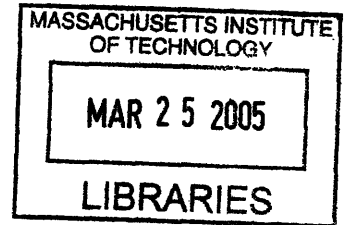# Three Essays in Finance and Macroeconomics

by

Stavros Panageas

Lic. oec. HSG (University of St. Gallen)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2005

Signature of Author . . . .
Department of Economics
30 September 2004

Certified by . . . . , ...
Ricardo J. Caballero
Ford International Professor of Economics
Thesis Supervisor

Accepted by . . . .
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Three Essays in Finance and Macroeconomics

by

Stavros Panageas

Submitted to the Department of Economics
on 30 September 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

In the first chapter I investigate whether firms' physical investments react to the speculative over-pricing of their securities. I introduce investment considerations in an infinite horizon continuous time model with short sale constraints and heterogeneous beliefs along the lines of Scheinkman and Xiong (2003) and obtain closed form solutions for all quantities involved. I show that market based q and investment are increased, even though such investment is not warranted on the basis of long run value maximization. I use a simple episode to test the hypothesis that investment reacts to over-pricing. With publicly available data on short sales during the 1920's, I examine both the price reaction and the investment behavior of a number of companies that were introduced into the "loan crowd" during the first half of 1926. In line with Jones and Lamont (2002), I interpret this as evidence of overpricing due to speculation. I find that investment by these companies follows both the increase and the decline in "q" before and after the introduction, suggesting that companies in this sample reacted to security over-pricing. In the next chapter of the thesis (co-authored with E. Farhi) we study optimal consumption and portfolio choice in a framework where investors save for early retirement. We assume that agents can adjust their labor supply only through an irreversible choice of their retirement time. We obtain closed form solutions and analyze the joint behavior of retirement time, portfolio choice, and consumption. In the final chapter of the thesis (co-authored with R. Caballero) we turn attention to hedging of sudden stops. We observe that even well managed emerging market economies are exposed to significant external risk, the bulk of which is financial. We focus on the optimal financial policy of such an economy under different imperfections and degrees of crowding out in its hedging opportunities.

Thesis Supervisor: Ricardo J. Caballero
Title: Ford International Professor of Economics

Contents

Acknowledgments

There are just not enough words to thank Ricardo Caballero, my principal advisor. Not only did he help improve this thesis considerably with his comments, but he set standards of excellence for research that will guide me for the rest of my academic career. I am equally unable to express appropriately my gratitude to Olivier Blanchard. His unique ability to encapsulate in a few words intuitions that would take me months to conceive, resulted in significant improvements especially in the first chapter of my thesis. I am also especially indebted to George Marios Angeletos. Besides giving me constant feedback on my ideas, Marios was always there during those times when I desperately needed encouragement.

During my graduate studies there was one professor, who was immeasurably influential in my education as an economist: Rudi Dornbusch. It is fair to say that I had no idea what economics was about, until I took Rudi's course. His socratic approach to teaching, his passion for economics, his ability to challenge common wisdom went far beyond just enhancing my knowledge of model X and Y. They were guidelines for what constitutes good economic thinking. It saddens me that I am completing this thesis without being able to give Rudi a copy, since he is no longer with us. I owe so much to him.

Throughout these four years a number of professors helped me with their feedback and comments. Particular thanks are due to Susan Athey, Victor Chernozhukov, Xavier Gabaix, Jerry Hausman, Leonid Kogan, S.P. Kothari, Guido Kuersteiner, Whitney Newey, Anna Pavlova, Jim Poterba, Roberto Rigobon, Steve Ross, Jiang Wang, Dimitri Vayanos, Ivan Werning.

I would also like to thank participants of seminars at MIT, Columbia GSB, Chicago GSB, Wharton, Kellogg, Yale SOM, Princeton, Stern School of Business, London Business School and the University of Maryland for helpful comments on the first chapter of this thesis, and participants of the 2004 AEA meetings and the 2003 UTDT Summer Camp for comments on the third chapter of the thesis.

I am particularly thankful to all my classmates and friends. Besides being the most

extensive discussants of every research idea, they also provided the quality of social life that is absolutely needed while writing a thesis and going through the academic job market.

By far the most important help came from my fiance Kedran Garrison. She was always there to remind me of things in life that are more important than economics. I cannot even imagine my life without her - she made these last years the nicest years of my life.

My family was always there for me. These years made me realize how much I miss Stelios, Stavros, Sandy and Vassia.

While this thesis was being written, my aunt Sandy and my uncle Yanni passed away. They were also my parents since my biological parents Dimitri and Kalliopi passed away when I was younger. I can only imagine how happy all four of them would be if they were able to see me graduate. Accordingly, this thesis is dedicated toall four of them.

Chapter 1

Introduction

In the last two decades financial markets experienced significant growth. Progress in information technology, along with wide ranging deregulation allowed an expansion of the financial markets in at least three important directions: first, a number of smaller more innovative firms were able to access stock and bond markets in order to obtain capital. Second, the supply of such capital increased by the presence of both -sometimes exuberant- optimism and because of the advent of vehicles like 401k's that gave workers an attractive alternative to save for retirement through the financial markets (especially the stock market). Third, financial markets crossed national boundaries offering investors and firms new investment opportunities.

This expansion made a number of issues concerning the interactions between the "real" side of the macroeconomy and the financial markets more apparent. First, the spectacular increase in stock values was accompanied by an increase in investment, which declined together with the stock market. Assuming that this behavior of the stock market was due to exuberant optimism, one is left to wonder why investment exhibited this behavior. Should the classical theory of Tobin's "q" hold, even if "q" does not reflect fundamentals? The theoretical and empirical implications of this question are the subject matter of the first paper: I investigate whether firms' physical investments react to the speculative over-pricing of their securities. I introduce investment considerations in an

8

infinite horizon continuous time model with short sale constraints and heterogeneous beliefs along the lines of Scheinkman and Xiong (2003). I obtain closed form solutions for all quantities involved. I show that market based q and investment are increased, even though such investment is not warranted on the basis of long run value maximization. Moreover, I show that investment amplifies the effects of speculation on prices through an increase in the value of "growth" options. In the empirical section of the paper, I use a simple episode to test the hypothesis that investment reacts to over-pricing. With publicly available data on short sales during the 1920's, I examine both the price reaction and the investment behavior of a number of companies that were introduced into the "loan crowd" during the first half of 1926. In line with Jones and Lamont (2002), I interpret this as evidence of overpricing due to speculation. I find that investment by these companies follows both the increase and the decline in "q" before and after the introduction, suggesting that companies in this sample reacted to security over-pricing.

In the next paper of the thesis (co-authored with E. Farhi) we develop a model in order to understand a different type of interaction between the "real" and the financial side of the economy. Namely, we study optimal consumption and portfolio choice in a framework where investors save for early retirement. The importance of these interactions became very apparent, especially after the collapse of the stock market in 2000, when retirement was postponed by many workers with funds invested in 401k's. In the model we assume that agents can adjust their labor supply only through an irreversible choice of their retirement time. We obtain closed form solutions and analyze the joint behavior of retirement time, portfolio choice, and consumption. We find that wealth plays a dual role: next to determining the resources available for future consumption, it controls the "distance" to early retirement. This introduces some new sources of wealth and horizon effects for optimal consumption and portfolio choice, that can be given an intuitive interpretation as option-type effects.

In the final paper (co-authored with R. Caballero) we turn attention to some implications of financial market expansion beyond national boundaries. This expansion made us

observe during the 80's and the 90's ,that even well managed emerging market economies are exposed to significant external risk, the bulk of which is financial. At a moment's notice, these economies may be required to reverse the capital inflows that have supported the preceding boom. Even if such a reversal does not take place, its anticipation often leads to costly precautionary measures and recessions. In this paper, we characterize the business cycle of an economy that on average needs to borrow but faces stochastic financial constraints. We focus on the optimal financial policy of such an economy under different imperfections and degrees of crowding out in its hedging opportunities. The model is simple enough to be analytically tractable but flexible and realistic enough to provide quantitative guidance.

The crossroads between Finance and Macroeconomics is an exciting area of research. This thesis was devoted to only three questions that I feel are important and representative of the main issues that arise at the intersection of the two fields. I hope that the papers in this thesis can contribute to a better understanding of these three very important issues.

Chapter 2

Speculation, Overpricing and Investment: Theory and Empirical Evidence

2.1 Introduction

Standard neoclassical theory predicts that investment is inherently tied with the stock market through Tobin's "q". The essence of "q" theory is the following argument: if the repurchase cost of capital is less than the net present value of additional profits it will bring at the margin, then the company should invest and vice versa. The only thing preventing the ratio of the two values (known as q) from being always equal to 1 is adjustment costs: it is expensive to install new capital and thus a deviation of q from 1 can exist, but it should diminish over time. The link between investment and the stock market follows: the value of a company is the net present value of its profits and thus[1] whenever one sees the stock market rising, one should simultaneously observe an increase

---

[1] Under Hayashi's (1982) conditions

in investment in order to bring the numerator and the denominator of the "q" ratio into line.

However, there is a concern with this line of reasoning. Namely, what happens if the stock market valuation at times does not reflect the net present value of profits but also contains terms that are unrelated to "fundamentals"? Will "q" theory continue to hold or will decision makers in companies be eclectic about the components of stock market valuation to which they will pay attention ?

This is the main question I take up in this paper. I start with an explicit reason for why assets can deviate from fundamentals. Then I introduce investment considerations and study investors' holding horizons, optimal investment, and the resulting equilibrium prices in a unified framework.

To be more specific, I use short sales constraints to derive positive deviations of prices from fundamentals. It is intuitive that the presence of a short sale constraint can cause the price of an asset to deviate from its fundamental value if market participants do not have homogenous beliefs. Agents who believe that the current price is above the net present value of dividends would have to go short in order to take advantage of what they perceive to be mispricing. However, they cannot do this because of the short sale constraint. Accordingly, for pricing purposes, it is as if they do not exist, and the price will only reflect the views of the most optimistic market participants.

This basic intuition was first expressed in a formal intertemporal model by Harrison and Kreps (1978). A number of papers extended the intuition into various directions. A partial listing includes Allen, Morris, and Postlewaite (1993), Detemple and Murthy (1997), Morris (1996) and most recently Scheinkman and Xiong (2003) and Hong and Stein (2002). All of these papers study an exchange setting without a role for investment.

The present paper extends this literature to allow for investment. In particular the model presented here is based on Scheinkman and Xiong (2003) with the difference that I allow firms to adjust their capital stock by investing. Because the model is set up in continuous time I can derive closed form solutions for prices, investment, trading

strategies and investors' horizons. First, I show that traditional "q" theory remains valid if investors have perfect access to financial markets, they are risk neutral and investment is determined in the best interest of current shareholders.[2] Whether the stock price is high because of fundamentals or resale premia is irrelevant. A shareholder value maximizing company will use the stock market valuation as a guide to how much investors can gain in the stock market by either holding the asset and reaping dividends or by reselling it to more optimistic investors. Second, I show that investment significantly amplifies the effects of speculation on the asset prices by affecting the value of growth options embedded in the company's price.[3] In the present framework, young dynamic companies with low adjustment costs and high disagreement associated with their underlying productivity can end up with high levels of q (low levels of book to market) and low expected returns. The closed form solution obtained for the price of the firm allows a quantification of these effects and a comparison with actual data.

It is possible however, to imagine circumstances where investment would not react to market based q and the above logic would fail. For instance, if a major shareholder owns a significant fraction of a firm and values control she would be unlikely to react to resale premia because they are irrelevant for her. Similarly, key investors might be afraid of selling their shares in large amounts because other investors might fear the presence of asymmetric information. In other words, resale premia are only relevant for investors that have short horizons and who can realize the full speculative gains associated with them. If they can't access the markets (or accessing the markets is costly) then the incentive to invest will be attenuated.[4] I derive optimal investment under this alternative and then discuss a set of observable implications.

Then I address the empirical question: which of the two theories is supported by the

---

[2] Risk neutrality is not essential if one is willing to make a specific assumption about the valuation of income streams by investors in incomplete markets. See the next section for details.

