D(a,l)-V(l,\mathbf{u})]\int_0^\infty e^{-\rho s}AF(a,s|x,0)\,ds. \end{aligned}$$

Substituting the backward Chapman-Kolmogorov equation (1.15) into the above expression yields

$$
\begin{aligned}
AV(x,\mathbf{u}) &= \int_0^\infty e^{-\rho s}\left[\int_a^b M(x(s)) f_s(x(s),s|x,0)\,dx(s)\right] ds \\
&\quad -\rho[B(b,u)-V(u,\mathbf{u})]\int_0^\infty e^{-\rho s}F_s(b,s|x,0)\,ds \\
&\quad -\rho[D(a,l)-V(l,\mathbf{u})]\int_0^\infty e^{-\rho s}F_s(a,s|x,0)\,ds. \\
&= \rho\int_a^b M(x(s))\left[\int_0^\infty e^{-\rho s}f(x(s),s|x,0)\,ds\right] dx(s) - M(x) \\
&\quad -\rho[B(b,u)-V(u,\mathbf{u})]\left[1-\rho\int_0^\infty e^{\rho s}F(b,s|x,0)\,ds\right] \\
&\quad -\rho[D(a,l)-V(l,\mathbf{u})]\left[1-\rho\int_0^\infty e^{\rho s}F(a,s|x,0)\,ds\right]. \qquad \text{(A.3)}
\end{aligned}
$$

Appendix B

Derivation of the solution of the transition density function in the forward Kolmogorov equation.

Form (1.23) we have

$$f_t(\xi,t|g,0) = -\alpha f_\xi(\xi,t|g,0) + \frac{1}{2}f_{\xi\xi}(\xi,t|g,0).$$

Using the method of separation of variables we obtain

$$f(\xi,t|g,0) = T(t)Q(\xi).$$

Substituting this into (1.23) yields

$$\frac{dT(t)}{dt}Q(\xi) = -\alpha T(t)\frac{dQ(\xi)}{d\xi} + \frac{1}{2}T(t)\frac{d^2Q(\xi)}{d\xi^2}.$$

It is clear from this expression that the solution to $T(t)$ and $Q(\xi)$ are of the form

$$T(t) = Be^{-\lambda^2 t},$$

where B is an arbitrary constant, and

$$Q(\xi) = D\sin[\sqrt{\lambda^2 - \alpha^2}(\xi + \varphi)]e^{\alpha\xi}.$$

The boundary condition $f(0, t|g, 0) = 0$ implies that $Q(0) = 0$. This is satisfied when $\varphi = 0$. We also have the other boundary condition $f(1, t|g, 0) = 0$. Setting $\xi = 1$ we obtain a value for λ in terms of the parameters of the differential equation; i.e.

$$\lambda = \sqrt{\alpha^2 + n^2\pi^2}.$$

This expression for λ gives us an Eigenfunction of the type

$$\sin(n\pi\xi)e^{\alpha\xi - \frac{1}{2}(\alpha^2+n^2\pi^2)t},$$

which is a solution to (1.23) satisfying its boundary conditions for all integers. Therefore, $f(\xi, t|g, 0)$ can be expressed as a solution to the forward Chapman-Kolmogorov equation in terms of a Fourier sine series. That is

$$f(y, s|x, 0) = \sum_{n=1}^{\infty} g_n \sin(n\pi\xi)e^{\alpha\xi - \frac{1}{2}(\alpha^2+n^2\pi^2)t}$$

where g_n is a constant. We can now from the initial condition calculate the coefficients of the series. We have

$$\begin{aligned} g_n &= 2\int_0^1 \delta(\xi - g)e^{-\alpha\xi}\sin(n\pi\xi)\,d\xi \\ &= 2e^{-\alpha g}\sin(n\pi g). \end{aligned}$$

Thus the Fourier series solution for the transition density function from the forward Chapman-Kolmogorov equation is

$$f(\xi, t|g, 0) = 2\sum_{n=1}^{\infty} e^{\alpha(\xi-g) - \frac{1}{2}(\alpha^2+n^2\pi^2)t}\sin(n\pi\xi)\sin(n\pi g)$$

The evaluation of the discounted value of the fluxes at the upper and lower boundaries over the infinite horizon

We have

$$\begin{aligned}\int_0^\infty e^{-\frac{\beta}{2}t} F_t(1,t|g,0)\,dt &= \int_0^\infty \pi \sum_{n=1}^\infty n \sin(n\pi(1-g)) e^{\alpha(1-g)-\frac{1}{2}(\beta+\alpha^2+n^2\pi^2)t}\,dt \\ &= \sum_{n=1}^\infty \frac{2n\pi \sin(n\pi(1-g)) e^{\alpha(1-g)}}{\beta+\alpha^2+n^2\pi^2}.\end{aligned}$$

We know as a fact that

$$\sum_{n=1}^\infty \frac{2n\pi \sin(n\pi(1-g))}{\beta+\alpha^2+n^2\pi^2} = \frac{\sinh \chi g}{\sinh \chi}.$$

Therefore we have

$$\int_0^\infty e^{-\frac{\beta}{2}t} F_s(1,t|g,0)\,dt = e^{\alpha(1-g)} \frac{\sinh \chi g}{\sinh \chi}$$

Similarly we can compute the rate at which the first stopping value is accrued on the lower boundary. We have

$$\begin{aligned}\int_0^\infty e^{-\frac{\beta}{2}t} F_t(0,s|g,0)\,dt &= \int_0^\infty \sum_{n=1}^\infty n\pi \sin(n\pi g) e^{-\alpha g-\frac{1}{2}(\beta+\alpha^2+n^2\pi^2)t}\,dt \\ &= \sum_{n=1}^\infty \frac{2n\pi \sin(n\pi g) e^{-\alpha g}}{\beta+\alpha^2+n^2\pi^2} \\ &= e^{-\alpha g} \frac{\sinh \chi(1-g)}{\sinh \chi}.\end{aligned}$$

The evaluation of the integral $\int_0^\infty e^{-\frac{\beta}{2}t} f(\xi,t|g,0)\,dt$

We have

$$\int_0^\infty e^{-\frac{\beta}{2}s} f(\xi,s|g,0)\,ds = \sum_{n=1}^\infty \frac{4\sin(n\pi\xi)\sin(n\pi g)}{\beta+\alpha^2+n\pi^2}.$$

We also know as a fact that

$$\sum_{n=1}^\infty \frac{2\sin(n\pi y)\sin(n\pi\xi)}{\beta+\alpha^2+n\pi^2} = \frac{\cosh\sqrt{\beta+\alpha^2}\omega - \cosh\sqrt{\beta+\alpha^2}\gamma}{\sqrt{\beta+\alpha^2}\sinh\sqrt{\beta+\alpha^2}},$$

where $\omega = 1 - |g - \xi|$, and $\gamma = 1 - g - \xi$. Now we can express the above integral as

$$\int_0^\infty e^{-\frac{\beta}{2}t} f(\xi, t|g, 0)\, dt = 2e^{\alpha(\xi-g)} \frac{\cosh \chi\omega - \cosh \chi\gamma}{\chi \sinh \chi}.$$

Appendix C

It is obvious that

$$V(0) = A + \frac{k\sigma^2}{\rho^2}.$$

substituting this into (2.1) yields

$$V(x) = \left\{ V(0) - \frac{k\sigma^2}{\rho^2} \right\} \cosh(\alpha x) + B \sinh(\alpha x) + \frac{kx^2}{\rho} + \frac{k\sigma^2}{\rho^2}. \qquad \text{(C.1)}$$

From the boundary condition on the upper boundary in (2.1) it is clear that

$$\left\{ V(0) - \frac{k\sigma^2}{\rho^2} \right\} (\cosh(\alpha b) - 1) + B \sinh(\alpha b) = g - \frac{kb^2}{\rho}. \qquad \text{(C.2)}$$

From the boundary condition in (2.1) for the lower boundary it is clear that

$$\left\{ V(0) - \frac{k\sigma^2}{\rho^2} \right\} (\cosh(\alpha a) - 1) + B \sinh(\alpha a) = g - \frac{ka^2}{\rho}. \qquad \text{(C.3)}$$

Multiplying (C.2) by $\sinh(\alpha a)$ and (C.3) by $\sinh(\alpha b)$ and then subtracting the latter from the former yields

$$\left\{ V(0) - \frac{k\sigma^2}{\rho^2} \right\} [\sinh(\alpha a)(\cosh(\alpha b) - 1) - \sinh(\alpha b)(\cosh(\alpha a) - 1)]$$
$$= \left(g - \frac{kb^2}{\rho} \right) \sinh(\alpha a) - \left(g - \frac{ka^2}{\rho} \right) \sinh(\alpha b).$$

Dividing through by $(\sinh(\alpha a) \sinh(\alpha b))$ gives

$$\left\{ V(0) - \frac{k\sigma^2}{\rho^2} \right\} \left[\tanh\left(\frac{\alpha b}{2}\right) - \tanh\left(\frac{\alpha a}{2}\right) \right]$$
$$= \left(g - \frac{kb^2}{\rho} \right) \frac{1}{\sinh(\alpha b)} - \left(g - \frac{ka^2}{\rho} \right) \frac{1}{\sinh(\alpha a)}.$$

This equation gives $V(0)$ as a function of the boundary values and the parameters of the problem. Similarly an expression for B as a function of the boundary values and the parameters of the problem can also be obtained. Dividing (C.2) by $(\cosh(\alpha b) - 1)$ and (C.3) by $(\cosh(\alpha a) - 1)$ and then subtracting the latter from the former yields

$$B\left(\coth\left(\frac{\alpha b}{2}\right) - \coth\left(\frac{\alpha a}{2}\right)\right) = \frac{g - \frac{kb^2}{\rho}}{\cosh(\alpha b) - 1} - \frac{g - \frac{ka^2}{\rho}}{\cosh(\alpha a) - 1}.$$

Simplifying the above two equations further, and substituting into (C.1) gives an expression for $V(x)$, i.e.:

$$\begin{aligned} V(x) &= \frac{kx^2}{\rho} + \frac{k\sigma^2}{\rho^2} + \frac{\left(g - \frac{kb^2}{\rho}\right)\frac{\cosh(\alpha x)}{\sinh(\alpha b)} - \left(g - \frac{ka^2}{\rho}\right)\frac{\cosh(\alpha x)}{\sinh(\alpha a)}}{\tanh\left(\frac{\alpha b}{2}\right) - \tanh\left(\frac{\alpha a}{2}\right)} \\ &\quad + \left[\frac{\left(g - \frac{kb^2}{\rho}\right)\sinh(\alpha x)}{\cosh(\alpha b) - 1} - \frac{\left(g - \frac{ka^2}{\rho}\right)\sinh(\alpha x)}{\cosh(\alpha a) - 1}\right] \\ &\quad \times \frac{\tanh\left(\frac{\alpha a}{2}\right)\tanh\left(\frac{\alpha b}{2}\right)}{\tanh\left(\frac{\alpha a}{2}\right) - \tanh\left(\frac{\alpha b}{2}\right)} \\ &= \frac{kx^2}{\rho} + \frac{k\sigma^2}{\rho^2} + \frac{\left[\frac{(g - \frac{kb^2}{\rho})\cosh(\alpha(x - \frac{a}{2}))}{2\sinh(\frac{\alpha b}{2})} - \frac{(g - \frac{ka^2}{\rho})\cosh(\alpha(x - \frac{b}{2}))}{2\sinh(\frac{\alpha a}{2})}\right]}{\sinh\left(\alpha\left(\frac{b-a}{2}\right)\right)} \end{aligned}$$

Dividing this solution for $V(x)$ by $g/2$ and substituting the non-dimensionalising parameters and variables in (2.3) yields

$$V(w) = \frac{2w^2}{\gamma} + \frac{1}{\gamma} + \frac{f(y)\cosh(2w - z) - f(z)\cosh(2w - y)}{\sinh(y - z)}.$$

Appendix D

Now that $z + y = 0$, and $f(y)$ is an odd function, we also must have z and y satisfying (8), i.e.

$$\frac{-2y}{\gamma}\sinh(2y) - f(y)\sinh(3y) - f(y)\sinh y = 0$$

Upon further simplification it is clear that

$$f(y)(\sinh y + \sinh 3y) + \frac{2y}{\gamma} = 0.$$

Therefore

$$2f(y)\sinh 2y\cosh y + \frac{2y}{\gamma}\sinh(\alpha y) = 0.$$

Dividing through by $(2\sinh(\alpha y))$ we obtain

$$\frac{y}{\gamma} + f(y)\cosh y = 0.$$

Substituting for $f(y)$ and simplifying further gives

$$y - \tanh(y) - \frac{\gamma}{y} = 0.$$

Appendix E

It is clear that

$$3\gamma - y^2\tanh^3(y) = 2y^2 + y^2\mathrm{sech}^2(y) - 3y\tanh(y).$$

From this it is obvious the when $y = 0$, the above expression is also zero.

Appendix F

The boundary conditions for the general price adjustment policy are

$$V(b) - V(u) = g \qquad \text{and} \qquad V(a) - V(l) = g,$$

where u and l are the upper and lower thresholds respectively, see Sivananthan and Lindsay (1996). Substituting these boundary conditions into (2.1) it is clear that

$$A(\cosh(\alpha b) - \cosh(\alpha u)) + B(\sinh(\alpha b) - \sinh(\alpha u)) + \frac{k}{\rho}(b^2 - u^2) = g, \quad (\text{F.1})$$

and

$$A(\cosh(\alpha a) - \cosh(\alpha l)) + B(\sinh(\alpha a) - \sinh(\alpha l)) + \frac{k}{\rho}(a^2 - l^2) = g. \quad (\text{F.2})$$

Multiplying the first equation by $(\sinh(\alpha a) - \sinh(\alpha l))$ and the second equation by $(\sinh(\alpha b) - \sinh(\alpha u))$ and then subtracting the latter from the former yields

$$\begin{aligned} A[(\cosh(\alpha b) - \cosh(\alpha u))&(\sinh(\alpha a) - \sinh(\alpha l)) \\ -(\cosh(\alpha a) - \cosh(\alpha l))&(\sinh(\alpha b) - \sinh(\alpha u))] \\ &= \left(g - \frac{k}{\rho}(b^2 - u^2)\right)(\sinh(\alpha a) - \sinh(\alpha l)) \\ &\quad - \left(g - \frac{k}{\rho}(a^2 - l^2)\right)(\sinh(\alpha b) - \sinh(\alpha u)). \end{aligned}$$

By simplifying the above equation further and making A the subject it can be seen that

$$A = \frac{\left(g - \frac{k}{\rho}(b^2 - u^2)\right) 2\sinh\left(\alpha\left(\frac{a-l}{2}\right)\right)\cosh\left(\alpha\left(\frac{a+l}{2}\right)\right)}{4A\sinh\left(\alpha\left(\frac{a-l}{2}\right)\right)\sinh\left(\alpha\left(\frac{u-b}{2}\right)\right)\sinh\left(\alpha\left(\frac{a+l-b-u}{2}\right)\right)}$$

$$
-\frac{\left(g-\frac{k}{\rho}(a^2-l^2)\right)2\sinh\left(\alpha\left(\frac{b-u}{2}\right)\right)\cosh\left(\alpha\left(\frac{b+u}{2}\right)\right)}{4A\sinh\left(\alpha\left(\frac{a-l}{2}\right)\right)\sinh\left(\alpha\left(\frac{u-b}{2}\right)\right)\sinh\left(\alpha\left(\frac{a+l-b-u}{2}\right)\right)}
$$

$$
= \frac{\frac{1}{2}\left[\left(g-\frac{k}{\rho}(b^2-u^2)\right)\frac{\cosh(\alpha(\frac{a+l}{2}))}{\sinh(\alpha(\frac{u-b}{2}))}+\left(g-\frac{k}{\rho}(a^2-l^2)\right)\frac{\cosh(\alpha(\frac{b+u}{2}))}{\sinh(\alpha(\frac{a-l}{2}))}\right]}{\sinh(\alpha(\frac{a+l-b-u}{2}))}.
$$

