Money, Reputation and Inventories Under Credit Market Imperfections

by

URS HAEGLER

London School of Economics and Political Science

Thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of London

Abstract

This thesis analyses the way in which credit market imperfections affect the behaviour of economic agents, and examines how a variety of tangible or intangible assets such as fiat money, reputation and inventories, facilitate bilateral exchange and influence investment decisions of firms under such circumstances.

The first chapter of the thesis deals with the role of fiat money as a medium of exchange in a model in which agents hold consumable goods or nonconsumable cash. The physical environment of pairwise random matching for bilateral trade, however, prevents them from issuing debt certificates. Unlike fiat money, consumables have uncertain quality characteristics, and agents can only detect the quality of a subset of goods. As a consequence, barter is plagued by asymmetric information, whereas monetary exchange involving generally recognisable legal tender is not. This suggests that it is because of, rather than despite, its intrinsic uselessness that, as a medium of exchange, fiat money is superior to goods or assets subject to some form of quality uncertainty.

The second chapter examines the effects of reputation and internal finance on a firm's investment incentives. An entrepreneur with unknown productivity finances risky production with a combination of internal finance and funds from external investors who, just like himself, are able to learn about his true productivity over time, a process that influences their willingness to lend. However, investment decisions taken by the entrepreneur, are not observable to outsiders. This information problem leads not only to underinvestment but also to premature liquidation. It is shown that the acquisition of reputation and internal funds may counteract such undesirable outcomes. On the other hand, it becomes clear that when assets are low, incentives to invest are disrupted because of a high probability of liquidation in the near future. Young firms appear to be particularly susceptible to effects of this type.

Finally, the third chapter studies inventory investment and internal-finance decisions of a financially constrained firm facing an uncertain demand process. The model gives an explanation for the stylised fact that production is more volatile than sales. Assuming that firms have limited access to capital markets they are forced to rely on internal finance. However, following a series of unfavourable sales realisations such funds possibly are so low that firms find themselves unable to re-establish the old inventory level in subsequent periods. Conversely, after a series of high sales the firm has a substantive amount of money to finance output quantities that may be in excess of sales.

Acknowledgments

In writing this thesis I enjoyed the privilege of having John Hardman Moore as my supervisor. I have benefitted tremendously from his guidance, and I would like to thank him for his support and encouragement as well as numerous inspiring discussions.

I am very grateful to David Webb for allowing me access to the LSE Financial Markets Group (FMG) which provided not only a stimulating research environment but also as a source of companionship. In particular, I would like to mention Alvaro Almeida and Sylvain Friederich with whom I have shared an office during the last three years.

I am also indebted to Martin Hellwig whose tuition had considerable influence well beyond my undergraduate years at the University of Basle.

During my studies at the LSE I had the opportunity to collaborate with and receive help from many researchers within the Economics Department and the FMG. The fourth chapter is a product of joint work with Ward Brown whose excellent co-operation I appreciated very much. This project benefitted from advice given by Nobuhiro Kiyotaki whose research has also helped to shape the second chapter of the thesis. And my discussions with Sven Rady on the intricacies of dynamic programming were of great help in writing the third chapter.

Financial support from the Swiss National Fund (SNF), the Janggen-Pöhn- Stiftung, St. Gall, and the Basler Kantonalbank, Basle, as well as from the Overseas Research Award Scheme, are also gratefully acknowledged.

Finally, I would like to thank my family for their financial and emotional support, and my partner, Muna, for her patience and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Walrasian theory of competitive equilibrium, developed in its modern form in the 1950s and 1960s by economists such as Arrow and Debreu, assumes that economic agents can enter mutually beneficial agreements of whatever complexity required, without this being reflected by an increase in transactions cost due to informational restrictions or co-ordination problems. As such costs seem particularly prevalent in credit arrangements, it is of little surprise that a large part of the literature dealing with deviations from the complete-market paradigm has focused particularly on credit market imperfections. A firm, for example, may find it difficult for a number of reasons to sell contingent claims on its future revenue streams. First of all, the firm may simply refuse to pay a promised amount of money to its creditors, and its owner-manager may deem it preferable to escape to some remote island with whatever cash there is. Alternatively, executives of the firm may use the funds given to them for nonproductive purposes, thereby more or less intentionally moving the business into a position of insolvency in the future. Secondly, even if the firm's management is less maverick, it may be impossible to enumerate all the future states of nature the firm could possibly encounter, or it is too costly to write a contract that is sufficiently complex to warrant an efficient outcome. Rational investors foreseeing such difficulties naturally become reluctant to lend funds to firms. This is particularly true when firms are young and when economic relationships are allowed to be of a short duration only, as uncertainty is higher and exertion of control over the firm becomes more difficult.

Uncertainty and costliness of information, which lie at the heart of these problems, come in a number of guises. For one thing, they may simply be due to the unforeseeability of future events. Another, more profound form of information shortage, stems from an uneven distribution of knowledge over economic agents. That is, in many situations there may prevail informational asymmetries between potential trading partners which prevent beneficial exchange of goods.

Numerous mechanisms to mitigate such market shortcomings have evolved over time, and this thesis mainly focuses on three of them. One such mechanism, money and in particular fiat money, is a social institution: For it to have a positive effect on trade it needs to be widely accepted as a means of payment. Even in a system of decentralised exchange, as opposed to fully co-ordinated trading of the Walrasian type, the existence of money may then reduce the amount of information individuals need in order to carry out certain transactions with a minimum of frictions.

The other two mechanisms can be considered as private measures to relieve the burden of uncertainty. The first is reputation acquisition which induces a certain amount of trust into the productivity of a firm or the ability of a worker, making potential creditors less reluctant to lend money. Reputation building, however, is not free as it requires agents to forego present utility in favour of future attractiveness on the credit market. The second is the holding of inventories, in the form of physical goods or financial assets. Underlying the desire for this type of insurance is a precautionary motive. If agents cannot rely on the credit market to act as a buffer against various income or revenue shocks, they are forced to provide their own cushioning to counter unforeseen events. However, just like reputation acquisition, this mechanism is generally not costless either.

There are two broad themes that underly the research presented in all the chapters of this thesis. The first is the identification of consequences that credit-market imperfections have for the behaviour of economic agents. The second is the attempt to analyse how and to what extent the above-mentioned mechanisms help to overcome inefficiencies created by those imperfections.

Chapter 2 deals with the role of fiat money as a medium of exchange in a model of decentralised exchange. The fact that money is a social institution requires the use of a general-equilibrium framework. However, in a decentralised economy barter trade is restricted by a double-coincidence-of-wants requirements Clower (1967).[1] A search model of the type put forward by Kiyotaki and Wright (1989) is employed here, as it not only incorporates those requirements but also provides a structure of extreme decentralisation which makes borrowing and lending virtually infeasible. A random-matching framework may appear too extreme a modelling device for decentralised exchange, as in reality trades are often mediated by middlemen. However, as Hahn (1988) points out, their services cannot be provided without incurring resource costs and they will be more cheaply provided if middlemen, too, can exchange their goods for money.

In this search framework there is a continuum of agents who meet randomly in pairs each period to bargain bilaterally. The probability of meeting the same agent twice is zero, and therefore it is almost impossible to ever honour a debt contract when agents are spatially separated in this way. In other words, a rational agent does not trust the promise of a counterparty to make a future payment. It has been recognised by many authors, (e.g. Goodhart (1989)) that absence or lack of trust in a trading partner's honesty is crucial for the existence of a means of immediate payment. Gale (1982) demonstrates that if there was no need for such trust all exchanges could be entirely based on credit.

However, for a good without intrinsic value, such as fiat money, to become a valued medium of exchange, agents need to have a sufficiently strong belief that it is acceptable to a sufficiently large number of sellers in possession of a good of intrinsic value

[1]More detailed discussions of the relevant literatures have been omitted from this introduction and relegated to the outset of the relevant chapters.

to them. It is clear that otherwise those agents themselves would refuse to accept the good in the first place. Consequently, fiat money may or may not have value in exchange. If it does, then it is due to some bootstrap argument, as reflected by the statement that fiat money has value because individuals expect it to be valued in the future.

Whilst *a priori* any commodity could be the object circulating in such a bootstrap equilibrium, it is obvious that some goods lend themselves more easily to play this role than others. One crucial distinguishing feature is storability. All other things equal, a highly perishable good, or equivalently, an asset with a very low rate of return, tends to be inadequate for transactions purposes. For an object to be a good medium of exchange it should also store value reasonably well. Note, however, that the reverse statement does not hold, as a high store-of-value performance is no guarantee for a good to qualify as a good medium of exchange. It is the tension between storability and acceptability in exchange that, e.g, Kiyotaki and Wright (1989) focus on.

In contrast, Chapter 2 of this thesis concentrates on another characteristic of goods, namely recognisability. To this end, trade is restricted not only by the double-coincidence-of-wants requirement but also by individuals' private information about the quality of intrinsically useful commodities, as opposed to fiat money which is assumed to be of uniform quality.

In the absence of fiat money agents are forced to use consumption and production goods for the purpose of indirect trade. There are three varieties of these physical goods each of which appears in either low or high quality. Only the latter can be consumed to yield utility whereas the former do not yield any utility in consumption but can, in principle, be used for trade purposes. There is specialisation in production and consumption in the sense that no agent can consume a commodity of the same type as that produced by himself, a feature which reflects gains from specialisation in production. However, different agents specialise in different consumption-production pairs. The difficulty that arises in the ensuing barter arrangement is that a typical agent can only recognise the quality of product varieties produced or consumed by himself but not of the third commodity when it is offered to him in trade. Thus, there is also specialisation in quality recognition.

It is first demonstrated that strategies which postulate to knowingly accept a good of low quality are weakly dominated. Eliminating them ensures that a situation of one-sided private information occurs only when an agent carrying his production good meets someone who wants to consume this good but offers a commodity whose quality cannot be detected by that agent.

Under these circumstances the economy exhibits a unique stationary equilibrium, which is characterised by uninformed parties strictly randomising over whether to accept or to reject a good whose quality they are unable to recognise. The probability of acceptance is decreasing in the share of low-quality goods present in the economy and the rate of time preference. This outcome is contrasted with a situation where such informational asymmetries are assumed away. In this case there are multiple equilibria, in one of which each good is accepted with certainty, and a higher welfare level is therefore achieved.

By introducing fiat money into the economy with private information agents obtain an additional object for indirect trade. It is different from the physical commodities in that it is always intrinsically useless and of uniform quality. The central result of the chapter is that there exists an equilibrium in this monetary economy such that money is accepted with probability one, and non-recognised goods are still accepted with a probability strictly within the unit interval. Hence, monetary and barter trade co-exist. As a consequence, fiat money can be introduced in such a way that welfare is increased in comparison to the nonmonetary economy with quality uncertainty, although, at each point in time, the output of high-quality consumable products is diminished by doing so.

Thus, in a world with private information about quality characteristics of goods, a medium of exchange such as fiat money is superior to consumption or production goods for trade purposes because of, and not despite, its intrinsic uselessness. If one extends the notion of quality uncertainty to the creditworthiness of issuers of financial instruments other than fiat money, such as, e.g. private debt, differing degrees of 'moneyness' of various assets can also be partly explained.

Unlike the first contribution Chapters 3 and 4 assume partial-equilibrium perspectives, as they address intertemporal decision-making problems of a single firm. The access to the credit market is limited in one case, and still excluded altogether in the other. The third chapter considers a dynamic-investment problem of a firm whose ability to raise external funds is reduced by investors' inability to observe the investment decisions taken by its owner-manager, a fact which may induce the latter to divert part of the external finance into private consumption rather than productive activities. In the light of this information shortage, other factors such as reputation and internal finance become crucial in determining the amount of funds raised.

Moral-hazard problems appear in many other areas. For instance, a firm buying labour services may not be able to observe directly the amount of labour supplied. As a consequence, wage payment will have to depend on a signal related to labour input. As such signals are usually also influenced by other factors, an inefficient risk allocation results under a contract that provides incentives to work. Holmström (1982) shows that time may have a favourable effect on a manager's incentives to exert effort, even when explicit long-term contracts are not feasible. If the manager is concerned about his reputation in the long run, he will be less tempted to shirk despite the fact that this would be in his short-term interest. The managerial labour market provides implicit incentives to work, which under certain circumstances may be particularly true for young managers.

In assessing the investment incentives of an entrepreneur who for some reason is not able to issue long-term debt contracts, one should therefore be able to draw similar conclusions about the provision of incentives by the credit market. Put differently, the fact that bad current outcomes may damage the entrepreneur's ability to raise funds in the future, may prevent him from diverting too many resources into consumption. The question are, however, whether these implicit control mechanisms can entirely remove inefficiencies stemming from moral-hazard problems, and to what extent they still work in the presence of an outside option as introduced by Jovanovic (1982), a study of industry selection in which reputational concerns do not matter.[2] If an entrepreneur no longer deems his business worthy of being kept alive he can always opt for liquidation and found a new firm or provide his services on the labour market.

As Chapter 3 will show, the presence of an outside options is of great significance. The true productivity of the entrepreneur who finances risky production with a combination of internal and external funds is known neither to the outside financiers nor to himself. All parties, however, are able to gradually find out more about that value over time through a process of Bayesian learning. This information directly determines external investors' willingness to lend.

The investment decisions taken by the entrepreneur, on the other hand, are not observable to outside parties. In equilibrium this leads not only to underinvestment but also to premature liquidation which is executed by the entrepreneur as soon as the outside option appears more valuable to him than the continuation of the firm's production activities. Whilst it is true that reputational concerns and the acquisition of internal funds may counteract undesirable aspects such as underinvestment it becomes clear that in cases of low asset positions and reputational distress investment incentives are disrupted because of a high probability of liquidation in the near future.

Young firms appear to be particularly susceptible to effects of this type, since the volatility of growth rates is inversely related to age. This is a consequence of the gradually decreasing sensitivity of the reputation-updating rule to new information.

[2] Reputation acquisition in debt markets has also been the focus of other analyses such as, e.g., Diamond (1989) and Diamond (1991). In these papers, the borrower chooses the riskiness of a project rather than investment size, and moreover, there exists a problem of adverse selection. Despite these structural differences, however, the conclusions about the way in which reputational concerns affect incentives are remarkably close to those drawn in Chapter 3.

As the firm grows older new data will have a diminishing impact on learning as it becomes relatively insignificant compared to a long history of outcomes. This implies that the death rate of firms decreases over time, which is consistent with casual observation.

On the other hand, it is shown that the investment rate does not generally decrease over time, in spite of the fact that Bayesian updating becomes less and less sensitive to new information. The explanation for this lies in the presence of the outside option which renders the entrepreneur cautious when putting resources into the risky technology. It may be preferable to keep operations on a lower level in the beginning in order to gather more information about his own productivity.

A business in reputational distress has little incentives to run its operations on a high level for at least two reasons. The first is the direct effect that a low reputation has as a signal to outside investors. A firm in such a position will find it relatively difficult to attract external funds. The second and possibly more important reason is that when a firm is close to the liquidation point it runs a considerable risk of not reaping substantive benefits from additional production inputs, as the probability of losing them through exit and liquidation cost is relatively high. A similar conclusion can be drawn for situations of financial distress, which is characterised by low levels of internal financial wealth.

The main conclusions of the analysis are concerned with the potential role of monitoring. The shortcomings of capital markets described above provide a rationale for close supervision of young or troubled firms by financial intermediaries. The framework of symmetric uncertainty about a firm's expected productivity is particularly appropriate for start-up entities, as the owners or managers of such new businesses often know little more about the probability of success of a new idea or technological development than outside parties. Therefore, the study may also contribute to an explanation why venture capitalists are usually heavily involved in new firms, not just financially but in almost any other business aspect. More generally, the possibility to obtain monitored funds may have a positive effect on young and distressed firms' incentives to invest and thereby reduce the number of socially undesirable liquidations. Furthermore, credit rating agencies also play an important role in filling some informational gaps of the kind discussed here.

The fourth chapter studies the implications of financing constraints on optimal intertemporal inventory-investment and internal-finance decisions of a firm facing an uncertain demand process.

The background for this exercise is the stylised fact from the empirical literature that the variance of production exceeds the variance of sales. Ample evidence for this has been found on firm-, industry- and macro-economic level (see, e.g., Blinder and Maccini (1991)). These findings have attracted considerable attention, as they contradict the predictions standard neoclassical, i.e. linear-quadratic, inventory-investment model. In the latter approach businesses attempt to smooth their output levels over time due to convex cost functions. Optimal inventory management supports them in the pursuit of that goal as stocks buffer unforeseen demand fluctuations. Correspondingly, inventory investment, defined as the change in target inventories (amount of goods made available for sale), is negatively correlated with sales in those models. Again, the empirical evidence tends to point in the opposite direction, albeit somewhat less clearly than with respect to the relative variances of output and sales.

Previous attempts to explain the excess variance of production can be put into three classes. In the first and most prominent approach a number of authors have modified the linear-quadratic model by introducing either nonconvex costs (e.g. Ramey (1991)) or cost shocks (e.g. Blinder (1986)). However, these models do not receive much empirical support.

The second approach is to use stockout-avoidance models with serially correlated demand processes or the possibility to backlog demand (e.g. Kahn (1987)). That serially correlated demand does exist and that demand backlogs are an important instrument for many firms and industries is not disputed here. However, demand itself, as opposed to realised sales, is mostly impossible to observe empirically, and when there are many competitors in an industry or close substitutes to a product, firms may find it very difficult to backlog demand.

Models in the second group have experienced very little empirical testing (see, e.g. Kahn (1992)), and the same is true for the third category (one exception being Mosser (1990)), the (s,S)-type inventory models. The latter are somewhat unsatisfactory for the purpose of explaining the two main puzzles at hand because they predict a zero, instead of a slightly positive, covariance between inventory investment and sales.

The model in chapter 4 is closest in spirit to the second approach because it also uses a stockout-avoidance framework. However, it distinguishes itself from that category because one of its main results is that even sales which do not contain information about expected future sales affect present production. Moreover, unlike (s,S) models it predicts a positive covariance between inventory investment and sales, matching empirical observations more closely.

The chapter considers a firm which produces an imperfectly storable good at a constant unit cost and attempts to sell its output each period at a constant price, such as to maximise the present discounted value of consumption (dividends) over an infinite lifetime. The business faces a sequence of independently and identically distributed demand shocks. Each period the stock of unsold goods depreciates partially, and the firm makes decisions about gross production expenditures, cash retention as well as consumption expenditures. These decisions are once again crucially constrained by the requirement that they be nonnegative, and that their sum must not exceed cash holdings in that period.

In a simple example where the demand distribution, the depreciation technology as well as production cost and output price are parametrised, explicit optimal policies in a stationary equilibrium can be found. From this one can derive the stationary distribution of production and sales distribution and show that the variance of production exceeds the variance of sales.

To show that the validity of the production counter-smoothing result does not hinge on the specific parameter values used in the simple example a general model is also examined. The outcome is contrasted with that of a model where one crucial financing constraints is absent, namely the nonnegativity constraint on consumption, in which case the variances for production and sales are identical. It follows as a corollary of the production counter-smoothing result that inventory investment is positively correlated with sales and, more generally, with the level of own assets such as money and goods.

Equally important as the above result, the equilibrium in this model exhibits the additional property of positive autocorrelation in sales. It does so even though the underlying demand process is specified to be serially uncorrelated. This result is very useful as demand is generally not empirically observable.

Chapter 2

Fiat Money and Quality Uncertainty

2.1 Introduction

This chapter studies the role of money as a medium of exchange when barter trade is restricted not only by the double-coincidence-of-wants requirement but also by private information about the quality of commodities. For this purpose the model of Kiyotaki and Wright (1989) is extended to include commodities of both high and low quality.

While this contribution is mainly directed towards a better understanding of the microeconomic foundations of monetary theory it also relates to important questions often raised in monetary policy debates. It is often said that a medium of exchange issued by authorities has many close substitutes such as IOUs or even barter commodities. To justify the claim that the medium of exchange supply is not entirely beyond monetary authorities' control and therefore such policy measures as open-market operations or regulations concerning the supply of bank deposits are not futile, one needs to demonstrate or even quantify imperfections in the substitutability between various media of exchange (Hellwig (1992)).

This problem is addressed here by studying a framework with bilateral exchanges and asymmetric information. In contrast, most of the monetary economics literature does not model trade as a decentralised process which is the main reason why satisfactory answers to the above questions do not exist.

Despite the work of classical (Smith (1776)) and early neo-classical (Jevons (1875)) economists, who have recognised that the central role of a object like money is that of a medium of exchange, Walrasian models have clearly dominated general equilibrium analysis for the last decades. These models are comparatively tractable and highly sophisticated. However, one of their drawbacks is that they use a completely centralised trading process (the Walrasian auctioneer). There is no role for a medium of exchange in general or fiat money in particular within such a framework, since every scarce commodity with a positive value in equilibrium can be exchanged for a positive amount of another commodity (see, e.g., Hahn (1965)).

That is true not only for the exchange process in the static Arrow-Debreu setting but also for its intertemporal versions, be it the sequential-equilibrium approach (e.g., Hahn (1973)) or the overlapping generations models (e.g., Wallace (1980)). Both of them introduce a device that allows individuals to break their lifetime budget constraints temporarily and thereby permits them to improve on the intertemporal allocation of their resources. Many authors call this device 'money', but Tobin (1980) points to a semantic problem of using this label. The means of payment referred to as money in those models bears little resemblance to the real-world object they try to study. The intertemporal reallocation device used in these contributions is merely a store of value, and numerous other assets (real or financial) or even social institutions could serve this purpose at least as well as (fiat) money. Therefore, even in these models there is no room for a medium of exchange.

Notable exceptions are papers by Gale (1978) or Grimes (1987) who consider money as a good that fills the gap left by an absence of complete trust in a trade relationship. Credit markets do not work perfectly as a seller may not always accept personal IOUs of a buyer, but usually he is willing to take cash as a means of payment such that trade occurs nevertheless. However, it still remains to be explained why an intrinsically valueless good is most appropriate to play this role.

There have been earlier attempts in the literature to incorporate the transactions role of money into the analysis of monetary economics. I only remind the reader of the well-known inventory-theoretical, partial-equilibrium models by Baumol (1952) and Tobin (1956), and the cash-constraint models by Clower (1967) which still represent the dominant paradigm for macroeconomic models with money. These lines of work have in common that they take the medium-of-exchange role of money as exogenous and are still based on centralised markets.

More recent studies try to endogenise the transactions role of money by explicitly analysing decentralised exchange processes that take the double-coincidence-of-wants problem seriously. Townsend (1980) introduces models of money with spatially separated agents that can be interpreted as special versions of overlapping-generations models in which agents live for two periods such that they can trade with an agent of another generation in one period only. Agents in his turnpike models do not issue IOUs since the particular structure rules out the possibility that two agents meet more than once. As a consequence no credit is given and there is a need for a device like money without any uncertain personal backing in order to overcome the lack of double coincidence of wants.

A quite different structure of bilateral trade relationships is examined in a series of papers by Kiyotaki and Wright (1989, 1990, 1991). These studies model trade as the outcome of an ongoing random matching process where the identity of future bilateral trading partners cannot be foreseen perfectly. The authors stress the importance of expected future trade possibilities when double coincidence fails to hold and storage of goods is costly.

Banerjee and Maskin (1990) suggest a framework that resembles closely the Walrasian setting. They avoid questions concerning the search for trading partners and the problem of co-ordination between several markets or trade relationships. Instead they merely concentrate on the double-coincidence-of-wants problem by introducing separate market clearing conditions for each barter market instead of just one goods market in the Walrasian sense. On a particular barter market only two goods can be exchanged whereas on a Walrasian commodity market a certain good can be traded against all other goods. However, proceeding in this way still leaves a very strong element of centralisation in the analysis, the only difference being that the auctioneer now has barter markets instead of commodity markets to clear.

Informational asymmetries are at the core of the present paper, too (see also Brunner and Meltzer (1971) or Alchian (1977) for less formal treatments); it aims at explaining the selection of a good like fiat money as the medium of exchange in bilateral trade under quality uncertainty.[1] Following Kiyotaki and Wright (1989) a model of sequential bilateral trading where agents meet randomly is considered. To explain how an intrinsically useless good like fiat money becomes a preferred medium of exchange is the goal of the analysis.

In contrast to Kiyotaki and Wright (1989) storage costs (rates of return) are identical for all goods in this paper. What is emphasised is the recognisability aspect of a medium of exchange. While a consumer or a producer of a good always detects its quality, agents who neither produce nor consume that good are not able to do so unless they own it. If the latter accept such a good, which with some probability is of low quality, they face the possibility of being matched next period with someone who may not want to trade because he immediately notices the quality of this good; as a result, they will be reluctant to accept it in the first place. In contrast, fiat money is a good whose quality is assumed to be uniform and to be recognised by all agents. It may therefore facilitate trade.