[3] Growth options are defined as the difference between the equilibrium price when investment is determined optimally and the equilibrium price when investment is set to 0 throughout.

[4] Similar points were made in Blanchard, Rhee and Summers (1993), Stein (1996) and Morck, Shleifer and Vishny (1990). All three papers emphasize the distinction between short and long horizons.

data? Answering this question is difficult because one has to identify a shock to resale premia but not to fundamentals. Only then can one study how investment reacts to the former type of shocks. Disentangling fundamental from non-fundamental deviations is a difficult task. The usual approach in the literature has been to try to find proxies for the two components. Such an approach is associated with the usual doubts on how successful one is in creating these proxies. Moreover, certain proxies that are often used are not obviously related to short selling costs and constraints alone, but capture asymmetric information or agency problems.

In this paper I take a direct approach: in the 1920's an entire market, known as the "loan crowd", was active for shorting stock. I use a dataset recently collected by Jones and Lamont (2002) based on daily coverage of this market by the *Wall Street Journal.* The list of companies in this market expanded in several waves. As Jones and Lamont (2002) argue, the introduction of a company into the "loan crowd" reflects a belief by investors that this company is particularly overpriced. I provide some additional evidence to that effect. The behavior of the stock price of the newly introduced companies indeed seems to confirm such an explanation. Stock prices show a marked runup for several quarters before the introduction and decrease dramatically thereafter. Not surprisingly, market-based q presents exactly the same behavior. To complement the dataset of Jones and Lamont (2002) with balance sheet data, I hand-collected financial data on a number of these companies from Moody's manuals and studied the behavior of investment in the years prior to their introduction and thereafter. I find that investment followed exactly the same behavior as market based "q".

The paper is related to a number of strands in the literature. There is a small number of papers that have addressed the same set of issues, mostly from an empirical angle. These include: Fischer and Merton (1984), Morck, Shleifer, and Vishny (1990), Blanchard, Rhee, and Summers (1993), Stein (1996), Chirinko and Schaller (1996) and more recently Polk and Sapienza (2002), Gilchrist, Himmelberg, and Huberman (2002). A central theme of this literature is the importance of investor's horizons. However,

the models developed in these papers do not explicitly characterize the optimal holding horizon (defined as the stopping time at which an investor finds it optimal to resell). Moreover, these models do not allow one to derive intertemporal implications for investment and stock prices jointly. For example this makes it difficult to determine why and when certain Euler relations should hold or fail, and thus is important from an empirical viewpoint.[5] The present paper models everything explicitly in an infinite horizon continuous time setting and thus one is able to model investor's horizons endogenously and derive testable implications about the relationship between investment and prices in an explicit way. The empirical approach to testing the theory is also more direct. Instead of using proxies to account for mispricing, I use the firms that were perceived to be as most overvalued at the time as evidenced by the fact that they were introduced into the loan crowd.[6]

The paper is also related to a literature in financial economics that uses insights from investment theory to address issues such as the predictability of returns, the role of book to market ratios, etc. A partial listing would include Cochrane (1991,1996), Naik (1994), Berk, Green, and Naik (1999), Lamont (2000). Berk, Green, and Naik (1999) in particular show how a model with investment can account for some apparent irregularities in asset pricing as e.g. the power of the Book to Market ratio to predict returns[7]. In this paper I obtain a closed form solution that decomposes the price into a component

---

[5]For instance Chirinko and Schaller (1996) derive a test for whether bubbles affect investment or not, by making the interesting assumption that bubbles lead to predictable returns. From that assumption they derive the result that if investment reacts to bubbles, certain Euler relations should fail to hold. However, not every source of predictability can be attributed to bubbles and bubbles will not necessarily lead to predictability. In the explicit framework of this paper, one can determine both the source of predictability and its implications for testing. This issue is explained in detail in the sections that follow.

[6]One direction that is not explored is the behaviour of investment, if decisionmakers are longtermist but the company is financially constrained. It can be conjectured that in this case investment could potentially react to market based q even if managers maximize long run performance. See e.g. Stein (1996), Baker, Stein, and Wurgler (2003). It is interesting to note that in the present paper one does not need to assume anything apart from shareholder value maximization to arrive at the result that investment reacts to market based q even in the absence of constraints. It is also conceivable that constraints could further amplify the result. I discuss this point in further detail in the conclusion of the paper.

[7]This fact is documented in the cross section by Fama and French (1992,1995) and in the time series dimension by Kothari and Shanken (1997) among others.

related to assets in place and "growth options" or "rents to the adjustment technology". Moreover, I can derive the effects of speculation on both components separately. I find that the "growth options" amplify significantly the effects of speculation. Returns are predictable and predictability of returns becomes strongest when both fundamentals and disagreement about fundamentals are high. This is in contrast to the pure exchange case where predictability only depends on disagreement. It also makes it easier for quantities like B/M or E/P to predict returns since the price of a company captures both fundamental and non-fundamental variations. In a quantitative exercise I show that q can become large even for small degrees of irrationality. This has the potential to explain quantitatively the very low book to market ratios that one observes during speculative episodes. Moreover, the model has the potential to produce reasonable levels of predictability of returns in quantitative terms as is shown by simulating an artificial CRSP dataset and re-running some Fama and MacBeth regressions of simulated monthly returns on Book to Market.

An interesting application of this paper concerns the relationship between return predictability and investment: the reaction of investment to speculative components in prices could potentially help in distinguishing rational and behavioral views of predictability. If investment only reacted to fundamental variations and was powerful at predicting returns, then this would be evidence that the variation in expected returns is due to variation in risk aversion and not to speculation motives and expectational errors. Lamont (2000) indeed documents the ability of investment plans to explain aggregate returns. However, to make the link between investment and variations in risk premia, one would need to establish that investment only reacts to variations in fundamentals and not to potentially irrationally optimistic beliefs. The empirical evidence that I provide in this paper, suggests that investment reacts to both fundamental and speculative terms. Thus, investment does not seem to be able to provide a clear way to distinguish between rational and behavioral theories.

The paper is also complementary to the strand in the macroeconomics literature that

16

models bubbles in the framework of overlapping generations models. Blanchard and Fischer (1989) present a textbook treatment, whereas Caballero and Hammour (2002) and Jacques (2000) are some recent contributions to this literature. This literature assumes short horizons, whereas in the present paper short horizons arise endogenously. Moreover, uncertainty is key in the present paper, whereas uncertainty typically plays a secondary (if any) role in the papers above. However, the simpler setup of overlapping generations allows one to address a richer set of issues (related e.g. to savings and fiscal policy) that would be hard in the present setup. In a sense, the model developed here provides a foundation for models in this literature.

The outline of the paper is as follows: Section 2.2 presents a simple three-period example that allows an easy presentation of most intuitions of the model. Section 2.3 presents the model in an infinite horizon continuous time setting with a richer set of dynamics for the beliefs of the agents. In this section I also discuss the properties of the model and its implications for testing. Section 2.4 presents the empirical evidence. Section 2.5 concludes. All proofs are given in the appendix.

2.2 A simple example

In this section I present a simple example that will help fix some ideas. I extend this example in the section that follows to a continuous time setting. I assume that the world lasts for three periods. There are two states of the world h and l and a single productive asset that pays out $f_t K_t$. f_t is 0 in state l and 1 in state h. K_t is the amount of capital available to the economy at time t. For simplicity I also assume that labor is not required to produce output and that the economy is small, i.e. the interest rate is taken as given and normalized to 0. To introduce heterogeneity of beliefs I assume that there are two types of agents, which I label agents A and B. There is a continuum of both types having infinite total wealth and /or an infinite ability to borrow.[8]

---

[8] This assumption is made by both Harrison and Kreps (1978) and Scheinkman and Xiong (2003) and is useful in order to drive values towards the reservation price.

17
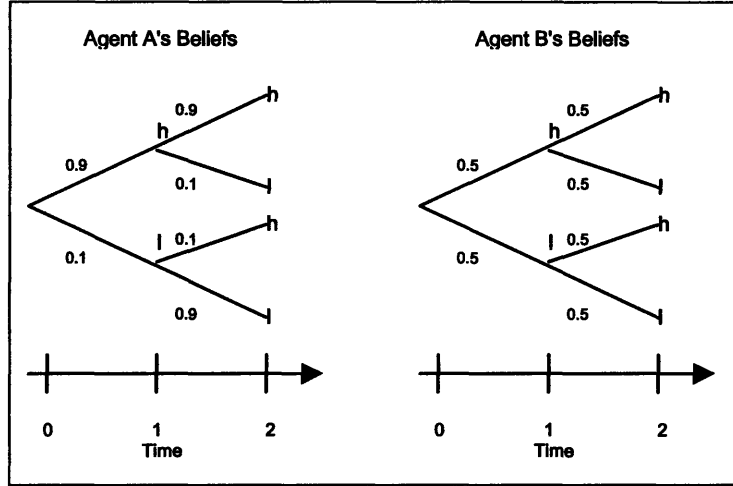
Figure 2-1: Transition probabilities as perceived by agents in groups A and B

Figure 2-1 depicts the transition probabilities that agents in each group assign to the transition from one state to the other. In particular, agents in group B perceive each state as equiprobable, while agents in group A are originally optimistic (they assign probability 0.9 to the high state occurring). If the high state occurs, then they continue to be optimistic about period 2, otherwise they become pessimistic (in the sense that they assign probability 0.9 to the low state occurring again). The only crucial feature of this setup is that agents do not agree on the transition probabilities. Agents cannot take short positions in the asset. I assume that at time 0 the economy is in state h. The setup is common knowledge to the agents who agree to disagree.

Suppose initially that there is no investment (i.e. I treat $K_t = 1$ as a constant). Moreover there is no depreciation. Equilibrium prices and trading strategies are determined by backwards induction. The joint assumptions of risk neutrality and infinite total wealth allow one to set the price equal to the reservation price of the person who values the asset most. In particular at time 1 and state h the agents who value the asset the most are agents in group A. The reservation price for agents in group A is given as $P_{1h}^A = 0.9 \mathrm{x} 1 + 0.1 \mathrm{x} 0 = 0.9$. The reservation price for agents in group B is given as

$P_{1h}^B = 0.5$x$1+0.5$x$0 = 0.5$. Even though agents in group B would be happy to short the asset in this state, they can't. Accordingly the price is given as $P_{1h} = \max \left[P_{1h}^A, P_{1h}^B \right] = 0.9$. Similarly in state $(1, l)$ the price will be given as $P_{1l} = 0.5$x$1 + 0.5$x$0 = 0.5$ since now agents of type A are less optimistic than agents in group B. At time 0 agents in group A value the asset at $P_0^A = 0.9$x$(1+0.9)+ 0.1$x$(0+0.5) = 1.76$. This is the relevant valuation for agents in group A because at the node $(1, l)$ they know that they will resell the asset to agents in group B. For group B the reservation price is $P_0^B = 0.5$x$1.9+ 0.5$x$0.5 = 1.2$, so that the equilibrium price of the asset will be given by $P_0 = \max \left[P_0^A, P_0^B \right] = 1.76$. A convenient way of summarizing the above discussion is in terms of the Harrison and Kreps (1978) formula:

$$P_t = \max_{o \in \{A,B\}} [P_t^o] = \max_{o \in \{A,B\}} \left[\sup_\tau E^o \left(\sum_{s=t+1}^\tau D_s + P_\tau \right) \right] \qquad (2.1)$$

where D_s are the dividends paid at the state-time pair s and τ is an optimally chosen stopping time at which an agent decides to sell the asset.

Another recursive relation that is true is[9]:

$$P_t = \max_{o \in \{A,B\}} [E^o (D_{t+1} + P_{t+1})]$$
$$P_T = 0$$

Interestingly, the price at node 0 is strictly higher than what either agent would be willing to pay if she didn't consider the possibility to resale the asset later on. In particular, if one prohibits agents from engaging in transactions at any point other than at time 0, the price $\widetilde{P_0}$ of the asset is given by:

$$\widetilde{P_0} = \max_{o \in \{A,B\}} \left[\widetilde{P_0^o} \right] = \max_{o \in \{A,B\}} \left[E^o \left(\sum_{s=1}^T D_s \right) \right]$$

---

[9]See Harrison and Kreps (1978)

The value has to be lower than the original price, since agents are deprived of the possibility to resell. This possibility is embodied in the optimization over stopping times in formula (2.1). That is: $P_0 \geq \widetilde{P_0}$. The difference in the two values is the option value ascribed to reselling the asset in the future. An additional implication of this formula concerns predictability of asset returns. If I assume that one of the market participants has the "right" beliefs then the prices are no longer martingales from her perspective. This investor views the asset as having potentially negative expected (excess) returns in certain states of the world. A further interesting interpretation of (2.1) is in terms of the investor's holding horizon. The interaction of heterogenous beliefs and short sales makes the optimal stopping time problem meaningful. In contrast, if one assumed homogenous beliefs then all stopping times would yield the same payoff and accordingly One could assume that the investor holds the asset until maturity.