Similar to how A was obtained, B can also be obtained. If (F.1) is multiplied by $(\cosh(\alpha a)-\cosh(\alpha l))$ and (F.2) by $(\cosh(\alpha b)-\cosh(\alpha u))$ and then subtract the latter from the former, it is obvious that

$$
\begin{aligned}
&B[(\cosh(\alpha a)-\cosh(\alpha l))(\sinh(\alpha b)-\sinh(\alpha u))\\
&\quad-(\cosh(\alpha b)-\cosh(\alpha u))(\sinh(\alpha a)-\sinh(\alpha l))]\\
&\qquad=\left(g-\frac{k}{\rho}(b^2-u^2)\right)(\cosh(\alpha a)-\cosh(\alpha l))\\
&\qquad\quad-\left(g-\frac{k}{\rho}(a^2-l^2)\right)(\cosh(\alpha b)-\cosh(\alpha u)).
\end{aligned}
$$

Using the same approach as was used for A, it can be shown that

$$
B=\frac{\frac{1}{2}\left[\left(g-\frac{k}{\rho}(b^2-u^2)\right)\frac{\sinh(\alpha(\frac{a+l}{2}))}{\sinh(\alpha(\frac{b-u}{2}))}-\left(g-\frac{k}{\rho}(a^2-l^2)\right)\frac{2\sinh(\alpha(\frac{b+u}{2}))}{\sinh(\alpha(\frac{a-l}{2}))}\right]}{\sinh\left(\alpha\left(\frac{a+l-b-u}{2}\right)\right)}.
$$

Now that expressions for the constants A and B have been obtained, the Bellman value function can be expressed as

$$
\begin{aligned}
V(x) &= \frac{kx^2}{\rho}+\frac{k\sigma^2}{\rho^2}\\
&\quad+\frac{1}{2}\left[\left(g-\frac{k}{\rho}(b^2-u^2)\right)\left(\sinh\left(\alpha\left(\frac{a+l}{2}\right)\right)\sinh(\alpha x)\right.\right.\\
&\quad\left.\left.-\cosh\left(\alpha\left(\frac{a+l}{2}\right)\right)\cosh(\alpha x)\right)\right]\\
&\qquad\div 2\sinh\left(\alpha\left(\frac{a+l-b-u}{2}\right)\right)\sinh\left(\alpha\left(\frac{b-u}{2}\right)\right)\\
&\quad+\frac{1}{2}\left[\left(g-\frac{k}{\rho}(a^2-l^2)\right)\left(\cosh\left(\alpha\left(\frac{b+u}{2}\right)\right)\cosh(\alpha x)\right.\right.\\
&\quad\left.\left.-\cosh\left(\alpha\left(\frac{b+u}{2}\right)\right)\sinh(\alpha x)\right)\right]
\end{aligned}
$$

$$\div 2\sinh\left(\alpha\Big(\frac{a+l-b-u}{2}\Big)\right)\sinh\left(\alpha\Big(\frac{a-l}{2}\Big)\right)$$

$$= \frac{kx^2}{\rho}+\frac{k\sigma^2}{\rho^2}+\Bigg[\frac{(g-\frac{k}{\rho}(a^2-l^2))\cosh\left(\alpha\Big(\frac{b+u-2x}{2}\Big)\right)}{\sinh(\alpha(\frac{a-l}{2}))}$$

$$-\frac{(g-\frac{k}{\rho}(b^2-u^2))\cosh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)}{\sinh(\alpha(\frac{b-u}{2}))}\Bigg]$$

$$\div\sinh\left(\alpha\Big(\frac{a+l-b-u}{2}\Big)\right)$$

Now write

$$\begin{aligned}\sinh\left(\alpha\Big(\frac{a+l-b-u}{2}\Big)\right) &= \sinh\left(\alpha\Big(\frac{[a+l-2x-(b+u-2x)]}{2}\Big)\right)\\ &= \sinh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)\cosh\left(\alpha\Big(\frac{b+u-2x}{2}\Big)\right)\\ &\quad -\cosh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)\sinh\left(\alpha\Big(\frac{b+u-2x}{2}\Big)\right).\end{aligned}$$

Therefore the value function can be expressed as

$$V(x)=\frac{kx^2}{\rho}+\frac{k\sigma^2}{\rho^2}+\frac{\left[\frac{(g-\frac{k}{\rho}(a^2-l^2))}{\sinh(\alpha(\frac{a-l}{2}))\cosh(\alpha(\frac{a+l-2x}{2}))}-\frac{(g-\frac{k}{\rho}(b^2-u^2))}{\sinh(\alpha(\frac{b-u}{2}))\cosh(\alpha(\frac{b+u-2x}{2}))}\right]}{\tanh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)-\tanh\left(\alpha\Big(\frac{b+u-2x}{2}\Big)\right)}.$$

Now define a function

$$\Phi(a,l)=\frac{\left(g-\frac{k}{\rho}(a^2-l^2)\right)}{\sinh\left(\alpha\Big(\frac{a-l}{2}\Big)\right)\cosh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)},$$

Then the value function will be

$$V(x)=\frac{kx^2}{\rho}+\frac{k\sigma^2}{\rho^2}+\frac{1}{2}\left[\frac{\Phi(a,l)-\Phi(b,u)}{\tanh\left(\alpha\Big(\frac{a+l-2x}{2}\Big)\right)-\tanh\left(\alpha\Big(\frac{b+u-2x}{2}\Big)\right)}\right].$$

Dividing $V(x)$ by $g/2$ and introducing the non-dimensional variables and parameters in (2.18) yields

$$V(v)=\frac{2v^2}{\gamma}+\frac{1}{\gamma}+\frac{\Phi(y,z)-\Phi(w,x)}{\tanh(z-2x)-\tanh(x-2v)},$$

where

$$\Phi(w,z)=\frac{\gamma-yz}{\sinh y\cosh(z-2v)}.$$

Appendix G

The Homogenous equation $V = M^\alpha$ is a solution to (4.8) where α satisfies

$$\frac{\sigma^2}{2}M^2\alpha(\alpha-1)M^{\alpha-2} + \mu M\alpha M^{\alpha-1} - \rho M^\alpha = 0.$$

Simplifying the above expression yields

$$\alpha^2 + \left(\frac{2\mu}{\sigma^2} - 1\right)\alpha - \frac{2\rho}{\sigma^2} = 0.$$

It can be seen that α has two solutions α_1 and α_2, i.e.

$$\alpha_1 = -\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right) + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} > 0,$$

and

$$\alpha_2 = -\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right) - \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} < 0.$$

Now let

$$V = M^{\alpha_1}\phi.$$

Then

$$V_M = M^{\alpha_1}\frac{\partial\phi}{\partial M} + \alpha_1 M^{\alpha_1 - 1}\phi.$$

and

$$V_{MM} = M^{\alpha_1}\frac{\partial^2\phi}{\partial M^2} + 2\alpha_1 M^{\alpha_1-1}\frac{\partial\phi}{\partial M} + \alpha_1(\alpha_1 - 1)M^{\alpha_1-2}\phi.$$

It can now be easily seen that the function ϕ satisfies

$$M^{\alpha_1+2}\frac{\partial^2\phi}{\partial M^2} + 2\alpha_1 M^{\alpha_1+1}\frac{\partial\phi}{\partial M} + \alpha_1(\alpha_1-1)M^{\alpha_1}\phi + \\ \frac{2\mu}{\sigma^2}\left(M^{\alpha_1+1}\frac{\partial\phi}{\partial M} + \alpha_1 M^{\alpha_1}\phi\right) - \frac{2\rho}{\sigma^2}M^{\alpha_1}\phi = -\frac{2U(M)}{\sigma^2}.$$

Grouping similar terms yields

$$M^{\alpha_1+2}\frac{d^2\phi}{dM^2} + \left(2\alpha_1 + \frac{2\mu}{\sigma^2}\right)M^{\alpha_1+1}\frac{d\phi}{dM} \\ + \left(\alpha_1^2 - \alpha_1 + \frac{2\mu\alpha_1}{\sigma^2} - \frac{2\rho}{\sigma^2}\right)M^{\alpha_1}\phi = -\frac{2U(M)}{\sigma^2}. \quad \text{(G.1)}$$

It can be easily shown that

$$\left(\alpha_1^2 - \alpha_1 + \frac{2\mu\alpha_1}{\sigma^2} - \frac{2\rho}{\sigma^2}\right) = 0.$$

Therefore (G.1) simplifies to

$$\frac{d^2\phi}{dM^2} + \left(2\alpha_1 + \frac{2\mu}{\sigma^2}\right)\frac{1}{M}\frac{d\phi}{dM} = -M^{-(\alpha_1+2)}\left(\frac{2U(M)}{\sigma^2}\right).$$

We know that $e^{\int(2\alpha_1+\frac{2\mu}{\sigma^2})\frac{1}{M}} = M^{(2\alpha_1+\frac{2\mu}{\sigma^2})}$. Therefore,

$$M^{(2\alpha_1+\frac{2\mu}{\sigma^2})}\phi_{MM} + \left(2\alpha_1 + \frac{2\mu}{\sigma^2}\right)M^{(2\alpha_1+\frac{2\mu}{\sigma^2}-1)}\phi_M = M^{(2\alpha_1+\frac{2\mu}{\sigma^2}-1)}\left(\frac{2U(M)}{\sigma^2}\right).$$

Hence

$$\frac{d}{dM}\left[\phi_M . M^{(2\alpha_1+\frac{2\mu}{\sigma^2})}\right] = M^{(2\alpha_1+\frac{2\mu}{\sigma^2}-1)}\left(\frac{2U(M)}{\sigma^2}\right). \tag{G.2}$$

It can be easily seen that

$$\begin{aligned}
\alpha_1 + \frac{2\mu}{\sigma^2} - 2 &= \frac{\mu}{\sigma} - \frac{1}{2} + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} - 1 \\
&= -(\alpha_1 + 1).
\end{aligned}$$

Furthermore

$$\begin{aligned}
2\alpha_1 + \frac{2\mu}{\sigma^2} &= \alpha_1 + \frac{2\mu}{\sigma^2} - \frac{\mu}{\sigma^2} + \frac{1}{2} + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} \qquad (G.3) \\
&= \alpha_1 + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right) + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^3} - \frac{1}{2}\right)^2} + 1 \\
&= \alpha_1 - \alpha_2 + 1.
\end{aligned}$$

And

$$\begin{aligned}
\alpha_1 + \frac{2\mu}{\sigma^2} - 2 &= -\frac{\mu}{\sigma^2} + \frac{1}{2} + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} - 2 + \frac{2\mu}{\sigma^2} \qquad (G.4) \\
&= \frac{\mu}{\sigma^2} - \frac{3}{2} + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2} \\
&= -(1 + \alpha_2).
\end{aligned}$$

Substituting (G.3) and (G.4) into (G.2) gives

$$\frac{d}{dM}\left[\phi_M M^{\alpha_1-\alpha_2+1}\right] = U(M)M^{-(1+\alpha_2)}.$$

Integrating the above equation yields

$$\phi_M M^{\alpha_1-\alpha_2+1} = -\frac{2}{\sigma^2}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy + A.$$

Dividing through by $M^{\alpha_1-\alpha_2+1}$ generates

$$\phi_M = -\frac{2M^{\alpha_1-\alpha_2+1}}{\sigma^2}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy + AM^{\alpha_1-\alpha_2+1}.$$

Integrating the above equation it is clear that

$$\begin{aligned}\int_M^\infty \phi_x\,dx &= -\frac{2}{\sigma^2}\int_M^\infty x^{\alpha_2-\alpha_1-1}\left(\int_0^x U(y)y^{-(1+\alpha_2)}\,dy\right)dx \\ &\quad +A\int_M^\infty x^{\alpha_2-\alpha_1-1}\,dx.\end{aligned}$$

This integral has a solution

$$\begin{aligned}\phi(\infty)-\phi &= -\frac{2}{\sigma^2}\left[\frac{x^{\alpha_2-\alpha_1}}{\alpha_2-\alpha_1}\int_0^\infty U(y)y^{-(1+\alpha_2)}\,dy\right]_M^\infty \\ &\quad +\frac{2}{\sigma^2}\int_M^\infty \frac{x^{\alpha_2-\alpha_1}}{\alpha_2-\alpha_1}U(x)x^{-(1+\alpha_2)}\,dx \\ &\quad +\frac{A}{\alpha_2-\alpha_1}\left[x^{\alpha_2-\alpha_1}\right]_M^\infty.\end{aligned}$$

Therefore

$$\begin{aligned}\phi(\infty)-\phi &= -\frac{2}{\sigma^2(\alpha_2-\alpha_1)}\lim_{x\to\infty}\left[\frac{\int_0^x U(y)y^{-(1+\alpha_2)}}{x^{\alpha_1-\alpha_2}}\right] \\ &\quad +\frac{2M^{\alpha_1-\alpha_2}}{\sigma^2(\alpha_2-\alpha_1)}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy \\ &\quad +\frac{2}{\sigma^2(\alpha_2-\alpha_1)}\int_M^\infty U(x)x^{-(1+\alpha_1)}\,dx+\frac{A}{\alpha_1-\alpha_2}\left[0-M^{\alpha_2-\alpha_1}\right].\end{aligned}$$

Integrating the first term of the R.H.S of the above equation by parts, it reduces to

$$-\frac{2}{\sigma^2(\alpha_2-\alpha_1)}\lim_{x\to\infty}\frac{U(x)x^{-(1+\alpha_2)}}{(\alpha_1-\alpha_2)x^{\alpha_1-\alpha_2-1}}+\frac{2M^{\alpha_2-\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy$$
$$+\frac{2}{\sigma^2(\alpha_2-\alpha_1)}\int_M^\infty U(x)x^{-(1+\alpha_2)}\,dx-\frac{AM^{\alpha_2-\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)}.$$

It can be easily seen that

$$\frac{2M^{\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)^2}\lim_{x\to\infty}\frac{U(x)}{x^{\alpha_1}}=0,$$

and also $\phi(\infty)=-B$, where B is a constant. Therefore it follows that

$$\begin{aligned}-V(M) &= AM^{\alpha_2}+BM^{\alpha_1}+\frac{2M^{\alpha_2}}{\sigma^2(\alpha_2-\alpha_1)}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy\\ &\quad+\frac{2M^{\alpha_1}}{\sigma^2(\alpha_2-\alpha_1)}\int_M^\infty U(x)x^{-(1+\alpha_1)}\,dx\end{aligned}$$

is the solution to (4.8).

Appendix H

The following properties naturally follow from the assumptions we make about utility functions.