[1] After writing the first draft of my paper on which the present chapter is based I became aware of a contribution by Steve Williamson and Randall Wright (now published in the American Economic Review [1994]) who also study bilateral exchange under private information. They proceed in a somewhat different framework. In their paper there is only one type of consumption good, i.e. there is no double-coincidence-of-wants problem, but agents can choose whether to produce the low or the high quality variant whereas in the present study there are three different types of goods with the proportion of low quality goods fixed exogenously.

The paper is organised as follows. 2.2 sets up the physical environment of the model and 2.3 presents some implications of the informational assumptions for the bargaining process. In the fourth section a benchmark economy devoid of quality uncertainty is analysed. Section 2.5 considers an economy with quality uncertainty in which there is no fiat money and in Section 2.6 the consequences of introducing fiat money are studied. Section 2.7 compares the welfare levels of the three economies and the final section provides some conclusive remarks.

2.2 The Physical Environment

2.2.1 The Model

The basic structure of the model is an extension of the one in Kiyotaki and Wright (1989). Consider an economy that evolves at an infinite sequence of periods. There are three varieties of indivisible goods i $(i = 1, 2, 3)$, and each variety can appear in two different qualities, high (H) and low (L). Let i^j $(i = 1, 2, 3;\ j = H, L)$ denote good i of quality j.

There is a continuum of infinitely-lived agents with unit mass which is equiproportionately divided into three types of agents k $(k = 1, 2, 3)$. The types differ with respect to their production skills and their consumption preferences. More precisely, agents of type $1, 2$, and 3 produce one unit of 2^H, 3^H, and 1^H, respectively, immediately after having consumed one unit of 1^H, 2^H, and 3^H, respectively, which is the only good they derive utility from. Low-quality goods, while being held by a positive proportion of agents, are neither produced nor consumed. All goods can be stored at zero costs, independent of quality, but in each period storage capacity is limited to one unit of one good only.

Consequently, the proportions of high and low quality goods, denoted by q and $(1 - q)$, respectively, $q \in (0, 1)$, remain constant throughout time. q is assumed to be identical for all three types of goods.

Let

$$U^k = E_0 \sum_{t=0}^{\infty} \beta^t u(i_t^j) \tag{2.1}$$

represent a type-k-agent's expected lifetime utility function ($k = 1, 2, 3$), discounted to the present, where $\beta \in (0, 1)$ is the discount factor and $u(i_t^j)$ is the instantaneous (net) utility for an agent who has i^j *immediately after bargaining* at time t.

$$u(i_t^j) = \begin{cases} u, & \text{if } i = k \text{ and } j = H \text{ at } t \\ 0, & \text{otherwise,} \end{cases} \tag{2.2}$$

$u > 0$ being the instantaneous net utility of consumption and immediate production.

The crucial feature of the model is the distribution of information about the goods' qualities. It is assumed to be common knowledge that each agent of some type always recognises the quality of his respective consumption and production good, but the quality of the good which is neither produced nor consumed (the third good, hereafter) can be observed if and only if he is in possession of it. He cannot tell apart high and the low quality of the third good as long as it is in the inventory of another agent.

Trade is organised as a stochastic matching process. More precisely, bilateral bargaining takes place between agents who meet at random each period. As for all three types the number of agents is infinitely large and drawings are independent, the cross-section distribution of pairwise meetings is almost surely constant.

Once a match has formed agents observe the *varieties* of goods in each other's inventory, but not necessarily their *qualities* and certainly not the *type* of an agent they are confronted with. They play the following simple bargaining game. After inspection of inventories agents decide whether to trade or not. Only if both agents are willing to do so inventories are swapped, whereas otherwise both agents hold on to their goods. Thereafter the pair separates, regardless of whether trade has occurred or not.

Throughout the analysis we will confine ourselves to symmetric equilibria which, at this stage, we roughly define as a profile of bargaining strategies such that each agent's utility U^k is maximised subject to the trading technology described above.

In the next section we demonstrate that the assumptions about the information structure have a number of implications which simplify the analysis considerably. Before that, however, a few words should be said about the use of a random-matching model.

2.2.2 Random Matching vs. Intermediation

It has been pointed out in the introduction that even though the decision-making in the Walrasian framework is decentralised the processes of finding market-clearing prices and physically exchanging goods are not. The concept of an auctioneer involves an extreme degree of centralisation. In contrast, the random-matching model used in this chapter is at the opposite end of the centralisation scale.

Some readers may argue that this other, fully decentralised extreme is just as unrealistic as the Arrow-Debreu framework. An individual who wants to buy a good usually does not wait until he bumps into someone who happens to be selling just that good, and naturally the same can be said for the seller himself. Instead, it is much more likely that a number of agents will decide to act as intermediaries by setting up trading posts whose geographical location is sufficiently well-known to both buyers and sellers.

However, one can counter this criticism by showing that the principle ideas of our search model with private information about goods qualities carry over straightforwardly to a framework with intermediated structures. The notion of division of labour is not confined to physical production but can be readily extended to the provision of intermediation services. Specialisation in trading just one particular type of good may be beneficial in many ways, such as lower unit cost of storage or transportation. More importantly, however, by focusing on one good only, a shop-keeper, for instance, is able to develop a high level of expertise in the quality assessment of that good. There is a possibility that an intermediary uses his superior knowledge to extract a rent from his less informed customers. However, in a world with many competitors and with a reputation for providing quality at stake, buyers may benefit from the presence of well-informed intermediaries not only through a reduction of search costs but also because of a reduced probability of acquiring an inferior product. In that case, specialisation in quality recognition improves Pareto efficiency.

However, in order for this type of specialisation to be feasible, the existence of an easily recognisable medium of exchange such as fiat money, e.g., is required.

2.3 Implications of the Information Structure

2.3.1 A Weak-Dominance Result

Since each variety appears both in high and low quality there are six different goods in the economy, all of which could be held by each type of agent. However, we apply a string of arguments to reduce the number of goods each rational agent may hold to three.

We begin by noting that a utility-maximiser does not store his consumption good since it is obviously preferable to consume it immediately and enter the next round with a newly produced good.

We will use our first proposition to rule out situations in which agents *knowingly* give up a high-quality good for a low-quality good, since it is at least weakly preferred not to do so.

PROPOSITION 1.1: *A strategy that prescribes to knowingly swap a good of high quality for a good of low quality is weakly dominated.*

Proof: To prove this proposition we will use a series of observations which allow the deletion of strictly and weakly dominated strategies.

(i) Not choosing 'trade with certainty' when a consumption good is offered is strictly dominated.

(ii) Consider a type-k-agent with a low-quality good recognisable to an agent of the same type. If the latter strictly prefers this good to some high-quality good he himself holds, there would not be trade because the former agent would want to hold on to his good (identical preferences).

(iii) Consider a type-k-agent with either his own production good. In any bargaining situation with an agent of type a ($a \neq k$) who is a consumer of k's production good, our type-k-agent can sell it with certainty (if he wishes to do so) because of (i); hence, the low-quality version of the same good cannot be more marketable. If he meets an agent of yet another type b, $b \neq a \neq k$, unable to distinguish between qualities, the two goods will be equally marketable. Overall, holding the production good is at least as beneficial for the agent as holding the low-quality version thereof. There is no gain from giving up the former for the latter.

(iv) Next compare a type-k-agent holding the high-quality version of his third good. In any bargaining situation with an agent of type b (consumer of the third good) our type-k-agent can sell it with certainty (if he wishes to do so) because of (i); hence, the low-quality version cannot be more marketable. If he meets an agent of type a (producer of the third good), we infer from (iii) that the high-quality version is at least as tradable as the low-quality version. The overall conclusion is again that our type-k-agent would not benefit by giving up the former for the latter.

(v) In order for consumption to occur at all the third good must be acquired by some agents in some periods. This together with (iv) implies that the third good of high quality is valued at least as highly as the production good (of high quality) and therefore at least as highly as the low-quality version of the production good (from (iii)).

(vi) Suppose our type-k-agent values his production good less than the low-quality version of the consumption good. That would be possible only if agents preferred the low-quality version of their production good to their production good (of high quality), which was ruled out by (iii). Hence, this leads to a contradiction. □

This proposition allows us to argue that in stationary symmetric equilibria agents will not hold the low quality-versions of their consumption or production goods because acquiring them would not benefit them.[2]

2.3.2 Possible Matches and Information Distribution

It follows from the previous subsection that each agent will hold either his production good (P hereafter), the high-quality version of his third good (H) or the low-quality version of his third good (L). Inventory distributions are triples (p, l, h), where p, l, and h represent the proportion of type-k-agents holding P, L, H, respectively, $k = 1, 2, 3$. Obviously, $p + l + h = 1$, and since the number of low quality goods remains constant throughout time, $p + h = q$ and $l = 1 - q$.

Define the set of commodities an agent can hold as $S := \{P, L, H\}$. Furthermore, denote with $y_{ss'} \in [0, 1]$ the probability that a specific agent holding s is willing to trade it for s', and with $Y_{s's} \in [0, 1]$ the probability of meeting an agent with s' willing to exchange it for s, $\forall s, s' \in S$.

Stationary inventory holdings are then governed by a Markov process with a time-invariant transition probability distribution function given by $y_{ss'}Y_{s's}$, $\forall\ s, s' \in S$.

An agent's optimisation problem is then equivalent to finding a function that solves the Bellman functional equation $\forall\ s \in S$,

$$V(s) = \max_{y_{ss'}} \beta \sum_{s' \in S} V(s') y_{ss'} Y_{s's}, \tag{2.3}$$

where $V(s)$ is an agent's valuation of good $s \in S$.

[2] Proposition 1.1 does not enable us to exclude the trivial equilibrium where all commodities are always accepted regardless of quality and where therefore trade always takes place. Note, however, that this equilibrium would not survive the introduction of a trembling-hand-perfection criterion.

We now analyse the distribution of information in all possible meetings. Consider an agent of type 1 carrying 2^H,3^L, or 3^H, as depicted at the top of Table 2.3.2. All possible matches with different type-good combinations are listed on the left. The cell entries indicate how information is distributed in the associated match (from the viewpoint of the type-1-agent), or, whenever such a statement can be made already, whether trade occurs or not.

No trade occurs (N) if one of the agents matched detects a low-quality good, or if one of them is offered to swap his H for what may turn out to be an L. Without loss of generality we can include situations in which two identical goods are offered. If our agent holds his P (2^H) and is offered a third good he is the uninformed party (U) in a relationship with asymmetric information; trade may or may not occur. If he is holding L or H (3^L or 3^H) he is never willing to trade if he is uninformed (N), but he may be as the privately informed side (I) when he meets agents of the same type holding P. For an agent in possession of H there are two situations in which both goods' high qualities are public information (C); trade may or may not take place. Finally, we can identify two occurrences of double coincidence of wants where trade takes place with certainty.

Corresponding matrices for type-2 and type-3-agents are presented in Tables 2.3.2 and 2.3.2, respectively.

Thus, trade evolves as follows. Iterated weak dominance has ruled out trade when at least on side detects a low quality good. This case occurs when L is offered to an agent who consumes or produces H.

Crucial to the analysis are situations with asymmetric information. Below we identify the conditions under which an agent of some type is willing to give up his P for a third good that can be used as a medium of exchange. Recall that this third good can be of low (L) or of high quality (H), each with positive probability.

Let x and $(1-x)$ denote the probabilities with which an uninformed agent accepts and refuses to trade, respectively, $x \in [0,1]$, in a situation where he is uninformed.

	1 2^H	1 3^L	1 3^H
$1(2^H)$	N	I	I
$1(3^L)$	U	N	N
$1(3^H)$	U	N	N
$2(3^H)$	U	N	N
$2(1^L)$	N	N	N
$2(1^H)$	D	N	C
$3(1^H)$	I	N	D
$3(2^L)$	N	N	N
$3(2^H)$	N	N	C

Table 2.1: Information structure in matches involving type 1

	2 3^H	2 1^L	2 1^H
$2(3^H)$	N	I	I
$2(1^L)$	U	N	N
$2(1^H)$	U	N	N
$3(1^H)$	U	N	N
$3(2^L)$	N	N	N
$3(2^H)$	D	N	C
$1(2^H)$	I	N	D
$1(3^L)$	N	N	N
$1(3^H)$	N	N	C

Table 2.2: Information structure in matches involving type 2

	3 1^H	3 2^L	3 2^H
$3(1^H)$	N	I	I
$3(2^L)$	U	N	N
$3(2^H)$	U	N	N
$1(2^H)$	U	N	N
$1(3^L)$	N	N	N
$1(3^H)$	D	N	C
$2(3^H)$	I	N	D
$2(1^L)$	N	N	N
$2(1^H)$	N	N	C

Table 2.3: Information structure in matches involving type 3

For an economy to be active it is necessary that a positive proportion of agents trades indirectly, i.e. via the good they do not produce nor consume and whose quality they do not know unless they are in possession of it; otherwise every agent would be forced to keep his production good or the low-quality good he was endowed with initially. In other words, there must be positive probability that a type-k-agent with his production good P is willing to give it up for a third good whose quality he can observe only after acquisition. For this to be optimal behaviour the expected valuation of holding the third good must be at least as high as the valuation assigned to holding P. To facilitate notation we define $v_s := V(s)$, $\forall\ s \in S$.

The following proposition shows that the valuation of P, v_P, is always strictly greater than the one of L, v_L, except for the trivial case of a completely inactive economy where both are equal to zero. For an uninformed agent to accept trade it is then necessary that v_H strictly dominates v_P.

PROPOSITION 1.2: *Given the informational assumptions either* $0 < v_L < v_P$ *or* $v_L = v_P = 0$.

Proof: Consider an agent of some type k, $k = 1, 2, 3$, holding L. Since agents of a different type recognise his L as being of low quality they will refuse to trade. Only for an agent of the same type holding P it might be preferable to accept trade that with a probability strictly between zero and one leaves him with L. Thus, with L the agent can never directly acquire his consumption good.

In contrast, this is possible with P if there is positive probability that double coincidence of wants occurs; taking into account that $\beta \in (0,1)$, it has therefore higher value than L. If the probability of acceptance is zero, $v_P = 0$; but then $v_L = 0$ since with L our agent can at best acquire P. □

For the discussion of active equilibria we have to keep in mind that agents being offered their P are willing to trade if they hold L, but will refuse to swap if they are in possession of H. The latter follows from Corollary 1.3:

COROLLARY 1.3: *For an equilibrium to be active, i.e. to have $x^* > 0$ for at least some agents, it is necessary that $v_P < v_H$.*

Proof: Activity implies $v_P > 0$. From Proposition 1.2 we know that then $v_L < v_P$. Suppose $v_P \geq v_H$. Then $x > 0$ cannot be an optimal strategy for any agent since this would lead to a negative expected gains from trade. □

2.4 The Model without Uncertainty about Quality

Before analysing the problems associated with quality uncertainty, it is useful to study a benchmark model where quality uncertainty is temporarily suspended such that every agent always recognises L as being of low quality. Since there are no problems of asymmetric information here, Corollary 1.3 does not have to hold. Let x and X now refer to the strategy adopted by agents holding P and being offered H whereas $\widetilde{x}$ and $\widetilde{X}$ refer to the reverse situation.

2.4.1 The Problem of a Typical Agent

Consider an agent immediately after production, having P in store. In the stochastic matching process he meets an agent of type 1, 2, or 3, each with a probability of 1/3.

With probability $p/3$ he meets an agent of the same type with P, and with probability $h/3$ he meets an agent with an identical good who is neither a consumer nor a producer of it. It is obvious that in these situations there is no scope for trade.

With probability $(1-q)$ he encounters agents with goods of low quality *all* of which he recognises in this benchmark case.

With probability $p/3$ he meets an agent who holds and has produced his H and who would like to swap since P is that agent's consumption good. In this case trade requires $v_H \geq v_P$. However, if the counterparty were an agent of the same type, the reverse inequality $v_H \leq v_P$ would be a necessary condition for trade to take place. These two inequalities imply that, in an active symmetric equilibrium where H is given up for P with positive probability, $v_H = v_P$.

If our agent bumps into someone with his consumption good, it is either a consumer of P (probability $h/3$), in which case there is double coincidence of wants and therefore trade, or it is a producer of his consumption good (probability $p/3$), in which case trade takes place with probability X.

For a given $(X, \widetilde{X})$ optimal choices of $(x, \tilde{x})$ have to satisfy our agent's valuation function for good P,

$$\begin{aligned} v_P = \quad & \frac{\beta}{3}\left[pv_P + lv_P + \left(hx\widetilde{X}v_H + h(1-x\widetilde{X})v_P\right)\right] \\ & +\frac{\beta}{3}\left[(pxv_H + p(1-x)v_P) + lv_P + h(u+v_P)\right] \\ & +\frac{\beta}{3}\left[(pX(u+v_P) + p(1-X)v_P) + lv_P + hv_P\right]. \end{aligned} \tag{2.4}$$

An agent unfortunate enough to be endowed with L in the beginning will be forced to retain it, since everyone else recognises the poor quality of the good and no other agent would ever accept it. This is reflected by the function

$$v_L = \frac{\beta}{3}(3v_L), \tag{2.5}$$

or $v_L = 0$.

Entering the period with H he will again refuse to participate in any trade activity with another agent having a good of variety 3 or with a low quality good in store. If he meets someone with his P, it will either be an agent of the same type (with probability $p/3$) or a consumer of H (with $h/3$). While the latter strictly favours trade, our agent is indifferent if and only if $v_H = v_P$, which must hold in an active symmetric equilibrium.

With probability $p/3$ our agent meets someone with his consumption good who is also its producer, which means that there is double coincidence of wants. If it is a producer of H trade takes place with probability $\widetilde{X}$. Thus, given $(X, \widetilde{X})$, optimal choices of $(x, \tilde{x})$ satisfy our agent's valuation function for H,

$$\begin{aligned} v_H = \quad & \frac{\beta}{3}\left[(p\tilde{x}Xv_P + p(1-\tilde{x}X)v_H) + lv_H + hv_H\right] \\ & + \frac{\beta}{3}\left[pv_H + lv_H + \left(h\widetilde{X}(u+v_P) + h(1-\widetilde{X})v_H\right)\right] \\ & + \frac{\beta}{3}\left[p(u+v_P) + lv_H + (h\tilde{x}v_P + h(1-\tilde{x})v_H)\right]. \end{aligned} \tag{2.6}$$

The system of equations (4), (5) and (6) can be rewritten in matrix form as

$$Tv = w \tag{2.7}$$

where, defining $r := \frac{3}{\beta} - 3$, the transition matrix

$$T := \begin{bmatrix} r + (p + h\widetilde{X})x & 0 & -(p + h\widetilde{X})x \\ 0 & r & 0 \\ -(1+\tilde{x})p - h(\widetilde{X} + \tilde{x}) & 0 & r + (1+\tilde{x})p + h(\widetilde{X} + \tilde{x}) \end{bmatrix},$$

$$v := \begin{bmatrix} v_P \\ v_L \\ v_H \end{bmatrix}, w := \begin{bmatrix} pXu + hu \\ 0 \\ pu + h\widetilde{X}u \end{bmatrix}.$$

2.4.2 Existence and Characterisation of a First-Best Efficient Equilibrium

Let $U^k(x, \tilde{x}; X, \widetilde{X})$ denote the expected lifetime utility of a type-k-agent playing $(x, \tilde{x})$, conditional on all others playing $(X, \widetilde{X})$.

DEFINITION 1.1: *A stationary symmetric Nash equilibrium in the game without uncertainty is a pair $(x^*, \tilde{x}^*)$, $x^*, \tilde{x}^* \in [0,1]$, which satisfies*

(i) $U^k(x^*, \tilde{x}^*; X^*, \widetilde{X}^*) \geq U^k(\hat{x}, \hat{\tilde{x}}; X^*, \widetilde{X}^*)$ *for all agents of type k ($k = 1, 2, 3$) and all $\hat{x}, \hat{\tilde{x}} \in [0,1]$ (optimality),*

(ii) $p[(p + h\widetilde{X}^*)x^*] = h[(1 + \tilde{x}^*)p + h(\widetilde{X}^* + \tilde{x}^*)]$ *(stationarity) and*

(iii) $x^* = X^*, \tilde{x}^* = \widetilde{X}^*$ *(symmetry).*

The following proposition can now be formulated.

PROPOSITION 1.4: *Any pair $(x^*, \tilde{x}^*)$ with*

$$x^* = \frac{1}{2}\left(\frac{q}{p} + \sqrt{9\frac{q^2}{p^2} - 16\frac{q}{p} + 4}\right)$$

and

$$\tilde{x}^* = 1 - \frac{p}{q-p}(1 - x^*)$$

is a stationary symmetric Nash equilibrium in the game without uncertainty, such that $v_P = v_H > 0$ *and* $v_L = 0$. *In particular there exists such an equilibrium with* $(x^*, \tilde{x}^*) = (1,1)$ *for which the stationary inventory distribution is given by* $\left(\frac{2q}{3}, 1-q, \frac{q}{3}\right)$.

Proof: Without loss of generality set $u = 1$ as well as $x = X$ and $\tilde{x} = \widetilde{X}$ (symmetry). Stationarity then requires

$$\tilde{x} = \frac{p(px - h)}{ph(1-x) + 2h^2},$$

and with $v_P = v_H > 0$ optimality implies equal probability to consume in the next period for P and H, i.e. $px + h = p + h\tilde{x}$, which can be expressed as

$$\tilde{x} = 1 - \frac{p}{h}(1-x).$$

Equating right-hand sides and rearranging terms we obtain the quadratic equation

$$p^2x^2 - p(p+h)x - (2h^2 - p^2) = 0.$$

The relevant solution for x is

$$x^* = \frac{1}{2}\left(\frac{q}{p} + \sqrt{\frac{q^2}{p} + 8\frac{h^2}{p^2} - 4}\right).$$

Using $h = q - p$, this can be simplified such that the expression for x^* given in the proposition. Substituting x^* for x in the optimality condition gives us $\tilde{x}^*$.

Finally, we set $x^* = \tilde{x}^* = 1$ in both the optimality condition and the stationarity equation to obtain the corresponding inventory distribution. □

In our setting a first best efficient outcome is implementable without money if we assume away all information asymmetries. Note that the agents with low quality goods are virtually excluded from the economy; they are forced to remain completely inactive.

2.5 The Model under Private Information without Fiat Money

Next we study the model with private information, i.e. we return to the uncertainty assumptions outlined in section 2.3. We raise the question whether under these circumstances media of exchange can still evolve and how powerful they are in facilitating trade. We identify an equilibrium for an active economy, i.e. an equilibrium in which

$$v_L < v_P < v_H. \tag{2.8}$$

As a consequence, $\widetilde{x}^* = \widetilde{X}^* = 0$.

2.5.1 Information Updating and Expected Gains from Trade

The crucial question is what happens in a bargaining situation with asymmetric information. Consider an agent with his P meeting someone willing to trade our agent's nonrecognisable third good. The latter agent must be either of the same type as he himself is and carry L, or a consumer of P. He cannot be of the same type and carry H because for such an agent giving up H for P implies a violation of (2.8). Therefore, our agent infers that with conditional probability $l/(l+p)$ he faces someone trying to get rid of L and with conditional probability $p/(l+p)$ someone offering H.

In both cases it becomes common knowledge that the agent with P has a high-quality good. If that was not the case both agents offering the third good would detect the low quality and refuse to trade.

The optimal strategy x of our agent maximises expected gains from trade

$$\frac{l}{l+p}\left[xv_L + (1-x)v_P\right] + \frac{p}{l+p}\left[xv_H + (1-x)v_P\right] - v_P \tag{2.9}$$

or, equivalently,

$$\frac{x}{l+p}\left[lv_L + pv_H - (l+p)v_P\right] \tag{2.10}$$

with respect to x given X.