Consider now the above example with investment. In particular, assume that there is a technology that allows agents to reduce todays' dividends by:

$$i_t + \chi \frac{i_t^2}{2}$$

in order to increase next period capital to

$$K_{t+1} = K_t + i_t$$

I will assume that investment is determined in the best interest of investors who are endowed with the stock at the beginning of the period[10]. Once again I start backwards in order to determine equilibrium outcomes. In state $(1, h)$ it is clear that agents of type A will end up holding the stock no matter which investment strategy is chosen and no matter which type of agent is endowed with the stock. This is so because for any investment

---

[10]In particular I assume (like Grossman and Hart (1979)) that investors arrive at the beginning of the period with a certain endowment of the stock, they determine the investment policy, dividends are paid and then they trade their shares in a Walrasian market.

decision their reservation value will be higher than the reservation value of agents of type B for the stock. Accordingly, the investment decision will be determined according to the beliefs of type A agents *independently of who is endowed with the company stock at the time-state pair* $(1, h)$. In mathematical terms:

$$1 + \chi i_{1,h}^* = E^A \left(f_2 | s_t = 1, h \right) \tag{2.2}$$

where f_2 denotes the productivity in period 2. This first order condition is obvious if agents of type A are endowed with the stock in the state-time pair $(1, h)$. It is also true however, if the the company stock belongs to agents of type B. To see this, notice that agents of type B have to balance two effects in making their investment decision. On the one hand they realize that by investing they reduce the dividends that they can obtain. On the other hand they increase the resale value of the asset since agents of type A will be willing to pay more for a company with a larger capital stock. Agents in group B understand that only agents of group A will matter for pricing purposes in state $(1, h)$. Thus they are led to understand that a marginal investment of Δi_t will change the price that they can gain for the asset by $E^A \left(f_2 | s_t = 1, h \right) \Delta i_t$ while reducing the current dividends by $(1 + \chi i_t) \Delta i_t$. Balancing out these two effects leads to the same first order condition as (2.2).

For the optimal investment rule derived from (2.2) one can determine time 1 dividends as:

$$D_1 = f_1 K_1 - i_1^* - \chi \frac{\left(i_1^* \right)^2}{2}$$

and period 2 capital as:

$$K_2 = K_1 + i_1^*$$

Working backwards by the same logic one can establish that independently of who controls the company at time 0 the optimal investment strategy is to invest until:

$$1 + \chi i_0^* = E^A \left(f_1 + q_1 \right)$$

where q_1 is given as $E^A\left(f_2|s_t=1,h\right)$ if the state h realizes and as $E^B\left(f_2|s_t=1,l\right)$ if the state l realizes. It is interesting to note that one can express the stochastic process for investment in terms of the recursive relations:

$$1+\chi i_0^* \;=\; q_t \tag{2.3}$$

$$q_t \;=\; \max_{j\in\{A,B\}} E^j\left[f_{t+1}+q_{t+1}\right] \tag{2.4}$$

$$q_T \;=\; 0$$

The second of these equations is exactly the same equation that was obtained for the price in the context of the simple exchange setting. The above discussion motivates the main concept of equilibrium that I will use in this paper. Namely, I will assume that the company is maximizing investor welfare and accordingly I will be searching for investment policies, selling/stopping times, and equilibrium pricing functions that satisfy the relation:

$$P_t(K_t) = \left(\max_{o\in\{A,B\}}\left[\sup_{i_s,\tau} E^o\left(\sum_{s=t+1}^{\tau} D_s(i_s,K_s)+P_\tau(K_\tau)\right)\right]\right)$$

where $i_{t..T}$ denotes the stochastic process of investment. Clearly, the analysis needs to be modified, if I assume that agents cannot retrade. For instance suppose that markets will only be open at time 0 and never thereafter. Then (2.4) should be replaced by the conventional q relationship

$$q_t \;=\; E^j\left[f_{t+1}+q_{t+1}\right] \tag{2.5}$$

$$q_T \;=\; 0$$

where j denotes the agent who will bid more for the company at time 0 (in this example agent A). It is evident by comparing (2.4) and (2.5) that investment will be necessarily higher in the presence of a resale premium on assets. Also, by the same argument as in the exchange setting marginal "q" no longer satisfies the usual martingale

type relationship under the beliefs of any agent. This means that every agent understands that in certain states of the world investment will be undertaken even though its expected (excess) return is negative from her perspective.[11][12]

2.3   A continuous time framework

The primary goal of this section is to derive testable implications of the hypothesis that investment is affected by resale premia and quantify the effects discussed in the last section. Moreover this section focuses on the effects of speculation on the value of a company's growth options and demonstrates how they magnify the effects of speculation on the stock price. I will expand the previous model to an infinite horizon continuous time setup with quadratic adjustment costs independent of the capital stock. Continuous time introduces some tractability into the problem. It allows one to determine closed form solutions for prices, investment and stopping policies. In particular, I will combine a framework proposed by Harrison and Kreps (1978) and Scheinkman and Xiong (2003) to

---

[11]One might wonder to what extent the conclusions of this section depend on risk neutrality. It can be shown that an extension of the ideas in Grossman and Hart (1979) can be used to address the risk aversion case. In particular, assuming "utility taking" behavior on the most optimistic agents allows one to generalize (2.4) to a setting with risk aversion by using the marginal utilities of the most optimistic agents at each node of the information tree to construct a pricing kernel. Details on this construction are available upon request.

[12]So far I assumed nothing about financing policies. This was done because the Modigliani Miller Theorem continues to hold in this setup despite the short selling constraint. (If one allowed for debt financing then one would also need to assume unlimited liability in light of the results in Hellwig (1981)). A proof of these claims can be given by arguments identical to DeMarzo (1988) and is available upon request. The intuition for why the MM Theorem holds is straightforward. In this setup the firm cannot do more by trading in its own stock (i.e. by issuing shares) than what the investor can do by selling her shares in the market. This is true because the only direction in which the investor is constrained is the short side. Accordingly, if financial policy could create value then it would have to be by promising to deliver a negative multiple of the company's dividends. This would effectively alleviate the investor's short sale constraint. Of course there is no financial policy that can do that, and accordingly financial policy cannot create value for the investor. This analysis also demonstrates one way to introduce active financial policy in this framework. Suppose for instance that accessing the financial markets directly is costly for existing investors due to e.g. asymmetric information or fears of nonlinear price impact. Then, an easy way for the investor to sell stock and realize speculative gains is by having the firm issue stock and not participating. In reality there seems to be a strong relationship between equity issuance and speculation as documented by Baker and Wurgler (2000), Baker Stein and Wurgler (2003).

study speculative premia on assets with a standard investment framework with quadratic adjustment costs along the lines of Abel (1983), Abel and Eberly (1994), (1998). I also discuss how one can generalize the basic predictions to a setup with an arbitrary number of groups of agents and arbitrary linear homogenous adjustment technologies.[13]

2.3.1 Setup

Company Profits and Investment

There is a single company and the goal will be to determine its value as part of the (partial) equilibrium solution of the model. The company's cumulative earnings process is given by:

$$dD_t = K_t f_t dt + K_t \sigma_D dZ_t^D \tag{2.6}$$

In units of installed capital this expression becomes:

$$\frac{dD_t}{K_t} = f_t dt + \sigma_D dZ_t^D$$

The first component captures a stochastic trend growth rate whereas the second term captures noise in the company earnings that prevents market participants from perfectly inferring the level of productivity f_t. K_t is the amount of physical capital installed in the company which "scales" both the trend growth rate and the "noise" in the cumulative earnings process. σ_D is a constant controlling the "noise", while dZ_t^D is a standard one dimensional Brownian Motion. The variable f_t is not observable and evolves according to an Ornstein Uhlenbeck process as:

$$df_t = -\lambda(f_t - \overline{f})dt + \sigma\sqrt{\frac{f_t}{\overline{f}}}dZ_t^f \tag{2.7}$$

where $\lambda > 0$ is a mean reversion parameter, $\overline{f} > 0$ is a long-run productivity rate, σ_f

---

[13]Unfortunately, in this case it appears very difficult to obtain closed form solutions for prices, investment etc.

is the volatility of the Ornstein Uhlenbeck process and dZ_t^f is a second Brownian motion that is independent of dZ_t^D. For simplicity I will also assume that the company is fully financed by equity and there is a finite number of shares of the company whose supply I normalize to 1. The company can invest in physical capital at the rate i_t. The evolution of the capital stock is accordingly given by:

$$dK_t = (-\delta K_t + i_t)\, dt$$

Investment is subject to quadratic adjustment costs so that the cumulative company earnings net of investment costs are given by:

$$d\Pi_t = dD_t - \left(pi_t + \frac{\chi}{2}(i_t^2)\right) dt$$

where χ is a constant controlling the significance of adjustment costs and p is the cost of capital. It will be useful to define p as a fraction ($\widetilde{p}$) of $\frac{\bar{I}}{r+\delta}$ so that $p = \widetilde{p}\frac{\bar{I}}{r+\delta}$. The assumption of adjustment costs that are independent of K_t has the benefit of allowing reasonably tractable solutions, however it comes at the cost of breaking down the equivalence between average and marginal "q". [14] In the appendix I show how one can generalize (at least qualitatively) the results of this section to a setup with linear homogenous adjustment cost technologies of the sort usually employed in the empirical literature.

Agents and Signals

There are two continuums of risk neutral agents that I will call type A and type B agents. Risk neutrality is convenient both in terms of simplifying the calculations and abstracting from considerations related to spanning etc. In addition to the earnings process (2.6)

---

[14] This assumption has been made by several authors in the literature. See e.g. Abel and Eberly (1994) and the references therein (especially footnote 19)

both agents observe two signals that I will denote signal s^A and signal s^B. These signals evolve according to:

$$ds_t^A = f_t dt + \sigma_s \phi dZ_t^f + \sigma_s \sqrt{1 - \phi^2} dZ_t^A$$
$$ds_t^B = f_t dt + \sigma_s dZ_t^B$$

where $(dZ_t^A, dZ_t^B, dZ_t^f, dZ_t^P)$ are standard mutually orthogonal Brownian motions.

Agents have heterogenous perceptions about the informativeness of the various signals. Agents in group A have the correct beliefs, while agents in group B assume that the innovations to the s_t^B process are more and the innovations to the s_t^A process less informative than they actually are. In particular they believe that the signals evolve according to:

$$ds_t^A = f_t dt + \sigma_s dZ_t^A$$
$$ds_t^B = f_t dt + \sigma_s \phi dZ_t^f + \sigma_s \sqrt{1 - \phi^2} dZ_t^B$$

This setup is meant to capture situations when there has been a regime shift in the economic environment and agents disagree about the informativeness of certain signals because they cannot use past data in order to measure the correlation between various signals with the underlying productivity process. For instance one could interpret the above informational setup as a situation where new signals (e.g. the amount of web site hits of a newly formed dot.com company) arise and analysts are unsure as to how important they are for future profitability.

Finally, as in the previous section, I assume that there is a continuum of agents of each type and the total wealth of each group is infinite.[15]

In the appendix I establish an approximate filter for this setup.[16] In particular I show

---

[15] This assumption is made by both Harrison and Kreps (1978) and Scheinkman and Xiong (2003) and is used to drive prices to the reservation value of each group.