We have

$$\begin{aligned}\lim_{M\to0}M^{\alpha_2}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy &= \lim_{M\to0}\frac{\int_0^M U(y)y^{-(1+\alpha_2)}\,dy}{M^{-\alpha_2}}\\ &= \lim_{M\to0}\frac{U(M)M^{-(1+\alpha_2)}}{-\alpha_2M^{-(1+\alpha_2)}}\\ &= -\frac{1}{\alpha_2}\lim_{M\to0}U(M)\\ &= -\frac{U(0)}{\alpha_2}.\end{aligned}$$

Now let $x = My$. Then

$$\begin{aligned}
\lim_{M\to 0} M^{\alpha_1} \int_M^\infty U(x)x^{-(1+\alpha_1)}\,dx &= \lim_{M\to 0} M^{\alpha_1} \int_1^\infty U(My)M^{-(1+\alpha_1)}y^{-(1+\alpha_1)}M\,dy \\
&= \lim_{M\to 0} \int_1^\infty U(My)y^{-(1+\alpha_1)}\,dy \\
&= U(0)\int_1^\infty x^{-(1+\alpha_1)}\,dy \\
&= \frac{U(0)}{\alpha_1}.
\end{aligned}$$

Using these two results and setting $A \equiv 0$, it is clear that

$$\begin{aligned}
\lim_{M\to 0} V(M) &= -\frac{2U(0)}{\sigma^2\alpha_2(\alpha_2-\alpha_1)} + \frac{2U(0)}{\sigma^2\alpha_1(\alpha_2-\alpha_1)} \\
&= \frac{2U(0)}{\sigma^2\alpha_1\alpha_2} \\
&= 0.
\end{aligned}$$

Appendix I

Differentiating (4.10) with respect to M it is clear that

$$\begin{aligned}
-V'(M) \quad &= \quad B\alpha_1 M^{\alpha_1-1} - \frac{2\alpha_2 M^{\alpha_2-1}}{\sigma^2(\alpha_2-\alpha_1)}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy \\
&\quad +\frac{2M^{\alpha_2}}{\sigma^2(\alpha_2-\alpha_1)}U(M)M^{-(1+\alpha_2} \\
&\quad +\frac{2\alpha_1 M^{\alpha_1-1}}{\sigma^2(\alpha_2-\alpha_1}\int_M^\infty U(y)y^{-(1+\alpha_1)}\,dy \\
&\quad \frac{2M^{\alpha_1}}{\sigma^2(\alpha_2-\alpha_1}\Big(-U(M)M^{-(1+\alpha_1)}.
\end{aligned}$$

This can be further simplified to yield

$$\begin{aligned}
-V'(M) \quad &= \quad B\alpha_1 M^{\alpha_1-1} + \frac{2\alpha_2 M^{\alpha_2-1}}{\sigma^2(\alpha_2-\alpha_1)}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy \\
&\quad +\frac{2\alpha_1 M^{\alpha_1-1}}{\sigma^2(\alpha_2-\alpha_1}\int_M^\infty U(y)y^{-(1+\alpha_1)}\,dy.
\end{aligned}$$

Appendix J

It is obvious that

$$\begin{aligned}
\lim_{M\to\infty} M^{\alpha_2-1}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy &= \lim_{M\to\infty}\frac{\int_0^M U(y)y^{-(1+\alpha_2)}\,dy}{M^{1-\alpha_2}}\\
&= \lim_{M\to\infty}\frac{U(M)M^{-(1+\alpha_2)}}{(1-\alpha_2)M^{-\alpha_2}}\\
&= \lim_{M\to\infty}\frac{U(M)}{(1-\alpha_2)M}\\
&= 0.
\end{aligned}$$

If we let $y = Mx$ it is also clear that

$$\begin{aligned}
\lim_{M\to\infty} M^{\alpha_1-1}\int_M^\infty U(y)y^{-(1+\alpha_1)}\,dy &= \lim_{M\to\infty} M^{\alpha_1-1}\int_M^\infty U(Mx)M^{-\alpha_1}x^{-(1+\alpha_1)}\,dx\\
&= \lim_{M\to\infty}\int_1^\infty \frac{U(Mx)}{M}x^{-(1+\alpha_1)}\,dx\\
&= \lim_{M\to\infty}\left(\left[\frac{U(Mx)x^{-\alpha_1}}{-\alpha_1 M}\right]_1^\infty\right.\\
&\quad \left.+\frac{1}{\alpha_1}\int_1^\infty U'(Mx)x^{-\alpha_1}\,dx\right)\\
&= \lim_{M\to\infty}\left(\frac{U(M)}{\alpha_1 M}+\frac{1}{\alpha_1}\int_1^\infty U'(Mx)x^{-\alpha_1}\,dx\right)
\end{aligned}$$

It can be easily seen that

$$\lim_{M\to\infty}\frac{U(M)}{\alpha_1 M} = 0.$$

Also

$$\lim_{M\to\infty} U'(Mx) = 0,$$

and therefore

$$\lim_{M\to\infty}\frac{1}{\alpha_1}\int_1^\infty U'(Mx)x^{-\alpha_1}\,dx = 0.$$

Appendix K

It is clear that

$$\begin{aligned}
\lim_{M\to\infty} M^{\alpha^2}\int_0^M U(y)y^{-(1+\alpha_2)}\,dy &= \lim_{M\to\infty}\frac{\int_0^M U(y)y^{-(1+\alpha_2)}\,dy}{M^{-\alpha^2}}\\
&= \lim_{M\to\infty}\frac{U(M)M^{-(1+\alpha_2)}}{-\alpha_2 M^{-(1+\alpha_2)}}\\
&= \lim_{M\to\infty}\frac{U(M)}{-\alpha_2}.
\end{aligned}$$

Self evidently $\lim_{M\to\infty} V(M) = \infty$ as $M \to \infty$.

Appendix L

It is easily see that letting $v = \frac{M}{z}$

$$\begin{aligned}
M\int_0^1 z^{\alpha_1-2}U'\left(\frac{M}{z}\right)dz &= \int_0^1 vU'(v)y^{\alpha_1-1}\,dy\\
&= \left[\frac{y^{\alpha_1}}{\alpha_1}vU'(v)\right]_0^1 - \int_0^1 \frac{y^{\alpha_1}}{\alpha_1}\Big(vU'(v)\Big)'\frac{M}{-y^2}\,dy\\
&= \frac{1}{\alpha_1}MU'(M) + \frac{M}{\alpha_1}\int_0^1 y^{\alpha_1-2}\Big(vU'(v)\Big)'\,dz.
\end{aligned}$$

Technical Annexures

Technical Annex 1-BFGS Method for Unconstrained Minimisation

Here the multi-dimensional optimisation technique BFGS which, is used to obtain the optimal values for a, l, u and b in Chapter 2, is discussed. First a simple explanation of multi-dimensional optimisation is provided by discussing finite-difference derivatives. Then Broyden's method on which BFGS is based is illustrated, and, finally the method of BFGS derived.

Finite-Difference Derivatives

In this section finite-difference approximations are derived by using first and second partial derivatives. Also some aspects of efficiency, convergence, numerical roundoff and mathematical accuracy are discussed. From elementary calculus it is clear that $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$$\lim_{h \to 0} \frac{F_i(x + he_j) - F_i(x)}{h} = \frac{\partial F_i(x)}{\partial x_j},$$

where e_j is the j-th column of the $n \times n$ identity matrix. This for obvious reasons is called the forward difference approximation and suggests the evaluation of the j-th column of the the Jacobian matrix $J(x_c)$ using

$$\Delta_j F(x_c, h) = \frac{1}{h_j}[F(x_c + h_j e_j) - F(x_c)] \tag{T1.5}$$

for a suitably chosen vector h.

Lemma 1 *Let $F \in Lip_\gamma(D)$, where the coordinates x_c and $x_c + h_j e_j$ for $j = 1, \cdots, n$ are also in D. Furthermore let $\|\cdot\|$ be the vector norm where $\|e_j\| = 1$. Then*

$$\|\Delta_j F(x_c, h) - J(x_c)e_j\| \leqq \frac{1}{2}\gamma|h_j|.$$

Also if $\|\cdot\|$ is the l_1 vector norm given by $\|v\|_1 = \sum_{j=1}^{n} |v_j|$ then in the l_1 operator norm it is clear that

$$\|A\|_1 = \max_{1 \leqq j \leqq n} \sum_{i=1}^{n} |a_{ij}|.$$

From this it trivially follows that

$$\|\Delta F(x_c, h) - J(x_c)\|_1 \leqq \frac{1}{2}\gamma\|h\|_\infty.$$

Proof. The proof is established using the remainder of a second order expansion of the Taylor series. Define $M_c^N(x_c + h_j e_j)$ as

$$M_c^N(x_c + h_j e_j) = F(x_c) + J(x_c)h_j e_j.$$

That is $M_c^N(x_c + h_j e_j)$ is a first order Taylor expansion of $F(x_c + h_j e_j)$ around x_c. Now it is clear that

$$\begin{aligned}
\|\Delta_j F(x_c, h) - J(x_c)\| &= |h_j|^{-1}\|\Delta_j F(x_c + h_j e_j) - F(x_c) - J(x_c)h_j e_j\| \\
&= |h_j|^{-1}\|\Delta_j F(x_c + h_j e_j) - M_c^N(x_c + h_j e_j)\| \\
&\leqq |h_j|^{-1}\frac{1}{2}\gamma\|h_j\|^2 = \frac{1}{2}\gamma|h_j|.
\end{aligned}$$

The proof in the l_1 operator norm trivially follows. Clearly

$$\begin{aligned}\|\Delta F(x_c,h)-J(x_c)\|_1 &= \max_{1\leqq j\leqq n}\|\Delta_j F(x_c,h)-J(x_c)e_j\|_1\\ &\leqq \max_{1\leqq j\leqq n}\frac{1}{2}\gamma|h_j|=\frac{1}{2}\gamma\|h\|_\infty.\end{aligned}$$

Although the forward difference method of evaluating the gradient is accurate enough, the central difference method may be preferable. Here it is defined as

$$\delta_i f(x_c,h)=\frac{f(x_c+h_iei)-f(x_c-h_ie_i)}{2h_i},$$

and

$$\delta f(x_c,h)=[\delta_1 f(x_c,h),\cdots,\delta_n f(x_c,h)].$$

Lemma 2 *Let $H\in Lip_\gamma(D)$, H being the Hessian matrix, where coordinates x_c and $x_c+h_je_j$ for $j=1,\cdots,n$ are also in D. Furthermore let $\|\cdot\|$ be the vector norm where $\|e_j\|=1$. Then the behaviour of $\delta_i f(x_c,h)$ is given by*

$$|\delta_i f(x_c,h)-\frac{\partial f(x_c)}{\partial x_i}|\leqq\frac{1}{6}\gamma|h_i|^2,$$

and

$$\|\delta f(x_c,h)-g(x_c)\|_\infty\leqq\frac{1}{6}\gamma|h_i|^2_\infty,$$

where $g(x_c)=\partial f(x_c)/2\partial x_i$.

Proof. Let $m_c^N(\cdot)$ be the first order Taylor expansion of $f(\cdot)$. Then

$$\begin{aligned}&[f(x_c+h_ie_i)-m_c^N(x_c+h_ie_i)]-[f(x_c-h_ie_i)-m_c^N-h_ie_i]\\ &\qquad=f(x_c+h_iei)-f(x_c-h_ie_i)-2h_i\frac{\partial f(x_c)}{\partial x_i}.\end{aligned}$$

From the triangle inequality it is clear that

$$|f(x_c+h_iei)-f(x_c-h_ie_i)-2h_i\frac{\partial f(x_c)}{\partial x_i}|=\frac{1}{3}\gamma|h_i|^3.$$

Thus it can be easily seen that the central difference gradient is more accurate than the forward difference gradient, however, it requires $2n$ evaluations rather than the n necessary for the forward difference. Let $H(x_c)$ be the second order term of the Taylor series expansion of a function $f(x_c+h_ie_i)$. If the gradient of a function can be derived analytically, but the Hessian matrix needs to be approximated, then (T1.5) can be used by applying it to $g(x)$ to obtain the approximation of $\Delta g(x_c, h)$. This approximation, however, will not yield a symmetric matrix, whereas $H(x_c)$ will. Here a sensible strategy is to use $B_c = \frac{1}{2}[\Delta g(x_c) + \Delta g(x_c)^T]$ as the approximation of the Hessian. This is justified by observing that the Frobenius norm projection of $\Delta g(x_c)$ into the subspace of all symmetric matrices is B_c. Using this property and the Pythagorean Theorem yields

$$\|H(x_c) - B_c\|_F \leqq \|H_c - \Delta g(x_c, h)\|_F$$

where

$$\|A\|_F \equiv \Big(\sum_{i,j} |a_{i,j}|^2\Big)^{\frac{1}{2}}$$

is the Frobenius norm.

Lemma 3 *Let $H \in Lip_\gamma(D)$, where the coordinates x_c and $x_c + h_je_j$ for $j = 1, \cdots, n$ are also in D. Furthermore let $\|\cdot\|$ be the vector norm where $\|e_j\| = 1$. Also let*

$$[H_c]_{i,j} \equiv \frac{f(x_c + h_ie_i + h_je_j) - f(x_c + h_ie_i) - f(x_c + h_je_j) + f(x_c)}{h_ih_j}.$$

Then

$$|[H_c]_{i,j} - [H(x_c)]_{i,j}| \leqq \frac{\gamma}{6}\Big(2\frac{|h_i|^2}{|h_j|} + 3|h_i| + 3|h_j| + 2\frac{|h_j|^2}{|h_i|}\Big).$$

In the l_∞, Frobenius or l_1 operator norm, it follows that

$$|[H_c]_{i,j} - [H(x_c)]_{i,j}| \leqq \frac{n\gamma}{6} \max_{i,j} \left(2\frac{|h_i|^2}{|h_j|} + 3|h_i| + 3|h_j| + 2\frac{|h_j|^2}{|h_i|} \right).$$

Proof. The proof follows from the previous Lemmas. If $s_i = h_i, s_j = h_j$, and $s_{ij} = s_i + s_j$, then

$$\begin{aligned}
&[f(x_c + s_{ij}) - m_c^N(x_c + s_{ij})] - [f(x_c + s_i) - m_c^N(x_c + s_i)] \\
&\quad -[f(x_c + s_j) - m_c^N(x_c + s_j)] \\
&\qquad = f(x_c + s_{ij}) - f(x_c + s_i) - f(x_c + s_j) + f(x_c) - h_i h_j [H(x_c)]_{ij}.
\end{aligned}$$

From the triangle inequality it is clear that

$$\begin{aligned}
|h_i h_j [H_c]_{ij} - h_i h_j [H(x_c)]_{ij}| &\leqq \frac{1}{6}\gamma[\|s_{ij}\|^3 + \|s_i\|^3 + \|s_j\|^3] \\
&\leqq \frac{1}{6}\gamma[(\|s_i\| + \|s_j\|)^3 + \|s_i\|^3 + \|s_j\|^3] \\
&\leqq \frac{1}{6}\gamma[(\|h_i\| + \|h_j\|)^3 + \|h_i\|^3 + \|h_j\|^3].
\end{aligned}$$

Now that some useful rules for evaluating derivatives and analysing their accuracy have been established, two Theorems which establish the rate of a convergence of finite difference approximation shall be stated.

Theorem 1 *Let $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be $C^1(x)$ in an open convex set D, $D \subset R$. Let there exist constants $r, \beta > 0$ for $x_* \in D$, and, $J(x_c) \in Lip_\gamma N(x_*, r)$, $\|J(x_*)\| \leqq \beta$, and $F(x_*) = 0$. Then there also exists an $\epsilon > 0$ for each $x_0 \in N(x_*, \epsilon)$ in the sequence of points $\{x_k\}$ generated by the steps*

$$x_{k+1} = x_k - J(x_k)^{-1} F(x_k), \qquad k = 0, 1, 2, \cdots$$

which is well defined and converges to x_* and is satisfied by

$$\|x_{k+1} - x_k\| \leqq \|x_k - x_*\|^2.$$

Proof. It follows trivially that Lipschitz continuity implies continuity of function. Furthermore, the determinant of a matrix is a continuous function of the entries of the matrix. Thus it is obvious that $J(x)$ is invertible, and also that $\|J(x)^{-1}\| \leqq 2\beta$, for $x \in N(x_*, r)$. If $x_k \in N(x_*, \epsilon)$ for $\epsilon \leqq \min\{r, (2\beta\gamma)^{-1}\}$, x_{k+1} exists and

$$\begin{aligned} x_{k+1} - x_k &= x_k - J(x_k)^{-1}F(x_k) - x_* + J(x_k)^{-1}F(x_*) \\ &= J(x_k)^{-1}[F(x_*) - F(x_k) - J(x_k)(x_* - x_k)]. \end{aligned}$$

This yields

$$\begin{aligned} \|x_{k+1} - x_k\| &= \|J(x_k)^{-1}[F(x_*) - F(x_k) - J(x_k)(x_* - x_k)]\| \\ &\leqq 2\beta\frac{\gamma}{2}\|x_k - x_*\|^2 \leqq \beta\gamma\|x_k - x_*\|^2 \\ &\leqq \beta\gamma\epsilon\|x_k - x_*\| \leqq \frac{1}{2}\|x_k - x_*\|. \end{aligned}$$

This establishes both convergence and quadratic convergence and thus concludes the proof.

Theorem 2 *If F and x_* obey the hypothesis in the above theorem, then in the l_1 operator norm there exists an $\epsilon, \eta > 0$ for a sequence $\{h_k\}$ in $\mathbb{R}^n$ where $0 \leqq \|h_k\| \leqq \eta$, and $x_0 \in N(x_*, \epsilon)$. Also the sequence $\{x_k\}$ generated by*

$$B_k e_j = \begin{cases} \frac{F(x_k + (h_k)e_j) - F(x_k)}{(h_k)_j}, & (h_k)_j \neq 0 \\ J(x_k)e_j & U(h_k)_j = 0 \end{cases}$$

$$x_{k+1} = x_k - B_k^{-1}F(x_k), \qquad k = 0, 1, 2, \cdots,$$

exists and converges to x_*.