2.5.2 The Problem of an Agent

Given the three possible characteristics of a bargaining situations, i.e. detection of low quality and immediate separation, double coincidence of wants and swap with certainty, or one-sided asymmetric information in the trade of a production good against a third good, we can write down an agent's valuations of holding P, L, and H as

$$\begin{aligned} v_P = \quad & \frac{\beta}{3}\left[pv_P + l\left(xv_L + (1-x)v_P\right) + hv_P\right] \\ & + \frac{\beta}{3}\left[p\left(xv_H + (1-x)v_P\right) + lv_P + h(u + v_P)\right] \\ & + \frac{\beta}{3}\left[p\left(X(u+v_P) + (1-X)v_P\right) + lv_P + hv_P\right], \end{aligned} \tag{2.11}$$

$$v_L = \frac{\beta}{3}\left[\left[p\left(Xv_P + (1-X)v_L\right) + lv_L + hv_L\right] + 2v_L\right], \tag{2.12}$$

$$v_H = \frac{\beta}{3}\left[2v_H + \left[p\left(u + v_P\right) + lv_H + hv_H\right]\right]. \tag{2.13}$$

As in the previous section we note them as a matrix equation

$$T'v' = w' \tag{2.14}$$

where $v' \equiv v$, the new transition matrix is given by

$$T' := \begin{bmatrix} r + lx + px & -lx & -px \\ -pX & r + pX & 0 \\ -p & 0 & r + p \end{bmatrix},$$

and the right-hand side is now

$$w' := \begin{bmatrix} pXu + hu \\ 0 \\ pu \end{bmatrix}.$$

Using Cramer's rule we calculate the valuations of v_P, v_L and v_H as

$$v_P = \frac{u}{|T'|}\left[(r + pX)(r + p)(pX + h) + p^2x(r + pX)\right], \tag{2.15}$$

$$v_L = \frac{u}{|T'|}\left[pX(r + p)(pX + h) + p^2x(pX)\right], \tag{2.16}$$

and

$$v_H = \frac{u}{|T'|}\left[(r + pX)p(pX + h) + p(r + pX)(r + px) + prlx\right], \tag{2.17}$$

respectively.

Substituting this into (2.10), the expression to be maximised with respect to x can now be written as

$$\frac{u}{|T'|}\frac{x}{(l + p)}\underbrace{[pr(r + pX)(p - pX - h) - lr(r + p)(pX + h)]}_{\Delta}. \tag{2.18}$$

Note that the value of Δ does not depend on the agent's strategy x but only on the other players' strategies X. Noting that $|T'| > 0$, it follows that his best-reply correspondence is given by

$$x(X) \in \begin{cases} \{0\} & \text{for } \Delta < 0 \\ [0,1] & \text{for } \Delta = 0 \\ \{1\} & \text{for } \Delta > 0. \end{cases}$$

2.5.3 Existence and Characterisation of an Unique Equilibrium

DEFINITION 1.2: *A stationary symmetric Nash equilibrium in the game with uncertainty is a value x^*, $x^* \in [0,1]$, which satisfies* *(i)* $U^k(x^*; X^*) \geq U^k(\hat{x}; X^*)$ *for all agents of type* k *(*$k = 1, 2, 3$*) and all* $\hat{x} \in [0, 1]$ *(optimality),*

(ii) $plx^* + p^2x^* = plX^* + ph$ *(stationarity) and*

(iii) $x^* = X^*$ *(symmetry).*

The next proposition states existence and uniqueness of such an equilibrium.

PROPOSITION 1.5: *There exists a unique stationary symmetric Nash equilibrium in the game with uncertainty,* x^* *, with* $0 < x^* < 1$.

Proof: Set $x = X$. Then stationarity requires $h = pX$, implying $p = q/(1+X)$ and $h = Xq/(1+X)$. Set $l \equiv 1-q$. After substitution into (2.18) $\Delta = 0$ if and only if

$$(r + pX)(p - 2pX) - 2(1-q)X(r+p) = 0, \tag{2.19}$$

where division by p and substituting for it yields

$$\left[r + \frac{Xq}{1+X}\right](1-2X) - 2(1-q)X\left[\frac{r(1+X)}{q} + 1\right] = 0. \tag{2.20}$$

Setting $X = 1$ the left hand side of (2.20) becomes negative, implying that $\Delta < 0$ for which the best reply is $x(1) = 0$, whereas for $X = 0$ the left hand side is positive and $\Delta > 0$ for which the best response is $x(0) = 1$. In both cases $x \neq X$ such that symmetry is violated. Since the left hand side is continuous and strictly monotonic in X there exists a unique fixed point X^* with $0 < X^* < 1$ for which (2.20) holds and $x(X^*) = X^*$ is a best reply. □

COROLLARY 1.6: *There is no efficient stationary symmetric Nash equilibrium in the game with uncertainty, i.e. one with* $x = 1$.

Proof: Follows directly from the proof of the previous proposition. □

Note that x^* solves the cubic equation

$$\frac{2(1-q)}{q}x^3 + \left[\frac{2(1-q)}{q} + 1 + (1/r)\right]x^2 + \left[\frac{2(1-q)}{q} + 1 - \frac{1-3(1-q)}{r}\right]x = 1$$

Two comparative-statics results are of particular interest. It is of little surprise that $\partial x^*/\partial q > 0$. With an increasing proportion of low quality goods it becomes more and more difficult to establish the third good as a medium of exchange since it is more likely to end up with a low quality good.

Moreover, it can be shown that $\partial x^*/\partial \beta > 0$, implying that the probability with which the third good becomes a medium of exchange is higher if individuals discount the future less strongly. The reason is that agents are less worried about the higher waiting cost associated with the acquisition of a low quality good.

Thus, we have established that the first-best result from the model without quality uncertainty cannot be implemented if we assume that a specific good's quality is not recognised by all agents. This result holds regardless of parameter values.

2.6 A Model with Quality Uncertainty and with Fiat Money

2.6.1 Introduction of Money and the Agent's Problem

The next step involves extending our model by introducing fiat money, M. Apart from having the usual properties of intrinsic uselessness and inconvertibility, fiat money in our model is characterised by the fact that it appears in uniform quality such that there is no problem of quality recognition for any agent.

The monetary authorities introduce m units, $m \in (0,1)$, of fiat money by buying from $q\alpha m$ agents with high-quality goods and from $m(1-q\alpha)$ agents with low-quality goods, where α is a parameter in the unit interval. Each of these agents receives exactly one unit of fiat money in exchange for the good he holds. Fiat money has the same storage properties as the other goods, ie. it is stored at zero cost and one unit exhausts the entire inventory capacity.

The crucial difference to the analysis of the previous model is that agents do not have to rely exclusively on the heterogeneous third good in order to accomplish indirect trade, ie. to prevent the economy from complete inactivity, because the homogeneous good fiat money can be used as a medium of exchange, too. In the present model we extend the agents' strategy space by adding a second dimension, with elements $\phi \in [0,1]$ representing the probabilities of swapping when money is involved in a bargaining situation. For simplicity we focus on identifying equilibria in which money is not traded for the third good, i.e. $v_H < v_M$ in an equilibrium where money is always accepted since good H cannot serve the purpose of being a medium of exchange as well as fiat money can.

The system of valuation functions for a representative agent is

$$T''v'' = w'' \tag{2.21}$$

where, after defining $l := 1 - q - m(1 - q\alpha)$ and the proportion of agents holding high-quality goods $\mathrm{A} := q(1 - \alpha m)$, the matrices are given by

$$T'' := \begin{bmatrix} r + lx + px + m\phi & -lx & -px & -m\phi \\ -pX & r + pX & 0 & 0 \\ -p & 0 & r + p + m\phi & -m\phi \\ -\mathrm{A}\Phi & 0 & 0 & r + \mathrm{A}\Phi \end{bmatrix},$$

$$v'' := \begin{bmatrix} v_P \\ v_L \\ v_H \\ v_M \end{bmatrix}, w'' := \begin{bmatrix} pXu + hu \\ 0 \\ pu \\ \mathrm{A}\Phi u \end{bmatrix}.$$

Φ denotes the probability with which other agents accept fiat money in trade.

2.6.2 Equilibrium: Existence and Characterisation

DEFINITION 1.3: *A stationary symmetric monetary Nash equilibrium in the game with uncertainty is a pair* (x^*, ϕ^*), x^*, $\phi^* \in [0,1]$, *which satisfies*

(i) $U^k(x*, \phi^*; X^*, \Phi^*) \geq U^k(\hat{x}, \hat{\phi}); X^*, \Phi^*)$ *for all agents of type* k *(*$k = 1,2,3$*) and all* $\hat{x}, \hat{\phi} \in [0,1]$ *(optimality),*

(ii) $plx^* + p^2x^* + pm\phi^* = plx^* + ph + \mathrm{A}m\Phi^*$ *(stationarity),*

(iii) $x^* = X^*$ *and* $\phi^* = \Phi^*$ *(symmetry) and*

(iv) $\phi^* > 0$ *(fiat money valued).*

PROPOSITION 1.7: *There exists a nonempty subset of the interval* $(0,1)$ *with the following property: For each element* m *of this subset there is a stationary symmetric monetary Nash equilibrium in the economy with quality uncertainty,* (x^*,ϕ^*) *with* $0 < x^* < 1$ *and* $\phi^* = 1$, *such that* $v_L < v_P < v_H < v_M$ *(consistency with transition matrix* T''*).*

Proof: Set $\phi = \Phi = 1$ and, without loss of generality $u = 1$. Stationarity is then equivalent to $p^2x = ph + mh$, and it can be checked that stationarity requires

$$p = \frac{q(1-\alpha m) - m + \sqrt{[q(1-\alpha m) - m]^2 + 4(1+x)mq(1-\alpha m)}}{2(1+x)}, \tag{2.22}$$

and $h = q(1 - \alpha m) - p$. First we prove the ranking of the valuations. From (4.18) we can see that

$$v_L = [pX/(r + pX)]v_P, \tag{2.23}$$

from which the first inequality in the proposition follows. Substituting this into the equilibrium condition (2.10), $p(v_H - v_P) + l(v_L - v_P) = 0$, yields

$$v_H = \frac{rl + p(r + px)}{p(r + px)} v_P, \tag{2.24}$$

from which $v_P < v_H$ immediately follows since r and l are bounded away from zero. Finally, system (4.18) reveals that $(r+p+m)v_H - mv_M = p(1+v_P)$. Since $1 + v_P = \frac{r+\mathrm{A}}{\mathrm{A}} v_M$ this is equivalent to

$$v_M = \frac{A(r+p)+Am}{p(r+A)+Am} v_H, \tag{2.25}$$

and therefore $v_H < v_M$.

Because of (2.23) and since

$$v_M = [A/(r+A])(1+v_P), \tag{2.26}$$

(4.18) can be simplified to yield the reduced system $\hat{T}''\hat{v}'' = \hat{w}''$ where

$$\hat{T}'' := \begin{bmatrix} r + \frac{lrx}{r+pX} + px + \frac{mr}{r+\mathrm{A}} & -px \\ -p - \frac{\mathrm{A}m}{r+\mathrm{A}} & r+p+m \end{bmatrix},$$

$$\hat{v}'' := \begin{bmatrix} v_P \\ v_H \end{bmatrix}, \hat{w}'' := \begin{bmatrix} pX + h + \frac{\mathrm{A}m}{r+\mathrm{A}} \\ p + \frac{\mathrm{A}m}{r+\mathrm{A}} \end{bmatrix}.$$

The solution to this system is given by $\hat{v}'' = (\hat{T}'')^{-1}\hat{w}''$, with

$$(\hat{T}'')^{-1} := \frac{1}{|\hat{T}''|} \begin{bmatrix} r+p+m & px \\ p + \frac{\mathrm{A}m}{r+\mathrm{A}} & r + \frac{lrx}{r+pX} + px + \frac{mr}{r+\mathrm{A}} \end{bmatrix}.$$

This yields

$$v_P = \frac{1}{|\hat{T}''|} \left[(r+p+m)\left(\mathrm{A} - (1-x)p + \frac{\mathrm{A}m}{r+\mathrm{A}}\right) + px\left(p + \frac{\mathrm{A}m}{r+\mathrm{A}}\right) \right]$$

and

$$v_H = \frac{1}{|\hat{T}''|} \left(p + \frac{\mathrm{A}m}{r+\mathrm{A}}\right) \left[(1-l) - (1-2x)p + \frac{r^2 + rpx + rlx}{r+px} \right]$$

(i) Set $X = x = 0$. (We can do so since, as in Proposition 1.5, the crucial expression below does not depend on x but on X only.) Then $p = \mathrm{A}$, and we get

$$v_P(0) = \frac{1}{|\hat{T}''(0)|}\left[(r + \mathrm{A} + m)\left(\frac{\mathrm{A}m}{r + \mathrm{A}}\right)\right]$$

and

$$v_H(0) = \frac{1}{|\hat{T}''(0)|}\left(\mathrm{A} + \frac{\mathrm{A}m}{r + \mathrm{A}}\right)(r + m).$$

Remember that for an equilibrium it is necessary that $\Delta = 0$. Note that $v_L(0) = 0$. Then $X = x = 0$ cannot be an equilibrium if $(1 - l - m)v_H(0) - (1 - m)v_P(0) > 0$. It can be checked that for each q and r there exists an $m > 0$ for which this inequality holds.

(ii) Set $X = x = 1$. This cannot be an equilibrium since substituting (2.24) into (2.10) for $m = 0$ yields, after some manipulations, $2pl > 0$, which is always the case given q is bounded away from 0 and 1. But then there exists a $m > 0$ such that this condition is still satisfied.

By applying a fixed-point theorem as in the proof of the previous proposition, we can find a unique symmetric mixed-strategy monetary equilibrium (with general money acceptance) for each value of $m > 0$ such that $X = x = 0$ and $X = x = 1$ do not constitute equilibria. □

Note that there are symmetric monetary equilibria with barter where $\phi^* < 1$. However, this type of equilibrium ceases to exist for values of ϕ such that $A\phi < p$ because then $v_M < v_H$ and money is no longer accepted by an agent with H in store. Here, we focus solely on the best possible outcome where money is always accepted with certainty.

Thus, we have found an equilibrium of the economy by introducing fiat money and letting agents trade their production goods for fiat money or for another medium of exchange.

2.7 Welfare Comparison

We measure welfare as the per-period consumption in the economy. More precisely, welfare W is given by

$$W(m) = p(px + h) + ph + mq(1 - \alpha m). \tag{2.27}$$

The welfare level of the nonmonetary economy without quality uncertainty equals q^2 and dominates the welfare levels of both the nonmonetary economy with quality uncertainty and the monetary economy. That is, the inefficiencies caused by asymmetric information cannot be removed by the introduction of fiat money. The introduction of fiat money into the economy without quality uncertainty would be detrimental to its welfare. This implies that in our models fiat money does not contribute anything to alleviate the double-coincidence-of-wants problem. The existing goods are sufficient to fulfil this role.

A more interesting question is concerned with the comparison between the economies with quality uncertainty. For this it is necessary to examine the effect of a slight increase in fiat money on the welfare of a previously nonmonetary economy.

PROPOSITION 6: *For each $q \in (0,1)$ there exist values of $m > 0$ and $\alpha \in (0,1]$ for which W is higher than at $m = 0$.*

Proof: Note that, due to the stationarity requirement, (2.27) can be rewritten as

$$W(m) = 3ph + m(2h + p). \tag{2.28}$$

Keeping in mind that p and h are functions of m, and that $h = q(1 - \alpha m) - p$, differentiation of (2.28) with respect to m at $m = 0$ yields

$$\frac{\partial W(0)}{\partial m} = 3(q - 2p)\frac{\partial p}{\partial m} - 3\alpha pq + 2h + p. \tag{2.29}$$

Totally differentiating $p^2x = h(p+m)$ (stationarity condition) with respect to m at $m = 0$, we obtain

$$p^2\frac{\partial x}{\partial m} + 2px\frac{\partial p}{\partial m} = (q-p)\left(1 + \frac{\partial p}{\partial m}\right) - \alpha qp. \tag{2.30}$$

We can also derive from the stationarity condition that

$$p^2\frac{\partial x}{\partial m} = h - q\frac{\partial p}{\partial m},$$

and by substituting this into (2.30) we have an equation that is linear in $\frac{\partial p}{\partial m}$, the solution to which is given by

$$\frac{\partial x}{\partial m} = \frac{\alpha qp}{q - ph}.$$

This allows us to rewrite (2.29) as

$$\frac{\partial W(0)}{\partial m} = -3\alpha qp^2\left(\frac{2-h}{q-ph}\right) + 2h + p. \tag{2.31}$$

For each $q \in (0,1)$ $2h + p > 0$. Thus, we can always find a value for α that is sufficiently small to ensure that $\frac{\partial W(0)}{\partial m}$ is strictly positive. □

Intuitively speaking, even though raising m means that there are less consumption goods around (if $\alpha > 0$), agents benefit through a shift of a positive proportion of the population from goods with a *relatively lower* to a good with a *relatively higher* valuation, ie. fiat money. This reflects the way in which money facilitates bilateral exchange.

As a consequence of the special assumption of limited storage capacity the choice of α is constrained. Due to this specific storage technology, money drives out high-quality goods such that their stocks at any moment in time are lower in the monetary economy than in the nonmonetary economy.

A substantial amount of indirect barter is still carried out in order to ensure occurrences of double-coincidence-of-wants events, which are fully efficient with respect to both the information problem (qualities of both goods are immediately recognised) and the waiting cost problem (no additional waiting cost for the agent with his production good in store).

Another interesting welfare aspect of the model revealed by simulation experiments is the behaviour of W when m is increased further. It can be shown for sufficiently high q that there is an optimal quantity of money at the point where m takes the value for which $v_H = v_M$ such that the agents are satiated with money balances. At this point $x^* = 0$, i.e. there is no more barter and the equilibrium collapses to a purely monetary one. If the quantity of money is expanded beyond this point, welfare decreases and becomes eventually lower than in the nonmonetary economy with uncertainty. This happens when the loss from less high-quality goods is just outweighed by the gains from facilitated transactions.

2.8 Summary and Conclusions

This chapter has demonstrated how under certain conditions rational agents choose a good without quality uncertainties and with good resale expectations to be a medium of exchange. The results have been derived under the assumption of risk-neutrality. With risk-aversion inefficiency would increase with or without fiat money. However, the relative gain from introducing fiat money into the economy would be larger than in the previous model. The study shows that fiat money may have positive value *because of* rather than *despite* its intrinsic futility once we emphasise its recognisability features.

One interesting conclusion can be drawn from the previous analysis if one associates the quality uncertainty with a range of financial assets (check-drawable deposits, bonds, stocks, etc.). While it is true that we have not explicitly modelled credit relationships, the kind of quality uncertainties encountered here are not completely unrelated to such assets. Their rates of return can usually not be foreseen perfectly and often there is the risk of the debtor defaulting. The model reinterpreted that way then explains why financial assets are not always accepted as medium of exchange but money is. Financial assets are linked with agents who promise to pay back but whose future income streams are uncertain or who may indulge in fraudulent activities. Such problems do not arise with unbacked currency.

Another implication of the model is that it seems not only very improbable that monetary economies are replaced by pure credit economies in the future, but also that it would be inefficient to do so. It is beneficial and necessary to have a standard good (paper money or some other appropriate commodity) upon which all assets are based.

Chapter 3

Reputation, Internal Finance and the Incentives to Invest

3.1 Introduction

This chapter studies the disruptive effects of reputational and financial distress on a firm's incentives to invest when it faces a set of external financiers who cannot directly observe its investment decisions. The unobservability of actions creates a moral-hazard problem in the relationship between the owner-manager (entrepreneur) of the firm and outside investors, which is shown to lead to inefficiently low levels of investment in a risky but productive technology. This in turn forces some, especially younger, businesses into liquidation, even though from an efficiency point of view they should continue to operate.

The study also suggests that the presence of an outside option for the borrower induces a positive effect of reputation, as measured by expected productivity, and internal finance on incentives to invest. The reason is that with a better reputation or more internal finance the liquidation probability is lowered and therefore the expected return on investment higher. Consequently, the closer a firm gets to the liquidation point the smaller such investment incentives become.

Both economists and policymakers have long suspected that many firms go into liquidation even when this is not desirable from a social planner's perspective. For some part this may be due to imperfections in the legal system which cause a suboptimal assignment of control rights over an insolvent business between different classes of claimants.[1] Even though the chapter is not without consequences for bankruptcy procedures, and some aspects of administration and restructuring procedures will be considered, the focus of this chapter is on credit market imperfections and the effects that the presence of an outside option has on incentive dynamics. In other words, we are interested in the nature of inefficiencies that may lead too many businesses into insolvency in the first place.

Moral hazard is very prominent in the literature as an explanation for financial constraints and their effects on investment, and it also plays a key role here. According to a number of empirical studies (see, e.g. Fazzari, Hubbard, and Petersen (1988)), financing constraints seem to be most severe for firms with little own assets. While their paper views internal assets primarily as net worth of tangibles[2] they may also encompass intangibles such as reputation. This study is an attempt to show how both insufficient internal financing capacity and the learning process through which a reputation is formed can diminish incentives to 'co-operate with lenders' in a way which makes liquidating the business privately preferable to continuation even when the social net present value of doing so is negative.

Such an outcome contrasts a study on the industry selection process by Jovanovic (1982). In his paper, firms are not financially constrained such that they can always choose output levels that maximise the one-period surplus subject to the expected value of a cost function parameter. The true value of this parameter is learnt through Bayesian updating based on observed realisations over time. But since (explicit) credit market agency problems are absent from that model it is no surprise that the

[1]See Aghion, Hart, and Moore (1993) for a discussion of insolvency practices.

[2]Bernanke and Gertler (1990) also analyse the effect of a low net worth on the agency costs of investment.

probabilities are first-best efficient.

As for the learning process, my contribution bears some similarity with a paper by Holmström (1982), which presents a model of a manager whose salary each period depends on his talent and effort as perceived by the owners of the firm.[3] He establishes that there are managerial incentives that stem from the possibility of building a reputation which lead to a reduction of the gap between second-best and first-best efficient efforts, and hence, inefficiency due to nonobservability of efforts is not as strong as it would be without such managerial career concerns. However, the beneficial effects of the desire to acquire a reputation are strongest at the beginning of a career and wear off gradually as time goes on. The driving forces behind this result are that only short-term contracts are feasible and that wage payments are made in advance, i.e. contingent on past performance data only. Learning on the expected managerial talent evolves according to a Bayesian updating process which is the more erratic the earlier an observation is made. Therefore, the possibility to manipulate the firm's learning becomes smaller over time such that efforts decline continually over time.

The present chapter, which deals with incentives in the context of external financing, shows that this monotonicity is not necessarily preserved if one introduces outside options. More specifically, it is not always the case that the younger the firm the lesser the extent of underinvestment. In fact, it turns out that firm age has an ambiguous effect on incentives to invest. As in Holmström (1982) it is true that attempts to manipulate the signal to lenders become less effective. But at the same time the liquidation probability may be lowered *ceteris paribus*, if the firm has experienced a series of mainly good productivity realisations, which makes it less likely that current investment does not bear any return in the future.

The entrepreneur is infinitely-lived and chooses the *size* of operation in a risky technology. As in Diamond (1989) the analysis is facilitated by the assumption that

[3]For a somewhat different approach to Bayesian learning and investment, see Tonks (1981).

in each period there is a completely new set of lenders, which excludes feasibility of long-term contracts. The entrepreneur cannot credibly commit to choose a certain investment level.[4] The reader may think of the entrepreneur as deciding about his own salary (or perquisites) after external funds have been provided, i.e. when outside investors no longer have direct control.[5]

The firm's entire output history is observable to the lenders at the beginning of each period, and so is the outcome realised at the end of the same period. The observability of output history enables each generation of external investors to update their beliefs about a firm's productivity, and the fact that current output is fully observable (and verifiable) suggests that the transfers specified by a contract should be fully contingent on outcomes, which will indeed be the case in this contribution.

It will be shown that a firm with a good reputation or high internal finance may be able reduce agency costs which not only leads to a higher total investment level but also reduces the exit probability. More precisely, the number of socially undesirable firm deaths is decreased as a consequence of the positive effect on incentives that is exerted by a good reputation or a relatively high level of own means.

The presence of financial constraints makes it more likely that, after a bad output realisation without immediate liquidation, a firms' assets are eroded to a point where continuation ceases to be worthwhile. These inefficiencies give rise to the introduction of monitoring or certain provisions in the bankruptcy law that have a potential to improve the outcome.

There is a number of studies on reputation acquisition in *debt markets.* Diamond (1989), e.g., introduces a model with moral hazard and a 'significant' amount of

---

[4]In this respect, the model is similar to Holmström (1982). Note that the assumption that funds are transferred to another party before any 'effort' has been exerted appears more adequate in a credit market context than in the one of managerial salary determination since such a sequence of events is in the nature of a borrower-lender relationship.