[16] In contrast to Scheinkman and Xiong (2003) I assume a square root process for f_t in (2.7) instead of a standard OU process in order to guarantee positivity of f_t. This allows one to put a lower bound

that the posterior mean $\widehat{f}_t^A$ of agent $A's$ beliefs about f, evolves approximately according to:

$$d\widehat{f}_t^A = -\lambda \left(\widehat{f}_t^A - \overline{f}\right) dt + \sqrt{\frac{\widehat{f}_t^A}{\overline{f}}}\sigma_f dB_t^A \tag{2.8}$$

where dB_t^A is an appropriate linear combination of the innovation processes $\left(ds_t^A - \widehat{f}_t^A dt\right)$, $\left(ds_t^B - \widehat{f}_t^A dt\right)$, $\left(\frac{dD_t}{K_t} - \widehat{f}_t^A dt\right)$ with the property that the volatility of dB_t^A is 1.Similarly for agent B:

$$d\widehat{f}_t^B = -\lambda \left(\widehat{f}_t^B - \overline{f}\right) dt + \sqrt{\frac{\widehat{f}_t^B}{\overline{f}}}\sigma_f dB_t^B \tag{2.9}$$

where dB_t^A is an appropriate linear combination of the innovation processes $\left(ds_t^A - \widehat{f}_t^B dt\right)$, $\left(ds_t^B - \widehat{f}_t^B dt\right)$, $\left(\frac{dD_t}{K_t} - \widehat{f}_t^B dt\right)$ with the property that the volatility of dB_t^B is 1.[17]

A quantity that will be central for what follows is the disagreement process. In the appendix I show that agent A perceives that the process:

$$g_t^A = \widehat{f}_t^B - \widehat{f}_t^A$$

which captures her disagreement with agent B can be approximated by a simple OU process:

$$dg_t^A = -\rho g_t^A dt + \sigma_g dW_g^A$$

with $< dB_t^A, dW_g^A >= 0$. The situation for agent B is symmetric. She perceives that the process:

$$g_t^B = -g_t^A$$

evolves approximately as an OU process with increments orthogonal to dB_t^B. Obviously, knowing $\widehat{f}_t^A, g_t^A$ allows one to compute $\widehat{f}_t^B = \widehat{f}_t^A + g_t^A$. Thus, if one is only interested in

---

on f_t which is convenient for some of the proofs in the appendix. The downside of this assumption is that filtering becomes much more involved and I have to settle for an approximate filter, the properties of which seem to be very good.

[17]Intuitively agent B will underweight signal A and overweight signal B and thus she will choose a different combination of the innovation processes.

posterior means, the pair $(\widehat{f}_t^A, g_t^A)$ summarizes the entire belief structure. The appendix presents these approximations in detail and discusses their accuracy. Conditional on these approximate dynamics for the belief processes the rest of the analysis is exact. It is also important to note that one could have chosen any belief structure dynamics as long as it implies disagreement between the agents in some states of the world. The present one was chosen only for tractability reasons.

2.3.2 Equilibrium Investment, Trading and Pricing

Homogenous Beliefs

I start with the simplest possible case[18] where every agent is of type A and accordingly everyone agrees on the interpretation of the signals. One could also think of the discussion in this section as the solution to the investment problem of a long termist risk neutral decision maker who will never resell her shares. The goal will be to maximize

$$P_t = \max_{i_s} E^A \int_t^\infty e^{-r(s-t)} d\Pi_s \tag{2.10}$$

which can be rewritten as[19]

$$P_t = \max_{i_s} E^A \int_t^\infty e^{-r(s-t)} \left(f_s K_s - p i_s - \frac{\chi}{2}(i_s^2) \right) ds$$

---

[18]These results in this section are fairly standard and the reader is refered for details to Abel and Eberly (1997).

[19]Throughout I will restrict attention to investment policies that satisfy the requirement:

$$E\left[\int_t^\infty e^{-r(s-t)} K_s dZ_s^D \right] = 0$$

which amounts to a standard square integrability condition on the allowed capital stock processes. Indeed in the present setup the capital stock turns out to be stationary and thus it is easy to verify this condition.

One can further rewrite the above objective as[20]

$$P_t = \max_{i_s} E^A \int_t^\infty e^{-r(s-t)} \left(\widehat{f}_s^A K_s - p i_s - \frac{\chi}{2}(i_s^2) \right) ds$$

This is a problem of exactly the same form as the ones considered in Abel and Eberly (1994),(1997). The solution to this problem (obtained in the appendix) is given by

Proposition 2.1 *The solution to (2.10) is given as:*

$$P_t \left(\widehat{f}_t^A, K_t \right) = \left(\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} \right) K_t + \tag{2.11}$$

$$+ \left(C_1 \left(\widehat{f}_t^A - \overline{f} \right)^2 + C_2 \left(\widehat{f}_t^A - \overline{f} \right) + C_3 \right) \tag{2.12}$$

for appropriate constants C_1, C_2, C_3 given in the appendix. Optimal investment is given by:

$$i_t = \frac{1}{\chi}(P_K - p) = \frac{1}{\chi} \left(\frac{\overline{f}(1-\widetilde{p})}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} \right) \tag{2.13}$$

For $0 \le \widetilde{p} < \overline{p}$ (where $\overline{p}$ is a constant given in the appendix) it can be shown that $P_f > 0$.

Exactly as in Abel and Eberly (1997), the equilibrium price is comprised of two components. The first is marginal "q" times the capital stock and the other term captures the rents to the adjustment technology or "growth options". The first term captures the expected net present value of profits that can be obtained with the existing capital stock i.e

$$P_K = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} = E \left(\int_t^\infty e^{-(r+\delta)(s-t)} \widehat{f}_s^A ds \right) \tag{2.14}$$

The second term captures the "rents to the adjustment technology" or "growht options", i.e. the value of being able to adjust the capital stock in the future:

$$\left(C_1 \left(\widehat{f}_t^A - \overline{f} \right)^2 + C_2 \left(\widehat{f}_t^A - \overline{f} \right) + C_3 \right) = \frac{1}{2\chi} E \left(\int_t^\infty e^{-r(s-t)} \left(P_K(s) - p \right)^2 ds \right) \tag{2.15}$$

---

[20]This is true since the objective is linear in the state and quadratic only in the control i_t. For details on such problems see Bertsekas (1995).

The first term is clearly increasing in $\widehat{f}_t^A$, while the second term is also increasing in $\widehat{f}_t^A$. This means that not only does a higher belief about current profitability increase the expected profits in the future, it also increases the value of growth options. This is because it becomes more likely that large investments will need to be undertaken in the future and thus the technology to adjust the capital stock becomes more valuable. Small adjustment costs (i.e. low values of χ) will tend to increase the value of the adjustment technology. This is intuitive: the less it costs to adjust the capital stock, the more a company is able to invest (disinvest) and take advantage of temporary increases (decreases) in fundamentals $\left(\widehat{f}_t^A\right)$.

Heterogenous beliefs: Optimal investment, trading, and equilibrium prices

This section discusses the recursion:

$$P = \max_{o \in \{A,B\}} \left(\sup_{i_s, \tau} E_t^o \left[\int_t^{t+\tau} e^{-r(s-t)} \left(dD_s - \left(p i_s + \frac{\chi}{2} i_s^2 \right) ds \right) + e^{-r(\tau-t)} P_{t+\tau} \right] \right) \quad (2.16)$$

As is shown in the appendix the crucial difficulty in dealing with this recursion is that it leads to a multidimensional optimal stopping problem. Fortunately, this problem can be solved explicitly. In the appendix I show the following result:

Proposition 2.2 *The solution to* $P\left(\widehat{f}_t^A, g_t^A, K_t\right)$ *is given by*[21]:

$$
\begin{aligned}
P\left(\widehat{f}_t^A, g_t^A, K_t\right) &= \left(\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + 1\{g_t^A > 0\} \frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-|g_t^A|) \right) K_t + \\
&\quad + C_1 \left(\widehat{f}_t^A + 1\{g_t^A > 0\} g_t^A - \overline{f} \right)^2 + [C_2 + n(-|g_t^A|)] \left(\widehat{f}_t^A + 1\{g_t^A > 0\} g_t^A - \overline{f} \right) \\
&\quad + d(-|g_t^A|) + C_3
\end{aligned}
$$

for functions $y_1(g_t^A), n(g_t^A)$ *and* $d(g_t^A)$ *and a constant* β. *The functions* $y_1(g_t^A), n(g_t^A)$ *and* $d(g_t^A)$ *are integrals and linear combinations of appropriate confluent hypergeometric*

---

[21] *Under some mild restrictions on the allowed parameters discussed in the appendix.*

functions and are given in the appendix. The constants C_1, C_2, C_3 are identical to the ones obtained in Proposition 2.1. The optimal investment rule is given by:

$$i_t = \frac{1}{\chi}(P_K - p) = \frac{1}{\chi}\left(\frac{\overline{f}(1-\widehat{p})}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + 1\{g_t^A > 0\}\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-\left|g_t^A\right|)\right)$$

and the optimal stopping time for each investor $o \in \{A, B\}$ is to resell the asset immediately once $\widehat{f}_t^o < \widehat{f}_t^{\overline{o}}$, where $\overline{o} = A$ if $o = B$ and vice versa.

I organize the discussion of the results in two subsections. I first discuss some properties of the derivative of the equilibrium price w.r.t. K_t (commonly called "marginal q"), i.e:

$$P_K = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + 1\{g_t^A > 0\}\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-\left|g_t^A\right|)$$

and then I discuss some properties of the rents to the adjustment technology, i.e.:

$$C_1\left(\widehat{f}_t^A + 1\{g_t^A > 0\}g_t^A - \overline{f}\right)^2 + [C_2 + n(-\left|g_t^A\right|)]\left(\widehat{f}_t^A + 1\{g_t^A > 0\}g_t^A - \overline{f}\right) + d(-\left|g_t^A\right|) + C_3 \tag{2.17}$$

Some observations about marginal "q"

As might be expected from the introductory example discussed in Section 2.2 investment is unambiguously higher in the presence of speculation. Comparing marginal q in the presence of speculation to the equivalent expression in the presence of homogenous beliefs one observes an extra term, namely:

$$b(g_t^A) = 1\{g_t^A > 0\}\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-\left|g_t^A\right|) \tag{2.18}$$

The term $\beta y_1(-\left|g_t^A\right|)$ is a positive term growing in expectation (instantaneously) at

the rate of interest plus the rate of depreciation.[22] I.e. it is a pure speculative "bubble" that arises endogenously. In contrast to "rational" bubbles that can grow indefinitely, this term is bounded. Moreover, one can determine its magnitude explicitly and speculative bubbles of this sort can exist even in finite horizon settings.[23]

Of course investment is inflated only if it is determined as part of shareholder value maximization. In this case investors are short termist and invest in order to increase resale value. It is interesting to see what would happen if investment only reacted to "long-run" fundamentals. Such a situation can arise if e.g. the company is run by a set of managers / shareholders who do not have frictionless access to the markets for whatever reason. For instance this group of managers / shareholders might be unwilling to sell its shares because it values control, or because it perceives that its shares might have a large non-linear effect on the price of the stock due to asymmetric information, or simply because there are vesting agreements that preclude sales of stock or finally for reasons related to capital gains taxes. If these managers/shareholders are of type A, then investment will continue to be given by (2.14). However, the stocks that are traded in the market will still contain speculative components and thus the link between "marginal q" (P_K) and investment will break down. I use this observation to develop tests in section 2.3.2.

A second observation is that marginal "q" (P_K) is now more volatile than the expression obtained in the case of homogenous beliefs. Applying Ito's Lemma to (2.18) and evaluating this expression at the stationary point ($g_t = 0$) one finds an increase in the

---

[22]Formally, for $g_t^A < 0$ and any $T > t$ this term satisfies

$$y_1\left(g_t^A\right) = E(e^{-(r+\delta)(T \wedge \tau - t)} y_1(g_{T \wedge \tau}^A))$$

where:

$$\tau = \inf\{t : g_t^A \geq 0\}$$

and similarly for g_t^B.

[23]This term is practically identical to the one obtained in Scheinkman and Xiong (2003) with the sole exception that the effective interest rate in the present setup is increased by the rate of depreciation. The reader is refered to that paper for a detailed discussion on the differences between speculative and rational bubbles.

volatility of q_t (compared to the homogenous beliefs case) of[24]:

$$\frac{\sigma_g}{2(r+\delta+\lambda)}$$

In other words the volatility in marginal "q" (P_K) due to the presence of short sale constraints and heterogenous beliefs is increasing in the volatility of the disagreement process and decreasing in the interest rate (r), the rate of depreciation (δ) and the rate of convergence to long run fundamentals (λ).