Broyden's Method

The secant method is an effective method for solving nonlinear equations in one dimension. It is a forward difference method in which the step size h_+ is used to construct the next iterate from x_+ to form the difference $(x_c - x_+)$. Thus resulting new local derivative B_+ will be $\frac{[F(x_+ + (x_c - x_+)) - F(x_+)]}{[x_c - x_+]}$. It is obvious from this that no extra function value will be needed to evaluate a new local model since $F(x_+ + h_+) = F(x_c)$. The secant method assumes that the first order Taylor expansion of $F(x_+ + d)$, $M_+(x_+ + d) = F(x_+) + B_+ d$, is evaluated by letting $M_+(x_c)$ tend towards $F(x_c)$. Thus B_+ is evaluated by

$$F(x_c) = M_+(x_+ + (x_c - x_+)) = F(x_+) + B_+(x_c - x_+).$$

This yields a system of linear equations

$$B_+ s_c = y_c$$

where $y_c = F(x_+) - F(x_c)$, and $s_c = x_+ - x_c$.

Lemma 4 *If* $s_c, y_c \in \mathbb{R}^n$, $s_c \neq 0$ *and* $B_c \in \mathbb{R}^n \times \mathbb{R}^n$, *Broyden's update*

$$B_+ = B_c + \frac{(y_c - B_c s_c) s_c^T}{s_c^T s_c}$$

gives the unique solution of

$$\min \|B - B_c\| \qquad \text{s.t.} \qquad B s_c = y_c.$$

Proof. The proof follows from the Lemmas established for forward differences. If $B s_c = y_c$, then $B_+ - B_c = [B - B_c] \frac{s_c s_c^T}{s_c^T s_c}$ we have

$$\begin{aligned} \|B_+ - B_c\|_F &\leqq \|B - B_c\|_F \cdot \left\| \frac{s_c s_c^T}{s_c^T s_c} \right\|_2 \\ &\leqq \|B - B_c\|_F \end{aligned}$$

BFGS Unconstrained Optimisation Technique

In the earlier section the secant approximation technique for choosing the Jacobian matrix was applied. Now this shall be adapted to the Hessian matrix. The analog of the previous section is simply

$$Bs_c = y_c = g(x_+) - g(x_c),$$

where B is the approximation of the Hessian matrix, $H(x_+)$. The above equation uses the second order Taylor expansion

$$m(x_+ + d) = f(x_+) + g(x_+)^T d + \frac{1}{2} d^T B_+ d$$

for the interpolation of $g(x_+), g(x_c)$, and $f(x_+)$. $H(x_+)$ is symmetric, however, B_+ or B_c are not. If B_+ is symmetric, it will approximate $H(x_+)$ more accurately since

$$\|\frac{1}{2}[B_+ + B_+^T] - H(x_+)\|_F \leqq \|B_+ - H(x_+)\|_F. \tag{T1.6}$$

If a projection of B_c on the intersection of matrices obeying the above equation with the subspace of symmetric matrices in $\mathbb{R}^{n\times n}$ is taken, B_+ is obtained in the form of the PSB(Powell symmetric Broyden) update

$$B_+ = B_c + \frac{(y_c - B_c s_c)s_c^T + s_c(y_c - B_c s_c)^T}{s_c^T s_c} - \frac{s_c(y_c - B_c s_c)s_c s_c^T}{(s_c^T s_c)^2}.$$

It is clear that B_+ will inherit its symmetry from B_c. This is clearly an effective update but has problems with poor scaling. Furthermore B_+ only inherits its positive definiteness from B_c and that too under conditions more restrictive than (T1.6). An obvious condition for B in (T1.6) to be positive definite is

$$s_c^T y_c = s_c^T B s_c > 0.$$

It is also possible to show that a positive definite and symmetric solution exists for B by constructing the BFGS method due to Broyden, Fletcher, Goldfarb, and Shanno. If B_c is assumed to be a positive definite and symmetric then it could be expressed by using Cholesky factor. That is

$$B_c = L_c L_c^T,$$

L_c being lower triangular, B_+ must be positive definite and symmetric where

$$B_c = J_+ J_+^T, \qquad \text{and} \qquad J_+ J_+^T s_c = y_c.$$

for J_+ nonsingular. Now let $v_c = J_+^T s_c$, such that $J_+ v_c = y_c$. If v_c is known then using Broyden's method J_+ could be evaluated as

$$J_+ = L_c + \frac{(y_c - L_c v_c) v_c^T}{v_c v_c^T}.$$

Transposing the above equation and multiplying the right and left hand sides by s_c it is clear that

$$v_c = J_c^T = L_c^T s_c + v_c \left(1 - \frac{v_c^T L_c^T s_c}{y_c^T s_c} \right).$$

If $y_c^T s_c > 0$ the above equation can be further simplified to yield

$$v_c = \left(\frac{y_c^T s_c}{s_c^T B_c s_c} \right)^{\frac{1}{2}} L_c^T s_c.$$

It is easy to see that the above two equations define J_+ such that $J_+ J_+^T = B_+$ which is the BFGS update. Alternatively the update could be expressed as

$$B_+ = B_c + \frac{y_c y_c^T}{y_c^T s_c} - \frac{B_c s_c s_c^T B_c}{s_c^T B_c s_c}.$$

This is the theoretical basis of the BFGS multi-dimensional optimisation technique. However, there exist problems in its implementation. Clearly if the last evaluation is near the minimum, then the update will arrive at it without any problem. If, however, it is further away, s_c may be sufficiently large such that it overshoots the minimum. The function evaluation could even explode. To eliminate this possibility a line minimisation routine is embedded within the algorithm with the following steps

1. Begin by choosing $x_0 \in \mathbb{R}^n$ and an $n \times n$ positive definite matrix B_0 where $B_0 = I$ and set $g_0 := g(x_0)$. For $k = 0, 1, \cdots$ obtain x_{k+1}, B_{k+1} from x_k, B_k using the following steps:

2. If g_k is zero, then stop; obviously because x_k is a stationary point. Else

3. compute $s_k := B_k^{-1} y_k$.

4. Choose the next coordinate

$$x_{k+1} = x_k - \lambda_k s_k$$

through the approximate minimisation

$$F(x_k) \approx \min\{F(x_k - \lambda s_k) | \lambda \geqq 0\},$$

and then set

$$g_{k+1} := g(x_{k+1}), \qquad s_k := x_{k+1} - x_k, \qquad y_k := g_{k+1} - g_k.$$

5. compute B_{k+1} according to the BFGS update described above.

Technical Annex 2-Simpson's Rule

If a function $f(x) \in [a, b]$ is $C^4(x)$, then it can be shown that

$$\int_a^b f(x)\,dx - \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] = -\frac{(b-a)^5}{2880} f^4(\psi), \qquad a < \psi < b,$$

where

$$\int_a^b f(x)\,dx \approx \frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

is the Simpson approximation. Thus Simpsons's rule is exact for all polynomials of degree three or less. However, Simpson's rule is most often applied in its compound form. The interval $[a,b]$ is divided into a number of intervals and Simpson's rule is applied to each. Let

$$a = x_0 < x_1 < \ldots < x_{2n-1} < x_{2n} = b$$

be a sequence of points on $[a,b]$ such that

$$x_{i+1} - x_i = h, \qquad i = 0, \cdots, 2n-1.$$

Then the compound Simpson's rule yields

$$\int_{x_0}^{x_{2n}} f(x)\,dx = \frac{h}{3}[f_0+4(f_1+\cdots+f_{2n-1})+2(f_2+f_4+\cdots+f_{2n-2})+f_{2n}]+E_n$$

where E_n is the remainder and is given by

$$E_n = -\frac{nh^5}{90}f^4(\psi), \qquad a < \psi < b.$$

Let N be the even number of sub-intervals of $[a,b]$. Then $N = 2n$ and h can be expressed as

$$h = \frac{(b-a)}{N},$$

such that

$$E_n = -\frac{(b-a)^5}{180n^4}f^4(\psi), \qquad a < \psi < b.$$

It can be shown that for functions that are $C^4(x)$, Simpson's rule converges to the actual value of the function with a velocity of $N{-}4$ at worst. Here an automatic Simpson's integrator is used where the limits of the integration is provided, along with a routine for computing $f(x)$, an error tolerance ϵ, and an upper bound on the number of function evaluations. The source code, once compiled then returns one of (a) the integral has been evaluated to the specified tolerance, (b) the interval of Integration has zero length, (c) the tolerance is either negative or not achievable, or (d) the error tolerance has not been met within the allowed iterations.

Technical Annex 3-Runge-Kutta Order Four

Consider a function $f(t,y)$ that is $C^{n+1}(t,y)$ on $D = \{(t,y)|a \leqq t \leqq b, c \leqq y \leqq d\}$. Let $(t_+, y_+) \in D$ such that

$$f(t,y) = P_n(t_+, y_+) + R_n(t_+, y_+),$$

where

$$\begin{aligned} P_n(t_+, y_+) \quad &= \quad f(t_+, y_+) + \left[(t-t_+)\frac{\partial f(t_+, y_+)}{\partial t} + (y-y_+)\frac{\partial f(t_+, y_+)}{\partial u}\right] \\ &\quad + \left[\frac{(t-t_+)^2}{2}\frac{\partial^2 f(t_+, y_+)}{\partial t^2} + \frac{(y-y_+)^2}{2}\frac{\partial^2 f(t_+, y_+)}{\partial y^2}\right] + \cdots \\ &\quad + \left[\frac{1}{n!}\sum_{j=0}^{n}\binom{n}{j}(t-t_+)^{n-j}(y-y_+)^j\frac{\partial^n f(t_+, y_+)}{\partial t^{n-j}\partial y^j}\right] \end{aligned}$$

and

$$R_n(t,y) = \left[\frac{1}{(n+1)!}\sum_{j=0}^{n}\binom{n+1}{j}(t-t_+)^{n+1-j}(y-y_+)^j\frac{\partial^{n+1} f(t_+, y_+)}{\partial t^{n+1-j}\partial y^j}\right]$$

where P_n is the Taylor polynomial resulting from the nth order Taylor series expansion of $f(t,y)$ around (t_+, y_+), and, R_n is the remainder resulting from this expansion. The Runge-Kutta scheme exploits this property to obtain a solution to a differential equation. Consider a differential equation of the form

$$\frac{dy}{dt} = f(t, y)$$

with the initial condition

$$y(t_0) = y_0.$$

The Runge-Kutta integrator order four integration scheme uses the formula

$$\begin{aligned}
y_0 &= y(t_0)\\
k_1 &= f(t_i, y_i)\\
k_2 &= f(t_i + \frac{h}{2}, y_i + \frac{hk_1}{2})\\
k_2 &= f(t_i + \frac{h}{2}, y_i + \frac{hk_2}{2})\\
k_4 &= f(t_i + h, y_i + hk_3)\\
y_{i+1} &= y_i + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)
\end{aligned}$$

for $i = 0, 1, 2, \cdots, N-1$. Here y_i is the computed value of the solution at t_i, where $t_{i+1} - t_i = h$. It can be easily seen that if $f(t, y) = g(x)$, then the above scheme reduces to

$$y_{i+1} = y_i + \frac{h}{6}\left[f(t_i) + 4f\left(t_i + \frac{h}{2}\right) + f(t_i + h)\right].$$

This method has a localised truncation error of order four, provided of course that $y(t)$ is $C^5(t)$. The Runge-Kutta scheme used here is a more refined automatic integrator allowing for more efficient forward steps to taken in the integration, based on rounding off errors obtained, with similar conditions to Simpsons rule. If the integrator is performing function evaluations and the specified accuracy is being met, then the evaluated function values are returned. However, if the accuracy is not being met, then it could return either (a) the actual function values but with an accompanying warning that the requested accuracy has not been met, or (b) more memory has been allocated than is necessary. If the integrator fails, then the reason for failure is returned. Which could be one of (c) the range of integration has been set to zero, (d) the effective range of integration is zero, (e) there has been a memory allocation failure, or finally (f) the order of equations has increased and memory needs to be reallocated.

Technical Annex 4-Chebyshev Polynomials

Integration

Chebyshev polynomials, $T_n(x)$ are cosine functions after a change in the independent variable, i.e.

$$T_n(x) = \cos n\theta \qquad \theta = \arccos x.$$

Preforming the transformation $x = \cos\theta$ allows numerous mathematical and spectral relationships to be used in a Chebyshev system. They also satisfy a three term recurrence relation

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \qquad n = 1, 2, \cdots$$

due to the trigonometric identity

$$\cos(n+1)\theta = 2\cos n\theta \cos\theta - \cos(n-1)\theta.$$

They are also bounded, i.e.

$$|T_n(x)| \leqq 1 \qquad \text{and} \qquad x \in (-1, 1),$$

which follows from the fact that T_n is a cosine.

It can be also shown that they satisfy the orthogonality property

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}} = \begin{cases} \pi, & m = n = 0, \\ \frac{\pi}{2}, & m = n \neq 0, \\ 0, & m \neq n. \end{cases}$$

Under mild conditions on a function $f(x)$; e.g. $f(x)$ is $C^1(x)$, $x \in [-1,1]$, $f(x)$ can be expressed in a uniformly convergent series of T's

$$\begin{aligned} f(x) &= \frac{1}{2}a_0 + a_1T_1(x) + a_2T_2(x) + \cdots \\ &= \sum_{i=0}^{\infty} a_iT_i(x). \end{aligned}$$

The constant coefficients a_i are referred to as "Fourier-Chebyshev" coefficients and are given by

$$a_0 = \frac{1}{\pi}\int_{-1}^{1} \frac{f(x)T_r(x)}{\sqrt{1-x^2}}\,dx.$$

and

$$a_i = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)T_r(x)}{\sqrt{1-x^2}}\,dx \qquad r \geqq 1.$$

Significantly a_0, a_1, $\cdots$ decay rapidly to zero. The partial sum $\frac{1}{2}a_0 + a_1T_1(x) + \cdots + a_nT_N(x)$ is polynomial of degree $\leqq N$, which is one of the most accurate estimations of $f(x)$ by a polynomial $p_N(x)$, the approximation being measured in the sense of $\max_{-1\leqq x \leqq 1}|f(x) - p_N(x)|$. Although three different quadratures could be used the preferred method is the *Chebyshev-Gauss-Labotto* quadrature of the form

$$x_j = \cos\frac{\pi j}{N} \qquad w_j = \begin{cases} \frac{\pi}{2N}, & j = 0, N \\ \frac{\pi}{N}, & 1 \leqq j \leqq N-1. \end{cases}$$

Differentiation

The derivative of a function $f(x)$ can be evaluated by the sum

$$f'(x) = \sum_{n=0}^{\infty} a_n^{(1)} T_n(x),$$

where

$$a_n^{(1)} = \sum_{j=n+1, j+n_{odd}}^{\infty} j a_j. \qquad \text{(T4.1)}$$

This is expression is derived from the trigonometric identity

$$2\sin(\theta)\cos(n\theta) = \sin((n+1)\theta) - \sin((n-1)\theta),$$

which enables us to express $T_n(x)$ in the form

$$2T_n(x) = \frac{T_{n+1}(x)}{(n+1)} - \frac{T_{n-1}(x)}{n-1}. \qquad n = 1, 2, \cdots.$$

In spectral space this specifies a relationship between the coefficients of the polynomial of the form

$$2na_n = a_{n-1}^{(1)} - a_{n+1}^{(1)}, \qquad n \geqq 1,$$

form which (T4.1) follows. This relationship suggests an efficient method by which Chebyshev polynomials could be differentiated in spectral space.

It clear by definition that $u_n = 0$ for $n \geqq N$. Thus one could arrange the above equation in a way by which the coefficients of the function derivative could be estimated from the the function coefficients through the recursive relationship

$$a_n^{(1)} = a_{n+1}^{(1)} + 2(k+1)a_{n+1} \qquad 0 \leqq k \leqq N-1.$$

The same methodology is applied to derive the k-th derivative in the form

$$a_n^{(n)} = a_{n+2}^{(n)} + 2(n+1)a_{n+1}^{(n-1)}.$$

The computational power required to perform this task could prove to be overpowering. Collocation differentiation offers a more efficient means of computing derivatives, since differentiation implies only a linear operation on the function values in physical space. The significant difference though is that we now need to use a Fast Fourier Transform (FFT) which requires equally spaced data points.