[5]Again, a particular aspect of the ongoing debate on corporate governance serves as an illustration of this point. The issue is whether companies should be forced to secure shareholders' approval for executive salary awards.

adverse selection. That reputation takes time to begin to work is true in that paper as well, i.e. the monotonicity result in Holmström (1982) does not necessarily hold there either, despite a rather different physical setup of the model.[6]

The presence of adverse selection ensures that in the beginning there may be a high proportion of 'bad' types in the population of borrowers. After a number of defaults by those borrowers adverse selection becomes less and less important, leading to a decrease in interest rates. It then becomes worthwhile to pick safe projects in order to preserve the good reputation acquired so far.

However, the assumption that no *new* borrowers which can select undesirable projects can enter seems somewhat restrictive. It is therefore important to note that the model here does not feature adverse selection effects, and therefore its results do not rely on changes in the composition of the set of borrowers. In other words, the destruction of incentives to 'co-operate' may occur even when there is only one borrower. The reason is that it is not only the age of the firm which determines incentives to invest but also the extent to which a firm is in (financial or reputational) distress.[7] The financial and reputational state of a firm determines how close it gets to the exit point. The higher the probability of switiching to the outside option the more relevant becomes its value. If this value is independent of current investment, the expected return on investment, and therefore the investment level, is lower.

The remainder of of the chapter is organised as follows. The next section outlines the framework of the model and establishes the occurrenceoccurence of underinvestment in each period. In the third section the entrepreneur's optimisation problem is formulated and, after establishing existence, uniqueness as well as differentiability of the value function, it is demonstrated that the liquidation probability is

[6]Note that while Holmström (1982) features a moral hazard problem it does not have adverse selection.

[7]Novaes and L. (1994) also deals with the relation between financial distress and the collapse of incentives. In their paper, however, it is the capital structure and the organisational form chosen by the managers to entrench themselves, and not the existence of an outside option, that drives their model.

ineffeciently high. In the section 3.4 we solve the entrepreneur's and analyse the comparative statics of the model. Section 3.5 considers the roles of monitoring and the bankruptcy law for liquidation rates, before the last section briefly summarises and concludes the chapter.

3.2 The Model

3.2.1 The Firm's Technology

Time evolves discretely as an infinite sequence of periods, $t = 0, 1, 2, ...$ The economy consists of two types of agents. There is a risk-neutral infinitely-lived entrepreneur and, following Diamond (1989), we assume that in each period there is a large number of lenders each of whom is alive for one period only. The entrepreneur is endowed with initial internal finance a_0 and with the access to a risky technology. Using this risky technology enables him to produce output

$$y_t = \tilde{\rho}_t + \pi \ln x_t \tag{3.1}$$

of some good in period t from input (investment) x_t. π is an exogenous productivity parameter, and the random variable $\tilde{\rho}_t = \rho(z_t)$ contributes to output in a time-dependent manner.[8] Its value is governed by the random variable $z_t \in Z = \mathbb{R}$, as defined by $z_t := \hat{\eta} + \epsilon_t$. $\hat{\eta} \in \mathbb{R}$, the true value of the firm's productivity parameter, is not known to any agent, but at $t = 0$ all are aware that it is drawn from a normal distribution with mean η_0 and variance $\frac{1}{h_0}$, with $h_0 \in \mathbb{R}_+$ denoting the precision of beliefs. The disturbance term ϵ_t is not observable either but known to be independently, identically and normally distributed with mean 0 and variance $\frac{1}{h_\epsilon}$ in each period.

[8]Instead of assuming an additive productivity shock one may choose to formulate the model with a multiplicative shock. This would entail a non-constant first-best efficient investment level, but the results of this chapter would qualitatively not be affected otherwise.

The function $\rho : \mathbb{R} \to [\underline{\rho}, \overline{\rho}]$ transforms z_t into a stochastic productivity parameter for each t.[9] It is positive, strictly increasing and continuous with $\lim_{z\to-\infty} \rho(z) = \underline{\rho} > 0$ and $\lim_{z\to\infty} \rho(z) = \overline{\rho} > \underline{\rho}$.

As described below, external finance, b_t, is determined on the credit market which is assumed to be perfectly competitive due to the large number of lenders. The latter have access to a safe storage technology which simply transforms one unit of input at the beginning of a period into R units of output at the end of the period.[10]

Given an amount of external finance, b_t, the entrepreneur has to allocate his total means at the beginning of t, $a_t + b_t$, with a_t representing internal finance, between investment x_t and consumption c_t. Whatever is not used as production input is consumed immediately by the entrepreneur, i.e. $c_t = a_t + b_t - x_t$, which yields current utility $u(c_t)$. $u : \mathbb{R}_+ \to \mathbb{R}$ is continuous, increasing and concave.

3.2.2 Equilibrium on the Capital Market

Let us now consider in more detail to what extent the entrepreneur is able to raise funds from the external investors. Since lenders are alive for only one period, long-term contracts are not feasible. However, all past information about the entrepreneur, i.e. his output levels and the total means he has disposed of in the beginning of each period, is publicly available to every generation.

Contracts are of the following type. At the beginning of the period some lenders give a total amount of b_t to the entrepreneur who then consumes and invests. We assume that he cannot commit to repay more than an exogenously determined share $1 - \mu$ of the revenues that accrue after production, $p_t y_t$, with $\mu \in [0, 1)$.[11] Assume also

---

[9] ρ may also capture effects of market structure on profitability, of developments on the demand side, or of product innovations.

[10] One may also think of the external investors as financial intermediaries with the possibility for refinancing at a time-independent net interest rate $R - 1$.

[11] Here one may think of a firm deciding not to pay out its entire profits as dividends but to retain a part of it.

that output demand is perfectly price-elastic in every period such that $p_t = p \equiv 1$ $\forall t$.[12] The share μ of the revenues is retained by the entrepreneur and serves as next period's internal finance. Thus, his financial assets evolve according to

$$a_{t+1} = \mu y_t. \tag{3.2}$$

This form of contract can be justified by the fact that output is observable whereas the sequence of investment rates, $\{x_t\}_{t=0}^{\infty}$, is not. An optimal contract should be based on the output realisations. Thus, we have a contingent contract rather than a standard debt contract as in Diamond (1989), where profits are non-observable. The contract we use here may not be optimal because we assume for simplicity that the entrepreneur does not optimise with respect to μ, the share of revenues he would credibly promise to repay.[13] We can therefore say that there are two sources of moral hazard, one being the non-contractibility of the investment shares and the other the commitment problem just mentioned.[14]

$\rho(\eta_t)$ is the expected value of the productivity parameter at t. We then state the arbitrage condition that determines the level of external funds as

$$Rb_t = (1 - \mu)[\rho(\eta_t) + \pi \ln x^*(a_t, \eta_t, h_t)], \tag{3.3}$$

where x_t^* is the value of x_t which the lenders expect an *optimising* entrepreneur to choose, given the current state (a_t, η_t, h_t). Thus, while x_t is not observable to the lenders, they are able to infer its value from solving the entrepreneur's optimisation problem (which is common knowledge). The condition simply says that the expected rate of return from lending to the firm must be equal to the opportunity cost R.

[12] Note that this makes the distinction between output and input goods redundant.

[13] Even if he could commit to a certain repayment the resulting contract would only be the most efficient in the class of linear sharing contracts, but still not necessarily optimal.

[14] Note that in our context the (second-best) optimal contract will generally not be achieved by setting $\mu = 0$. Even though this maximises the lenders' input, it minimises internal finance (setting it to zero in all but the initial period) such that the total means available to the entrepreneur at the beginning of a period may be higher with a value $\mu > 0$.

For convenience we define $\gamma := \frac{1}{R}$, and in order to make external finance attractive to the firm we assume $\gamma > \beta$, where $\beta \in (0,1)$ is the entrepreneur's discount factor.

At any point in time the entrepreneur may decide to liquidate his firm, in order to work as an employee in another firm for the remainder of his life. The associated wage payments have a present discounted value of q, in terms of consumption utility. The two activities are mutually exclusive.

The firm can be viewed as a collection of assets that give the entrepreneur access to the risky technology. However, he does not receive a revenue from the liquidation of those assets as they are worthless without his specific input of human capital. Moreover, we assume the presence of substantial legal costs which fully erode any financial assets left over after liquidation. This ensures that the entrepreneur's utility after liquidation is independent of his investment behaviour before.

3.2.3 Updating of Beliefs

Although x_t is not observable to the lenders, they are able to infer its value from solving the entrepreneur's optimisation problem (which is common knowledge). Together with the general observability of y_t, this implies that the realisation of z_t can be fully inferred by them.

The information that is relevant for updating beliefs is therefore the same for all agents. Bayesian updating is a straightforward procedure under our assumptions of independently, identically and normally distributed disturbance terms.[15] With the prior distribution of the mean of z_t at t being $\mathcal{N}\left(\eta_t, \frac{1}{h_t}\right)$ it can be shown that the posterior at $t+1$ (after a realisation of z_t) is again normal with mean

$$\eta_{t+1} = \frac{h_t}{h_t + h_\epsilon}\eta_t + \frac{h_\epsilon}{h_t + h_\epsilon}z_t, \tag{3.4}$$

[15]See, e.g., Berger (1985) for Bayesian updating under the normality assumption.

and variance

$$\frac{1}{h_{t+1}} = \frac{1}{h_t + h_\epsilon}, \tag{3.5}$$

which declines monotonically to zero over time, such that in the limit η will become fully known.

From (3.1) we see that what lenders observe at t, call it $\hat{z}_t$, is given by

$$\hat{z}_t = \rho^{-1}(y_t - \pi \ln x_t^*) = \rho^{-1}(\rho(z_t) + \pi \ln x_t - \pi \ln x_t^*). \tag{3.6}$$

If the entrepreneur actually behaves optimally and sets $x_t = x_t^*$, all agents always have the same information about the firm's productivity. There is a potential for the entrepreneur to manipulate the signal but this will be taken into account by the lenders such that in equilibrium $\hat{z}_t = z_t$.

We see from (3.4) and (3.5) that the impact of a given realisation of z_t on the revision of beliefs becomes less and less important as time goes by, since $\frac{h_\epsilon}{h_t + h_\epsilon}$ goes to zero. Each observation has the same weight, and each new observation is divided by an ever-increasing number of old observations.

3.2.4 First-best Efficiency

We end this section by analysing the investment problem of a social planner who is interested in overall welfare. The solution to this is fairly simple as all he needed to do is tell the entrepreneur to invest in each period up to the point where marginal social return is equal to marginal social cost, which with the simple technology used here would imply setting $x^{**} = \gamma\pi$. Note that this could be implemented by the lenders themselves if they could observe and contract upon the investment level chosen by the entrepreneur.

3.3 The Problem of the Entrepreneur

3.3.1 The Sequence Problem

The problem of the entrepreneur consists of choosing (i) a sequence of investment levels, and (ii) to decide when to switch to working as an employee.

More precisely, he has to find a feasible plan for the sequence of investment levels, $x_t(z^{t-1})$, $t = 0, 1, 2, ...$ which maximises the entrepreneur's sum of expected discounted returns, given a discount rate $\beta \in (0,1)$ and the history of shocks $z^{t-1} = (z_0, ..., z_{t-1})$, contained in the set of histories up to $t-1$, Z^{t-1}, *conditional on* its value not being too low to justify continuation.

In each period we can define a cutoff point $\underline{z}_t$ representing the minimum value the random variable must attain in order to prevent liquidation. $\underline{z}_t$ is determined by the state variables' values in t, and therefore ultimately by the history of shocks $\{z_\tau\}_{\tau=0}^{\tau=t-1}$, *conditional* on no element z_τ of this sequence being below the critical level $\underline{z}_\tau$ relevant for that period. Define by $\tilde{Z}^t := \{(\tilde{z}_0, ..., \tilde{z}_t) | \tilde{z}_\tau > \underline{z}_\tau, \tau = 1, ..., t\}$ the set of sequences of realisations for which this condition holds, and as $\backslash\tilde{Z}^t$ the set of sequences of realisations for which this condition holds only up to the period $t-1$, but not in period t.

Denote with v^* the function which attains the supremum in the corresponding sequence problem to maximise the expected discounted sum of per-period utilities, i.e.

$$v^*(\cdot) = \sup_{\{x_t\}_{t=0}^{\infty}} \left\{ u(c_0) + \sum_{t=1}^{\infty} \beta^t \int_{\tilde{Z}^{t-1}} u(c_t) \phi^t(z^{t-1}; \eta_{t-1}, h_{t-1}) dz^{t-1} \right. \\ \left. + q \sum_{t=1}^{\infty} \beta^t \int_{\backslash\tilde{Z}^{t-1}} \phi^t(z^{t-1}; \eta_{t-1}, h_{t-1}) dz^{t-1} \right\}. \tag{3.7}$$

The second term of the right-hand side is the expected continuation value, and the second term the discounted labour income q weighted by the liquidation probability.

Let the feasibility constraints on x_t be described by the correspondence $\Gamma : S \to \mathbb{R}$, where $S := \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ is the set of the state variables (a, η, h). Γ is nonempty-valued, its graph is measurable, and it has a measurable selection. Moreover, the law of motion associated with (3.2), (3.4) and (3.5) (denoted by m in what follows) is continuous and therefore also (Borel) measurable.

In what follows we assume the return function u to be linear.[16] Furthermore, we can show that the optimal plan is such that returns are bounded. To see this, we note that the marginal cost of investing is constant at unity, whereas, in equilibrium, its marginal benefit (in the form of higher internal funds) converges to zero as we let x go to infinity. Hence, as $\rho(z)$ is bounded, the value of internal finance a_t is always finite. Thus, the return function is bounded and it follows that v^* is bounded as well.

3.3.2 The Bellman Functional Equation

In this subsection we transform the sequence problem into the Bellman functional equation

$$\begin{aligned} v(a,\eta,h) = \quad & \max_x \quad a + \gamma(1-\mu)[\rho(\eta) + \pi \ln x^*(a,\eta,h)] - x \\ & + \beta \int \max[q, v(a',\eta',h')]\phi(z|\eta,h)dz, \end{aligned} \tag{3.8}$$

subject to

$$\begin{aligned} a' &= \mu[\rho(z) + \pi \ln x], && (3.9) \\ \eta' &= \frac{h}{h+h_\epsilon}\eta + \frac{h_\epsilon}{h+h_\epsilon}\rho^{-1}\left[\rho(z) + \pi \ln x - \pi \ln x^*\right], && (3.10) \\ h' &= h + h_\epsilon. && (3.11) \end{aligned}$$

[16]As it is continuous, the return function is clearly (Borel) measurable as well.

Given the properties of Γ, m and u, boundedness of v^* implies that the solution to the functional equation, v, is equal to v^*, and that any plan generated by the policy correspondence associated with (3.8) is optimal (see Stokey, Lucas, and Prescott (1989)). Since the return function is continuous and bounded it is Riemann-integrable with respect to ϕ. This and the non-emptiness of Γ imply that v^* is well-defined. For any feasible sequence $\{x_t\}_{t=0}^{\infty}$ given *any* $s_0 := (a_0, \eta_0, h_0) \in S$, a solution to (3.8) must satisfy

$$v(s_0) \geq u[s_0, x_0] + \beta\Phi(\underline{z}_0|s_0)q + \beta \int_{\tilde{Z}^0} v[s_1(x_0, \tilde{z}_0)]\phi(\tilde{z}_0|s_0)d\tilde{z}_0,$$

which can be restated as

$$\begin{aligned} v(s_0) \quad \geq \quad & u[s_0, x_0] + \beta\Phi(\underline{z}_0|s_0)q + \beta \int_{\tilde{Z}^0} \Big\{ u[s_1(\tilde{z}^0), x_1(\tilde{z}^0)] \\ & + \beta\Phi[\underline{z}_1(\tilde{z}^0|s_0)q + \beta \int_{\underline{z}_1(\tilde{z}^0)} v[s_2(\tilde{z}^1), x_2]\phi(\tilde{z}_1|s_1)d\tilde{z}_1 \Big\} \phi(\tilde{z}_0|s_0)d\tilde{z}_0. \end{aligned}$$

Further rearranging of the right hand side yields

$$\begin{aligned} v(s_0) \quad \geq \quad & u[s_0, x_0] + \beta\Phi(\underline{z}_0|s_0)q + \beta \int_{\tilde{Z}^0} \Big(u[s_1(\tilde{z}^0), x_1(\tilde{z}^0)] \\ & + \beta^2\Phi[\underline{z}_1(\tilde{z}^0)]q \Big) \phi(\tilde{z}_0|s_0)d\tilde{z}_0 + \beta^2 \int_{\tilde{Z}^1} v[s_2(\tilde{z}^1)]\phi^1(\tilde{z}_2|s_1)d\tilde{z}^1. \end{aligned}$$

It follows by induction that

$$\begin{aligned} v(s_0) \quad \geq \quad & u[s_0, x_0] + \beta\Phi(\underline{z}_0|s_0)q + \sum_{t=1}^{n} \int_{\tilde{Z}^t} \beta^t \Big(u[s_t(\tilde{z}^t), x_t(\tilde{z}^t)] \\ & + \beta^t\Phi[\underline{z}_t(\tilde{z}^{t-1})]q \Big) + \beta^{n+1} \int_{\tilde{Z}^n} v[s_{n+1}(\tilde{z}^n)]\phi(\tilde{z}_{n+1}|.)d\tilde{z}_n, \end{aligned}$$

$n = 1, 2, \ldots$. As we let n go to infinity boundedness of v is a sufficient condition for the last term to tend to zero in the limit. The equation then becomes

$$v(s_0) \geq u[s_0, x_0] + \beta\Phi(\underline{z}_0|s_0)q + \sum_{t=1}^{\infty} \int_{\tilde{Z}^t} \beta^t \Big(u[s_t(\tilde{z}^t), x_t(\tilde{z}^t)] + \beta^t\Phi[\underline{z}_t(\tilde{z}^{t-1})]q \Big)$$

with the right-hand side simply being the expected discounted value of the infinite return stream under $\{x_t\}_{t=0}^{\infty}$.

Next take a policy $\{x_t^*\}_{t=0}^{\infty}$ generated by the policy correspondence associated with (3.8),

$$G(s_0) = \{x \in \Gamma(s_0) : v(s_0) = u[s_0, x] + \beta \int v[m(s_0, x, z), z]\phi(z|.)dz\},$$

which is nonempty and permits a measurable selection. With this policy the above holds as an equality, and we note that the right hand side is then, in the limit, identical to the definition of the supremum function $v^*(s_0)$. Since s_0 was arbitrary, this establishes that $v = v^*$.

As explained above, q is the present discounted value of the outside option, to which the entrepreneur will switch immediately if the value of v drops below q. Lemma 1 demonstrates that a solution to (3.8) exists, and that it is unique and bounded.

LEMMA 3.1: *There exists a unique, continuous and bounded function v which solves (3.8). The function is strictly increasing in a and η.*

Proof: Define an operator T that yields v as a fixed point of the equation $v = Tv$, i.e.

$$(Tv)(a, \eta, h) = \max_x a + b(a, \eta, h) + \beta \int \max[q, v(a', \eta', h')]\phi(z|\eta, h)dz.$$

Note that the transition function $\phi(z|\eta, h)$ has the Feller property (Stokey and Lucas [1989], p.220): As the distributions of η and $\{\epsilon_t\}_{t=0}^{\infty}$ are normal, boundedness and continuity of a function f are preserved by the operation $\int_Z f\phi(z|\eta, h)dz$ on this function.

Lemma 12.14 in Stokey and Lucas (1989) demonstrates that under these conditions the operator T also preserves continuity and boundedness of v, which is a consequence of the continuity and the boundedness of the return function. Since for two bounded functions f and g, $f(x) \leq g(x)$, $\forall x \in X$, $(Tf)(x) \leq (Tg)(x)$, $\forall x \in X$ (monotonicity), and since $[T(f + a)](x) \leq (Tf)(x) + \beta a$, $\beta \in (0, 1)$ for all bounded functions f, $a \geq 0$, $x \in X$ (discounting), Blackwell's sufficient conditions for a contraction are satisfied, and accordingly the Banach fixed point theorem ensures uniqueness.

It follows as a corollary to the Contraction Mapping Theorem (Stokey and Lucas [1989], p. 50) that if the operator T maps a nondecreasing, bounded and continuous function into the set of strictly increasing, bounded and continuous functions, then the fixed point v associated with T must itself be an element of this set.

That v is increasing in a follows from the fact that if we raise a by some amount, the entrepreneur can take the same investment decisions, leaving expected future rewards unchanged, and consume more in the current period, increasing overall utility.

Next note that any monotonicity property of a function is weakly preserved by the contraction operator T and that $v = \lim_{n\to\infty} T^n f$. We can pick any function f that is nondecreasing in η. Then f is nondecreasing in z because the transition functions for these state variables are, which in turn implies that $\max[q, f(a', \eta', h')]$ is nondecreasing in z. As an increase in η shifts distributional weight from points with lower functional values to points with (weakly) higher functional values, the same applies to $\int \max[q, v(a', \eta', h')]\phi(z|\eta, h)dz$. As the return function u is strictly increasing in η, v is is strictly increasing in η. □

According to (3.8) the entrepreneur uses the risky technology as long as $v(a, \eta, h) \geq q$. The state of his firm allows him to raise sufficient funds to ensure the risky activity is more profitable. Once the value of $v(a, \eta, h)$ drops below q he will turn to the labour market to offer his services as an employee. He will stay there ever after, i.e. there is no return to the original business because the state variables will never again assume values that would warrant a return to entrepreneurial activities.

Lemma 3.1 implies that we can define a function $\underline{z} : S \rightarrow \mathbb{R}$ that assigns to each point in the state space a cutoff level for the realisation of z, denoted by $\underline{z}(a, \eta, h)$, for which next periods value of v, $v(a', \eta', h')$ is just equal to q.

3.3.3 Growth Rates and Exit Probabilities

The fact that v is increasing in η provides the basis for another result, which is summarised in the following proposition.

PROPOSITION 3.2: *For a given a level of reputation, η, the volatility of the firm's growth rates and its exit probability are decreasing in its age (as measured by h).*

Proof: The decrease in volatility over time follows directly from the weakening of each new observation's impact on the revision of beliefs about productivity. Furthermore, according to Lemma 1, v is increasing in η. Thus, conditional on having survived so far, next period's reputation, η', would have to be lower than some cutoff level $\underline{\eta}' < \hat{\eta}$ to provoke exit of the firm. Considering the law of motion of the reputation variable in (3.8), the critical realisation of z required to reach this cutoff level, given any pre-posterior mean η, decreases in h. But the probability that the actual realisation exceeds this critical value (survival) is then increasing in h. □

Proposition 3.2 implies that, all other things equal, a young firm is more likely to exit than an older firm. Conversely, young firms with good outcomes grow more quickly than older firms with the same *a priori* expected value of productivity and the same realisation of the random variable z. More generally, the volatility of growth rates decreases in age. This is due to the fact that the updating of the expected productivity parameter is more responsive to new observations when the firm is young and relatively little information has been gathered about it.

These findings *per se* are by no means new, as they match those in Jovanovic (1982). However, whereas in the latter contribution the selection mechanism is first-best efficient, this is not the case here. We will demonstrate in the next section that under normal circumstances the entrepreneur's optimal investment is smaller than the first-best efficient level established in Section 3.2. Thus, in the present model an entrepreneur may shut down his firm operations even when this is not efficient from a social planner's point of view.

That firms exit even when continuation would be justified is a consequence of underinvestment, which itself is due to the agency problems stemming from the nonobservability of investment. If an entrepreneur whose reputation is just slightly below the cutoff level could commit to a higher investment level x (or if he possessed more internal finance a), it would be possible for him to obtain higher total funds. Thus, inputs used in production would unambiguously increase and get closer to the socially optimal level without affecting the lenders' welfare. At the same time current and expected future consumption could be augmented such that our entrepreneur would be better off holding on to the business.

The intuition for this is as follows. As the assets become worthless in the case of liquidation, the entrepreneur does not expect to reap the entire benefits from his investment. Whatever is invested in the firm in a particular period will be lost to the entrepreneur if the productivity shock is sufficiently bad to render the outside option more attractive than continuation of the project. Benefits from higher investment by the entrepreneur are not lost, however, to the economy as a whole in the case of liquidation, since creditors or lawyers may obtain them. This observation implies that, compared to the first-best efficient outcome there will be underinvestment in the present model.

The next section focuses on the choice of an optimal policy. In particular, we are interested in the effect of reputation (expected productivity) and the level of own finance on the entrepreneur's investment behaviour. Furthermore, we will attempt to shed some light on how the optimal policy is affected by the age of the firm.

3.4 The Roles of Reputation and Internal Finance

3.4.1 Differentiability of the Value Function

To find the optimal policy by deriving first-order and envelope conditions we need to show that the value function v is differentiable. Using arguments from Blume, Easley, and O'Hara (1982), we first state and then prove the following lemma which ensures differentiability of the value function of any desired order.