A third observation concerns predictability. Marginal "q" no longer satisfies the relation:

$$q_t = E^A\left[\int_t^\infty e^{-(r+\delta)(s-t)}\widehat{f}_s^A ds\,|\mathcal{F}_t\right]$$

and more importantly, it will no longer be the case that:

$$q_t = E^A\left[\int_t^{t+\Delta} e^{-(r+\delta)(s-t)}\widehat{f}_s^A ds + e^{-(r+\delta)\Delta}q_{t+\Delta}\,|\mathcal{F}_t\right]$$

In the appendix I show that:

Proposition 2.3 *q_t satisfies the relationship:*

$$q_t \;=\; E^A\left[\int_t^{t+\Delta} e^{-(r+\delta)(s-t)}\widehat{f}_s^A ds + e^{-(r+\delta)\Delta}q_{t+\Delta}\,|\mathcal{F}_t\right] + \tag{2.19}$$

$$E^A\left[\int_t^{t+\Delta} e^{-(r+\delta)(s-t)}\frac{(r+\delta+\rho)}{r+\delta+\lambda}g_s^A 1\{g_s^A > 0\}ds\,|\mathcal{F}_t\right] \tag{2.20}$$

Defining:

$$Z(g_t^A;\sigma_g) = E^A\left[\int_t^{t+\Delta} e^{-(r+\delta)(s-t)}\frac{(r+\delta+\rho)}{r+\delta+\lambda}g_s^A 1\{g_s^A > 0\}ds\,|\mathcal{F}_t\right]$$

---

[24]To derive this, apply Ito's Lemma to the expression $1\{g_t^A > 0\}\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-|g_t^A|)$ keeping terms that multiply the martingale parts. In the appendix I show that $\beta = \frac{1}{2(r+\delta+\lambda)y_1'(0)}$ and so $b(g_t^A)$ is differentiable everywhere, and accordingly there are no terms involving the local time of the process at 0. This in turn is a consequence of smooth pasting.

it can be shown that $Z_g > 0, Z_{\sigma_g} > 0$

These properties of q_t deserve some comment. The first term in (2.19) is the usual expression one obtains for marginal "q" in the traditional infinite horizon setting. It can easily be derived from formula (2.14). The second term (Z) is capturing the fact that returns in a setup with heterogenous beliefs and short sale constraints are predictable. The properties $Z_g > 0, Z_{\sigma_g} > 0$ suggest that this predictability will be strongest when the disagreement process is temporarily high and / or when the volatility in the disagreement process increases.

Some observations on growth options, stock prices and returns

The rents to the adjustment technology present a richer set of interactions between speculation, fundamentals and investment. This is to be expected. The ability to adjust the capital stock becomes more valuable when investment is increased due to speculation. This effect becomes magnified, when one takes into account that the differences in beliefs about fundamentals also affect the value of the adjustment technology. As a result investors speculate not only on the ability of the existing capital stock to generate profits in the future, but also on the ability of the company to leverage its value in the future by further increasing its capital stock. As is demonstrated in the quantitative exercises that follow, the effect of these "growth options" on prices can be large.

Applying Ito's Lemma to (2.17) one can establish the analogs of the results discussed for the case of marginal "q", i.e. excess volatility and predictability. However, it is more interesting to analyze the stock price directly. The following result is proved in the appendix:

34

Proposition 2.4 *The equilibrium price satisfies:*

$$P_t = E^A\left[\int_t^{t+\Delta} e^{-r(s-t)}\left(\widehat{f}_s^A K_s - p i_s - \chi\frac{i_s^2}{2}\right)ds + e^{-r\Delta}P_{t+\Delta}|\mathcal{F}_t\right] + \quad (2.21)$$

$$+E^A\left[\int_t^{t+\Delta} e^{-r(s-t)}\frac{(r+\delta+\rho)}{r+\delta+\lambda}g_s^A K_s 1\{g_s^A > 0\}ds|\mathcal{F}_t\right]$$

$$+E^A\left[\int_t^{t+\Delta} e^{-r(s-t)}\left(\xi(g_s^A) + Cg_s^A\left(\widehat{f}_s^A - \overline{f}\right)\right)1\{g_s^A > 0\}ds|\mathcal{F}_t\right]$$

for an appropriate function $\xi(g_s^A)$ and a constant $C > 0$. Denoting

$$\Xi = E^A\left[\int_t^{t+\Delta} e^{-r(s-t)}\left(\xi(g_s^A) + Cg_s^A\left(\widehat{f}_s^A - \overline{f}\right)\right)1\{g_s^A > 0\}ds|\mathcal{F}_t\right]$$

one can show that $\Xi_f, \Xi_{fg}, \Xi_{f\sigma_g} > 0$.

The first term in (2.21) is the standard recursive relation that connects profits and the price next period to the current price. The second term is the predictability due to marginal "q" that was analyzed in the previous section. The final term is the predictability due to speculation on the value of growth options. Both the second and third terms are positive. The last term is increasing in both g_t^A and $\widehat{f}_t^A$ and moreover the cross partial derivative of the third term with respect to $\widehat{f}_t^A$ and g_t^A is positive. This demonstrates the interaction between beliefs about "fundamentals" i.e. $\left(\widehat{f}_t^A\right)$ and the differences in beliefs $\left(g_t^A\right)$ which arises in the presence of investment. Predictability can be expected to be strongest in the present setup when *both* fundamentals and the divergence in beliefs are large. By contrast in the absence of investment the extent of predictability is independent of fundamentals. This makes it easier to link predictable variation in returns to variables that react to both fundamentals and speculation like the B/M. These issues are analyzed further in a quantification exercise that follows.

Testing if bubbles affect investment

In this section I will use the theory developed previously in order to derive the properties of some tests concerning bubbles and investment. I will derive the implications of the theory for certain standard statistical tests under alternative hypotheses.

The analysis will be focused mostly on marginal q and its relationship to investment. In the presence of bubbles marginal q is given by:

$$P_K = q_t = \left(\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + 1\{g_t^A \geq 0\} g_t^A + \beta y_1(-\left|g_t^A\right|) \right)$$

Where the two theories differ is to what extent investment reacts to P_K or not. According to H_0, q theory is valid even in the presence of speculative premia and accordingly:

$$i_t = \frac{1}{\chi}(q_t - p) = \frac{1}{\chi}(P_K - p) \tag{2.22}$$

According to the alternative

$$i_t = \frac{1}{\chi}(q_t^F - p) \tag{2.23}$$

where $q_t^F = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} < q_t$ captures a rational "long run" valuation of marginal profits. One intuitive and straightforward test of the two theories is the following. Suppose one starts with a firm where beliefs are homogenous so that short sale constraints are initially irrelevant. Then, suppose that differences in beliefs arise so that the short sale constraints lead to the creation of speculative premia on the asset. If the company decides to conform with (2.22) the basic investment- P_K relationship will continue to hold, whereas under (2.23) investment should be overpredicted.

This idea is effectively behind Blanchard Rhee and Summers (1993). They examine whether an investment-q relationship estimated over roughly 90 years tends to produce negative residuals around periods when the stock market is most likely driven by a bubble. Moreover, they test if positive residuals are observed after these bubbles crash.

36

This idea is simple and intuitive. The main identifying assumption behind it, is that the researcher is able to identify periods of time or specific stocks where the prices are more likely to be driven by speculative components and not fundamentals. I employ such an empirical strategy in the next section.

An alternative approach is to take advantage of the excess volatility in P_K in the presence of speculation. As demonstrated in section 2.3.2, P_K is excessively volatile compared to long run fundamental marginal "q" as perceived by agent A. Compared to q_t^F, P_K is more volatile by:

$$\sigma_g \beta y_1'(g_t^A)$$

which-evaluated at $g_t^A = 0$ gives:

$$\frac{\sigma_g}{2(r + \delta + \lambda)}$$

This would introduce classical measurement error in a regression of investment on P_K. Accordingly, for companies whose stock contains speculative components, one should expect a biased downward estimate of q compared to companies without speculation. Actually, one can compute the magnitude of this bias. If one decomposes the variance of P_K into a fundamental and a nonfundamental component then the attenuation bias due to classical measurement error would be equal to:

$$\frac{\frac{\sigma_{q_f}^2}{(r+\delta+\lambda)^2}}{\frac{\sigma_{q_f}^2}{(r+\delta+\lambda)^2} + \left(\frac{\sigma_g}{2(r+\delta+\lambda)}\right)^2} = \frac{1}{1 + \frac{1}{4}\left(\frac{\sigma_g}{\sigma_{q_f}}\right)^2}$$

The attenuation bias increases with the ratio of the volatility in the disagreement process relative to the variability in q_t^F. As one approaches homogenous beliefs this volatility goes to 0 and the attenuation bias disappears.[25]

---

[25]This basic idea is behind a number of papers that blame the poor performance of q theory on excessively volatile stock prices relative to some notion of long run fundamentals. For instance see Bond and Cummins (2001), or the survey of Chirinko (1993). Of course, this attenuation bias is only present

Another straightforward test of the theory is to create some measure of q_t^F and compare its performance in a "horse" race with P_K. Such a method is devised in Abel and Blanchard (1986) and also used in Blanchard Rhee and Summers (1993). I use such an approach in section 2.4.3 as one of the robustness checks that I perform.

A final methodology is based on Euler equations.[26] This approach takes advantage of the predictability introduced into P_K by the relation (2.19). This methodology is very appealing from a theoretical viewpoint because it does not require a lot of assumptions apart from predictability in the variation of marginal "q" which is true for the model discussed. In particular, following essentially the same ideas as in Chirinko (1993) I show in the appendix that the following relationship holds if investment reacts to fundamentals only (irrespective of whether there are nonfundamental components in the price):

$$E\left[I_t - e^{-(r+\delta)}I_{t+1} - \frac{1}{\widetilde{\chi}}\Pi_t + C|\mathcal{F}_{t-1}\right] = 0 \qquad (2.24)$$

I_t denotes the change in the capital stock between $t-1$ and t, Π_t are the observed profits between $t-1$ and t divided by the capital stock, and $\widetilde{\chi}, C$ are constants determined in the appendix. If -by contrast- investment reacts to speculation one can use the predictability of returns derived in the previous sections to show that:

$$E\left[I_t - e^{-(r+\delta)}I_{t+1} - \frac{1}{\widetilde{\chi}}\Pi_t + C|\mathcal{F}_{t-1}\right] \geq 0$$

I give an explicit derivation in the appendix. An interesting implication of the results in section 2.3.2 is that one can make additional statements about the strength of the predictability as a function of the properties of the disagreement process. Moreover, one

---

if companies react only to long run fundamentals.

[26]Similar ideas to the ones discussed here are proposed by Chirinko and Schaller (1996). However in this paper predictability is assumed rather than derived from first principles. This makes it hard to say with confidence whether predictability is due to rational variations in the discount factor. At a theoretical level, without an explanation for why bubbles exist it becomes difficult to connect them to predictability as Chirinko and Schaller (1996) explain in Appendix A of their paper when they study rational bubbles.

can pin down its sign. This gives additional predictions that can be fruitfully used in the cross section.

The appendix to this section also shows how to generalize the findings to arbitrary linear homogenous adjustment technologies and an arbitrary number of investor groups. The main advantage of doing so is that a) marginal and average q become equal, allowing one to obtain a measure of marginal q (P_K) through average q and b) the investment to capital ratio (in contrast to absolute investment) becomes a function of q. Moreover the observations about equation (2.24) continue to hold with $\frac{I_t}{K_{t-1}}$ replacing I_t. However, it seems difficult to obtain closed form solutions for prices in this case.