The FFT is a recursive algorithm for obtaining a Discrete Fourier Transform, such as described above, and its inverse. The FFT is conventionally expressed for the evaluation of

$$a_n = \sum_{j=0}^{N-1} f(x_j) e^{\frac{2\pi i j n}{N}} \qquad k = 0, 1, \cdots, N-1 \tag{T4.2}$$

$$a_n = \sum_{j=0}^{N-1} f(x_j) e^{\frac{-2\pi i j n}{N}} \qquad k = 0, 1, \cdots, N-1, \tag{T4.3}$$

where $f(x_j)$, $j = 0, 1, \cdots, N-1$ are a set of complex data. See Cooley and Tukey (1965) for a description of the FFT algorithm.

We know form previous sub-section that the Chebyshev transformation of a function, based on a Gaus-Labotto quadrature are given by

$$a_n = \sum_{j=0}^{N} f(x_n) \cos\left(\frac{\pi j k}{N}\right) \qquad n = 0, 1. \cdots, N \tag{T4.4}$$

where a_0 and a_N are halved, and

$$f(x) = \sum_{j=0}^{N} a_n \cos\left(\frac{\pi j k}{N}\right) \qquad n = 0, 1. \cdots, N. \tag{T4.5}$$

Assume that a transformation of the form is required for two sets of real data

a_n^1 and a_n^2. To do this, first define complex data of the form

$$z_n = \begin{cases} a_n^1 + i a_n^2 & n = 0, 1, \cdots, N \\ v_{2N-j} & n = N+1, N+2, \cdots, 2N-1 \end{cases}$$

Now define $\hat{z}_n$, $n = 0, 1, \cdots, N$ by (T4.4) and $\hat{Z}_n$, $n = 0, 1, \cdots, 2N-1$ by (T4.2) with N being replaced by $2N$. Then it follows that

$$\hat{Z}_n = \frac{z_n}{N}, \qquad n = 0, 1, \cdots, N,$$

and

$$\hat{Z}_k = \sum_{j=0}^{N-1} z_{2j} e^{\frac{2\pi i n j}{N}} + e^{\frac{\pi i n}{N}} \sum_{j=0}^{N-1} z_{2j+1} e^{\frac{2\pi i n j}{N}}.$$

See Burden and Faires (1993), Davis and Rabinowitz (1984), and Canuto et. al. (1988) for details of this. Let q_j be defined by

$$q_j = z_{2j} + i(z_{2j+1} - z_{2j-1}) \qquad j = 0, 1, \cdots, N-1,$$

and estimate $\hat{q}_n$ through the complex FFT given by (T4.2). This yields

$$\hat{q}_n = \sum_{j=0}^{N-1} z_{2j} e^{\frac{2\pi i n j}{N}} + i(1 - e^{\frac{2\pi i n}{N}}) \sum_{j=0}^{N-1} e^{\frac{2\pi i n j}{N}}$$

and

$$\hat{q}_{N-k} = \sum_{j=0}^{N-1} z_{2j} e^{\frac{2\pi i n j}{N}} - i(1 - e^{\frac{2\pi i n}{N}}) \sum_{j=0}^{N-1} e^{\frac{2\pi i n j}{N}}$$

See Burden and Faires (1993), Davis and Rabinowitz (1984), and Canuto et. al. (1988) for details of this. As a result

$$\hat{z}_0 = \frac{1}{N} \sum_{j=0}^{N} z_j,$$

$$\hat{z}_n = \frac{1}{N} \left[\left(\frac{1}{2} + \frac{1}{4 \sin(\frac{\pi n}{N})} \right) \hat{q}_n + \left(\frac{1}{2} - \frac{1}{4 \sin(\frac{\pi n}{N})} \right) \hat{q}_{N-n} \right]$$

and

$$\hat{z}_N = \frac{1}{N}\sum_{j=0}^{N}(-1)^j z_j.$$

Based on these evaluations, the derivative coefficients can be evaluated more efficiently using the Differentiation methods listed above. The resulting collocation matrix has the following points for the differentiation matrix (D_N)

$$(D_N)_{jn} = \begin{cases} \frac{c_j(-1)^{j+n}}{c_n(x_l - x_j} & j \neq n \\ \frac{-x_n}{2(1-x_n^2} & 1 \leqq l = j \leqq N-1 \\ \frac{2N^2+1}{6} & 1 = j = 1 \\ -\frac{2N^2+1}{6} & 1 = j = N \end{cases}$$

where

$$c_l, c_j = \begin{cases} 2 & j = 0, N \\ 1 & 1 \leqq j \leqq N-1 \end{cases}$$

See Canuto et. al. (1988) for details of this. These points are obtained by differentiating the Lagrange interpolating polynomial Φ in the FFT

$$\Phi_n(x) = \frac{(-1)^{n+1}(1-x^2)T_N'(x)}{c_n N^2(x - x_j)}.$$

Technical Annex 5-The Source Code Chapter 2

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <malloc.h>


/****************************************************************************
     In defining the characteristic root of the ODE, we set SIGMA=0.1
        and RHO=0.05, therefore ALPHA_1=sqrt(2*RHO)/SIGMA=sqrt(10.0).
****************************************************************************/


#define SIGMA       0.1
#define RHO         0.05
#define ALPHA_1     sqrt(2.0*RHO)/SIGMA
#define K           0.5
#define G           0.1
#define NITER       200
#define NV          4


double xval;     /*  xval initial position of state  */


void main( void )
{
    void parms4( double *, double *, double *, double *, double *);
    void parms2( double *, double *, double *);
    double  value_funct(double *), min_value, h, dixit_funct(double *),
            tol=5.0e-10, diag_hessian[NV], a, l, u, b;
    double   p[NV]={  0.5,  0.5,  1.0,  1.0},
            pl[NV]={ 0.01, 0.01, 0.01, 0.01}, ppl[2]={ 0.01, 0.01},
            pu[NV]={ 0.99, 0.99, 3.99, 3.99}, ppu[2]={ 3.99, 3.99};
    int     iter, i;
    void        bfgs(int, double *, double *, double *, double , double *,
                int *, double *, double (*func)(double *));
```

```
    char     filename[13];
    FILE     *output;

/*******************Create filename appropriate to application*************/
    filename[0] = 'C';
    filename[1] = '3';
    filename[2] = 'R';
    filename[3] = 'E';
    filename[4] = 'S';
    filename[5] = 'U';
    filename[6] = 'L';
    filename[7] = 'T';
    filename[8] = '.';
    filename[9] = 'D';
    filename[10] = 'A';
    filename[11] = 'T';
    printf("\nFilename in %12s",filename);

    for ( h=1.0e-5 ; f(h)>0.0 ; h +=1.0e-5 );

    iter = NITER;

    for( i=0 ; i<=40 ; i++ ) {
        output=fopen(filename,"a");
        printf("\n Implementing iteration %3d",i);
        xval = -2.0+0.1*((double) i);
        fprintf(output,"\n%10.6lf\t", xval);
        p[0] = 0.5;
        p[1] = 0.5;
        p[2] = 1.0;
        p[3] = 1.0;
        iter = NITER;
        bfgs(NV,p,pl,pu,tol,diag_hessian,&iter,&min_value,value_funct);
        parms4( p, &a, &l, &u, &b);
        fprintf(output,"%15.6lf %10.6lf  %10.6lf %10.6lf %10.6lf",
        min_value, a, l, u, b);
        p[0] = 1.0;
        p[1] = 1.0;
        iter = NITER;
        bfgs(2,p,ppl,ppu,tol,diag_hessian,&iter,&min_value,dixit_funct);
        parms2( p, &a , &b);
```

```
        fprintf(output,"%15.6lf %10.6lf  %10.6lf", min_value, a, b);
        fclose(output);
    }

    exit(0);
}


/*****************************************************************************
                            The Value Function
*****************************************************************************/

double value_funct( double *p)
{
    double ang1, ang2, a, l, u, b, temp;
    void parms4( double *, double *, double *, double *, double *);

    parms4( p, &a, &l, &u, &b);

    ang1 = 0.5*ALPHA_1*(b+u-2.0*xval);
    ang2 = 0.5*ALPHA_1*(a-l);
    temp = (G-(K/RHO)*(pow(a,2)-pow(l,2)))*cosh(ang1)/sinh(ang2);
    ang1 = 0.5*ALPHA_1*(a+l-2.0*xval);
    ang2 = 0.5*ALPHA_1*(b-u);
    temp -= (G-(K/RHO)*(pow(b,2)-pow(u,2)))*cosh(ang1)/sinh(ang2);
    ang1 = 0.5*ALPHA_1*(a+l-b-u);
    temp = 0.5*temp/sinh(ang1)+K*(pow(xval,2)/RHO+pow(SIGMA/RHO,2));

    return temp;
}

/*****************************************************************************
                        Parameter conversion function

        k_1=p[0]        k_2=p[1]          c_3=p[2]          c_4=p[3]

          a = xval-p[2]
          l = xval+p[0]*p[3]-(1.0-p[0])*p[2]
          u = xval+(1.0-p[1])*p[3]-p[1]*p[2]
          b = xval+p[3]
*****************************************************************************/
```

```
void parms4( double *p, double *a, double *l, double *u, double *b)
{
    *a = xval-p[2];
    *l = xval+p[0]*p[3]-(1.0-p[0])*p[2];
    *u = xval+(1.0-p[1])*p[3]-p[1]*p[2];
    *b = xval+p[3];

    return;
}


/*****************************************************************************
                        Parameter conversion function

                  a = xval-p[0]            b = xval+p[1]
*****************************************************************************/

void parms2( double *p, double *a, double *b)
{
    *a = ( xval-p[0]<-0.01 ) ? xval-p[0] : -0.01;
    *b = ( xval+p[1]>0.01 ) ? xval+p[1] : 0.01;

    return;
}


/*****************************************************************************
                Our version of Dixit's solution for the ODE
*****************************************************************************/

double dixit_funct( double *p)
{
    void    parms2( double *, double *, double *);
    double  a, b, temp, ang1, ang2;

    parms2( p, &a, &b);
    ang1 = 0.5*ALPHA_1*(a-2.0*xval);
    ang2 = 0.5*ALPHA_1*b;
    temp = (G-(K/RHO)*pow(b,2))*cosh(ang1)/sinh(ang2);
    ang1 = 0.5*ALPHA_1*(b-2.0*xval);
    ang2 = 0.5*ALPHA_1*a;
```

```
    temp -= (G-(K/RHO)*pow(a,2))*cosh(ang1)/sinh(ang2);
    ang1 = 0.5*ALPHA_1*(b-a);

    return 0.5*temp/sinh(ang1)+K*(pow(xval,2)/RHO+pow(SIGMA/RHO,2));
}


/*****************************************************************************
                              BFGS Optimising technique
*****************************************************************************/

#define     ALPHA   1.0e-4
#define  STEP_MAX   100.0

void bfgs(int n, double *x, double *xl, double *xu, double tol, double
          *diag_hessin, int *iter, double *y, double (*func)(double *))
{
    int its, i, j, k, iterm;
    short int start;
    double *scale, *g, *dg, *xn, *xi, *ptr, **hessin, **htmp;
    double tol_g, tol_h, tol_hh, tol_x, amp_x, amp_g, rnderr, stpmax,
           slope, temp, temp1, temp2, fac, fad, fae, sumdg, sumxi,
           h, fp, fm, xstore, get_step( double, double);
 /*
 **  Declaration of function prototypes
 */

    void runtime_error(char *);
    void dfcn( int, double *, double *, double, double *,
               double (*func)(double *));
    short int pd_fail( int, double *, double **, double **);
    short int line_search( int, double *, double *, double *, double *,
         double *, double *, int *, double, double, double, double *,
         double (*func)(double *));
    double grad( int, double *, double, double (*func)(double *));

 /*
 **  Step 1. ... Check parameter ranges on entry
 */

    for ( i=0 ; i<n ; i++ ) {
```

```
        if ( x[i]<xl[i] || x[i]>xu[i] )
        runtime_error("\nStarting values out of range...fatal error!\n");
    }

    if ( *iter<=0 )
    runtime_error("\nNumber of iterations unspecified...fatal error!\n");

/*
**  Step 2. ... Acquire memory for holding vectorial quantities
**
**    scale - Holds scale factors of original input
**        g - Gradient of surface at x[ ]
**       xn - Next estimate of the minimum point
**       dg - Initially gradient of surface at xn[ ] - latterly
**            difference in gradients
**       xi - Downhill slope at x[ ]
**     htmp - Temporary storage for Choleski decomposition
*/

    scale = (double *) calloc((size_t) n, (size_t) sizeof(double));
    g = (double *) calloc((size_t) n, (size_t) sizeof(double));
    xn = (double *) calloc((size_t) n, (size_t) sizeof(double));
    dg = (double *) calloc((size_t) n, (size_t) sizeof(double));
    xi = (double *) calloc((size_t) n, (size_t) sizeof(double));
    hessin = (double **) calloc((size_t) n, (size_t) sizeof(double *));
    htmp = (double **) calloc((size_t) n, (size_t) sizeof(double *));
    for ( i=0 ; i<n ; i++ ) {
        hessin[i] = (double *) calloc((size_t) n, (size_t) sizeof(double));
        htmp[i] = (double *) calloc((size_t) n, (size_t) sizeof(double));
    }
    if ( !htmp[n-1] ) runtime_error("\nMemory acquisition problem\n");

/*
**  Step 3. ... Scale variables based on parameter range
*/

    for ( i=0 ; i<n ; i++ ) {
        scale[i] =
              ( (temp1=fabs(xu[i]))>(temp2=fabs(xl[i])) ) ? temp1 : temp2;
        x[i] /= scale[i];
        xl[i] /= scale[i];
```

```
        xu[i] /= scale[i];
    }

/*
**  Step 4. ... Compute rounding error
*/

    rnderr = 1.0;
    while ( rnderr+1.0!=1.0 ) rnderr *= 0.5;
    rnderr *= 2.0;

/*
**  Step 5. ... Initialise test parameters and tolerances
*/

    tol_h = pow(tol,0.66);
    tol_hh = pow(tol,0.33);
    tol_g = tol_h;
    tol_x = 4.0*tol;
    start = 1;
    iterm = -1;
    its = 0;

/*
**  Step 6. ... Start the iteration phase
*/

    while ( its<*iter ) {
        if ( start || pd_fail(n, diag_hessin, hessin, htmp) ) {

/*
**  Calculate initialisation function value and downhill gradient.
**  Initialise the inverse hessian at the identity.
**  Compute norms for position x[ ] and gradient g[ ].
*/

            for ( i=0 ; i<n ; i++ ) x[i] *= scale[i];
            *y = func(x);
            for ( i=0 ; i<n ; i++ ) x[i] /= scale[i];
            dfcn(n, x, g, tol_h, scale, func);
            for ( amp_x=0.0,amp_g=0.0,i=0 ; i <n ; i++ ) {
```

```
                amp_x += x[i]*x[i];
                amp_g += g[i]*g[i];
                for ( j=0 ; j<n ; j++ ) hessin[i][j] = 0.0;
                hessin[i][i] = 1.0;
                xi[i] = -g[i];
            }
            if ( iterm!=-1 ) xi[iterm] = 0.0;
            amp_x = sqrt(amp_x);
            amp_g = sqrt(amp_g);
            stpmax = (( amp_x>(double )n ) ? amp_x : (double )n )*STEP_MAX;
            start = 0;
        }

        its++;
        if ( amp_g>stpmax )
                for ( fac=stpmax/amp_g,i=0 ; i<n ; i++ ) xi[i] *= fac;
        for ( slope=0.0,i=0 ; i<n ; i++ ) slope += g[i]*xi[i];
        start = line_search( n, x, xl, xu, xi, xn, y,
                                &iterm, slope, rnderr, tol_x, scale, func);

/*
**  Step 7. ... Check exit condition on parameter convergence
*/

        if ( start==0 ) {
            for ( temp=0.0,i=0 ; i<n ; i++ ) {
                temp1 = ( (temp2=fabs(x[i])) > 1.0 ) ? temp2 : 1.0;
                if ( temp < (temp2=fabs(xi[i])/temp1) ) temp = temp2;
            }
            if ( temp<=tol_x ) {
                *iter = its;
                for ( i=0 ; i<n ; i++ ) {
                    temp = (temp1=fabs(x[i]-xl[i])) <
                            (temp2=fabs(x[i]-xu[i])) ? temp1 : temp2;
                    if ( temp<=10.0*rnderr ) {
                        printf("\nOptimal parameter values appear
                                to be on the boundary!\n");
                        break;
                    }
                }
            }
```