LEMMA 3.3: *Given independence of the distribution of stochastic shocks, if the return function and the law of motion are differentiable p times, then the value function is differentiable p-1 times*

Proof. The assumption of independently distributed shocks allows us to transform the original functional equation. In particular, we can change variables such that next period's state s' becomes the variable with respect to which we integrate.

First define the function $w(s) := v(s) - q$ and write (3.8) as

$$w(s) = -(1 - \beta\Phi[\underline{z}|s])q + u(s, x^*) + \beta \int_{\underline{z}(s)}^{\infty} w[s'(s, z)]d\Phi(z),$$

where $\underline{z}(s)$ is now defined as the value at which $w(s') = 0$.

Due to the invertibility of the law of motion for the state variable $s' = m(s, x, z')$ we can define the inverse function for z' by $z' = z^I(s, x, \tilde{s}')$ which is also p times differentiable. The probability that next period's state variable $\tilde{s}'$ (conditional on continued viability) is in a Borel subset of S, say $\tilde{S}$, given current state s and action x, is

$$\int_{\tilde{S}} w(\tilde{s}')\phi[z^I(s, x, \tilde{S})]\frac{\partial z^I(s, x, \tilde{s}')}{\partial \tilde{s}'}d\tilde{s}'.$$

Thus, one can rewrite the modified value function as

$$w(s) = -(1 - \beta\Phi[.])q + u(s, x^*) + \beta \int_{\tilde{S}} w(\tilde{s}')\phi[z^I(s, x, \tilde{s}')]\frac{\partial z^I(s, x, \tilde{s}')}{\partial \tilde{s}'}d\tilde{s}'.$$

On the right hand side s does not appear as an argument of w, but only of functions which by assumption are differentiable. Obviously, the integrand is then differentiable $p - 1$ times. □

We rewrite our functional equation (3.8) as

$$\begin{aligned} v(a,\eta,h) = \quad & \max_x \;\; a + \gamma(1-\mu)[\rho(\eta) + \pi \ln x^*(a,\eta,h)] - x + \beta\Phi[\underline{z}|\eta,h)]q \\ & + \beta \int_{\underline{z}(a,\eta,h)}^{\infty} v(a',\eta',h')\phi(z|\eta,h)dz, \end{aligned} \tag{3.12}$$

where the constraints (laws of motion) are the same as in (3.9), (3.10) and (3.11).

3.4.2 The Euler Equation

If there were no outside option and no external investors at all, such that the entrepreneur would have to finance his operations entirely by himself (through negative consumption), his optimal per-period investment choice in the case of continuation would be $x = \beta\pi$, which is obviously less than the first-best optimum of $\gamma\pi$. One of the questions addressed here is how the presence of an imperfectly operating credit market and an outside option affects the entrepreneur's choice in relation to the first-best.

As the return function is bounded the transversality condition associated with the above problem is satisfied. Hence, the first-order condition, combined with the envelope condition are necessary and sufficient for an investment plan to be optimal. Lemma 3.3 allows us to derive both first-order and envelope conditions.

The first-order condition is

$$\begin{aligned} -1 \;\; + \;\; & \beta \int_{\underline{z}}^{\infty} \left[\frac{\mu\pi}{x} \frac{\partial v(a',\eta',h')}{\partial a'} \right. \\ + \;\; & \left. \frac{h_\epsilon}{h+h_\epsilon} \frac{d\rho^{-1}[\rho(z) + \pi \ln x - \pi \ln x^*]}{d\tilde{\rho} \frac{\pi}{x} \frac{\partial v(a',\eta',h')}{\partial \eta'}} \right] \phi(z|\eta,h) = 0, \end{aligned} \tag{3.13}$$

and the envelope conditions with respect to a' and η' are

$$\frac{\partial v(a',\eta',h')}{\partial a'} = 1 + \gamma(1-\mu)\Delta_a(\cdot) \tag{3.14}$$

and

$$\frac{\partial v(a',\eta',h')}{\partial \eta'} = \gamma(1-\mu)[\rho_z(H\eta+(1-H)z)+\Delta_\eta(\cdot)], \tag{3.15}$$

respectively.

$$\Delta_a(\cdot) \equiv \frac{x^*_a(\mu\rho(z)lnx, H\eta+(1-H)z, h+h_\epsilon)}{x^*(\mu\rho(z)lnx, H\eta+(1-H)z, h+h_\epsilon)}$$

and

$$\Delta_\eta(\cdot) \equiv \frac{x^*_\eta(\mu\rho(z)lnx, H\eta+(1-H)z, h+h_\epsilon)}{x^*(\mu\rho(z)lnx, H\eta+(1-H)z, h+h_\epsilon)}$$

are the percentage changes in next period's investment in response to an infinitesimal increase in the current period's internal finance and reputation, respectively. We have defined $H \equiv \frac{h}{h+h_\epsilon}$, and hence $1-H = \frac{h_\epsilon}{h+h_\epsilon}$.

It follows from the inverse function theorem that $\frac{d\rho^{-1}(\cdot)}{d\tilde{\rho}} = \frac{1}{d\rho(\cdot)/dz} \equiv [\rho_z(\cdot)]^{-1}$. Replacing (3.14) and (3.15) in (3.13) and using the equilibrium condition $x = x^*$, we obtain an Euler equation, the solution to which gives is the entrepreneur's optimal investment decision:

$$\begin{aligned} x^* \;=\; & \beta\int_{\underline{z}(a,\eta,h)}^{\infty} \pi\mu[1+\gamma(1-\mu)\Delta_a(\cdot)] \qquad (3.16)\\ & + \frac{h_\epsilon}{h+h_\epsilon}[\rho_z(z)]^{-1}\pi\gamma(1-\mu)[\rho_z(H\eta+(1-H)z)+\Delta_\eta(\cdot)]\Big]\,\phi(z|\eta,h)dz. \end{aligned}$$

Unfortunately, one cannot obtain an explicit solution for x^* as both Δ_a and Δ_η are functions of next period's internal-finance level and therefore of x^*. In the light of the diminishing marginal returns exhibited by entrepreneur's technology it seems plausible to assume that both these function are at least nonincreasing in internal finance. But then there exists a unique solution to (3.16).

Note that the only place where a appears in this equation is in the the lower bound of the integration interval, as the functions Δ_a and Δ_η do not depend on current internal finance. Hence, if there were no outside option such that $\underline{z}(a,\eta,h)$ goes to $-\infty$ the equilibrium choice x^* would be completely independent of a. Even in the presence of an outside option the responsiveness of x^* to changes in a is limited if the precision h_0 of the initial signal for the true productivity value $\hat{\eta}$ is sufficiently high relative to h_ϵ, the (constant) precision of the distribution of the disturbance term.

We use this observation to prove the next proposition which states the underinvestment result mentioned already in the previous section.

PROPOSITION 3.4: *For sufficiently small values of $\frac{h_\epsilon}{h_0}$ the entrepreneur invests less than the first-best optimum in each period. This result holds even when there is no outside option.*

Proof: As we lower $\frac{h_\epsilon}{h_0}$, Δ_a decreases as well because that diminishes the impact of investment on the external investors' updating. For one thing, this has the effect of depressing the first term in the integrand of (3.16) below π.[17] Moreover, the second term can be made arbitrarily close to zero such that the overall value of the right-hand side is smaller than the first-best optimum $\gamma\pi$, even when let $\underline{z}(a, \eta, h) \to -\infty$. □

Proposition 3.4 leads us directly to one of the central results of this chapter, namely the inefficiently high death rate of firms.

COROLLARY 3.5: *Under the assumption of Proposition 3.4 the liquidation rate of firms is inefficiently high.*

Proof: If the firm could commit to increase its investment level slightly there would be room for Pareto improvement, because at the private optimum (where we have underinvestment according to Proposition 3.4) marginal return on investment his higher than at the social optimum. □

REMARK: For high values of $\frac{h_\epsilon}{h_0}$ we cannot exclude the possibility that the firm overinvests relative to the social optimum, at least for a number of periods in the beginning. It is conceivable that the incentive to manipulate the output signal to

[17]In fact, for this statement to be true regardless of the value of μ, the value of Δ_a simply must not exceed 2.

the external investors is so strong that an entrepreneur chooses to invest more than $\gamma\pi$. However, this incentive will wear off eventually as h_t grows over time. In the case of overinvestment the liquidation rate is obviously to small from a welfare point of view.

3.4.3 Comparative Statics

The comparative statics of the investment decision with respect to the exogenous parameters are very straightforward. Investment is strictly increasing in the productivity parameter π, the discount factor β, and the inverse of the lender's cost of refinancing, γ.

The effect of the state variables is less clear-cut, as only the impact of internal finance is unambiguous. Remember that due to Lemma 3.1 the function $\underline{z} : X \to \mathbb{R}$ assigns, for a given (η, h), a shock cutoff level to each value of internal finance. More precisely, $\underline{z}(a, \eta, h)$ is decreasing in a because the probability of exit is diminished by a higher level of finance. As already mentioned above the integrand on the right-hand side of (3.16) does not depend on a, and hence, the entrepreneur's chosen investment increases in the level of own finance he disposes of.

The effect of reputation (expected productivity) on investment is more complex, as a change in that state variable affects not only the support of the truncated distribution of shocks, i.e. $\underline{z}$, but also the distribution ϕ itself, as well as the integrand in (3.16). The combined impact of the latter two, which may be called 'productivity effects', is very complex and remains ambiguous even if one could make precise statements about how the functions Δ_a and Δ_η depend on η. If the integrand responds positively to increases in η and z the expected *marginal* value of reputation (in terms of financing capacity) increases in reputation, which means that distributional weight is shifted from lower to higher values of the integrand. Hence, this partial effect of reputation on investment is positive. If, however, the expected *marginal* value of reputation (in terms of financing capacity) decreases in the firm's reputation, which means that distributional weight is shifted from higher to lower values of the integrand, the opposite will occur and the productivity effect on investment is negative.

What we are primarily interested in this subsection, however, is how the presence of an outside option affects investment in response to a change in reputation and how this is related to the probability of liquidation. Here the results are more clear-cut. It follows from Lemma 3.1 that, given (a, h), $\underline{z}$ is decreasing in η. In other words, an improved reputation diminishes the probability of the firm's exit as a consequence of which investment is augmented. Nevertheless, it is not possible to make an unambiguous statement about the overall effect of reputation on investment.

Comparative statics with regards to the age of the firm, as measured by the state variable h, turns out to be even more difficult. In particular, we cannot tell whether $\underline{z}$ is decreasing in h throughout, as we do not know whether $\max[q, v]$ is convex or not. If $\underline{z}$ *is* decreasing in h then the exit probability is lowered by age. This counteracts the negative effect on investment stemming from a reduction in the incentives to build a reputation (a reduction in $\frac{h_\epsilon}{h+h_\epsilon}$ over time). Otherwise the negative incentive effect is reinforced by an increased exit probability.

However, equation (3.16) does reveal something about investment behaviour in the limit as the firm grows infinitely old.[18] Note that as h goes to infinity the second term of the integrand goes to zero and $\underline{z}(a, \eta, h)$ converges to $-\infty$ and $\Delta_a(\cdot)$ to the zero function. Conditional on surviving that long, the true productivity of the firm is learnt perfectly, implying that there is no shock bad enough to trigger exit of the firm. Thus, in the limit we may write

$$x^* = \beta \int_{-\infty}^{\infty} \pi \phi(z|\eta, h_\epsilon) dz = \beta\pi. \tag{3.17}$$

[18]Strictly speaking, in order to converge to such a limiting behaviour it would be required that the firm never experiences a shock that is bad enough to warrant liquidation. This, however, is a zero-probability event.

If the firm were to survive infinitely long, the firm is no longer able to credibly commit to an investment above $\beta\pi$. The reason is that there are no more incentives to acquire a reputation and therefore external investors are no longer willing to lend funds beyond that level.

We summarise the most important comparative-statics results in the following proposition.

PROPOSITION 3.6: *The equilibrium investment level, x^*, is*

(i)increasing in internal finance a,

(ii) increasing in reputation if and only if the lower exit probability is not offset by the productivity effect,

(iii) increasing in age (precision of the productivity signal) if and only if a lower exit probability is not offset by the reputation-building disincentive caused by age.

Proposition 3.6 and the preceding discussion underline the significance of the outside option for investment behaviour. It establishes a positive relationship between reputation and, possibly, age on one hand and incentives to invest on the other. The lower bound of the set of viable shock realisation diminishes with an increase in η. This represents a lower exit probability for the following period, making it more worthwhile to invest in the risky technology. A low current reputation implies a high likelihood of becoming an employee in the next period, and the level of current investment is irrelevant for the value of that new career. This exit-probability effect reinforces or counteracts the productivity if the latter is positive (substitution dominates income effect) or negative (income dominates substitution effect), respectively.

Similarly, *if* higher age of a firm means lower exit probability, it may, *ceteris paribus* become more lucrative for an older firm to invest than for a younger firm, even though reputation-building incentives work in the opposite direction. Again, a lower exit probability implies that states with a zero marginal return on investment become less likely.

The decline of reputation-building incentives stems from the decrease in the responsiveness of the updating rule to good outcomes. Trying to manipulate the signal received by the lenders becomes more and more costly. This effect has been identified by Holmström (1982). In his paper, however, the manager incentives to exert effort exhibit an unambiguously monotonic decrease, whereas here it cannot be ruled out that this negative effect of age is outweighed by a potentially positive effect of higher survival probability which implies a higher expected marginal rate of return on investment.

Thus, due to the outside option it may well be the case that the investment incentives for a young firm are not as high as implied by a model of the Holmström (1982) type where such an outside option is absent. In other words, it may take time for incentives to begin to work. Young firms might not be willing to invest too much in the beginning because the probability of early exit and hence of not reaping any benefits is very high. They may prefer to keep inputs at a low level in early periods in order to obtain more information about their own productivity.

The detrimental direct effect of this kind of uncertainty on investment is amplified by the indirect effect it has on next period's internal finance and therefore on the level of external funds provided, as they are linked to each other. Thus, if in reality we observe external investors putting in rather small amounts into a starting business, this may be a consequence not only of attempting to gain more information about the quality of the management but also of responding to the firm's own low willingness to invest.

The possible non-monotonicity of the incentive schedule over time can also be observed in Diamond (1989), but for quite different reasons. While Diamond (1989) relies on the presence of asymmetric information and changes in the structure of credit applicant types, the present model is entirely independent of the industry's cross-section. It is the existence of an outside option that is at the core of the comparative-statics results here. *Per se* the desire to build a reputation reduces some of the capital market inefficiency created by moral hazard, especially in early stages of a firm's life. However, this beneficial effect is at least partially destroyed when the owner-manager of a firm has alternative ways of making a living.

Whenever investment responds negatively to age, as predicted by a model without outside option, another interesting result of our analysis is a possible asymmetry between upward and downward mobility of growth rates. A bad outcome leads to a downgrading of the reputation measure, and at the same time the precision h of its distribution, increases. The overall effect is then one of reducing the investment rate. A good outcome, however, may trigger two opposing effects. Improved reputation should cause the investment rate to increase. But h is still increasing and its impact may at least partially offset the first effect. Thus, the reaction to a (moderately) good outcome may be ambiguous.

3.4.4 Evolution of Internal Funds

Finally, we consider the evolution of the firm's internal funds. The expected value of a', given today's assets a, can be written as

$$E[a'] = \int_{\underline{z}(a,\eta,h)}^{\infty} [\rho(z) + \pi \ln x^*(a,\eta,h)]\phi(z|\eta,h)dz. \tag{3.18}$$

As the set of values that a takes on equilibrium paths is bounded, $E[a'] = a$ has a fixed point. To examine the dynamics of internal funds two cases need to be differentiated. Firstly, if $E[a']/a$ is decreasing in a, i.e. when the positive effect of internal finance on exit-avoidance probability (and therefore on investment) is weak or moderate, the system is stable. For low values of a, $E[a']/a > 1$, such that the firm's assets are driven up towards the fixed point, and for high values the converse holds. This first case is more likely to occur if the capital market imperfections are rather weak, as it implies that the influence of internal funds on investment is not very strong.

In contrast, the second case occurs when investment is very responsive to changes in a, or, more formally, when $E[a']/a$ is increasing in a. This reflects a situation in which capital market imperfections are very prominent such that a small variation in internal finance has substantial impact on investment and exit-avoidance probability. Unsuccessful firms have little internal funds, such that $E[a']/a < 1$. This raises expectations of being driven out of business in the near future. Conversely, successful firms are expected to grow even more, since for them $E[a']/a$ exceeds unity. Thus, the stronger capital market imperfections the more divergent growth rates become.

3.5 Monitored Finance and Administration Procedures

In Section 3.3 it was shown that due to agency problems too many firms are liquidated. As the entrepreneur cannot credibly commit to a higher investment rate he cannot attract external funds to an extent that would allow him to produce at a more profitable level. This section considers two mechanisms, monitoring and administration procedures in bankruptcy laws, which may improve on this situation.

Without explicitly modelling financial intermediation, our framework can also provide a rationale for lenders' monitoring of some borrowers. Monitoring can be thought of as an activity which either directly allows the entrepreneur to commit to a higher investment rate, i.e. increases x^*, or is appropriate to raise reputation η by some amount δ. The former effect increases expected internal finance next period and therefore the willingness of external investors to lend money in the future. The latter, which can be thought of as stemming from the introduction of additional business expertise which raises expected productivity (or *effective* reputation) to $\eta + \delta$.[19] The two interpretations are qualitatively more or less equivalent in our setting.

Assume that our firm can hire one of the lenders as a monitor at a certain cost

[19]Note that in many companies, especially those operating in financial systems which are strongly bank-oriented, financial intermediaries have a substantive influence on business decisions.

c_m. This may be particularly profitable for a firm exhibiting low incentives to invest, which, as the previous section has demonstrated, may be due to reputational distress (low η), young age (low precision h), or both. It may be desirable to incur the monitoring cost in order to increase the value of the business via such a commitment device. It may be a less profitable action for firms with a sufficiently high η or h, as for those firms incentive problems are not as important. In other words, it is rather the young or distressed firms that rely more heavily on monitored funds than others, whereas older, well-positioned are more likely to seek direct external finance.[20]

PROPOSITION 3.7: *Given a sufficiently small monitoring cost c_m, there exists, for each precision h, a critical level of reputation, $\hat{\eta}_m$, such that a firm with $\eta < \eta_m$ wishes to be monitored whereas a firm with $\eta \geq \eta_m$ does not. Hiring a monitor can prevent some firms from inefficient liqudation.*

The presence of monitoring financial intermediaries such as banks with their monitoring abilities geared towards firms in reputational distress, or venture capitalists with their more specific knowledge about new and risky projects may prevent some undesirable (inefficient) liquidations.

We now turn to the second potential mechanism to alleviate the undesirable effects of excessive liquidation. Assume that the bankruptcy law gives legal authorities an instrument which allows them to exert control over the actions taken by the entrepreneur at a cost $c_a > 0$. Such a mechanism may be referred to as 'administration procedures'. Usually, this entirely removes control from the entrepreneur, thereby more or less credibly creating the prospect of a higher investment level benefitting external investors.[21] This would increase external funds provided and therefore the profitability of the firm.

Instead of simply liquidating the business as before, an entrepreneur in reputational

[20]Similar results have emerged from a number of other papers (e.g., Diamond (1991)).

[21]One alternative activity usually associated with 'administration procedures' is the restructuring of a firm in financial difficulties. Restructuring may have a similar beneficial effect on the willingness of outsiders to invest.

or financial turmoil can concede his control rights to an administrator. This may increase his expected discounted utility to an extent which exceeds the administration cost c_a (and potential nonpecuniary costs from temporarily losing control over the firm).

3.6 Concluding Remarks

We have studied the investment behaviour of an intertemporally optimising entrepreneur whose assets consist of a reputation (expected productivity) and of internal finance. It is these values and the age of his firm that decide to what extent he is able to raise external funds.

We have shown that, due to the nonobservability of his investment choice the borrowing capacity falls short of the socially efficient level. As a consequence of these agency costs in the relationship between the entrepreneur and external investors, the exit probability is inefficiently high, in particular for young firms.

Moreover, there is not just a negative response of external investors to a low reputation or little internal funds *per se*, but a firm with such characteristics also has poor incentives to invest (or to exert high effort). This effect, which is due to the presence of an outside option, is foreseen by external investors, thereby aggravating the firm's situation even more.

Giving such firms the possibility to obtain monitored funds may improve their incentives and reduce the number of socially undesirable liquidations. Thus, one tentative policy conclusion of the chapter is that there exists a potential for enhancement of capital market performance by banks in general, and by specialised providers of venture capital in particular. Furthermore, the possibility of a firm in severe distress to undergo restructuring first rather than going into receivership immediately could prevent some undesirable (inefficient) liquidations.

A possible extension of the framework used here includes a stationary version of the model, similar to that in Holmström (1982), where the true value of the reputation parameter itself follows a random walk. In such an environment learning has to start over and over again, such that reputation building remains equally worthwhile throughout time. The erosion of incentives to invest would then solely be due to the presence of an outside option. A further modification would involve allowing for changes of the reputation parameter in any period with a certain (small) probability only as, e.g., in Kiyotaki (1990). This may give rise to other interesting dynamic patterns of the incentive structure.

Finally, explicitly considering financial intermediaries or other capital market institutions, may be a fruitful way to go.

Chapter 4

Financing Constraints and Inventories

4.1 Introduction

A widely reported fact is that the variance of production exceeds the variance of sales.[1] This contradicts the standard linear-quadratic model of inventory investment (e.g. Holt, Modigliani, Muth, and Simon (1960)), which predicts that in order to minimise costs firms will smooth production over time using inventories as a buffer against demand shocks. For this reason the fact is often referred to as the 'excess variance of production'.

In this chapter we build a model of inventory investment and impose constraints on the firm's access to external sources of finance. It is found that the presence of these financing constraints can explain the excess variance of production in a model which otherwise would not deliver this result. In addition, the model with financing constraints predicts that inventory investment and sales covary positively. Moreover,

[1] This fact has been found using data of all levels of aggregation. Examples of studies which report this finding in U.S. data are Blinder and Maccini (1991), Blanchard (1983), and Blinder (1981). For U.K. data see Guariglia and Schiantarelli (1995). For evidence to the contrary see Fair (1989) and Krane and Braun (1991).

even though the stochastic process governing the demand for the firm's output is specified to be independently and identically distributed (i.i.d.), the endogenous sales process exhibits positive serial correlation.[2]

Evidence that financial factors influence inventory investment has been documented by a number of recent empirical studies. This chapter provides a theoretical link between the evidence presented in these studies and the fact that production varies more than sales. Gertler and Gilchrist (1994) find evidence that small manufacturing firms draw down their inventory stocks heavily following a monetary contraction, whereas large firms appear to borrow in order to smooth the impact of a downturn on their inventory behaviour. Using U.S. panel data, Carpenter, Fazzari, and Petersen (1994) find that the inventory investment of small firms is more sensitive to cash flow than the inventory investment of large firms. Kashyap, Stein, and Wilcox (1993) include the ratio of bank loans to commercial paper in several structural models of inventory investment and find the it has a significant effect. In an examination of the 1982 U.S. recession, Kashyap, Lamont, and Stein (1994) find that the ratio of liquid to total assets is significant in explaining the inventory investment of firms without bond ratings, but is not significant for those firms with bond ratings. Taking different structural models to UK panel data, Guariglia (1996) and Guariglia and Schiantarelli (1995) find that financial factors have an important effect on the inventory behaviour of only those firms which may be in financial distress as indicated by a low coverage ratio (the ratio of cash flow to total interest expense). The latter study presents evidence that firms with high coverage ratios are more likely to smooth production.

From a macroeconomic perspective, inventories are a crucial component of fluctuations in aggregate output. For example, Blanchard and Fischer (1989) report that while the stock of inventories makes up only 1% of US GNP, declines in the stock account for 50% of the drop in output in recessions.[3] Clearly any model of fluc-

[2] In models where there is the possibility that the firm may stock out in a given period sales and demand are generally not identical. Such models are referred to as stockout-avoidance models (e.g. Abel (1985), Kahn (1987)).

[3] Similarly, for the UK Sensier (1996) reports figures of 3% and 30%, respectively.

tuations in aggregate output must be able to explain the behaviour of firm level inventory investment.