A basic quantification exercise

In this section I examine the ability of the model to produce quantitatively plausible magnitudes for q and the extent of predictability. The model has a number of parameters that can be classified in two categories: a) parameters that are mainly related to the underlying productive and adjustment technologies, b) parameters that are related to the beliefs of the rational agent and c) parameters that are related to the beliefs of the irrational agent. The main parameter of interest is the degree of disagreement which is controlled by ϕ. Accordingly, the results are reported as a function of ϕ. The rest of the parameters are used in order to produce sensible first and second time series moments of returns, marginal q, average q, the investment to capital ratio and the dividends to price ratio in the absence of speculation. I chose the values $\delta = 0.1$, $r = 0.05$, $\lambda = 0.1$, $\bar{f} = r + \delta$, $\sigma = 0.25\bar{f}$, $\chi = 2$, $p = 0.6$, $\sigma_D = 0.5\sigma$, $\sigma_s = \sigma$.[27]

---

[27]With these parameters I simulated the model to determine prices, investment and capital if all agents are rational and $\phi = 0$. I simulated 80 years of data dropping 10 years in order to enforce that initial values are drawn from the stationary distribution. The results are given in the following table

	B/M	Marg. Q	D/P	Returns
Mean	0.599	0.995	0.038	0.054
Std. Dev.	0.148	0.264	0.021	0.181

To compare, the study of Kothari and Shanken (1997) reports an average for B/M of 0.69 with a standard deviation of 0.22 whereas the dividend yield is given as 0.036 with a standard deviation of 0.014.
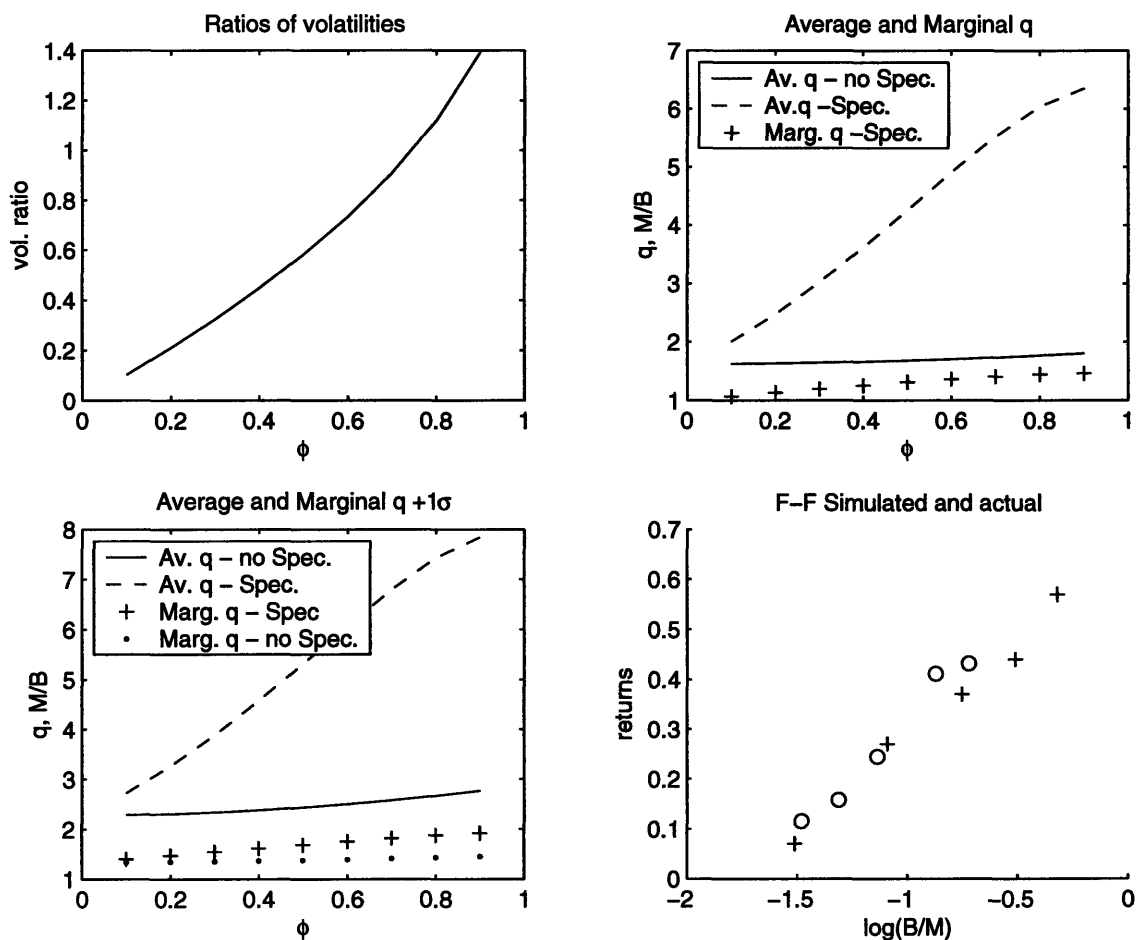
Figure 2-2: Behaviour of various quantities in the model: The top left panel depicts the "disagreement" ratio as a function of ϕ. The numerator of this ratio is the standard deviation of the stationary distribution of the disagreement process g_t. The denominator is the (average) standard deviation of the beliefs of the rational agent. The top right panel depicts average and marginal q in the presence and absence of speculation. In all cases the capital stock is fixed at its stationary value in the absence of speculation. The fundamentals (f) and the disagreement process (g) are evaluated at the stationary values $f = \bar{f}, g = 0$. The bottom left panel repeats the same exercise as the top right panel with the sole exception that f is evaluated at one positive standard deviation above its mean $\bar{f}$. The bottom right figure simulates a sample of 2300 companies over 27 years to have a similar setup as Fama and French (1992). The crosses denote the 5 F-F portfolios with the lowest B/M as reported in p.442 of their paper adjusted for an annual inflation rate of 7.2% between 1963 and 1990. The circles indicate simulated values. For 75% of the companies $\phi = 0$ whereas for the rest $\phi = 0.9$. The rest of the parameters are: $\delta = 0.1$, $r = 0.05$, $\lambda = 0.1$, $\bar{f} = r + \delta$, $\sigma = 0.25\bar{f}$, $\chi = 2$, $p = 0.6$, $\sigma_D = 0.5\sigma$, $\sigma_s = \sigma$

Figure 2-2 depicts various quantities of interest. The top left panel allows one to "translate" levels of ϕ in terms of the disagreement ratio between the rational and the irrational agent. The disagreement ratio is constructed as the ratio of the standard deviation of the stationary distribution of g_t^A to the average standard deviation of the posterior belief distribution of the rational agent. For instance a ratio of 1 means that the standard deviation of the stationary distribution of the disagreement process is equal in magnitude with one standard deviation of the posterior beliefs of the rational agent. [28] The top right panel reports results of the following exercise: fixing the capital stock at its steady state value in the absence of disagreement, I compute average q and marginal q for various level of ϕ. I also report average q in the absence of speculation (i.e. if all agents are of type A) for comparison.[29] All quantities are evaluated at $g = 0, f = \overline{f}$, so that marginal q is equal to 1. As can be observed, the presence of disagreement increases both marginal "q" and average q or "Market to Book". The increase in marginal q is identical to the effect documented in Scheinkman and Xiong (2003) with the sole exception that r is replaced by $r + \delta$. The second effect is the significant increase in average q or market to Book. In this example, if all agents share homogenous rational beliefs, marginal q is 1 and average q is about 1.6. When heterogenous beliefs and speculation enter the picture marginal q is increased mildly (not more than 50 percent) but the rents to the adjustment

---

The average value of marginal q and the average value of returns are predetermined by construction at 1 and 0.05 by the choices of r and $\overline{f}$. The simulations unsurprisingly produce values very close to these parameters. Kothari and Shanken (1997) report an equal weighted return of 15.3 and a value weighted return of 9.4 % with standard deviations of 38.8 and 23.9 % respectively. However these returns are not real returns. Adjusting for an average annual inflation of 3.27% from 1926 to 1990 and taking into account the volatility of inflation produces real returns close to the ones reported in the simulation. Moreover the chosen values imply that regressions of the investment to capital ratio on average q return a value of 0.078 which corresponds to the values that are obtained in the empirical section.

[28] In practical terms this implies that the rational agent will not be able to tell with 95% confidence that the irrational agent is wrong "most" of the time.

[29] One might be puzzled why average q increases with ϕ in the absence of speculation. This is because ϕ - besides controlling disagreement- also tightens the confidence intervals of the rational agent. However, this effect is of second order. Moreover it can be completely avoided if one were to also modify σ_s with ϕ in order to keep the variance of posterior beliefs of the rational agent constant.

technology (or growth options) are increased substantially. [30] [31]. The bottom left panel repeats the above exercise when fundamentals are at one positive standard deviation, i.e. $f = \bar{f} + \sigma^{st}$ where σ^{st} is the stationary standard deviation of f. This picture demonstrates that effects are amplified when fundamentals are strong. In summary, growth options form a non-negligible source of valuations in the presence of speculation. The values produced are in line with the relatively large values of market to book found in the data during speculative episodes.

The bottom right panel reports results on the ability of the model to produce both reasonable book to market ratios and predictability. I simulated paths of 2300 companies over 27 years assuming that all companies are identical, except for ϕ[32].The returns of these companies and the Book to Market ratios were simulated under the assumption that for 75% of the companies there is no disagreement ($\phi = 0$) whereas for 25% agents disagree with $\phi = 0.9$. With these assumptions I calculated equal weighted returns for 10 portfolios formed on Book to market as described in Fama and French (1992). The bottom right panel of figure 2-2 plots the resulting returns and compares them to the results reported in Fama and French (1992)[33]. I focused only on the portfolios with the 5 lowest B/M ratios since this paper is concerned with overpricing. The results suggest that the present model can produce degrees of predictability very similar to the ones observed in the data. Fama-MacBeth regressions produce coefficients of roughly 0.38 compared to 0.5 reported in Fama and French (1992). Moreover, a number of alternative parameter values seem to suggest that one needs to assume that only a small number of companies needs to be overpriced in order to explain the data. However the disagreement in these companies needs to be relatively large.

---

[30]It can be shown that this picture is independent of the level of χ, since I normalize by the steady state capital stock.

[31]Of course as time passes average q will fall because the capital stock will start to increase

[32]Once again a number of initial years (prior to the 37 that form the simulation study) was dropped to make sure that initial capital stocks, fundamentals and disagreement are drown from the stationary distribution.

[33]To compare the results I subtracted a 7.2% annual (or 0.6% monthly) from the results in Fama anf French (1992) in order to compute real returns

42

2.4 Empirical Evidence

2.4.1 Overview

In this section of the paper I use the theoretical results obtained in order to test the most central predictions of the model. The presence of a short sale constraint should increase valuations for the underlying assets, while the behavior of investment will depend on the shareholders' ability and willingness to sell their shares and take advantage of the speculative components in asset prices. The H_0 hypothesis in this section will be that investment reacts to both "long run" fundamental components and short run speculative components. The alternative (H_1) is that it reacts only to the former.

What is difficult, in order to operationalize this notion is to disentangle shocks to fundamental marginal "q" and shocks to the resale premium. If one can identify a negative shock to the resale premium (due for instance to a relaxation of the short sale constraint) then under H_0 the basic "q"-type relation should be able to accurately predict a drop in investment. Under (H_1) there should be no drop in investment and the "q" relationship would falsely predict one. Similarly, during the buildup of speculative components in prices the basic "q" type relationship should overstate the increase in investment under (H_1).

Various studies have used proxies to disentangle fundamental from non-fundamental sources of valuation, such as breadth of ownership, discretionary accruals, equity issuance, etc.[34] A problem with this approach is that most of these indications of mispricing could be explained in an alternative way that is not related to the speculative component of prices. They provide indirect ways of controlling for mispricing.

In this paper I adopt a more direct approach to identifying shocks to the speculative component of stock prices. In particular, I test "q" theory on a set of companies for which data on the existence of a market and the costs to market participants of short

---

[34]See e.g. Polk and Sapienza (2002), Gilchirst et. al. (2002)

selling a company's stock is publicly available.

The study focuses on the 1920's, because short selling was done via a public market, and data on short selling was available in the Wall Street Journal. This data set was collected by Jones and Lamont (2002). I describe this data set in more detail in the next section. The interest will be focused on an episode during the beginning of 1926 when 32 industrial companies were added to what was called "the loan crowd"[35], i.e. a market for borrowing and lending stock. As is explained in Jones and Lamont (2002) the most likely reason for the introduction into the loan crowd was that market participants considered these companies as particularly overpriced compared to their fundamental value.