```
/*
**  Step 8. ... Calculate new gradient and check gradient exit condition
*/

        ptr = dg;
        dg = g;
        g = ptr;
        dfcn(n, x, g, tol_h, scale, func);
        temp1 = (temp2=fabs(*y)) > 1.0 ? temp2 : 1.0;
        for ( temp=0.0,i=0 ; i<n ; i++ ) {
            if ( i!=iterm ) {
                temp2 =
                 fabs(g[i])*(fabs(x[i]) > 1.0 ? fabs(x[i]) : 1.0)/temp1;
                if ( temp<temp2 ) temp = temp2;
            }
        }
        if ( temp<=tol_g ) {
            *iter = its;
            for ( i=0 ; i<n ; i++ ) {
                temp =  (temp1=fabs(x[i]-xl[i])) <
                        (temp2=fabs(x[i]-xu[i])) ? temp1 : temp2;
                if ( temp<=10.0*rnderr ) {
                    printf("\nOptimal parameter values appear to be on
                            the boundary!\n");
                    break;
                }
            }
        }

/*
**  Step 9. ... No convergence and so continue with update procedure
**              recognising that xn[ ] can now be used as temporary
**              storage
*/

        for ( i=0 ; i<n ; i++ ) dg[i] = g[i]-dg[i];
        for ( i=0 ; i<n ; i++ ) {
            for ( xn[i]=0.0,j=0 ; j<n ; j++ )
            xn[i] += hessin[i][j]*dg[j];
        }
```

```
/*
**  Step 10. ... Calculate dot products for denominators
*/

          for ( fac=fae=sumdg=sumxi=0.0,i=0 ; i<n ; i++ ) {
               fac += dg[i]*xi[i];
               fae += dg[i]*xn[i];
               sumdg += dg[i]*dg[i];
               sumxi += xi[i]*xi[i];
          }

/*
**  Step 11. ... Test fac for size - skip update if too small
*/

          if ( fac*fac>rnderr*sumdg*sumxi ) {
               fac = 1.0/fac;
               fad = 1.0/fae;
               for ( i=0 ; i<n ; i++ ) dg[i] = fac*xi[i]-fad*xn[i];
               for ( i=0 ; i<n ; i++ ) {
                    for ( j=0 ; j<n ; j++ ) {
                         hessin[i][j] += fac*xi[i]*xi[j]-
                                        fad*xn[i]*xn[j]+fae*dg[i]*dg[j];
                    }
               }
          }

/*
**  Step 12. ... Compute next search direction
*/

          for ( amp_g=0,i=0 ; i<n ; i++ ) {
               amp_g += g[i]*g[i];
               for ( xi[i]=0.0,j=0 ; j<n ; j++ ) xi[i] -= hessin[i][j]*g[j];
          }
          amp_g = sqrt(amp_g);
     }
  }

/*
```

```
**  Step 13. ... Free temporary vector memory - rescale to true values
*/

   free(g);
   free(dg);
   free(xn);
   free(xi);
   for ( i=0 ; i<n ; i++ ) {
       x[i] *= scale[i];
       xl[i] *= scale[i];
       xu[i] *= scale[i];
   }

/*
**  Step 14. ... Compute true Hessian matrix at minimum
*/

   for ( i=0 ; i<n ; i++ ) {
       for ( j=0 ; j<n ; j++ ) {
           h = ( (h=tol_h*fabs(x[j]))>tol_hh ) ? h : tol_hh;
           xstore = x[j];
           h = get_step(xstore, h);
           x[j] = xstore+h;
           fp = grad(i, x, tol_hh, func);
           x[j] = xstore-h;
           fm = grad(i, x, tol_hh, func);
           x[j] = xstore;
           htmp[i][j] = 0.5*(fp-fm)/h;
       }
   }

/*
**  Step 15. ... Test Hessian matrix for positive definiteness
**               and compute inverse when positive definite
*/

   if ( pd_fail( n, diag_hessin, htmp, hessin) ) {
       fprintf(stderr,"\nWARNING - Irregular exit from BFGS ....");
       fprintf(stderr,"\nThe Hessian is not positive definite!\n");
       for ( i=0 ; i<n ; i++ ) {
           free(hessin[i]);
```

```
            free(htmp[i]);
        }
    } else {
        for ( i=0 ; i<n ; i++ ) {
            hessin[i][i] = 1.0/diag_hessin[i];
            for ( j=i+1 ; j<n ; j++ ) {
                for ( temp=0.0,k=i ; k<j ; k++ )
                    temp -= hessin[j][k]*hessin[k][i];
                hessin[j][i] = temp/diag_hessin[j];
            }
        }
        for ( i=0 ; i<n ; i++ ) {
            for ( temp=0.0,j=0 ; j<=i ; j++ ) temp += pow(hessin[i][j],2);
            diag_hessin[i] = sqrt(temp);
            free(hessin[i]);
            free(htmp[i]);
        }
    }

    free(scale);
    free(hessin);
    free(htmp);
    return;
}


double grad(int j, double *x, double tol_h, double (*func)(double *))
{
    double h, fp, fm, xstore, get_step( double, double);

    h = ( (h=tol_h*fabs(x[j]))>tol_h ) ? h : tol_h;
    xstore = x[j];
    h = get_step(xstore, h);
    x[j] = xstore+h;
    fp = func(x);
    x[j] = xstore-h;
    fm = func(x);
    x[j] = xstore;

    return 0.5*(fp-fm)/h;
}
```

```
/******************************************************************************
          Function to signal run_time errors and exit to system.
******************************************************************************/

void runtime_error(char *error_text)
{
     fprintf(stderr,"Run-time error...\n");
     fprintf(stderr,"%s\n",error_text);
     fprintf(stderr,"...now exiting to system...\n");

     exit(1);
}


/******************************************************************************

               Function which performs line minimisation.

                 VARIABLES USED IN LINE_SEARCH FUNCTION

   ON ENTRY:
     n                - number of independent parameters
     x[ ]             - current estimate of parameters at minimum
     xl[ ]            - vector holding lower bounds on parameters
     xu[ ]            - vector holding upper bounds on parameters
     xi[ ]            - scaled downhill search direction
     xn[ ]            - contains temporary estimates of x[ ]
     *y               - current estimate of minimum ( *y=func(x) )
     *iterm           - variable previously against boundary
     slope            - measures cosine of angle between search direction
                        and gradient at point x[ ]
     rnderr           - compiler rounding error
     tol_x            - tolerance on components of x[ ]
     func(double *)   - pointer to a scalar function of a vector
                        (user supplied)

   ON EXIT:
     n                - unchanged on exit
     x[ ]             - new estimate of parameters at minimum
```

```
    xl[ ]          - unchanged on exit
    xu[ ]          - unchanged on exit
    xi[ ]          - unchanged on exit
    xn[ ]          - contains a copy of x[ ]
    *y             - new estimate of minimum ( *y=func(x) )
    *iterm         - variable now against boundary (if any)
    slope          - unchanged on exit
    rnderr         - unchanged on exit
    tol_x          - unchanged on exit

****************************************************************************/

short int line_search(int n, double *x, double *xl,
               double *xu, double *xi, double *xn, double *y,
               int *iterm,  double slope, double rnderr,
               double tol_x, double *scale,
               double (*func)(double *))
{
    int i, j;
    double temp1, temp2, temp, vl_min, vl_max, vl, vll, yn, ynn, a , b, disc;
    void runtime_error(char *);

    for ( temp1=0.0,i=0 ; i<n ; i++ ) {
        temp2 = (temp=fabs(x[i])) > 1.0 ? temp : 1.0;
        if ( temp1 < (temp=fabs(xi[i])/temp2)) temp1 = temp;
    }
    vl_min = tol_x/temp1;
    vl_max = 1.0;

 /*
 **  Fix maximum vl as the smaller of unity and a value determined
 **  by the upper and lower bounds. Near the minimum, vl=1.0
 */

    for ( *iterm=-1,i=0 ; i<n ; i++ ) {
        if ( xi[i]>0.0 && vl_max>(temp=(xu[i]-x[i])/xi[i]) ) {
            vl_max = temp;
            *iterm = i;
        }
        if ( xi[i]<0.0 && vl_max>(temp=(xl[i]-x[i])/xi[i]) ) {
            vl_max = temp;
```

```
            *iterm = i;
        }
    }
    if ( vl_max<=vl_min) {
        for ( i=0 ; i<n ; i++ ) x[i] += vl_max*xi[i];
        return 1;
    } else {
        vl = 1.0;
        vl_min /= vl_max;
        for ( i=0 ; i<n ; i++ ) xi[i] *= vl_max;
    }

/*
**  Take a Newton step ...
*/

    for ( i=0 ; i<n ; i++ ) xn[i] = x[i]+vl*xi[i];
    for ( j=0 ; j<n ; j++ ) xn[j] *= scale[j];
    yn = func(xn);
    for ( j=0 ; j<n ; j++ ) xn[j] /= scale[j];
    temp = yn-*y-ALPHA*vl*slope*vl_max;
    while ( vl>vl_min && yn>*y+ALPHA*vl*slope*vl_max ) {
        if ( fabs(1.0-vl)<=rnderr ) {
            temp = -0.5*slope/(yn-*y-slope);
        } else {
            temp1 = yn-*y-vl*slope;
            temp2 = ynn-*y-vll*slope;
            a = (temp1/(vl*vl)-temp2/(vll*vll))/((vl-vll)*pow(vl_max,3));
            b = (-vll*temp1/(vl*vl)+
                vl*temp2/(vll*vll))/((vl-vll)*pow(vl_max,2));
            if ( fabs(a)<=rnderr ) {
                temp =-0.5*slope/b;
            } else {
                disc = b*b-3.0*a*slope;
                if ( disc<0.0 )
                runtime_error("\nRoundoff problems in line search\n");
                temp = (-b+sqrt(disc))/(3.0*a);
                if ( temp>(temp1=0.5*vl) ) temp = temp1;
            }
        }
        vll = vl;
```

```
        ynn = yn;
        vl = ( temp>(temp1=0.1*vl) ) ? temp : temp1;
        for ( i=0 ; i<n ; i++ ) xn[i] = x[i]+vl*xi[i];
        for ( j=0 ; j<n ; j++ ) xn[j] *= scale[j];
        yn = func(xn);
        for ( j=0 ; j<n ; j++ ) xn[j] /= scale[j];
    }
    for ( *y=yn,i=0 ; i<n ; i++ ) {
        x[i] = xn[i];
        xi[i] *= vl;
    }

    return 0;
}


/****************************************************************************
        Function calculates numerical gradients of func(x[ ]) at x[ ]
****************************************************************************/

void dfcn(int n, double *x, double *g, double tol_h, double *scale,
          double (*func)(double *))
{
    int i, j;
    double h, fp, fm, xstore, get_step( double, double);

    for ( i=0 ; i<n ; i++ ) {
        h = ( (h=tol_h*fabs(x[i]))>tol_h ) ? h : tol_h;
        xstore = x[i];
        h = get_step(xstore, h);
        x[i] = xstore+h;
        for ( j=0 ; j<n ; j++ ) x[j] *= scale[j];
        fp = func(x);
        for ( j=0 ; j<n ; j++ ) x[j] /= scale[j];
        x[i] = xstore-h;
        for ( j=0 ; j<n ; j++ ) x[j] *= scale[j];
        fm = func(x);
        for ( j=0 ; j<n ; j++ ) x[j] /= scale[j];
        x[i] = xstore;
        g[i] = 0.5*(fp-fm)/h;
    }
```

```
    return;
}


/***************************************************************************
      Function to ensure stepsize in gradient is machine representable.
      Provided as a separate function to deceive optimising compilers.
***************************************************************************/

double get_step(double x, double h)
{
    double temp;
    temp = x+h;

    return temp-x;
}


/***************************************************************************
      Function to check if updated Hessian is positive definite.
      Tries to perform a Choleski decomposition on hessin[ ][ ].
***************************************************************************/

short int pd_fail( int n, double *p, double **hessin, double **bb )
{
    int i, j, k;
    double sigma;

    for ( i=0 ; i<n ; i++ ) {
        for ( j=0 ; j<n ; j++ ) bb[i][j] = hessin[i][j];
    }
    for ( i=0 ; i<n ; i++ ) {
        for ( j=i ; j<n ; j++ ) {
            for ( sigma=bb[i][j],k=i-1 ; k>=0 ; k-- )
                 sigma -= bb[i][k]*bb[j][k];
            if ( i==j ) {
                if ( sigma<=0.0 ) return 1;
                p[i] = sqrt(sigma);
            } else {
                bb[j][i] = sigma/p[i];
```

```
            }
        }
    }

    return 0;
}
```

The source code evaluating Dixit (1991a)'s optimal h and cost function

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <malloc.h>


/**************************************************************************
     In defining the characteristic root of the ODE, we set SIGMA=0.1
        and RHO=0.05, therefore BETA=sqrt(2*RHO)/SIGMA=sqrt(10.0).
**************************************************************************/

#define SIGMA       0.1
#define RHO         0.05
#define BETA        sqrt(2.0*RHO)/SIGMA
#define K           0.5
#define G           0.1
#define GAMMA       (G*pow(RHO,2))/(2.0*K*pow(SIGMA,2))


double xval;     /*  xval initial position of state  */

void main( void )
{
    double  func(double), dixit(double), h;
    int     i;

    for ( h=1.0e-6 ; func(h)>0.0 ; h+=1.0e-6 );
    for( i=0; i<=3; i++){
         xval = 0.1*((double) i);
```

```
        printf("\n H is %15.6lf\t DIXITS VAL %10.6lf\t", h, dixit(h));
    }

    exit(0);
}


/***************************************************************************
       Function providing the 'optimal' value for Dixit (1991a)'s h
***************************************************************************/

double func(double h)
{
     return (G*RHO-K*pow(h,2))*BETA*sinh(BETA*h)+
               2.0*K*h*(cosh(BETA*h)-1.0);
}

/***************************************************************************
       The solution to the HJB(ODE) equation obtained Dixit (1991a)
***************************************************************************/

double dixit(double h)
{
    double temp;

    temp = -2.0*K*h/(RHO*BETA*sinh(BETA*h));
    temp *= cosh(BETA*xval);
    temp += K*pow(xval,2)/RHO;
    temp += K*pow((SIGMA/RHO),2);

    return temp;
}
```

Technical Annex 6-The Source Code Chapter 4

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <malloc.h>


#define  N              20     /***No. of Nodes used in Chebyshev Polyn.***/
#define  VAR        2.5e-3     /***Variance of Geometric Brownian Motion***/
#define  MU            0.0     /***Mean of Geometric Brownian Motion***/
#define  RHO          0.05     /***Constant Continuous Discount Rate***/
#define  WEALTH        1.0     /***Initial Level of Wealth***/
#define  G          1.0e-3     /***Lump Sum Transaction Cost***/
#define  OMEGA       0.010     /***Prop. Trans. Cost at Upper Boundary***/
#define  DELTA       0.030     /***Prop. Trans. Cost at Lower Boundary***/
#define  ODRATE     2.0e-2     /***Overdraft Premium is 2%***/


double  mval, alpha_p, alpha_m, trigs[N+1][N+1], diff[N+1][N+1], cash,
        p[2];


main(void)
{
    int     i, j;
    double  pi_by_n, func_1(double *);
    char    filename[13];
    FILE    *output;

/*  Create filename appropriate to application */
    filename[0] = 'M';
    filename[1] = 'U';
    filename[2] = 'O';
    filename[3] = 'R';
    filename[4] = 'E';
    filename[5] = 'S';
    filename[6] = 'U';
```

```
    filename[7] = 'L';
    filename[8] = 'T';
    filename[9] = 'S';
    filename[10] = '.';
    filename[11] = 'D';
    filename[12] = 'A';
    filename[13] = 'T';
    printf("\nFilename in %12s",filename);

    pi_by_n = PI/((double) N);

    for( j=0; j<=N; j++)
        for( i=0; i<=N; i++)
            trigs[i][j] = cos(pi_by_n*((double) i*j));

    cash = 0.0;

    p[0] = -0.02;
    for ( j=1; j<=40 ; j++){
        output=fopen(filename,"a");
        p[1] = cash+0.001*((double) j);
        fprintf(output,"%10.6lf %10.6lf %10.6lf\n", func_1(p), p[0], p[1]);
        fclose(output);
    }

    exit(0);
}


/**************************************************************************
    Function To Be Optimised Which Also Contains The Initial Value Problem
**************************************************************************/

double func_1( double *pval )
{
    static int  start = 1;
    int         ifail, integrate( double, double, double, double *,
                double (*func)(double)), i, j;
    double      fcn( double), temp, m[N+1], v[N+1], value, r_in, r_out;
    static double  theta, tmp1, tmp2, pi_by_n, tol=1.0e-11;
    void        fprime(double, double *, double *);
```