One of the shortcomings shared by many theoretical models of aggregate fluctuations is the weakness of their propagation mechanisms.[4] To strengthen this mechanism many recent papers have considered the effects of capital market imperfections.[5] In the light of this development of the theoretical literature, it seems natural to ask the following question. Would imposing financing constraints on a partial equilibrium model of firm inventory investment explain the excess variance of production? This chapter provides an answer to that question.

Existing attempts to explain the excess variance of production can be put into three classes: (i) the linear-quadratic model modified with the introduction of either non-convex costs or cost shocks (e.g. Blinder (1986), Eichenbaum (1989), Ramey (1991), Hall (1996), Bresnahan and Ramey (1994)), (ii) the stockout-avoidance model with a demand process that exhibits serial autocorrelation (e.g. Kahn (1987)), and (iii) models of the (s,S) type (e.g. Blinder (1981), Caplin (1985)). Although much of the empirical research has been directed towards the first class of models, satisfactory evidence in support of these models is scarce. For example, with the exception of Ramey (1991), most authors have estimated marginal costs to be upward sloping (e.g. Blanchard (1983), Eichenbaum (1989), West (1986)) suggesting that increasing returns are not a source of nonconvex costs.[6] Similarly, little evidence has been found that shocks to observable costs have a significant effect on inventory investment (e.g. Blinder and Maccini (1991), Miron and Zeldes (1988)).[7]

---

[4]This is particularly true of Real-Business-Cycle models. See Cogley and Nason (1995).

[5]See, for example, Kiyotaki and Moore (1995), Greenwald and Stiglitz (1993), Gertler (1992), Calomiris and Hubbard (1990), Bernanke and Gertler (1989).

[6]Moreover, Bils and Kahn (1996) show that if marginal costs are decreasing, firm behaviour which minimises quadratic costs produces the counterfactual prediction that the ratio of inventories to sales is procyclical. More promising models which incorporate non-convex costs while retaining increasing marginal costs have only been tested with data from the automobile industry (e.g. Hall (1996), Bresnahan and Ramey (1994)).

[7]However, Eichenbaum (1989) found no evidence against the version of the linear-quadratic model in which *unobservable* cost shocks are incorporated.

In contrast, much less testing has been performed on models in the other two classes.[8] However, the model developed here delivers testable predictions which distinguish it from all three categories above. In the second class of model, for instance, sales which do not contain information about expected future sales do not affect production. This is not true in our model, since such sales change the amount of internal funds available to finance production. Similarly, the (s,S) model predicts that the covariance between inventory investment and sales is zero, whereas the model here predicts that this covariance should be positive.

The model is presented in the next section of the chapter. To illustrate the intuition behind our results, a small example and its solution is described in the third section. Then, in section 4.4, we discuss properties of the value and the policy function in the general model. Section 4.5 gives sufficient conditions for the production-variance result to hold in the general model discusses further predictions of the model, before a number of conclusions are drawn and suggestions for future research are given in section 4.6.

4.2 The Model

Consider a firm which produces an (imperfectly) storable good at a constant unit cost (which we normalise to 1 without loss of generality). It attempts to sell its output each period at price p_t. We assume that $p_t = p > 1\ \forall t$. The firm enters period $t \geq 1$ with a stock of goods G_t (those goods not sold in the previous period) and a stock of a liquid asset, M_t, which for convenience we will call money here. The timeline of events in each period is shown in Figure 4.1.

At the beginning of each period the stock of goods depreciates according to a depreciation technology represented by a nondecreasing and convex function $\delta : \mathbb{R}_+ \to \mathbb{R}_+$, $G \mapsto \delta(G)$. We impose $\delta(0) = 0$ and $\delta(G) < G$, i.e. the firm never loses all unsold

[8] Kahn (1992) tests the stockout-avoidance model, and Mosser (1990) tests the (s,S) model.

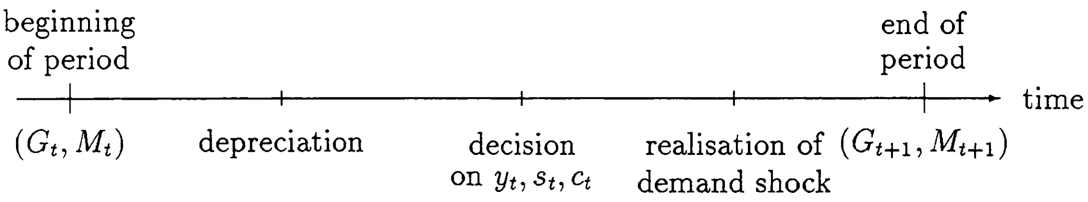

Figure 4.1: Timing of events

goods through depreciation. Thus, after depreciation in period t the firm is left with $G_t - \delta(G_t) > 0$ goods.[9]

The firm then makes its *gross* production decision, y_t, its savings decision, s_t, and its consumption decision, c_t subject to the following financing constraints,

$$c_t \geq 0 \tag{4.1}$$

$$s_t \geq 0 \tag{4.2}$$

$$y_t \geq 0 \tag{4.3}$$

$$c_t + s_t + y_t \leq M_t \tag{4.4}$$

The non-negativity constraint on consumption can be interpreted as preventing the firm from raising equity capital from shareholders by issuing negative dividends.[10] The non-negativity constraint on savings is simply a borrowing constraint. The non-negativity constraint on gross production implies that the firm cannot 'reverse-engineer' and thereby consume (or save) out of its beginning-of-period stock of goods. The final inequality is the budget constraint.

Once production has taken place the demand realisation, z_{t+1}, occurs. As usual in stockout-avoidance models we impose the following non-negativity constraint on the

[9] An equivalent assumption would be to require the firm to pay a pecuniary storage cost $\delta(G)$, financed by its money holdings. One then has to allow the firm to be able to 'reverse-engineer' finished goods (generating one unit of money per unit) in order to pay the storage cost in those periods when the stock of money is insufficient to cover them.

[10] Alternatively, it could be interpreted as restricting the firm's access to trade credit.

stock of goods (i.e. the firm is not allowed to sell short output):

$$x_{t+1} = \min[z_{t+1}, n_t], \tag{4.5}$$

where n_t is the total amount of goods the firm makes available for sale,

$$n_t = G_t - \delta(G_t) + y_t. \tag{4.6}$$

The demand realisation z_{t+1} is identically and independently distributed in each period t on a subset $\mathcal{Z} \subseteq I\!R_+$, according to the probability density function $\phi : \mathcal{Z} \to I\!R$. The associated cumulative distribution function is denoted by $\Phi : \mathcal{Z} \to [0,1]$. For convenience we define $\underline{z} = \inf \mathcal{Z} \geq 0$, and $\bar{z} = \sup \mathcal{Z}$ (which may be negative infinity or infinity, respectively).

The stock of money next period is the sum of this periods savings, s_t, and the revenue generated from sales. Thus, the law of motion for the money stock is[11]

$$M_{t+1} = s_t + p x_{t+1}, \tag{4.7}$$

The law of motion for the stock of goods is

$$G_{t+1} = G_t - \delta(G_t) + y_t - x_{t+1}, \tag{4.8}$$

which can be rewritten as $G_{t+1} = n_t - x_{t+1}$.

The objective of the firm is to choose gross production and savings maximise the present discounted value of consumption, subject to the above non-negativity constraints and laws of motion. More formally, the firm's problem can be written as

$$\max E_0 \sum_{t=0}^{\infty} \beta^t c_t \tag{4.9}$$

[11]For simplicity we have implicitly assumed the gross interest rate on money to be equal to 1. The results are not affected by a gross interest rate R different from unity, as long as $\beta R < 1$.

subject to (4.1), (4.2), (4.3), (4.4), (4.5), (4.7), (4.8), with the discount factor $\beta \in (0,1)$.

Since production is not directly observable in most firm data, it is defined empirically to be the change in inventories stocks plus sales.[12] Using the law of motion for the stock of goods, (4.8), observable production, q_t, is defined to be

$$q_t = G_{t+1} - G_t + x_{t+1} = y_t - \delta(G_t) \tag{4.10}$$

Thus, observable production is gross production, y_t, net of depreciation. Our primary interest is to compare the distribution of observable production q_t with that of sales x_t.

4.3 An Example

The main features of the general model can be illustrated with a simple example in which all variables are restricted to be integers. In any given period the number of goods demanded is either 0, 1, or 2, with probabilities ϕ_0, ϕ_1, and ϕ_2, respectively, such that $\phi_0 + \phi_1 + \phi_2 = 1$. The firm can store only one unit of inventory between periods. If the firm has unsold inventories in excess of this storage capacity, they depreciate completely (i.e. the depreciation function is parametrised as $\delta(G) = \max\{0, G-1\}$).

We first postulate and then, for a set of given parameter values, demonstrate the optimality of the following policy. The firm provides two units of goods for sale whenever this is feasible, and one otherwise. If after setting n to 2 there is money left over, first one unit is saved, ie. $s = 1$. If there are still remaining funds they are spent on consumption.

---

[12]See for example the discussion in Blinder and Maccini (1991), p.77.

(G, M)	x	q	n	s	c	u	x'	(G', M')
							0	(1,0)
(1,0)	0	0	1	0	0	0	1	(0,2)
							1	(0,2)
							0	(1,0)
(2,0)	0	−1	1	0	0	0	1	(0,2)
							1	(0,2)
							0	(2,0)
(2,1)	0	0	2	0	0	0	1	(1,2)
							2	(0,4)
							0	(2,0)
(0,2)	1	2	2	0	0	0	1	(1,2)
							2	(0,4)
							0	(2,1)
(1,2)	1	1	2	1	0	0	1	(1,3)
							2	(0,5)
							0	(2,1)
(1,3)	1	1	2	1	1	1	1	(1,3)
							2	(0,5)
							0	(2,1)
(0,4)	2	2	2	1	1	1	1	(1,3)
							2	(0,5)
							0	(2,1)
(0,5)	2	2	2	1	2	2	1	(1,3)
							2	(0,5)

Table 4.1: The postulated policy and its implications

Table 4.1 summarises this information. In a stationary equilibrium the firm can assume eight possible states, (G, M), which are listed in the first column. The second column is the (unique) sales realisation, x, that has brought the firm into that state. Columns four to six list the policies we have postulated for each state. For each state an action is a triplet (n, s, c). For example, in state $(1, 3)$ the firm puts up two goods for sale, saves one unit of money, and uses one unit for consumption (the action is (2,1,1)). Column three lists the net production decisions, q, implied by these actions. Column seven shows the flow utility enjoyed by the firm in that state. Finally, the last two columns show the possible sales realisations, x', and next period's state, (G', M'), associated with each of the three realisations.

Our goal is to solve for the distributions of sales, x, and net production, q. To do so on has to derive conditions under which the policy we have postulated is optimal. Then we choose a parameter set $\Gamma = \{p, \beta, \phi_0, \phi_1, \phi_2\}$ for which these conditions are satisfied. Finally, given this parameter vector one can solve for the stationary joint distribution of the state variables, and compute distributions for x and q from there.

4.3.1 Optimality Conditions

To ensure that it is optimal to provide $n = 2$ in the first best case when the financing constraint (4.1) is not imposed the chosen vector of parameter values must satisfy the following conditions:

$$
\begin{aligned}
2\phi_2 + \phi_1 - \phi_0 &\geq \beta^{-1}, \\
\phi_2 - (\phi_1 + \phi_0) &\leq \beta^{-1}.
\end{aligned}
$$

The first inequality ensures that it is optimal to provide two units for sale at the margin. The second ensures that it is never optimal to provide three units.

To analyse the optimality conditions when the financing constraint is imposed first compare $v(1, 0)$ with $v(2, 0)$. In both states flow utility $u = 0$, and the expected discounted value of next period is $\beta(\phi_1 + \phi_2)v(0,2)$. Thus $v(1,0) = v(2,0)$, and we call the set $\{(1,0),(2,0)\}$ an equivalence class of states. For simplicity we refer to the value of this equivalence class as v_1. Using the same reasoning we get the four equivalence classes

$$\{(0,2),(2,1)\}, \quad \{(1,2)\}, \quad \{(1,3),(0,4)\}, \quad \{(0,5)\}.$$

We refer to the values of these equivalence classes as v_2, v_3, v_4, and v_5, respectively. Note that $v_5 - v_4 = v_4 - v_3 = 1$. In other words, the value function is linear over those states where consumption is positive.

Summing across states there are 40 feasible actions.[13] First we compute the difference in value between the equivalence classes of states. The difference between the first two classes is given by

$$v_2 - v_1 = \beta[(1-\phi_0)(v_3 - v_2) + \phi_2], \tag{4.11}$$

A demand of 1 or 2, which happens with probability $(1-\phi_0)$, brings a firm from equivalence class v_2 to class v_3, but a firm from v_1 only into v_2. Moreover, the former enjoys one unit of consumption if demand is 2 (probability ϕ_2).

For the difference between the second and the third class we have

$$v_3 - v_2 = \beta[\phi_0(v_2 - v_1) + (1-\phi_0)]. \tag{4.12}$$

A demand of 1 or 2, which happens with probability $(1-\phi_0)$, yields a firm from equivalence class v_3 one consumption unit more than a firm from v_2. Moreover, with zero demand (probability ϕ_2), the former ends up in v_2 whereas the latter is thrown back to v_1.

[13]Due to the non-negativity restriction on gross production, y, (1,0,0) is the only feasible action for states (1,0) and (2,0).

Thus, we obtain a system of two linear equations in the two unknowns $v_2 - v_1$ and $v_3 - v_2$, the solution to which is given by

$$v_2 - v_1 = \frac{\beta^2(1-\phi_0)^2 + \beta\phi_2}{1 - \beta^2\phi_0(1-\phi_0)}, \tag{4.13}$$

and

$$v_3 - v_2 = \frac{\beta^2\phi_0\phi_2 + \beta(1-\phi_0)}{1 - \beta^2\phi_0(1-\phi_0)}. \tag{4.14}$$

Let us assume parameter values $\{p, \beta, \phi_0, \phi_1, \phi_2\} = \{2, 0.96, 0.15, 0.25, 0.6\}$. Substituting them into (4.13) and (4.14) yields

$$v_2 - v_1 = 1.4072$$

and

$$v_3 - v_2 = 1.0266,$$

respectively. Stepping into the next higher equivalence class, v_4, simply means an additional unit of consumption, and therefore $v_4 - v_3 = 1 < 1.0266 < 1.4072$. The same is true for the difference between v_5 and v_4. Thus, the value increments are strictly decreasing over the first four equivalence classes and remain constant thereafter. This illustrate the strict concavity of the value function in the range of assets for which the constraint $c \geq 0$ is binding and its linearity beyond that region.

For the same reasons as in the unconstrained model, it is never optimal to choose $n > 2$ or $n = 0$. However, due to the additional constraint it is no longer preferable to always set $s = 0$.

That it is still optimal to provide $n = 2$ whenever possible is shown in the following arguments. Consuming the resource unit freed up by providing only $n = 1$ instead is not optimal if and only if

$$1 < \beta[(1-\phi_0)(v_3 - v_2) + \phi_2]. \tag{4.15}$$

Since $1 < 1.4137$, this is clearly the case with our parameter values.

Due to the decreasing increments in value established above the *change* in expected valuation from saving it (rather than investing it in production) is bounded by $\beta[\phi_0(v_2 - v_1) + \phi_1(v_3 - v_3) + \phi_2(v_3 - v_4)]$. The second term is, of course, zero, and stated here only for expositional reasons. If the demand shock is equal to 1 the firm will end up in state $(1,2)$ regardless of its decisions about saving and investment.

Since $v_3 - v_4 = -1$, this expression is negative if and only if

$$\phi_2 > \phi_0(v_2 - v_1). \tag{4.16}$$

For our parameter values this translates to $0.6 > 0.2111$. Thus, it is optimal to set $n^*(N,S) = 2 \ \forall \ (G,M)$ such that this is feasible. It can be shown that, as a consequence, 29 out of the 40 feasible actions can be excluded from the optimal set on those grounds.

Next consider states $(0,4)$ and $(1,3)$, where $n = 2$ is clearly feasible and will therefore be chosen. Instead of setting $(n,s,c) = (2,1,1)$ the firm's decision could choose $(n,s,c) = (2,2,0)$. This is dominated by the former choice if and only if

$$\beta[\phi_0(v_3 - v_2) + \phi_1 + \phi_2] < 1,$$

or

$$\beta\phi_0(v_3 - v_2) < 1 - \beta(1 - \phi_0). \tag{4.17}$$

which is equivalent to $0.9638 < 1$, given Γ.

For $(2,0,2)$ not to be better than $(2,1,1)$,

$$\beta[\phi_0(v_2 - v_1) + \phi_1 + \phi_2] > 1,$$

or

$$\beta\phi_0(v_2 - v_1) > 1 - \beta(1 - \phi_0). \tag{4.18}$$

must hold $(1.0186 > 1)$.

Inequality (4.17) also ensures that in state $(0,5)$ the choice $(2,1,2)$ is preferred to $(2,2,1)$; and since it implies that

$$\beta[\phi_0(1+v_3-v_2)+2\phi_1+2\phi_2] < 2,$$

$(2,1,2)$ will not be dominated by $(2,3,0)$ either. (The additional unit saved has expected marginal valuation of $\beta < 1$ as it is used for consumption with certainty next period.)

Inequality (4.18) is required for $(2,0,3)$ not to be preferred to $(2,1,2)$. Thus, we have disposed of another 5 actions.

In $(G,M) = (1,2)$, again only choices s.t $n = 2$ need to be considered. The choice of $(n,s,c) = (2,1,0)$ over $(n,s,c) = (2,0,1)$ requires (4.18) to hold, which we know to be true. Thus, the latter can be eliminated.

That the postulated action $(2,0,0)$ is optimal in the states $(0,2)$ and $(2,1)$ is true because all other possibilities involve $n < 2$ (ie. 6 actions excluded).

In the remaining two states, $(2,0)$ and $(1,0)$, $(1,0,0)$ is the only feasible action and therefore trivially optimal.

In total we have excluded $29+5+1 = 35$ suboptimal actions, leaving us with 5 optimal ones across all states.

The four independent optimality conditions are summarised in Table 4.2.

Table 4.2 should be read as follows. For example, in states (0,4) and (1,3) the action (2,1,1) is preferred to the action (2,2,0) if the condition $[1-\beta(1-\phi_0)] > \beta\phi_0(v_3-v_2)$ holds (i.e. consuming the marginal unit of money yields greater value than saving it). This condition is derived by simply comparing the values of each state under the alternative actions.

states	action preference	condition
$(0,4)$,$(1,3)$	$(2,1,1) \succ (2,2,0)$	$1-\beta(1-\phi_0) > \beta\phi_0(v_3 - v_2)$
$(1,2)$	$(2,1,0) \succ (1,2,0)$	$1-\beta(1-\phi_0) < \beta\phi_0(v_2 - v_1)$
$(0,2)$,$(2,1)$	$(2,0,0) \succ (1,1,0)$	$\phi_2 > \phi_0(v_2 - v_1)$
$(0,2)$,$(2,1)$	$(2,0,0) \succ (1,0,1)$	$1 < \beta(1-\phi_0)(v_3 - v_2)$

Table 4.2: Independent optimality conditions

4.3.2 Stationary distributions

Define f_s, $s \in S := \{(1,0),(2,0),(2,1),(0,2),(1,2),(1,3),(0,4),(0,5)\}$, $\sum_{s\in S} f_s = 1$, as the probability with which the firm finds itself in state s at any point in time.

To obtain stationarity one has to ensure that, given the optimal policy, in each period the marginal probability of leaving any state matches the marginal probability of coming into it. In other words, the following system of equations has to be satisfied.

$$
\begin{aligned}
\phi_0 f_{20} &= (1-\phi_0) f_{10}, \\
\phi_0(f_{02} + f_{21} &= f_{20}, \\
\phi_0(f_{12} + f_{13} + f_{04} + f_{05}) &= f_{21}, \\
(1-\phi_0)(f_{10} + f_{20}) &= f_{02}, \\
\phi_1(f_{21} + f_{02} &= f_{12}, \\
\phi_1(f_{12} + f_{04} + f_{05} &= (\phi_0 + \phi_2) f_{13}, \\
\phi_2(f_{21} + f_{02}) &= f_{04}, \\
\phi_2(f_{12} + f_{13} + f_{04}) &= (1-\phi_2) f_{05}.
\end{aligned}
$$

f_{10}	f_{20}	f_{21}	f_{02}	f_{12}	f_{13}	f_{04}	f_{05}
0.0039	0.0219	0.1242	0.0219	0.0365	0.2070	0.0877	0.4969

Table 4.3: Stationary distribution

Since these equations are not linearly independent, this system of equations needs to be 'pinned down' by the adding up constraint

$$\sum_{s \in S} f_s = 1.$$

The solution to the system is given by

$$\begin{aligned}
f_{20} &= \left[\frac{2-\phi_0}{1-\phi_0} + \frac{1-\phi_0}{\phi_0}\left(1+\frac{1}{\phi_0}\right)\right]^{-1}, \\
f_{10} &= \frac{\phi_0}{1-\phi_0} f_{20}, \\
f_{21} &= \frac{1-\phi_0}{\phi_0} f_{20}, \\
f_{02} &= f_{20}, \\
f_{12} &= \frac{\phi_1}{\phi_0} f_{20}, \\
f_{13} &= \frac{\phi_1(1-\phi_0)}{\phi_0^2}, \\
f_{04} &= \frac{\phi_1}{\phi_0} f_{20}, \\
f_{05} &= \frac{\phi_2(1-\phi_0)}{\phi_0^2} f_{20}.
\end{aligned}$$

We assume that the set of parameters Γ takes on the values $\{p, \beta, \phi_0, \phi_1, \phi_2\} = \{2, 0.96, 0.15, 0.25, 0.6\}$.

The resulting stationary joint distribution for the state variables are presented in Table 4.3, and the implied distributions of (net) production and sales are given in Table 4.4.

$y=-1$	$y=0$	$y=1$	$y=2$
0.0219	0.1281	0.2436	0.6064

$x=0$	$x=1$	$x=2$
0.1500	0.2655	0.5845

Table 4.4: Distributions of net production and sales

The mean of production and sales is $\bar{q} = \bar{x} = 1.4345$, and the variances are $var(q) = 0.6334 > 0.5447 = var(x)$. Thus, in this example the firm exhibits production counter-smoothing.

4.3.3 Financing Constraints and the Excess Variance of Production

It is the presence of the financing constraints in this model which delivers the excess variance of production result. If capital markets were perfect, then the firm's optimal policy for production would be such that $var(q) = var(x)$. Each period the firm would simply replace what had be sold and what had depreciated. This implies that net production, q, is set equal to sales, x, in each period. This implies that the number of goods put up for sale each period is constant.

When the firm is financially constrained such a policy is not possible. After particularly low sales realisations depreciation is so high that the firm does not have sufficient funds available to replace what had been sold and what had depreciated. In these circumstances net production, q, will be less than sales, x, and the amount of goods put up for sale next period drops below the unconstrained amount. We say that in this case the firm 'underproduces'. Now suppose that after a low sale the firm experiences a relatively high sale. In this case it will have more than a sufficient amount of cash to replace all that had been sold and depreciated. It will now be in a position to rebuild its inventory back up to its unconstrained level or at least close to it. Under these circumstances the firm will 'overproduce'. What drives the variance of the firm's production above that of sales is the association of underproduction with low sales realisations and overproduction with medium sales realisations.

Table 4.1 reveals that there is a discrepancy between x and q only for states $(2,0)$ and $(0,2)$. In $(2,0)$, where the sales realisation $(x=0)$ is associated with production $(y=-1)$, underproduction of one unit occurs. In this state the firm would like to make two units available for sale, but is prevented from doing so by the financing constraint. This, in turn, causes the firm to overproduce by one unit in state $(0,2)$. Figure 4.2 helps to illustrate the intuition behind this result. In effect, the financing constraints cause a low sales realisation to be mapped into an even lower production realisation, and a medium sales realisation to be mapped into a high production realisation. If we compare this to the distribution of sales and net production when the nonnegativity constraint on consumption is absent, as depicted in Figure 4.3, we see this effect clearly.

4.3.4 The Nature of Financing Constraints

To what extent do the results in this paper depend on the type of financing constraints used? Here we offer only informal arguments that our results will not alter under different assumptions about the financing constraints. The basis for this is that regardless of their nature , financing constraints will be binding primarily after low sales realisations. Thus, underproduction will still be associated only with low sales realisations, which is the central feature of the excess-variance-of-production result.

More specifically, consider two different ways to model the financing constraints. For one thing, we may assume that there is a perfectly elastic supply of external finance, but that it is more costly than internal finance. For the other alternative we may think of the firm being able to enter information-constrained insurance contracts.