This introduction can be interpreted as a relaxation of the short sale constraint. Accordingly, in line with the theory developed, one should expect to observe a drop in the stock price of the companies after their introduction into the "loan crowd", independent of whether H_0 or the alternative holds. Results of this nature were established in Jones and Lamont (2002). I reconfirm their results for the subsample that I consider and provide additional evidence concerning the "q" ratio of these companies.

Then I study investment. The drop in the price that is observed for most of the companies in the subsample can be reasonably interpreted as the effect of a correction to "overpricing". I then proceed to compare the behavior of investment for these companies. There are at least two easily testable implications. First, I run standard regressions of the form:

$$\frac{I_{i,t}}{K_{i,t-1}} = \alpha_i + \delta_t + \beta q_{i,t-1} + \varepsilon_{i,t} \tag{2.25}$$

for "control" companies that have been in the loan crowd for some time and the cost of short selling them is low[36]. I compare the results of these regressions to the equivalent regressions for the companies of the "treatment" group. What one should observe under H_0 is that the coefficients of β are the same up to sampling error. Otherwise, the

---

[35]Even though there were additions later on to this list most of them came after the August of 1930, a period where the U.S. enters the great depression.

[36]Companies in the control group are comprised of all companies that were in the loan crowd at least 2 years before 1926 and their rebate rates were at least 2% in February 1926.

44

coefficient β should be biased downward because of the measurement error type problem analyzed in subsection 2.3.2.

Another simple observation is that under (H_1) one should expect the residuals in regressions of the type (2.25) to be negative on average immediately prior to inclusion and significantly positive thereafter. These intuitive and simple implications of the theory are tested in detail in the sections that follow. Sections (2.4.3) and (2.4.4) run the tests described in section 2.3.2 to check if the long-termist hypothesis H_1 can be rejected.

2.4.2 Data

The data for the empirical study come from various sources. Data for the loan crowd market are from Jones and Lamont (2002). [37] They collected data from end of the month *Wall Street Journals (WSJ)*. The data collected provide information on rebate rates from 1919-1933. The list of companies that were on the WSJ list was very small in 1919 (less than 20 industrial companies) and expanded in 4 waves described in Jones and Lamont (2002). The first wave occurred in 1926 when 32 industrial companies were added to the list along with a number of railroad companies that I ignore in this study. The other waves came after August 1930, a period during which the U.S. economy was going into a deep recession.

The *Wall Street Journal* reports the names of the companies along with the so-called rebate rates. The difference between a rebate rate and the prevailing interest rate is the cost of short selling. This is illustrated by an example given in Jones and Lamont (2002): suppose A lends shares to B and B sells the stock short. When the sale is made the proceeds go to A and not to B. A is effectively using collateral to borrow and thus must pay interest to B. At the end of the loan A repays cash to B and B returns the shares to A. The rate of interest received by B is called the rebate rate or "loan" rate. Accordingly, stocks with 0 rebate rates are the most expensive to short whereas stocks with positive and high rebate rates are relatively inexpensive to short. In other words, the rebate rate

---

[37] For details of this data set the reader is refered to that paper.

is the price that brings the loan market back to equilibrium.

To form a control group for the study I selected only companies that were in the loan crowd before 1926 and were trading at rebate rates above 2% in February of 1926. This yielded 15 companies that form the control group. For these companies I assume that short sales were possible and relatively inexpensive since 1919.

The "treatment" group is comprised of companies that enter the loan crowd from January 1926 to June 1926 with the vast majority entering at the end of February. There are 32 industrial companies that meet these criteria[38]. Virtually all of these companies enter at a rebate rate of 0 which captures either very high shorting demand or limited supply of short selling. Conceivably, it also captures conservatism with the creation of a new market. In either case I adopt the interpretation given in Jones and Lamont (2002) namely, that these are stocks that were considered as particularly overvalued and thus the demand for shorting the stock exceeded the amount the normal broker could accommodate "in house"[39].

To produce measurements of "q" I could not rely on standard sources of data like COMPUSTAT, since there is no widely available, electronic source of balance sheet data going back to 1918. Accordingly, data was hand-collected from *Moody's Manuals of Investments* for the years 1918-29.[40] A particularly difficult problem with balance sheet data from the 20's is that companies did not have to comply with any particular form

---

[38]Some of the companies were dropped for one of the following reasons: a) Data could not be found for at least 3 years prior to February 1926 b) the fiscal year ended more than 3 months before or after December 31. c) the company was a pure holding company d) there was an important merger e) Most of the company's balance sheet was undepreciated goodwill. With these selection criteria I tried to address issues related to IPO's, non-synchronous data, issues related to governance and measurement error in q. In contrast to common practice I did not winsorize the data in any way, (i.e. by truncating q) because it is precisely the large variations in q caused by speculation that form the object of this study. The final sample consisted of 25 companies. For 3 of them I was able to construct q but could not find profit data for some of the years 1922-26. To safeguard that the results do not capture IPO related issues, I ran all of the main regressions on the subset of companies that I had data reaching back to at least 1918. The results were unaltered.

[39]I.e. by using the accounts of one customer who is long the stock to lend it to another who wants to short sell.

[40]I am indebted to Tom Nicholas for providing a data set that contained balance sheet data on some of the companies investigated.

of data reporting. Especially detailed profit and loss data are typically unavailable. An additional problem is that most companies did not start reporting depreciation and accumulated depreciation reserves until 1926. This introduces measurement error in the investment data which -fortunately- is the left hand side variable. To create the time series for "q" I used the same procedure as Nicholas (2003). This procedure is basically the standard Lindenberg and Ross (1981) procedure adapted to the typical balance sheet data of the 1920's. q is computed as the product of common shares outstanding times the price of common shares plus the market value of preferred stock[41] plus the (book) value of debt. The replacement cost of capital is determined by the usual Lindenberg and Ross (1981) type recursion:

$$k_{i,t}^{rc} = k_{i,t-1}^{rc} \left[\frac{1 + \dot{p}_t}{(1 + \rho)(1 + \delta)} \right] + \left(NAV_{i,t}^{BV} - NAV_{i,t-1}^{BV} \right)$$

where NAV_t^{BV} is the net asset value of physical capital (Plant, Equipment, and Property).[42] This was the only variable related to physical capital consistently provided for all companies. $\dot{p}_t$ is the inflation rate obtained from the *Historical Statistics of the United States: 1790-1950*. ρ and δ are the rate of technological obsolescence and the depreciation rate respectively and were set to 0. This choice was dictated by the fact that NAV_t^{BV} already includes depreciation. The inventories were computed at book value whereas liquid assets were computed as the difference of total assets and the sum of the book value of plant, equipment and property and inventories.

Investment was hard to compute accurately for many of the firms under consideration.

---

[41] I follow Tom Nicholas (2003) here and determine the market value of preferred stock as if it were a perpetuity discounted with Moody's Average yield. This approach is dictated by data availability. To check if this introduces any significant measurement error, I looked at the price of preferred stock for a few companies that I could find data on preferred stock and computed q with actual prices for preferred stock. The estimate of q was practically unaffected.

[42] The algorithm was initialized with $k_{1918}^{rc} = NAV_{1918}$, or setting NAV equal to the first available observation year if data could not be found for 1918.

By a basic accounting identity it is the case that:

$$I_{i,t} = D_{i,t} + NAV_{i,t}^{BV} - NAV_{i,t-1}^{BV}$$

where $D_{i,t}$ is the accounting depreciation of the assets during year t and $I_{i,t}$ is gross investment. The above relationship can be rewritten as

$$\frac{I_{i,t}}{NAV_{i,t-1}^{BV}} = \frac{D_{i,t}}{NAV_{i,t-1}^{BV}} + \frac{NAV_{i,t}^{BV} - NAV_{i,t-1}^{BV}}{NAV_{i,t-1}^{BV}}$$

As long as $\frac{D_{i,t}}{NAV_{t-1}^{BV}}$ is given as a company specific constant plus some error that is orthogonal to q, i.e.

$$\frac{D_{i,t}}{NAV_{t-1}^{BV}} = c + \varepsilon_{i,t}, \quad E\left(\varepsilon_{i,t} | q_{1...T}\right) = 0$$

then this induces classical measurement error. However, investment is the left hand side variable so that consistency of the estimated parameters is not affected, only their confidence intervals.[43]

For some regressions a variable that I label profits is also used. This variable refers to accounting profits after interest and depreciation, $\Pi_{i,t}$, that were reported consistently for most companies. Unfortunately, cash flow variables could not be constructed because depreciation was not reported for most companies. The variable that I call profit rate is defined as $\pi_{i,t} = \frac{\Pi_{i,t}}{K_{i,t-1}}$. [44]

Stock price and capitalization data were obtained from *CRSP* for the months following December of 1925 whereas the *Commercial and Financial Chronicle* was used for stock price data prior to December 1925.

---

[43] To check the influence of measurement error on the results, I ran the investment regressions on a subset of companies where depreciation rates were available and so I could compute investment accurately. The results were practically identical, suggesting that the measurement error is indeed classical, i.e. orthogonal to the regressors.

[44] Unfortunately, separate sales and cost data were not reported for most companies and as a result I cannot address effects of imperfect competition in the usual way that this is done in the literature. Measurement error in the profit rate is partially taken care of in the section on Euler equations by estimating everything with instruments and allowing for fixed effects.

Control Group

	Obs	Mean	S.D.	5%	10%	25%	50%	75%	90%	95%
investment	152	0.076	0.221	-0.077	-0.046	-0.007	0.024	0.078	0.213	0.434
q	167	1.075	1.197	0.377	0.436	0.567	0.683	0.914	2.212	3.875
profits	150	0.207	0.335	-0.030	0.020	0.045	0.092	0.197	0.599	0.858

Treatment Group

	Obs	Mean	S.D.	5%	10%	25%	50%	75%	90%	95%
investment	228	0.080	0.213	-0.078	-0.053	-0.013	0.022	0.116	0.264	0.535
q	254	1.250	0.914	0.382	0.477	0.708	1.012	1.499	2.373	2.839
profits	219	0.392	0.585	-0.010	0.022	0.093	0.195	0.455	0.907	1.811

Table 2.1: Table of Summary Statistics sorted by treatment and control group. 5%, 25% etc. correspond to the respective quantiles of the distribution.

2.4.3 Results

Summary Statistics

Table 1 gives some summary statistics of the data. The profit rates of the companies have different distributional properties. Companies in the control group have relatively less dispersed profit rates with a lower mean than the companies in the treatment group. The companies in the treatment group also have a higher and more volatile q compared to the ones in the control group. Both of these observations conform well with the setup of the theoretical model: one would expect a higher variability in the profit rate to leave room for diverging opinions and accordingly cause average q to be more volatile. At first glance there are no obvious differences in the distribution of the investment to capital ratio.

The companies under consideration are relatively large. Companies in the control group belong to the two highest capitalization deciles of CRSP, whereas companies in the treatment group are slightly smaller with the median company in the 7th CRSP capitalization decile.

	(1)	(2)	(3)
	V-weight	Eq-weight	$R - R_{size}$
NEWQ(1)	-0.039	-0.023	-0.032
	(0.014)	(0.013)	(0.012)
NEWQ(2)	-0.023	-0.015	-0.014
	(0.006)	(0.007)	(0.007)
Observations	1147	1147	1147

Table 2.2: Monthly abnormal returns for companies in the treatment group. NEWQ(1) and NEWQ(2) are dummies that take the value 1 if the return is observed in the first quarter of introduction to the loan crowd and 0 otherwise. NEWQ(2) is defined similarly for the second quarter. A separate beta type model is estimated for each stock with 32 monthly returns. The first column contains results when the index is taken to be the Value weighted CRSP index and the second column contains results for the Equal weighted CRSP index. The third column matches stocks by CRSP capitalization decile and contains results from a regression of this difference on a constant and the two dummies described above. The standard errors are computed with a heteroskedasticity robust covariance matrix that allows for clustering by month.

The behavior of q and excess returns

A central prediction of the theory developed earlier is that the presence of short sale constraints will lead to "overpricing" (irrespective of whether investment reacts to it or not).