```
    void        coeffs( int, double *, double *);
    void        fsolve(int, double, double *, double, double *, void (*fvalue)
                (double, double *, double *));


/****Compute Values for alpha_p and alpha_m for the Initial Value Problem****/
    if( start ){
        tmp1 = MU/VAR-0.5;
        tmp2 = 2.0*RHO/VAR;
        alpha_p = -tmp1+sqrt(tmp2+pow(tmp1,2));
        alpha_m = -tmp1-sqrt(tmp2+pow(tmp1,2));
        pi_by_n = PI/((double) N);
        start = 0;
    }
/**********************Solution of Initial Value Problem********************/
    for ( i=0 ; i<=N ; i++ ) {
        theta = 0.5*pi_by_n*((double) i);
        mval = p[0]+(p[1]-p[0])*pow(cos(theta),2);
        m[i] = mval;
        mval += WEALTH;
        ifail = integrate( 0.0, 1.0, tol, &value, fcn);
        if ( ifail != 0 ) {
            printf("Integration problem ! \n\n");
            return;
        } else {
            temp  = mval/(VAR*sqrt(tmp2+pow(tmp1,2)));
            v[i]  = temp*(value+(1.0/pow(alpha_p,2))/(1.0+mval));
            v[i] += log(1.0+mval)/RHO;
        }
    }
    r_in = 0.0;
    r_out = 4.5e-2;
/********Solution Being Obtained To The Entire Problem To Be Optimised*******/
    fsolve( N+1, tol, &r_in, r_out, v, fprime);
    coeffs( N+1, v, m);
    value = 2.0*(cash - p[0])/(p[1] - p[0]) - 1.0;
/**************Value is the Spectrally Transformed Value of Cash************/
    for( temp=m[0], j=1 ; j<=N ; j++)
       temp += m[j]*cos(j*acos(value));

    return temp;
```

```
}
void fprime( double r, double *v, double *dv)
{
    double          temp, temp_2, cj, ck, vv[N+1], rate(double);
    int             k, j;
    static int      start=1;
    static double   pi_by_n, pi_by_2n, con_1[N+1], con_2[N+1], con_3,
                    con_4, f[N+1];
    void            coeffs( int, double *, double *);

    if ( start ) {
/*************Delivers Chebychev-Lobatto differentiation matrix.***********/
        con_3 = RHO;
        con_4 = -1.0;
        pi_by_n = PI/((double) N);
        pi_by_2n = 0.5*PI/((double) N);
        for ( j=0 ; j<=N ; j++ ) {
            cj = 1.0;
            if ( j==0 || j==N ) cj = 2.0;
            for ( k=0 ; k<=N ; k++ ) {
                ck = 1.0;
                if ( k==0 || k==N ) ck = 2.0;
                if ( j!=k ) {
                    temp = 2.0*sin(pi_by_2n*((double) k+j))*
                               sin(pi_by_2n*((double) k-j));
                    diff[j][k] = (cj/ck)*pow(-1.0,j+k)/temp;
                } else if ( j==0 ) {
                    diff[0][0] = (2.0*pow(((double) N),2)+1.0)/6.0;
                } else if ( j==N ) {
                    diff[N][N] = -(2.0*pow(((double) N),2)+1.0)/6.0;
                } else {
                    temp = 2.0*pi_by_2n*((double) j);
                    diff[k][j] = -0.5/(tan(temp)*sin(temp));
                }
            }
        }
        start = 0;
    }
/*****************Treatement of Stationary Condition********************/

    for ( j=0 ; j<=N ; j++ ) {
```

```
        temp = cos(pi_by_n*((double) j));
        temp_2 = WEALTH+p[0]+0.5*(p[1]-p[0])*(1.0+temp);
        if(WEALTH <= temp_2){
            f[j] = con_4*log(1.0+temp_2);
        }else {
            f[j] = con_4*log(1.0+
            WEALTH+(1.0-(ODRATE+r))*(p[0]+0.5*(p[1]-p[0])*(1.0+temp)));
        }
        con_1[j] = 0.5*VAR*con_4*pow(temp,2);
        con_2[j] = MU*con_4*temp;
/*        printf("\n A is  %20.18lf\t CON_2 is %20.18lf\t", p[0], con_2[2]);
        getchar();   */
    }


    if ( rate(r) == 0.0 ){
          for( j=0 ; j<=N ; j++ ) dv[j] = 0.0;
        return;
    }
/******************Application of Boundary Conditions*********************/
    if ( r > 0.0 ) {
        do{
            temp = v[0];
            cj = v[N];
            coeffs(N+1,v,vv);
            for( ck=vv[0], j=1 ; j<=N ; j++)
                ck += vv[k]*cos(j*acos((p[1]+p[0])/(p[0]-p[1])));
            v[0] = ck+log(1.0+G+OMEGA*fabs(p[1]));
            v[N] = ck-log(1.0+G+DELTA*fabs(p[0]));
        }while( fabs(temp-v[0]) > 1.0e-12 && fabs(cj-v[N]) > 1.0e-12 );
    }

    for( j=0 ; j<=N ; j++ ) {
        for( temp=0.0,k=0 ; k<=N ; k++ ) temp += diff[j][k]*v[k];
        vv[j] = temp;
    }

    for ( j=0 ; j<=N ; j++ ) {
        for ( temp=0.0,k=0 ; k<=N ; k++ ) temp += diff[j][k]*vv[k];
        dv[j] = (f[j]+con_1[j]*temp+con_2[j]*vv[j]+con_3*v[j])/(rate(r));
    }
```

```
/*  printf("\n Value of V(a) is %20.18lf\t", v[0]);
    getchar();  */
    return;
}


/****************************************************************************
                       The Function to be integrated
****************************************************************************/

double fcn( double x)
{
    double tmp1, tmp2;

    tmp1 = pow(x,-alpha_m)/(1.0+mval*x);
    tmp2 = pow(x,alpha_p)/pow(mval+x,2);

    return tmp1/alpha_m+tmp2/pow(alpha_p,2);
}


/****************************************************************************
                  Adaptive Simpson's rule integrator
****************************************************************************/

#define MAXSPL  30
#define MINSPL  5
#define RNDERR  5.e-16

int integrate(double a, double b, double eps, double *quad,
              double (*func)(double))
 /*
 **                         Return codes
 **                         ------------
 **
 **     Return  0 ... Regular exit.
 **     Return -1 ... Interval of Integration has zero length
 **     Return -2 ... Tolerance is either negative or unachievable
 **     Return -3 ... Error tolerance has not been met within the
 **                   allowed iterations.
```

```
 */
{
    double area=0.0, valold=0.0, hnow, tol, tolerr, vlower, valnew,
           valdif, x[5], f[5], v[MAXSPL], xstore[3][MAXSPL],
           fstore[3][MAXSPL];
    int finish=1, j, nsplit;
    long number;
    if ( fabs(b-a) <= RNDERR ) return -1;
    if ( eps <= RNDERR ) return -2;
    number=pow(2,MINSPL);
    tol = (30.0*eps)/(b-a);
    *quad = 0.0;
    x[0] = a;
    x[2] = 0.5*(a+b);
    x[4] = b;
    f[0] = (*func)(x[0]);
    f[2] = (*func)(x[2]);
    f[4] = (*func)(x[4]);
    for ( nsplit=0 ; nsplit<MINSPL ; nsplit++ ) {
        x[1] = 0.5*(x[0]+x[2]);
        x[3] = 0.5*(x[2]+x[4]);
        f[1] = (*func)(x[1]);
        f[3] = (*func)(x[3]);
        hnow = (x[4]-x[0])/12.0;
        vlower = hnow*(f[0]+4.0*f[1]+f[2]);
        v[nsplit] = hnow*(f[2]+4.0*f[3]+f[4]);
        valnew = vlower+v[nsplit];
        valdif = valnew-valold;
        area = area+valdif;
        for ( j=0 ; j<=2 ; j++ ) {
            xstore[j][nsplit] = x[j+2];
            fstore[j][nsplit] = f[j+2];
        }
        valold = vlower;
        x[4] = x[2];
        f[4] = f[2];
        x[2] = x[1];
        f[2] = f[1];
    }
    nsplit--;
    while ( finish==1 ) {
```

```
x[1] = 0.5*(x[0]+x[2]);
f[1] = (*func)(x[1]);
x[3] = 0.5*(x[2]+x[4]);
f[3] = (*func)(x[3]);
hnow = (x[4]-x[0])/12.0;
vlower = hnow*(f[0]+4.0*f[1]+f[2]);
v[nsplit+1] = hnow*(f[2]+4.0*f[3]+f[4]);
valnew = vlower+v[nsplit+1];
valdif = valnew-valold;
area = area+valdif;
tolerr = tol*hnow*fabs(area);
tolerr = (tolerr > tol) ? tolerr : tol;
if ( (nsplit+1)>=MAXSPL ) {
    return -3;
} else if ( fabs(valdif)<=tolerr ) {
    *quad += (valnew+valdif/15.0);
    while ( number != (2*(number/2)) ) {
        number = number/2;
        nsplit--;
    }
    number++;
    if ( nsplit<0 ) {
        finish = 0;
    } else {
        valold = v[nsplit];
        x[0] = x[4];
        f[0] = f[4];
        for ( j=0 ; j<=2 ; j++ ) {
            f[2*j] = fstore[j][nsplit];
            x[2*j] = xstore[j][nsplit];
        }
    }
} else {
    number *= 2;
    nsplit++;
    for ( j=0 ; j<=2 ; j++ ) {
        xstore[j][nsplit] = x[j+2];
        fstore[j][nsplit] = f[j+2];
    }
    valold = vlower;
    x[4] = x[2];
```

```
            f[4] = f[2];
            x[2] = x[1];
            f[2] = f[1];
        }
    }

    return 0;
}


/***************************************************************************
                          FUNCTION INTEGRATING PDE
***************************************************************************/

void fsolve(int n, double tol, double *astart, double aend, double *y,
            void (*fvalue)(double, double *, double *) )
{
    double d1, hmin, hmax, tolerr, temp, hnow, range, errmax, errest, ain;
    double size(int, double *);
    double rnderr, **w;
    int reduce, finish=0, returnval=0;
    int odeint(int, double *, double *, double *, double **, double,
         double, double, void (*fvalue)(double, double *, double *));
    void fsolve_err(int);

    w = (double **) malloc( 7*sizeof(double *) );
    if (!w) fsolve_err(-3);
    for ( reduce=0 ; reduce<7 ; reduce++ ) {
        w[reduce] = (double *) malloc( n*sizeof(double) );
        if (!w[reduce]) fsolve_err(-3);
    }
    rnderr = 1.0;
    while ( rnderr+1.0!=1.0 ) rnderr *= 0.5;
    rnderr *= 2.0;
    range = aend-*astart;
    hmax = fabs(range);
    if ( hmax<=rnderr ) fsolve_err(-1);
    if ( tol==0.0 ) {
       tolerr = rnderr;
    } else {
       tolerr = tol;
```

```
}
hmin = fabs(*astart);
d1 = fabs(aend);
hmin = (hmin > d1) ? hmin : d1;
d1 = (hmin > 1.0) ? hmin : 1.0;
hmin = d1 * pow(rnderr,0.33);
if (hmin >= hmax) fsolve_err(-2);
d1 = tol * size(n, y);
errest = (d1 > tol) ? d1 : tol;
(*fvalue)(*astart,y,w[0]);
errmax = size(n,w[0]);
temp = errmax*pow(hmax,5);
hnow = hmax;
if (errest < temp) {
    d1 = fabs(*astart);
    hnow = (hmax > d1) ? hmax : d1;
    hnow = tol * hnow;
    d1 = errest/errmax;
    temp = pow(d1,0.2);
    hnow = (hnow > temp) ? hnow : temp;
}
d1 = (range > 0.0) ? 1.0 : -1.0;
hnow *= d1;
ain = *astart;
while ( !finish ) {
    if ((ain+hnow-aend)*range >= 0.0) {
        hnow = aend-ain;
        finish = 1;
    }
    reduce = odeint(n, &ain, &hnow, y, w, hmin, hmax, tolerr, fvalue);
    if ( fabs(hnow)==hmin ) returnval = 1;
    if ( !reduce && finish ) {
        *astart = aend;
    } else {
        finish = 0;
    }
}
if ( returnval==1 ) fsolve_err(1);
for ( reduce=0 ; reduce<7 ; reduce++ ) free(w[reduce]);
free(w);
```

```
    return;
}


void fsolve_err(int error_code)
{
    if ( error_code>0 ) {
        printf("\n\nWARNING error in fsolve");
        if ( error_code==1 )
           printf("\nIntegration completed but requested accuracy not met!");
        if ( error_code==2 )
           printf("\nMore memory allocated than is necessary!");

        return;
    } else {
        printf("\n\nFATAL execution error in fsolve");
        if ( error_code==-1 )
             printf("\nZero range of integration!");
        if ( error_code==-2 )
             printf("\nEffective range of integration is zero!");
        if ( error_code==-3 )
             printf("\nMemory allocation failure!");
        if ( error_code==-4 )
             printf("\nOrder of equations increased - reallocation memory!");

        exit(1);
    }
}


int odeint(int n, double *a, double *h, double *y, double **w, double hmin,
        double hmax, double tolerr, void (*fcn)(double, double *, double *))
{
    double d1, d2, hval, tolest;
    int j, i;
    void rkck(int, double, double, double *, double **,
             void (*fcn)( double, double *, double *));

    rkck(n, *a, *h, y, w, fcn);
    tolest = 0.0;
    for (j = 0; j < n; j++ ) {
```

```
        d1 = fabs(w[1][j]);
        d2 = fabs(w[6][j]);
        if ( d1>1.0) d2 = d2/d1;
        tolest = ( tolest > d2 ) ? tolest : d2;
    }
    if (tolest >= tolerr) {
        if (tolest >= tolerr * 59049.0) {
            *h *= 0.1;
        } else {
            d1 = tolest / tolerr;
            *h *= 0.9 / pow(d1,0.2);
        }
        if (fabs(*h) <= hmin) {
            *a += *h;
            for (j = 0; j < n; j++ ) y[j] = w[1][j];
            d1 = (*h > 0.0 ) ? 1.0 : -1.0;
            *h = d1*hmin;
            return 0;
        }

        return 1;
    }
    *a += *h;
    for (j = 0; j < n; j++ ) y[j] = w[1][j];
    if (tolest <= tolerr * 1.889568e-4) {
        *h *= 5.0;
    } else {
        d1 = tolerr / tolest;
        *h *= 0.9 * pow(d1,0.2);
    }
    d1 = fabs(*h);
    hval = (hmin > d1) ? hmin : d1;
    hval = (hmax > hval) ? hval : hmax;
    d1 = (*h > 0.0) ? 1.0 : -1.0;
    *h = d1*hval;

    return 0;
}


double size(int n, double *y)
```

```
{
    double vsize, d1;
    int j;

    vsize = fabs(y[0]);
    for (j = 1; j < n; j++ ) {
        d1 = fabs(y[j]);
        vsize = (vsize > d1 ) ? vsize : d1;
    }

    return vsize;
}


void rkck(int n, double a, double h, double *y, double **w,
          void (*fprime)(double, double *, double *))
{
    static double b1=-11.0/54.0,b2=2.5,b3=-70.0/27.0,b4=35.0/27.0,
                  c1=1631.0/55296.0,c2=175.0/512.0,c3=575.0/13824.0,
                  c4=44275.0/110592.0,c5=253.0/4096.0,
                  d1=37.0/378.0,d2=250.0/621.0,d3=125.0/594.0,d4=512.0/1771.0,
                  e1=-277.0/64512.0,e2=6925.0/370944.0,e3=-6925.0/202752.0,
                  e4=-277.0/14336.0,e5=277.0/7084.0;
    int j;
    double tmp1, tmp2, tmp3, tmp4, tmp5, tmp6;