The first case, where the firm faces a hierarchy of finance, is closer to our model. In fact, the financing constraints in our model could be interpreted as representing a finance hierarchy in which the premium on external funds is so high that it is never optimal for the firm to use external finance. Suppose instead that the premium

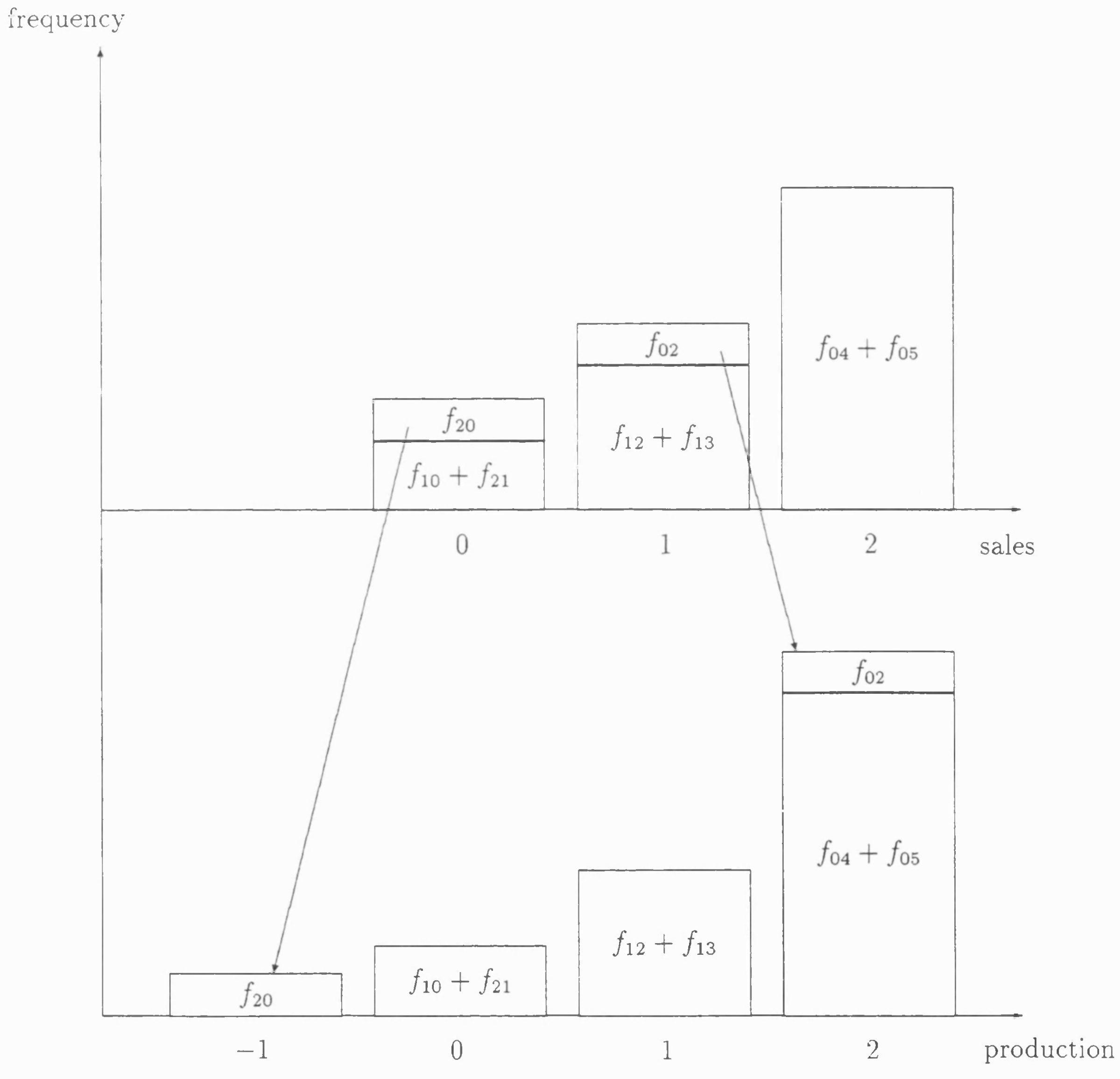

Figure 4.2: Distributions of sales and net production under financing constraints

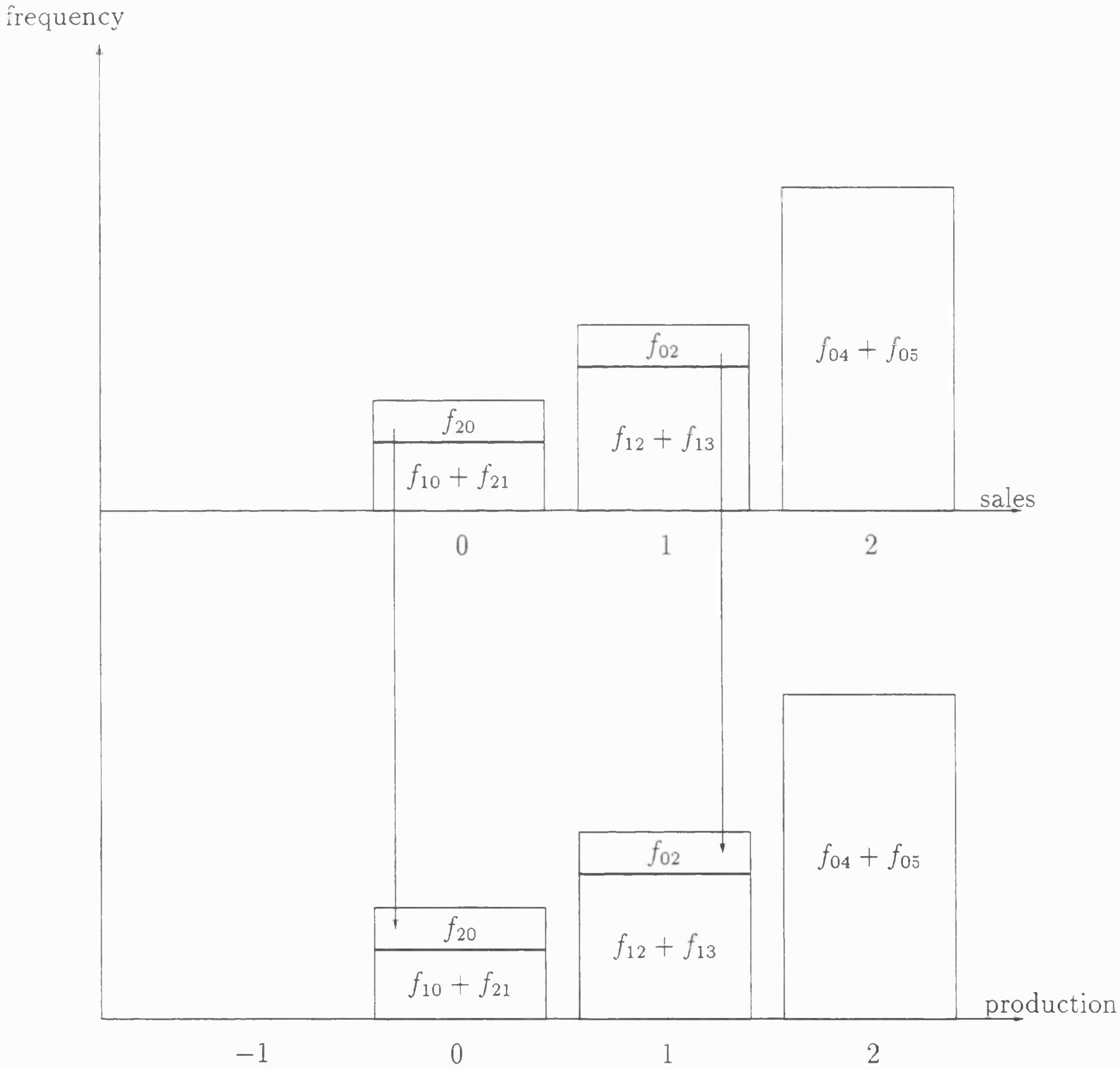

Figure 4.3: Distribution of sales and net production without the financing constraint

were low enough to make the use of external finance attractive in some situations. In this case the firm can find itself in three qualitatively different regions. In the first region, internal funds will be so low that the benefit of the marginal good put up for sale is high enough to warrant the use of external finance. In the second region the firm is still constrained, but the value of the marginal good put up for sale is not high enough to justify the use of external funds. Here the firm behaves exactly as it would in our model. In the third region the firm is unconstrained. Note that, because external finance is more costly, the optimal amount of goods put up for sale will be lower when the marginal unit is financed externally than when the marginal unit is financed with internal funds. Thus, if the firm begins with the unconstrained amount of goods for sale and has a low sales realisation it will underproduce. In other words, underproduction will still be associated with low sales realisations.

For the second case, which is somewhat further from our model, our argument is only suggestive. Suppose that lenders can observe inventories at only two points in time: after sales *and* depreciation, and just before sales. In other words, lenders can observe the firm's gross production decision, y_t. However, lenders cannot observe sales and depreciation.[14] In this setup there is an incentive for the firm to report a bad sales realisation when in fact there had been a good sales realisation. Typically, the optimal contract in such a setup would place restrictions on the observable decision variable in order to ensure that the borrower truthfully reports good outcomes. This usually implies that in bad outcomes the level of the decision variable is lower than it would be if all variables were observable, in order to introduce a cost to misreporting a good outcome as a bad outcome. We know that if all variables are observable, then the optimal policy of the firm is to set gross production such that it equals the sale plus depreciation. In the information-constrained framework, an optimal contract would force the firm to set gross production lower than this for low sales realisations. Thus, this informal argument suggests that the association

[14]This argument is consistent with depreciation taking the form of a fixed storage capacity as in the example.

of low sales with underproduction would remain, and that the excess-variance-of-production result is preserved.

4.4 The General Model

We can specify either gross production, y_t, or, equivalently, the amount of goods made available for sale, n_t, as a choice variable since $n_t = G_t - \delta(G_t) + y_t$. Therefore, taking n_t and s_t as the choice variables, the Bellman equation for the problem outlined in section 4.2 is

$$\begin{aligned} v(G_t, M_t) &= \max_{n_t, s_t} \Big\{ M_t - s_t - (n_t - G_t + \delta(G_t)) + \beta \int_{\underline{z}}^{n_t} v(G_{t+1}, M_{t+1})\, \phi(dz_{t+1}) \\ &\quad + \beta \int_{n_t}^{\bar{z}} v(0, M_{t+1})\, \phi(dz_{t+1}) \Big\} \end{aligned} \tag{4.19}$$

subject to (4.1), (4.2), (4.3),(4.7), and (4.8).

4.4.1 The Unconstrained Problem

Consider removing the non-negativity constraint on consumption (4.1) from the problem above.[15] In this case the following policy would be optimal. Optimal savings $s^\star(G, M) = 0$ in each period since $\beta R < 1$. The optimal choice of n_t will always be an interior solution so that we can differentiate the Bellman equation with respect to n_t. Assuming that δ is differentiable (denoting with δ' its first derivative)

[15] We refer to this problem as the unconstrained problem, and to the non-negativity constraint on consumption as the financing constraint. Relaxing the latter allows the firm to perfectly insure its desired production expenditure against negative demand shocks through negative consumption in those periods where it does not have sufficient cash on hand. It is precisely the availability of this insurance we wish to remove with the constraint on consumption. If, instead, we had relaxed the non-negativity constraint on savings (4.2), the firm's optimal policy would have been simply to borrow as much as possible in the first period, due to the linear utility function and $\beta R < 1$.

and using the envelope conditions, $v_{M_t} = 1$ and $v_{G_t} = (1 - \delta'(G_t))$, we obtain

$$\int_{\underline{z}}^{n_t} (1 - \delta'(G_{t+1}))\, \phi(dz_{t+1}) + \int_{n_t}^{\infty} p\, \phi(dz_{t+1}) = \beta^{-1} \tag{4.20}$$

as the Euler equation. Since equation (4.20) involves neither M_t nor G_t the optimal policy for the firm is to set $n^*(G, M) = N$, a constant. This implies that in each period net production, q_t, is set equal to sales, x_t. Therefore, if this financing constraint is relaxed $var(q) = var(x)$.

4.4.2 The Constrained Problem

When the non-negativity constraint on consumption is imposed we can no longer guarantee the differentiability of the value function since there is no way to guarantee that the optimal policies will be interior solutions. Thus we cannot simply analyse a version of equation (4.20). However, the following properties of the solution can be established, and they are sufficient to prove that the variance of production exceeds the variance of sales when all of the financing constraints are imposed.

Property 1. The Bellman operator defined in (4.19) is a contraction mapping with a unique fixed point.

Due to the non-negativity constraints (4.2) and (4.3), the state space for this problem is bounded below by zero. We assume arbitrary upper bounds to the state space, $\bar{M}$ and $\bar{G}$.[16] With this assumption it is straightforward to show that the conditions for Lemma 9.5 and Theorem 9.6 in Stokey, Lucas, and Prescott (1989) (pp.261-64) are satisfied. Thus, the Bellman operator defined on the right-hand side of (4.19) is a contraction mapping with a unique fixed point.

Property 2. v is increasing in M and G.

Because of the law of motion for money (4.7), the one-period return to the firm,

[16] Below we show that solution to the problem is characterised by upper bounds on n and s, which implies that M and G are bounded, too.

$M - s - (n - G + \delta(G))$, and the feasibility constraints are all increasing in the state variable, M, v is also increasing in M by theorem 9.7 in Stokey, Lucas, and Prescott (1989) (p.264). Similarly, since the law of motion for goods (4.8), and the firm's one-period return are also increasing in G, v is increasing in G as well.

Property 3. v is concave and strictly concave in the region of constrained states.

Proof: That v is concave is standard and stems from the fact that the Bellman functional equation $v = Tv$ is a contraction mapping, with an operator T that preserves concavity. This is because integration in (4.19) preserves concavity and the return function is concave as well (see Stokey, Lucas, and Prescott (1989) [p.265] for details).

To show strict concavity in the constrained region we adopt the method of a concavity proof in Abel (1985).[17]

Consider two firms, A and B in constrained states (G_A, M_A) and (G_B, M_B), respectively, such that $G_B - \delta(G_B) + M_B > G_A - \delta(G_A) + M_A$, with optimal policy choices (n_A, s_A) and (n_B, s_B), respectively. Assume that $n_B > n_A$. A convex combination of the two firms is in state $(\theta G_A + (1-\theta)G_B, \theta M_A + (1-\theta)M_B)$, and a policy choice $(\theta n_A + (1-\theta)n_B, \theta s_A + (1-\theta)s_B)$ is feasible due to the convexity of δ.

If this period's demand shock z is greater than both n_A and n_B, both firms stock out which leads to next period's states $(G'_A, M'_A) = (0, s_A + pn_A)$ and $(G'_B, M'_B) = (0, s_B + pn_B)$, respectively. The same is true, however, for the convex combination, whose new state would then be $(G'_\theta, M'_\theta) = (0, \theta(s_A + pn_A) + (1-\theta)(s_B + pn_A))$. As v is concave the value of the convex-combination firm next (and hence in the current) period, denoted by v_θ, is therefore not smaller than the convex combination of the two firms, given by $\theta v_A + (1-\theta)v_B$. Note that the production-savings choice made by the convex-combination firm is not necessarily optimal.

[17]In that paper production occurs with a lag. Consequently, only the stock of inventories which the firm has at the beginning of the period are available for sale. Moreover, the stock of inventories is the only state variable.

Similarly, if the demand shock z lies below both n_A and n_B, none of the firms, including the convex combination, will stock out. Next period's states are then $(G'_A, M'_A) = (n_A - z, s_A + pz)$ and $(G'_B, M'_B) = (n_B - z, s_B + pz)$ for firms A and B, respectively, and $(G'_\theta, M'_\theta) = (\theta n_A + (1-\theta)n_B - z, \theta s_A + (1-\theta)s_B + pz)$ for the convex-combination firm. Its state is just the convex combination of (G'_A, M'_A) and (G'_A, M'_A). for the convex combination. Because of the concavity of v, $v_\theta \geq \theta v_A + (1-\theta)v_B$ in this case as well.

If the demand shock z lies between n_A and n_B, firm A stocks out but firm B does not. Their states next period are then $(G'_A, M'_A) = (0, s_A + pn_A)$ and $(G'_B, M'_B) = (n_B - z, s_B + pz)$, respectively. If $z < \theta n_A + (1-\theta)n_B$, $(G'_\theta, M'_\theta) = (\theta n_A + (1-\theta)n_B - z, \theta s_A + (1-\theta)s_B + pz)$, and if $z \geq \theta n_A + (1-\theta)n_B$, $(G'_\theta, M'_\theta) = (0, \theta(s_A + pn_A) + (1-\theta)(s_B + pn_A))$. In either case we can imagine that, in a first stage, shares θ and $1-\theta$ of demand z are covered by divisions a and b, respectively, of the convex-combination firm, where division a corresponds a proportion θ and division b a proportion $1-\theta$ of the firm. Its value would then just be the same as the convex combination of the values of firm A and B. However, in a second stage, excess demand from division a can be shifted to division b and (partially) satisfied there. This leads not only to higher overall revenues but also to a reduction in goods depreciation. Hence, with this demand constellation, $v_\theta > \theta v_A + (1-\theta)v_B$. In expected terms, the valuation of the convex-combination firm is therefore strictly higher than the convex combination of the values of A and B. □

In the example of section (4.3) we were able to compute a stationary distribution of states. Showing that an ergodic distribution exists in our general setting is a rather formidable task. We therefore simply assume existence of stationarity in the general model.

The following propositions and corollaries characterise optimal policy functions and are used to derive a number of statements about the distribution of sales and net production in the next section.

PROPOSITION 4.1: *There exist upper bounds to the optimal policies, n and s, referred to as $\bar{n}, \bar{s}$. If $M - \bar{s} > \bar{n} - G + \delta(G)$ the remaining money is used for consumption.*

Proof: Suppose the firm never consumes. Since $v(G, M)$ is then strictly concave in both its arguments , the marginal returns to savings and to production will be decreasing in both state variables. The value of an infinitesimal unit of money would tend to β as $M \to \infty$, since the probability of encountering a sequence of sales realisations in which the non-negativity constraint on consumption will be binding tends to zero (effectively, the firm becomes unconstrained). Since $\beta < 1$, there will be a level of savings, $\bar{s}$, beyond which the firm will prefer consumption to further savings. Similarly, as $G \to \infty$ the marginal value of inventories tends to a value below one, since the probability of selling the marginal unit goes to zero. Thus, there also exists a level of goods the firm puts up for sale, $\bar{n}$, beyond which the firm prefers consumption to further production.[18] □

COROLLARY 4.2: *There exist optimal policy functions, $s^{\star}(G, M)$ and $n^{\star}(G, M)$, which map each element of the state space into the space of feasible actions.*

The one-period return function is concave in the state variables, and the feasibility set for the choice variables is convex. Over the region where consumption is zero v is strictly concave. It follows that, in this region, the maximum in (4.19) is attained by unique choices of s_t and n_t. From Proposition 4.1. it follows that when consumption is positive, $s_t^* = \bar{s}$ and $n_t^* = \bar{n}$. □

COROLLARY 4.3: *The mean of production, $\hat{q}$, equals the mean of sales, $\hat{x}$.*

Proof: Suppose $\hat{q} > \hat{x}$. Then the firm would accumulate inventories indefinitely, contradicting the existence of $\bar{n}$. Conversely, if $\hat{q} < \hat{x}$ were true, then $G \to 0$ as $t \to \infty$, which cannot be optimal since $v(G, M)$ is increasing in G. □

Intuitively speaking, this corollary holds because each unit of output that the firm produces is either sold, or depreciates which is accounted for as negative production.

[18] The existence of $\bar{s}$ and $\bar{n}$ implies the existence of endogenous upper bounds to the state space, $\bar{M}$ and $\bar{G}$

PROPOSITION 4.4: *The optimal policy functions $n^*(G, M)$ and $s^*(G, M)$ are non-decreasing in total funds available to the firm.*

Proof: We will first present the proof for $n^*(G, M)$. The statement clearly holds for states in which the non-negativity constraint on consumption does not bind, as for those states additional money holdings are simply used for consumption, without changing the amount $\bar{n}$ made available for sale. For states in which the non-negativity constraint on consumption is binding we prove the statement by contradiction. Note that in those states the return function is equal to zero, and we can therefore write

$$\begin{aligned} v(G, M) &= \beta \int_{\underline{z}}^{n^*(G,M)} v(n^*(G, M) - z, s^*(G, M) + pz)\phi(z)dz \\ &\quad + \beta[1 - \Phi(n^*(G, M))]v(0, s^*(G, M) + pn^*(G, M)). \end{aligned} \tag{4.21}$$

Assume that in some (G, M) the optimal policy prescribes to choose values (n, s). Also suppose that, given some small $\epsilon > 0$, the optimal action in $(G, M + \epsilon)$ is $(n - \mu, s + \epsilon + \mu)$, for some small $\mu > 0$, which implies

$$\begin{aligned} &\int_{\underline{z}}^{n-\mu} v(n - \mu - z, s + \epsilon + \mu + pz)\phi(z)dz + [1 - \Phi(n - \mu)]v(0, s + \epsilon - (p - 1)\mu + pn) > \\ &\qquad \int_{\underline{z}}^{n} v(n - z, s + \epsilon + pz)\phi(z)dz + [1 - \Phi(n)]v(0, s + \epsilon + pn). \end{aligned} \tag{4.22}$$

But then the firm could do better by shifting funds μ from production to savings in (G, M) as well, i.e.

$$\begin{aligned} &\int_{\underline{z}}^{n-\mu} v(n - \mu - z, s + \mu + pz)\phi(z)dz + [1 - \Phi(n - \mu)]v(0, s + (p - 1)\mu + pn) > \\ &\qquad \int_{\underline{z}}^{n} v(n - z, s + pz)\phi(z)dz + [1 - \Phi(n)]v(0, s + pn). \end{aligned} \tag{4.23}$$

Hence, (n, s) cannot be an optimal choice in state (G, M), which contradicts the initial assumption.

To show that the previous inequality holds, note first that for very small μ, the decrease of the first term of the left-hand side due to the change μ in the (upper) integral boundary is offset by the increase of the second term of the left-hand side due to the same change μ in the (lower) integral boundary.

Since v is strictly concave where the non-negativity constraint on consumption is binding, the stock-out terms (without probabilities) of the two inequalities compare as follows:

$$v(0, s+\epsilon+(p-1)\mu+pn)-v(0, s+\epsilon+pn) \le v(0, s+(p-1)\mu+pn)-v(0, s+pn). \quad (4.24)$$

Moreover, conditional on not stocking out, reducing n in favour of s is at least as valuable at lower savings as it is at higher savings (because of depreciation). Hence, $\forall\ z \le n - \mu$,

$$v(n-\mu-z, s+\epsilon+\mu+pz)-v(n-z, s+\epsilon+pz) \le v(n-\mu-z, s+\mu+pz)-v(n-z, s+pz). \quad (4.25)$$

This is true because, by approximating both sides of the inequality arbitrarily closely, we can rewrite it as

$$-[v(n-z, s+\epsilon+pz)-v(n-\mu-z, s+\epsilon+pz)]\mu+[v(n-z, s+\epsilon+\mu+pz)-v(n-z, s+\epsilon+pz)]\mu$$

$$\le -[v(n-z, s+pz)-v(n-\mu-z, s+pz)]\mu+[v(n-z, s+\mu+pz)-v(n-z, s+pz)]\mu, \quad (4.26)$$

or

$$[v(n-z, s+\epsilon+\mu+pz)-v(n-z, s+\epsilon+pz)]-[v(n-z, s+\mu+pz)-v(n-z, s+pz)]$$

$$\le [v(n-z, s+\epsilon+pz)-v(n-\mu-z, s+\epsilon+pz)]-[v(n-z, s+pz)-v(n-\mu-z, s+pz)]. \quad (4.27)$$

Due to strict concavity of v, both sides of the inequality measure the extent to which the value gain, due to the funds being larger by ϵ, is reduced by an increase in savings and production, respectively. More precisely, the left-hand side of the inequality represents the *decrease in the value gain* (implied by an ϵ-increase) that stems from adding even more funds, μ, to savings. The right-hand side is equal to the *decrease in the value gain* (implied by an ϵ-increase) due to an increase μ of the target inventory. The former decrease is larger in absolute terms since the gross unit return on additional savings μ is 1 $\forall$ z, whereas the gross unit return on an additional goods provision of μ is strictly smaller than 1 for at least some z that do not lead to a stockout. An analogous argument can be made when there is a jump from state (G, M) to a state $(G + \epsilon, M)$, as this would be equivalent to leaving G unchanged and increasing M by some fraction of ϵ (which depends on the depreciation technology).