    (*fprime)(a,y,w[0]);
    tmp1 = h*0.2;
    for ( j=0 ; j<n ; j++ ) w[6][j] = y[j]+w[0][j]*tmp1;
    (*fprime)(a+h*0.2,w[6],w[1]);
    tmp1 = 0.075*h;
    tmp2 = 0.225*h;
    for ( j=0 ; j<n ; j++ ) w[6][j] = y[j]+w[0][j]*tmp1+w[1][j]*tmp2;
    (*fprime)(a+0.3*h,w[6],w[2]);
    tmp1 = 0.3*h;
    tmp2 = -0.9*h;
    tmp3 = 1.2*h;
    for ( j=0 ; j<n ; j++ ) {
        w[6][j] = y[j]+w[0][j]*tmp1+w[1][j]*tmp2+w[2][j]*tmp3;
    }
    (*fprime)(a+0.6*h,w[6],w[3]);
```

```
    tmp1 = h*b1;
    tmp2 = h*b2;
    tmp3 = h*b3;
    tmp4 = h*b4;
    for ( j=0 ; j<n ; j++ ) {
        w[6][j] = y[j]+w[0][j]*tmp1+w[1][j]*tmp2+w[2][j]*tmp3+w[3][j]*tmp4;
    }
    (*fprime)(a+h,w[6],w[4]);
    tmp1 = h*c1;
    tmp2 = h*c2;
    tmp3 = h*c3;
    tmp4 = h*c4;
    tmp5 = h*c5;
    for ( j=0 ; j<n ; j++ ) {
        w[6][j] = y[j]+w[0][j]*tmp1+w[1][j]*tmp2+w[2][j]*tmp3
                  +w[3][j]*tmp4+w[4][j]*tmp5;
    }
    (*fprime)(a+0.875*h,w[6],w[5]);
    tmp1 = h*d1;
    tmp2 = h*d2;
    tmp3 = h*d3;
    tmp4 = h*d4;
    for (j = 0 ; j<n ; j++ ) {
        w[1][j] = y[j]+w[0][j]*tmp1+w[2][j]*tmp2+w[3][j]*tmp3+w[5][j]*tmp4;
    }
    tmp1 = h*e1;
    tmp2 = h*e2;
    tmp3 = h*e3;
    tmp4 = h*e4;
    tmp5 = h*e5;
    for (j=0 ; j<n ; j++ ) {
        w[6][j] = w[0][j]*tmp1+w[2][j]*tmp2+w[3][j]*tmp3+w[4][j]*tmp4
                  +w[5][j]*tmp5;
    }

    return;
}


/****************************************************************************
                 Sub-Routine evaluating Spectral Coefficients
```

```
*****************************************************************************/

void coeffs( int n, double *v, double *v_hat )
{
unsigned k, j;
double sum;
for ( k=0; k<=n; k++ ){
for ( sum=0.0, j=1; j<n; j++ )
            sum += v[j]*trigs[j][k];
        sum += 0.5*(v[0] + v[n]*cos(PI*(double) k));
        v_hat[k] = 2.0*sum/((double) N);
}
    v_hat[0] *= 0.5;
    v_hat[n] *= 0.5;  /* we halve the nth coefficient because we
                         are using a Gauss-Labboto quadrature */

    return;
}


/*****************************************************************************
                 Coefficient function for opportunity rate
*****************************************************************************/

double rate(double r)
{
/*      return cos(PI*r/0.20)-sin(PI*r/0.20);  */
        return -1.0;
}
```

References

Akerlof, G. and Yellen, J. (1985), "A Near-Rational Model of the Business Cycle with Wage and Price Inertia", *Quarterly Journal of Economics*, 100, 823-838.

Artis, M. J. and Lewis, M. K. (1976), "The Demand for Money in the U.K. 1963-73", *Manchester School*, 7, 69-83.

Artis, M. J. and Lewis, M. K. (1981) *Monetary Control in the UK* (Oxford: Phillip Allan).

Artis, M. J. and Lewis, M. K. (1984), "How Unstable is the Money in the United Kingdom?", *Economica*, 51, 473-6.

Baba, Y., Hendry, D. F. and Starr, R. M. (1988), "US money demand 1960-1984", *Disscussion Paper 27*, Nuffield College, Oxgord.

Baumol, W. J. (1952), "The Transactions Demand for Cash: an Inventory Theoretic Approach", *Quarterly Journal of Economics*, 66, 545-56.

Bensoussan, A. and Lions, J. L. (1975a), "Novelles Methodes en Controle Impulsionnel", *Applied Mathematical Optimization*, 1, 289-312.

Bensoussan, A. and Lions, J. L. (1975b), "Problems de Temp D'erral Optimaux et de Perturbations Singuilieres Dans les Inequations Variationelles", Proc. Internat. Symp. on Control, Numerical Methods and Computer Systems Modelling, IRIA, June 1974, *Springer Lecture Notes in Econ. and Math. Systems*, 107.

Blanchrd, O. J. (1987), "Individual and Aggregate Price Adjustment", *Brookings Papers on Economic Activity*, 1, 57-122.

Blanchard, O. J. and Kiyotaki, N. (1987), "Monopolistic Competition and the Effects of Aggregate Demand", *American Economic Review*, 77, 647-666.

Borch, K. (1969), "A Note of Uncertainty and Indifference Curves", *Review of Economic Studies*, 36, 1-4.

Bratley, P., Fox, B. L. and Scharge, E. L. (1983), *A Guide to Simulation*

(New York: Springer-Verlag).

Bray, M. (1982), "Learning Estimation and the Stability of Rational Expectations Models", *Journal of Economic Theory*, 26, 318-339.

Caplin, A., and Spulber, D. (1987), "Menu Costs and the Neutrality of Money", *Quarterly Journal of Economic*, 102, 703-726.

Canuto, C., Hussaini, M. Y., Quateroni, A., and Zang, T. A. (1988), *Spectral Methods in Fluid Dynamics* (Berlin Heidelberg: Springer-Verlag)

Constantinides, G. M. and Richard, S. F. (1978), "Existence of Optimal Simple Policies for Discounted Cost Inventory and Cash Management in Continuous Time", *Operations Research*, 26, 620-636.

Cooley, J. W. and Tukey, J. W. (1965), "An Algorithm for the Machine Calculation of Complex Fourier Series" *Mathematics of Computation*, 19, 297-301.

Courakis, A. S. (1988), "Modelling Portfolio Selection", *Economic Journal*, 98, 619-42.

Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes* (London: Methuen.)

Cuthbertson, K. (1985) *The Supply and Demand for Money* (Oxford: Basil Blackwell).

Cuthbertson, K. (1986), "Price Expectations and Lags in the Demand for Money", *Scottish Journal of Political Economy*, 33, 334-54.

Cuthbertson, K. and Taylor, M. P. (1986), "Monetary Anticipations and the Demand for Money in the UK: Testing Rationality in the Shock Absorber Hypothesis", *Journal of Applied Econometrics*, 1 (2), 1-11.

Dalal, A. J. (1983), "Comparitive Statics and Asset Substitutability/Complemenatry in a Portfolio Model: a Dual Approach", *Review of Economic Studies*, 50, 355-67.

Davis, P. J. and Rabinowitz, P. (1984), *Methods of Numerical Integration* 2nd Ed. (London: Academic Press)

De Canio, S. (1979), "Rational Expectations and Learning From Experience", *Quarterly Journal of Economics*, 93, 47-58.

Dennis Jnr, J. E. and Schnabel, R. B. (1989), "A View of Unconstrained Optimisation", in Nemhauser, G. L. et al, Eds, *Handbooks in Operations Research and Management Science, Vol. 1* (Amsterdam: Elsevier Science Publishers B.V. North-Holland), 1-66.

Dixit, A. (1991a), "Analytical Approximations in Models of Hysteresis", *Review of Economic Studies*, 58, 141-151.

Dixit, A. (1991b), "A Simplified Treatment of the Theory of Optimal Control of Brownian Motion", *Journal of Economics Dynamics and Control*, 15, 657-673.

Dixit, A. (1992), "Investment and Hysteresis", *Journal of Economic Perspectives*, 6, 107-132.

Dixit, A. (1994), *Investment Under Uncertainty* (Princeton: Princeton University Press).

Dumas, B. (1991), "Super Contact and Related Optimality Conditions", *Journal of Economic Dynamics and Control*, 15, 675-685.

Dutta, P., Majumdar, M. K. and Sundaram, R. K. (1994), "Parametric Continuity in Dynamic Programming Problems", *Journal of Economic Dynamics and Control*, 18, 1063-1092.

Engle, R. F. and Granger, C.W.J. (1987) "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 391-407.

Evans, G. (1985), "Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models", *Quaterly Journal of Economics*, 100, 1217-1233.

Feldstien, M. S. (1969), "Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection", *Review of Economic Studies*, 36, 5-12.

Feller, W. (1952), "The Parabolic Differential Equations and the Associated Semigroups of Transformations", *Annals of Mathematics*, 55, 468-519.

Feller, W. (1954), "Diffusion Processes in One Dimension", *Transactions of the American Mathematical Society*, 77, 1-31.

Feller, W. (1957), "On Boundaries and Lateral Conditions for the Kolmogorov Differential Equations", *Annals of Mathematics*, 65, 527-570.

Fleming, W. H. and Rishel, R. W. (1975), *Deterministic and Stochastic Optimal Control* (Berlin: Springer-Verlag).

Flood, R. P. and Garber, P. M. (1991), "The Linkage Between Speculative Attack and Target Zone Models of Exchange Rate Rules", *Quarterly Journal of Economics*, 106, 1367-72.

Flood, R. P. and Garber, P. M. (1991), "The Linkage Between Speculative Attack and Target Zone Models of Exchange Rate Rules: Some Extended Results" in Krugman, P., and Miller, M. eds., *Exchange Rate Targets and Currency Bands* (Camdridge: Cambridge University Press).

Frenkel, J. A. and Jovanovic, B. (1980), "On Transactions and Precautionary Demand for Money", *Quarterly Journal of Economics*, 95, 25-43.

Gihman, I. I. and Skorohod, A. V. (1972), *Stochastic Differential Equations* (New York: Springer-Verlag).

Goldfeld, S. M. (1976), "The Case of the Missing Money", *Brookings Papers on Economic Activity*, 3, 683-730.

Gordon, R. J. (1984), "The Short-Run Deand for Money: A Reconsideration", *Journal of Money, Credit and Banking*, 16, 403-404.

Grossman, S. J. and Laroque, G. (1990), "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods", *Econmetrica*, 58, 25-52.

Harrison, J. M., Selke, T.M. and Taylor, A.J. (1983), "Impulse Control of Brownian Motion", *Mathematics of Operations Research*, 8, 454-466.

Haussmann, U. G. and Lepeltier, P. (1990), "On the Existence of Optimal Controls", *SIAM Journal of Control and Optimisation*, 28, 851-902.

Hendry, D. F. and Ericsson, N. R. (1983), "Assertion Without Emperical Basis: an Econometric Appraisal of Friedman and Schwartz 'Money Trends in...United Kingdom' ", Panel Papper 22, Bank of England, Panel of Academic Consultants, October 1978.

Hendry, D. F. and Ericsson, N. R. (1988), "An Econometric Analysis of UK Money Demand" in *Monetary Trends in the United States and the United Kingdom* by Friedman, M. and Schwartz. Mimeo, Nuffield College, Oxford, November.

Hendry, D. F. (1979), "Predictive Failure in Economic Modelling in Macroeconomics: the Transactions Demand for Money" in P. Omerod (ed.), *Economic Modelling* (London: Heinemann).

Hendry, D. F. (1985), "Monetary Economic Myth and Reality", *Oxford Review of Economic Policy* (Oxford: Oxford University Press)

Hurn, A. S. and Muscatelli, V. A. (1991a) "Modelling the Demand for M0 and M4 in the UK", *University of Glasgow Discussion Papers in Economics*, 9114.

Hurn, A. S. and Muscatelli, V. A. (1991b) "Testing Superexogeneity: The Demand for Broad Money in the UK", *University of Glasgow Discussion Papers in Economics*, 9115.

Karlin, S. and McGregor, J. L. (1957), "The Differential Equations of Birth-and-Death Processes and the Steltjes Moment Problem", *Transaction of the American Mathematical Society*, 87, 489-546.

Karlin, S. and McGregor, J. L. (1958), "The Classification of Birth Death Processes", *Transaction of the American Mathematical Society*, 86, 366-400.

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money* (London: Macmillan).

Knuth, D. E. (1981), *Seminumerical Algorithms*, 2nd ed., vol. 2 of *The Art of Computer Programming* (Reading, MA: Addison-Wesley).

Krugman, P. (1988), "Target zones and exchange rate dynamics" *Working paper no.281* (National Bureau of Economic Research, Cambridge, MA).

Krugman, P. (1991), "Target Zones and Exchange Rate Dynamics", *Quaterely Journal of Economics*, 106, 669-81.

Laidler, D. E. W. (1980) "The Demand for Money in the United State-yet again", in Brunner, K. and Metzler, A. (eds.), *On the State of Macroeconomics*, Carnegie-Rochester Conference Series on Public Policy, 12 (Amsterdam: North-Holland).

Lee, K. C., Pesaran, M. H. and Pierse R. G. (1990), "Testing for Bias in Linear Models", *Economic Journal*, 100 (Conference 1990), 137-150.

Mankiw, N. G. (1985), "Small Menu Costs and Large Business Cycles", *Quarterly Journal of Economics*, 100, 529-539.

Markowitz, H. (1952), "Portfolio Selection", *Journal of Finance*, 7, 77-91.

Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investment* (New York: Wiley).

Merton, R. C. (1973), "An intertemporal asset pricing model", *Econometrica*, 41, 867-89.

Milbourne, R. D. (1983), "Price Expectations and the Demand for Money: Resolution of a Paradox", *Review of Economics and Statistics*, 65, 633-8.

Milbourne, R. D., Buckholtz, P., and Wasan, M. T. (1983), "A Theoretical Derivation of the Functional Form of Short Run Money Holdings", *Review of Economic Studies*, 50, 531-541.

Miller, M. and Orr, D. (1966), "A model of the demand for money by firms", *Quarterly Journal of Economics*, 109, 68-72.

Muscatelli, V. A. (1989), "Cointegration and Long-Run Properties of the Demand for Money in the UK", *University of Glasgow Discussion Papers in*

Economics, 8909.

Peng, S. (1992), "Stochastic Hamilton-Jacobi-Bellman Equations", *SIAM Journal Control and Optimisation*, 30, 284-304.

Pesaran, M. H., Pierse, R. G. and Kumar, M. S. (1989), "Econometric Analysis of Aggregation in the Context of Linear Prediction Models", *Econometric*, 57, 861-888.

Pindyck, R. S. (1988), "Irreversible investment, capacity choice, and the value of the firm", *American Economic Association*, 78, 969-985.

Pindyck, R. S. (1991), "Irreversibility, Uncertainty, and Investment", *Journal of Economic Literature*, 29, 1110-1152.

Richard, S.F. (1977), "Optimal Impulse Control of a Diffusion Process with Both Fixed and Proportional Costs of Control", *SIAM Journal of Control and Optimisation*, 15, 79-91.

Rotemberg, J. (1983), "Aggregate Consequances of Fixed Costs of Price Adjustment", *American Economic Review*, 73, 343-346.

Scarf, H. (1960), "The Optimality os (S, s) Policies in the Dynamic Inventory Problem", in Arrow K. J., Karlin S., and Suppes P., eds., *Mathematical Methods of Social Sienecs* (Stanford University Press, Stanford, CA)

Sivananthan, A. and Lindsay, K. A. (1996), "Zero Threshold Restoration of an Infinite Horizon 'Impulse' Controlled Stochastic System", *University of Glasgow Discussion Papers in Economics*, 9607.

Sivananthan, A. and Lindsay, K. A. (1996), "Boundary Conditions for Optimal 'Impulse' Control of a One Dimensional Wiener Process", *University of Glasgow Discussion Papers in Economics*, 9607.

Smith, G. W. (1989), "Transactions Demand for Money with a Stochastic, Time-Varying Interest Rate", *Review of Economic Studies*, 56, 623-633.

Stoer, J. and Bulirsch, S. (1980), *Introduction to Numerical Analysis* (New York: Springer-Verlag).

Tobin, J. (1956), "The Interest Elasticity of Transactions Demand for Cash", *Review of Economics and Statistics*, 38, 241-247.

Tobin, J. (1958), "Liquidity Preference as Behaviour Towards Risk", *Review* *of Economic Studies*, 25, 65-86.