As for the function $s^*(G, M)$ assume that there is a range of states over which the function is decreasing. The only way s^* could be decreasing in the constrained region is if there was an overproportionate increase of n in response to a small increase of funds available (be it in the form of more money or of higher goods inventories). Due to strict concavity of the value function, however, the negative effect on the expected value of v next period due to a decrease in s would more than offset the positive effect of an increase in n. We conclude that such a policy cannot be optimal. □

COROLLARY 4.5: *Each state (G, M) has a unique sales realisation which moves the firm from other states into this state. In other words, there exists a function, call it $x^*(G, M) : (G, M) \mapsto x$ which maps each element of the state space into the set of feasible sales realisations.*

Proof: Suppose that the sale x' moves the firm to the state (G, M) from the post-production pair (n', s'), and that a different sale x'' moves the firm to the same state $(\tilde{G}, \tilde{M})$ from a different post-production pair (n'', s''). Then

$$G = n' - x' = n'' - x'' \qquad \Rightarrow \qquad x'' - x' = n'' - n'$$

$$M = s' + px' = s'' + px'' \qquad \Rightarrow \qquad x'' - x' = \frac{s' - s''}{p}$$

However, from Proposition 4.2 we know that if $n'' > n'$ then $s'' \geq s'$. Thus the above equations cannot be satisfied simultaneously when $x' \neq x''$. □

The next proposition turns out to be very useful in obtain in the main results of the chapter (section 4.5). The first part states that, in the range of sales that do not lead to stock-out, an increase in sales x ($x < k$ where k is the level of the previous-period target inventory) does not lead to a reduction in the firm's overproduction, defined as $\pi(x, k) := q(x, k) - x$. Put differently, net production q grows by at least as much, in the response to the increase in x, as x itself.

The second part implies that as the firm gets wealthier (in terms of post-depreciation goods and money held at the beginning of a period) the share of additional wealth used for production financing does not increase.

PROPOSITION 4.6: Conditional on not stocking out, i.e. for all values $n > x$, π is nondecreasing in x. Moreover, π is concave in x.

Proof: The first part of the statement is a consequence of Proposition 4.2. Increase the sales of a firm by one unit. The firm will then produce at least one unit more to replace this extra unit sold, because even after having done so it has goods inventories at least as high, and money holdings that are higher (by $p - 1$) than without the sales increase. Therefore $q - x$ is at least as high as before. For the second part of the proposition note that for both assets (goods and money) to be held in positive quantities optimality requires that the marginal valuation of both assets be equalised at each level of wealth (goods after depreciation plus money holdings, both before production).

As the probability of not stocking out increases in n, investing additional wealth in goods inventories becomes *relatively* less attractive, whereas retaining additional wealth as cash becomes *relatively* more attractive. Whilst the marginal valuations on both assets are decreasing in wealth (which is due to to the strict concavity of

v in both state variables), that of goods diminishes at a higher rate than that of money. □

4.5 Main Results

In this section we will present not only the main result that production is more volatile than sales but also a number of other findings. In particular, it will be shown that the sales process exhibits first-order autocorrelation although the demand shocks are i.i.d.

First, we have to make a statement about the expected overproduction as a function of the sales realisation, which will then serve as a sufficient condition for the main proposition.

Let $Q_k = \{(n_{-1}, s_{-1}) \mid n_{-1} = k\}$, and $g(k) = \int_{Q_k} f(n_{-1}, \tilde{s}_{-1}) d\tilde{s}_{-1}$, where f is the joint probability density function of n and s, and recall that $\pi(x, k) = q(x, k) - x$. (The subscript -1 indicates that the previous-period value of a variable is referred to.) Also denote with K_x the union of all Q_k such that $k > x$. Moreover, define as ρ the inverse of the demand distribution's hazard rate $\frac{\phi}{1-\Phi}$.

We can then write the expected difference between production and sales, conditional on selling x, as

$$h(x) = \phi(x) \int_{K_x} g(k)\pi(x, k)dk + [1 - \Phi(x)]g(x)\pi(x, x,).$$

For the proof it is useful to divide this expression by $\frac{\phi(x)}{\pi(x,x)}$, which yields

$$H(x) = \frac{\int_{K_x} g(k)\pi(x, k)dk}{\pi(x, x)} + \rho(x)g(x).$$

It is clear that $H(x)$ always carries the same sign as $h(x)$. We make the following assumptions about the demand distribution ϕ.

Assumption A.1: In the range where x is such that $H(x) \leq 0$ the demand probability density function ϕ exhibits a monotone hazard rate, i.e. ρ is nonincreasing, and it generates a stationary distribution g such that $\frac{1-\rho'}{\rho} \geq \frac{g'}{g} \geq 0$.

Assumption A.2: In the range where x is such that $H(x) \leq 0$, $-\frac{\rho''}{\rho} \geq \frac{g''}{g}$.

LEMMA 4.7: *Under Assumptions A.1 and A.2 there is a unique value of x in the interior of $[\underline{z}, \bar{n}]$ such that $h(x) = 0$.*

Proof: To begin with we note the following three observations: Firstly, note that $h(\underline{z}) < 0$ since at the lower bound of the sales interval net production is always lower than sales. Otherwise the firm would have to lower cash balances to return to or exceed the target inventory before the sale realisation; however, we know from Proposition 4.2 that this is suboptimal. Secondly, $h(\bar{n}) = 0$ because stocking out at the maximum target-inventory level implies an increase in money holdings. Therefore, the firm after production will simply have returned to $\bar{n}$. Thirdly, we can find an $\epsilon > 0$ small enough such that, for a sale $x = \bar{n} - \epsilon$, $h(\bar{n} - \epsilon) > 0$; otherwise $\bar{n}$ would never be reached.

It follows that there is at least one value of x, such that (i) $h(x) = H(x) = 0$, and (ii) both h and H are increasing at that point. Denote the highest such value of x by x_0. We will now show that x_0 is the only point satisfying these two conditions by arguing that H is concave for values of x below x_0, which means that once H is negative it never assumes positive values as we lower x.

To keep notation simple we assume that h (and H) are twice differentiable.[19] By differentiating $H(x)$ twice and rearranging terms we obtain

$$\begin{aligned} H''(x) &= -g'(x) + \frac{\int_{K_x} \pi''(x,k)g(k)dk}{\pi(x,x)} - \frac{\pi''(x,x)\int_{K_x} \pi(x,k)g(k)dk}{\pi(x,x)^2} \\ &\quad - \frac{2\pi'(x,x)\int_{K_x} \pi'(x,k)g(k)dk}{\pi(x,x)^2} + \frac{2\pi'(x,x)^2\int_{K_x} \pi(x,k)g(k)dk}{\pi(x,x)^3} \\ &\quad + \rho''(x)g(x) + 2\rho'(x)g'(x) + \rho(x)g''(x). \end{aligned}$$

[19]Arguments analogous to those here could be made by using discrete increments for the variables as in the proof of Proposition 4.4

Under A.1 the first term is non-positive. Next note that for all $x \leq x_0$

$$\int_{K_x} \pi(x,k)g(k)dk \leq -\rho(x)g(x)\pi(x,x) \leq 0,$$

which, together with Proposition 4.4 means that the second and the third term are non-positive.

At x_0 H has positive slope, which implies

$$\frac{\int_{K_{x_0}} \pi'(x_0,k)g(k)dk}{\pi(x_0,x_0)^2} > \frac{[1-\rho'(x_0)]g(x_0)}{\pi(x_0,x_0)} - \frac{\rho(x_0)g'(x_0)}{\pi(x_0,x_0)} + \frac{\pi'(x_0,x_0)\int_{K_{x_0}} \pi(x_0,k)g(k)dk}{\pi(x_0,x_0)^3}.$$

It follows that under A.1

$$-2\frac{\pi'(x_0,x_0)\int_{K_{x_0}} \pi'(x_0,k)g(k)dk}{\pi(x_0,x_0)^2} + 2\frac{\pi'(x_0,x_0)^2\int_{K_{x_0}} \pi(x_0,k)g(k)dk}{\pi(x_0,x_0)^3} < 0.$$

Finally, due to assumptions A.1 and A.2, the last three terms are non-positive either, such that $H''(x) \leq 0$ for all $x \leq x_0$. □

We are now in a position to state and prove two central results of the chapter, as summarised in Proposition 5.2.

PROPOSITION 4.8: *The variance of production exceeds the variance of sales. The covariance of production and sales also exceeds the variance of sales.*

Proof: We know from the proof of Lemma 4.7 that $h(\underline{z}) < 0$ and $h(\bar{n}) = 0$. Lemma 5.1 itself tells us therefore that there is a value of x, call it $\hat{x}$ such that $h(x) > 0$ $\forall$ $x \in (\hat{x}, \hat{n})$ and $h(x) < 0$ $\forall$ $x \in (\underline{z}, \hat{x})$. As a consequence,

$$\int_{\underline{z}}^{\bar{n}} h(x)xdx > 0. \tag{4.28}$$

This is true because from Corollary 4.3 we know that the means of production and sales are equal, implying

$$\int_{\underline{z}}^{\bar{n}} h(x)dx = 0. \tag{4.29}$$

In conjunction with Lemma 5.1. this implies that in (4.28) the high sales outcomes are weighted with the same absolute amount of positive mass as the low sales outcome are with negative mass, from which observation the inequality follows.

We can express the variance of net production, $var(q)$, as

$$\begin{aligned} var(q) &= \int_{\underline{z}}^{\bar{n}} (h(x) + x - \mu)^2 dx \\ &= \int_{\underline{z}}^{\bar{n}} (x^2 + \mu^2 - 2\mu x)dx + \int_{\underline{z}}^{\bar{n}} h(x)^2\, dx - 2\mu \int_{\underline{z}}^{\bar{n}} h(x)\, dx + 2\int_{\underline{z}}^{\bar{n}} h(x)\, x\, dx \\ &= var(x) + \int_{\underline{z}}^{\bar{n}} h(x)^2\, dx + 2\int_{\underline{z}}^{\bar{n}} h(x)\, x\, dx \end{aligned}$$

Thus, (4.28) is a sufficient condition for $var(q) > var(x)$.

We can write out a similar expression for the covariance of sales and net production, $cov(x, q)$.

$$\begin{aligned} cov(x,q) &= \int_{\underline{z}}^{\bar{n}} (h(x) + x)(x - \mu^2)dx \\ &= \int_{\underline{z}}^{\bar{n}} h(x)x - \mu \int_{\underline{z}}^{\bar{n}} h(x)dx + \int_{\underline{z}}^{\bar{n}} x^2 dx - \mu \int_{\underline{z}}^{\bar{n}} x dx \\ &= \int_{\underline{z}}^{\bar{n}} h(x)x dx + var(x) \end{aligned}$$

Condition (4.28) is both necessary and sufficient for $cov(x, q) > var(x)$ to hold. □

This result has straightforward implications for the covariance of sales and changes. The finding that it is positive is once again conforms to empirical observations.

COROLLARY 4.9: *Changes in the stock of inventory put up for sale, Δn_t, covary positively with sales, x_t.*

Proof: Writing out the expression for $cov(\Delta n, x)$ we obtain

$$\begin{aligned} cov(\Delta n, x) &= cov(n, x) - cov(n_{-1}, x) \\ &= cov(q + G, x) - cov(n_{-1}, x) \\ &= cov(q, x) + cov(n_{-1} - x, x) - cov(n_{-1}, x) \\ &= cov(q, x) - var(x)\,, \end{aligned}$$

which is positive from Proposition 4.8. □

Another important result of the present contribution is that the model generates positive serial correlation in sales despite the fact that the underlying demand process i.i.d.

PROPOSITION 4.10: *Sales exhibit positive first-order autocorrelation.*

Proof: With each pair of current and next-period sales realisations (x, x') we can associate a set $K_{x'}(x)$, which is the union of sets of pairs (n, s) such that $n > x'$ (see also the definition at the beginning of this section). We can then write the expected difference of next period's sale from the unconditional mean, $\hat{x}$, conditional on this period's sale x as

$$m(x) = \phi(x') \int_{K_{x'}(x)} g(k)(x' - \hat{x})dk + [1 - \Phi(x')]g(x')(x' - \hat{x}). \tag{4.30}$$

The covariance of sales with its first lag is then

$$\begin{aligned} cov(x', x) &= \int_{\underline{z}}^{\bar{n}} m(x)(x - \hat{x})dx \\ &= \int_{\underline{z}}^{\bar{n}} [m(x)x - m(x)\hat{x}]dx \\ &= \int_{\underline{z}}^{\bar{n}} m(x)xdx, \end{aligned} \tag{4.31}$$

where the third line follows from the fact that

$$\int_{\underline{z}}^{\bar{n}} m(x)dx = 0, \tag{4.32}$$

which is equivalent to saying that by integrating the conditional mean of next period's sales over all possible sales realisations this period we obtain the unconditional mean of sales $\hat{x}$.

From Proposition 4.4 it follows that an increase in the sales realisation leads to a target inventory which is at least as high as that without the increase. Therefore, the conditional mean $m(x)$ is nondecreasing in x. Moreover, since the firm does not always put up $\bar{n}$ for sale, it must be true that $m(\bar{n}) > 0$. This implies that there is a sale, $\hat{w} \in [\underline{z}, \bar{n}]$ such that for all $x > \hat{w}$ the conditional expectation of next period's differential between sales and the unconditional mean of sales is positive, i.e. $m(\bar{n}) > 0$. Similarly to the proof of Proposition 4.8, positive weight is attached to higher values of x and the same negative weight (in absolute terms) to lower values of x. Hence, (4.31) must be positive. □

4.6 Conclusions and Suggestions for Further Research

This chapter has presented a model of inventory investment in which the firm's access to external sources of finance is constrained by a borrowing constraint and, in particular a nonnegativity constraint on consumption. This model can explain the excess variance of production, and yields the prediction of positive co-variation between inventory investment and sales. Equally important, the endogenous sales process exhibits positive serial correlation even though the underlying (and unobservable) demand process is exogenously specified to be serially uncorrelated.

It was also shown that as a constrained firm's internal wealth increases, it puts more resources in the production of goods, but at the same time also increase its savings (cash holdings). Thus, the two forms of investment can be regarded as complements rather than substitutes. The more a firm offers for sale, the higher the probability that it is left with a certain unsold amount of goods. As after a bad sales realisation the the low revenue and the ensuing depreciation erode the firm's asset position, it may not be able to produce the desired amount of goods in the following period, unless its money holdings are sufficiently high.

One possible extension of the model would involve the relaxation of the constant-price assumption. If the firm has some form of monopoly power it may want to use the price as an additional instrument to respond to unforeseen shocks by influencing expected future demand in certain ways. It would be particularly interesting to examine how a monopolist reacts to a very bad sales shock that leaves him both with very few goods after depreciation and with little money. On one hand the firm then has very few goods to offer which, *ceteris paribus*, makes the probability of stocking out very high, which may induce the firm to set a rather high price. On the other hand, however, the constrained firm may be desperate enough for cash that it will not want to risk a low sales realisation due to a price that is too high.

Another promising avenue for future research is to study a competitive-equilibrium framework with the features presented here. Even though firms would then be price-takers as they are in this chapter, the price would be endogenously determined in equilibrium. The presence of financing constraints may lead to interesting price processes and aggregate output fluctuations.

Bibliography

ABEL, A. B. (1985): "Inventories, Stock-Outs and Production Smoothing," *Review of Economic Studies*, LII, 283–293.

AGHION, P., O. HART, and J. MOORE (1993): "A Proposal for Bankruptcy Reform in the U.K.," *FMG Discussion Paper, London School of Economics*, 167.

ALCHIAN, A. (1977): "Why Money?," *Journal of Money, Credit and Banking*, 9, 133–140.

BANERJEE, A. V., and E. S. MASKIN (1990): "A Semi-Walrasian Theory of Barter and Money: I," mimeo.

BAUMOL, W. (1952): "The Transactions Demand for Cash: An Inventory-Theoretic Approach," *Quarterly Journal of Economics*, 66, 545–556.

BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

BERNANKE, B., and M. GERTLER (1989): "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79, 14–31.

BERNANKE, B., and M. GERTLER (1990): "Financial Fragility and Economic Performance," *Quarterly Journal of Economics*, 55, 87–114.

BILS, M., and J. A. KAHN (1996): "What Inventory Behaviour Tells Us about Business Cycles," *mimeo.*

BLANCHARD, O. J. (1983): "The Production and Inventory Behaviour of the American Automobile Industry," *Journal of Political Economy*, 91, 365–400.

BLANCHARD, O. J., and S. FISCHER (1989): *Lectures on Macroeconomics.* MIT Press, Cambridge, Massachusetts.

BLINDER, A. S. (1981): "Real Inventory Behaviour and Business Cycles," *Brookings Papers on Economic Activity*, 2, 443–505.

BLINDER, A. S. (1986): "Can the Production Smoothing Model of Inventory Behaviour be Saved?," *Quarterly Journal of Economics*, CI, 431–453.

BLINDER, A. S., and L. J. MACCINI (1991): "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives*, 5, 73–96.

BLUME, L., C. EASLEY, and M. O'HARA (1982): "Characterisation of Optimal Plans for Stochastic Dynamic Programs," *Journal of Economic Theory*, 28, 221–234.

BRESNAHAN, T., and V. RAMEY (1994): "Output Fluctuations at the Plant Level," *Quarterly Journal of Economics*, 109, 593–624.

BRUNNER, K., and A. H. MELTZER (1971): "The Uses of Money: Money in the Theory of an Exchange Economy," *American Economic Review*, 61, 784–805.

CALOMIRIS, C., and R. G. HUBBARD (1990): "Firm Heterogeneity, Internal Finance and Credit Rationing," *Economic Journal*, 100, 90–104.

CAPLIN, A. S. (1985): "The Variability of Aggregate Demand with (s,S) Inventory Policies," *Econometrica*, 53, 1395–1409.

CARPENTER, R. E., S. M. FAZZARI, and B. C. PETERSEN (1994): "Inventory Investment, Internal-Finance Fluctuations, and the Business Cycle," *Brookings Papers on Economic Activity*, pp. 75–138.

CLOWER, R. W. (1967): "A Reconsideration of the Microfoundations of Monetary Theory," *Western Economic Journal*, 6, 1–8.

COGLEY, T., and J. M. NASON (1995): "Output Dynamics in Real-Business-Cycle Models," *American Economic Review*, 85, 492–511.

DIAMOND, D. W. (1989): "Reputation Acquisition and Debt Markets," *Journal of Political Economy*, 97, 828–862.

DIAMOND, D. W. (1991): "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy*, 99, 689–721.

EICHENBAUM, M. S. (1989): "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," *American Economic Review*, 79, 853–864.

FAIR, R. C. (1989): "The Production Smoothing Model is Alive and Well," *Journal of Monetary Economics*, 24, 353–370.

FAZZARI, S., G. HUBBARD, and B. PETERSEN (1988): "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1, 141–195.

GALE, D. (1978): "The Core of a Monetary Economy without Trust," *Journal of Economic Theory*, 19, 456–491.

GALE, D. (1982): *Money: In Equilibrium.* Cambridge University Press, Cambridge.

GERTLER, M. (1992): "Financial Capacity and Output Fluctuations in an Economy with Multiperiod Financial Relationships," *Review of Economic Studies*, 59, 455–472.

GERTLER, M., and S. GILCHRIST (1994): "Monetary Policy, Business Cycles, and the Behaviour of Small Manufacturing Firms," *Quarterly Journal of Economics*, 95, 309–340.

GOODHART, C. A. E. (1989): *Money, Information and Uncertainty (2nd Edition).* The Macmillan Press Ltd., London.

GREENWALD, B., and J. STIGLITZ (1993): "Financial Market Imperfections and Business Cycles," *Quarterly Journal of Economics*, 108, 77–114.

GRIMES, A. (1987): "Bargaining, Trust and the Role of Money," *Scandinavian Journal of Economics*, 92, 605–612.

GUARIGLIA, A. (1996): "The Effects of Finacial Constraints on Inventory Investment: Evidence form a Panel of UK Firms," Discussion Paper, University of Essex.

GUARIGLIA, A., and F. SCHIANTARELLI (1995): "Production Smoothing, Firm Heterogeneity and Financial Constraints: Evidence from a Panel of UK Firms," mimeo, University of Essex.

HAHN, F. H. (1965): "On some Problems of Proving the Existence of an Equilibrium in a Monetary Economy," in *The Theory of Interest Rates*, ed. by F. H. Hahn, and F. P. R. Brechling, pp. 126–135. London:Macmillan.

HAHN, F. H. (1973): "On Transactions Costs, Inessential Sequence Economies and Money," *Review of Economic Studies*, 40, 449–461.

HAHN, F. H. (1988): "On Monetary Theory," *Economic Journal*, 98, 957–973.

HALL, G. J. (1996): "Non-convex Costs of Capital Utilization: A Study of Production and Inventories at Automobile Plants," mimeo.

HELLWIG, M. F. (1992): "The Challenge of Monetary Theory," WWZ-Discussion Paper, University of Basel.

HOLMSTRÖM, B. (1982): "Managerial Incentive Problems," in *Essays in Economics and Management in Honor of Lars Wahlbeck*. Swedish School of Economics, Helsinki.

HOLT, C. C., F. MODIGLIANI, J. F. MUTH, and H. SIMON (1960): *Planning Production, Inventories and Work Force*. Prentice-Hall, Englewood-Cliffs.

JEVONS, W. S. (1875): *Money and the Mechanisms of Exchange*. Appleton, London.

JOVANOVIC, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50, 649–670.

KAHN, J. A. (1987): "Inventories and the Volatility of Production," *American Economic Review*, 77, 667–679.

KAHN, J. A. (1992): "Why is Production more Volatile than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory Holding," *Quarterly Journal of Economics*, 107, 481–510.

KASHYAP, A. K., O. A. LAMONT, and J. C. STEIN (1994): "Credit Conditions and the Cyclical Behaviour of Inventories," *Quarterly Journal of Economics*, 109, 565–591.

KASHYAP, A. K., J. STEIN, and D. W. WILCOX (1993): "Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance," *American Economic Review*, 83, 78–98.

KIYOTAKI, N. (1990): "Learning and the Value of the Firm," *FMG Discussion Paper, London School of Economics*, 99.

KIYOTAKI, N., and J. MOORE (1995): "Credit Cycles," Discussion Paper.

KIYOTAKI, N., and R. WRIGHT (1989): "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, 927–954.

KRANE, S., and S. BRAUN (1991): "Production Smoothing Evidence from Physical Production Data," *Journal of Political Economy*, 99, 558–81.

MIRON, J. A., and S. P. ZELDES (1988): "Seasonality, Cost Shocks and the Production Smoothing Model of Inventories," *Econometrica*, 56, 877–908.

MOSSER, P. (1990): "Trade Inventories and (s,S)," *Quarterly Journal of Economics*, pp. 1267–1286.

NOVAES, W., and Z. L. (1994): "Financial Distress as a Collapse of Incentives," mimeo.

RAMEY, V. A. (1991): "Nonconvex Costs and the Behaviour of Inventories," *Journal of Political Economy*, 99, 306–334.

SENSIER, M. (1996): "Inventories and Asymmetric Business Fluctuations," mimeo.

SMITH, A. (1776): *An Inquiry into The Nature and Causes of the Wealth of Nations.* Modern Library (1937), New York.

STOKEY, N. L., R. E. LUCAS, and E. PRESCOTT (1989): *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge.

TOBIN, J. (1956): "The Interest Elasticity of the Transactions Demand for Cash," *Review of Economics and Statistics*, 38, 612–616.

TOBIN, J. (1980): "Discussion of Overlapping Generations Models," in *Models of Monetary Economies*, ed. by J. Kareken, and N. Wallace, pp. 83–90. Federal Reserve Bank of Minneapolis.

TONKS, I. (1981): "Bayesian Learning and the Optimal Investment Decision of the Firm," mimeo.

TOWNSEND, R. (1980): "Models of Money with Spatially Separated Agents," in *Models of Monetary Economies*, ed. by J. Kareken, and N. Wallace, pp. 265–303. Federal Reserve Bank of Minneapolis.

WALLACE, N. (1980): "The Overlapping Generations Models of Fiat Money," in *Models of Monetary Economies*, ed. by J. Kareken, and N. Wallace, pp. 49–82. Federal Reserve Bank of Minneapolis.

WEST, K. D. (1986): "A Variance Bounds Test of the Linerar Quadratic Inventory Model," *Journal of Political Economy*, 94, 374–401.