Figure 2-3 gives a visual impression of such an effect. It depicts the average first difference in "q" year by year for companies in the treatment and the control group. The average first difference in q is identical for both types of companies until 1923. From that point on, companies in the treatment group start having large positive first differences compared to companies in the control group. 1926 presents a structural break. Companies in the treatment group have a large negative adjustment. To the contrary, companies in the control group have a positive first difference in q in 1926. The difference between groups of the yearly difference in q is -0.314 with a standard error of 0.138 for the year 1926. This drop is large both in economic and statistical terms.

Diff.Q–Treatment and control Group
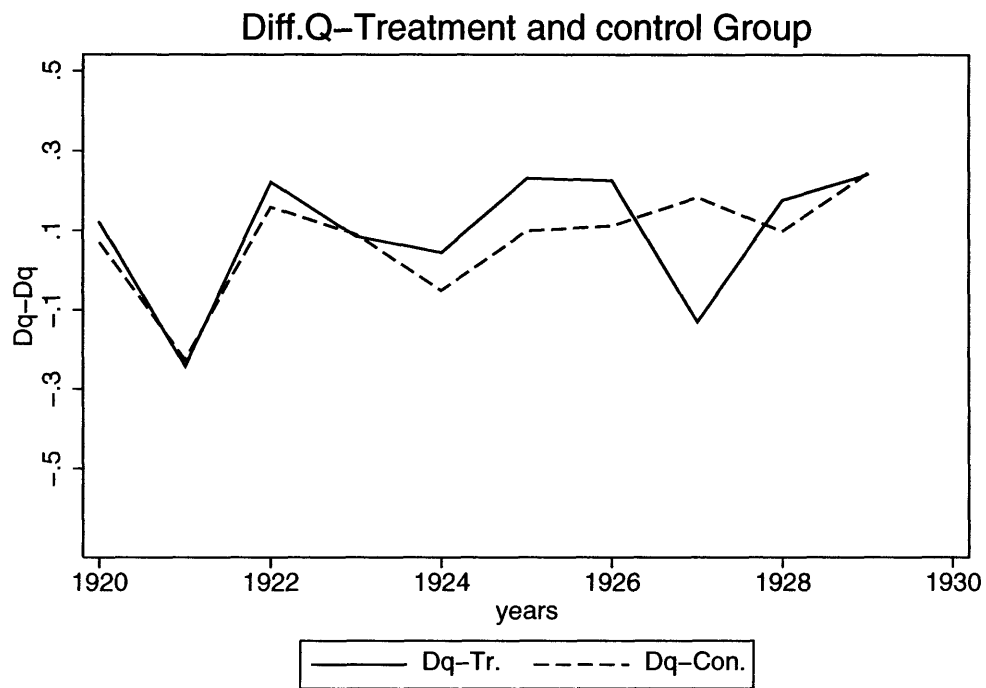
Figure 2-3: Plot of average first differences in "q" for companies in the treatment and the control group. The solid line denotes companies in the treatment group and the dashed line denotes companies in the control group. q is evaluated at the beginning of each period.

51

Excess returns can help in testing the overpricing hypothesis statistically. The usual case-study methodology of studying the excess returns of stocks around a particular "event" presents one major difficulty. First and most importantly, CRSP starts in January of 1926, so that one cannot run regressions to determine the "betas" of the stocks on the market before their introduction and I am forced to estimate these "betas" from subsequent observations. Table 2.2 presents regression results for the model:

$$R_{it} - r_t = a_i + \beta_i(R_{Mt} - r_t) + \gamma 1\{NEQ1\} + \delta 1 \{NEWQ2\} + \varepsilon_{it}$$

where $(R_{Mt} - r_t)$ is the (excess) return on a market wide index, $R_{it} - r_t$ is the excess return of security i at time t and the dummy variable $1\{NEQ1\}$ is 1 if the observation belongs to the first quarter in which the stock has been introduced into the market, and $1\{NEWQ2\}$ if the observation belongs to the second quarter. In other words these dummies are capturing the average abnormal return in the months following the introduction of the stocks into the "loan crowd". Columns (1) and (2) show an economically very significant drop in the holding period return in the 2 quarters following the introduction. In column (1) I use the CRSP Value weighted index in order to control for market-wide effects whereas in column (2) I use the equal weighted index. After a stock is introduced into the loan crowd an average –3.9% (monthly) abnormal return can be expected in the first quarter and a -2.3% in the subsequent quarter. To make sure that this is not just a size-related effect column (3) matches the returns of the companies in the sample with the portfolio returns of the CRSP capitalization decile in which they belong. In other words, I construct $R_{it} - R_{Cap_{it}}$ and regress this magnitude on a constant and the dummies described above. In all cases the results are very similar varying only in the strength of the effect.

Investment and q

This section studies the relationship between investment and q. Figure 2-4 depicts the comovement between average first differences in (beginning of period) q and investment for companies in both the treatment and the control group. The only thing that can be said is that the link between investment and q is not apparently different in any way between the two groups. Investment seems to follow both the upturns and downturns of q for companies in the treatment group. Moreover q co-moves with investment even during periods where one would suspect that the stock prices are driven primarily by non-fundamental forces.

Tables 2.3 and 2.4 present some formal econometric tests. Table 2.3 shows results of simple regressions of investment on "q" for various subgroups. Column (1) estimates a regression of investment on beginning of period q allowing for an individual fixed effect and a time fixed effect. The first column runs this regression on all the data in the sample whereas the second column restricts attention to companies in the control group. The third and fourth columns run the same regressions on companies in the treatment group pre and post 1926. The first two rows correspond to different methods of removing individual fixed effects. The first row eliminates individual effects by estimating them out (fixed effects regression) while the second row eliminates fixed effects by first differencing. The third row estimates a fixed effects median regression.

The fixed effects and first differences estimator produce similar results for all the subgroups suggesting that measurement error in q[45] (due to e.g. mismeasurement of the replacement cost of capital) is not very important.[46] [47] The standard errors are wide since

---

[45]See e.g. the results in Grilliches and Hausman (1985) on measurement error in panel data.

[46]Moreover it suggests that the errors satisfy a strict exogeneity condition $(E(\varepsilon_{it}|q_{i1...T}) = 0)$ not just a sequential exogeneity assumption $(E(\varepsilon_{it}|q_{i1...t-1}) = 0)$. I tested for this directly by including one lead of q in the fixed effects specification. The coefficient was both economically and statistically insignificant. This suggests an interpretation of the errors in the investment regression as adjustment cost shocks. See e.g. Chirinko (1993) and Chirinko and Schaller (1996).

[47]One caveat is in order. If there is correlation in the measurement error then first differences and fixed effects could be producing the same answer even though measurement error is present and as a

Q-Investment Diagram: Treatment

Q-Investment Diagram: Control
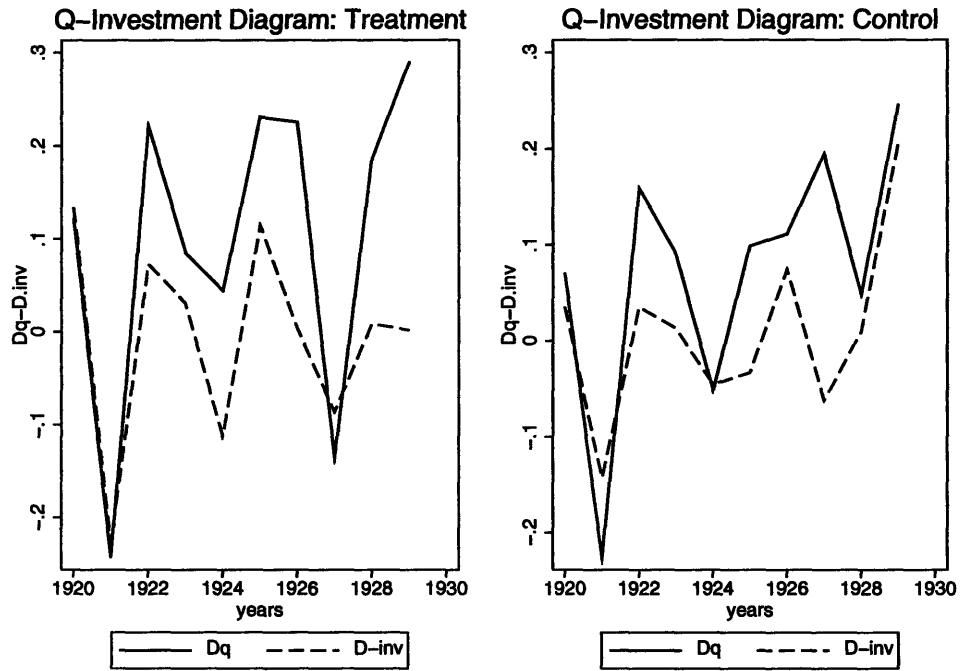
Figure 2-4: The left panel is a plot of average first differences in investment and q for companies in the treatment group. The solid line denotes average first differences in q and the dashed line average first differences in investment. The right panel depicts the same magnitudes for companies in the control group. q is evaluated at the beginning of the period.

the amount of data is very limited. q is significant in the first differences specification if one includes the entire set of companies and is also significant (for both specifications) for the companies in the treatment group prior to 1926. The point estimates are somewhat surprising. They are substantially larger than in the usual Compustat sample.[48] One potential explanation for this is that most companies are large, industrial stocks so that problems related to financial constraints, intangibles etc. become less prevalent. The point estimates are very large for companies in the treatment group prior to 1926. This suggests that companies do not distinguish between the sources of variation in q. Else, the estimates on q in this regression should be downward biased.

Row (3) in Table 2.3 and Figure 2-5 demonstrate some distributional properties of the error term. Row (3) estimates median regressions for subgroups. The estimates are roughly comparable for all subgroups, suggesting that the large estimates of "q" in the fixed effects (or the first differences) specification for companies in the treatment group are driven by a skewed error distribution. This is confirmed by a look at Figure 2-5 which plots residuals of the fixed effects regression for the two subgroups. This picture reveals two patterns. First the median residual is roughly the same for companies in the treatment and the control group. Second, the distribution of the error term is shifted to the right for companies in the treatment group for the years 1924 and 1925. This suggests that a number of companies adjusted to market based "q" and possibly in a non-linear way, not captured completely by the simple linear q model.

Table 2.4 contains results on interactions of q with year and treatment effects. Under H_0 one should expect all columns to not be significantly different than 0. Under the alternative the first two columns should be significantly negative. No matter how they are estimated, the interaction effects in the first two columns are positive, suggesting that

result coefficients are downward biased. To address this I also estimated the adjustment cost parameter using Euler relations in section 2.4.4 which produced similar results to the ones reported here.

[48] Abel and Eberly (2003) give estimates of 0.03 and 0.02 for the fixed effects and the first difference regression. The highest estimates for the linear model are produced by using analyst forecasts as instruments. The number they obtain for this specification is 0.11 very close to the numbers reported here.

one cannot reject the base hypothesis that investment reacts to both fundamental and nonfundamental sources of "q". If instead of interactions of q with treatment and year dummies one uses simple interactions of year and treatment dummies the coefficients prior to 1926 remain positive and become negative thereafter for companies in the treatment group, which again supports H_0.

I also estimated the model on companies in the treatment group that could be characterized as representative of the "high-tech" companies of the time[49] (mostly automobile related companies). The motivation behind this estimation is simple: there is increased uncertainty (and hence room for disagreement) about the fundamentals of companies in emerging sectors making short selling constraints more relevant and overvaluation more likely. In addition the automobile sector of the time was characterized as "speculative" by most financial publications[50]. Accordingly, under H_1 this should be the sector in which one would expect to see a heavily downward biased q. Running investment-q regressions in first differences in this subset confirms the previous findings since the estimate on q remains at 0.19, well above the estimate for the control group.[51]

Fundamental q, profits and investment

In this section and the next I run some robustness checks. In particular, I investigate whether the alternative hypothesis (H_1) can be rejected, (namely that the results obtained are attributable to profits and "fundamental" q). By constructing a measure for fundamental q one can also indirectly test the identifying assumption of the previous section, namely that most of the variation in q for companies in the treatment group comes from non-fundamental sources. Ideally, one would like to obtain some measure

---

[49] I chose American Brake Shoe and Foundry, Simmons Co., Nash Motor Cars, Hudson Motor Cars, Mack Trucks and American Locomotive as a sample of companies that were active in the emerging industries of the time.

[50] (such as the Standard Trade Statistics, a predecessor of S&P)

[51] Moreover, to safeguard that the results on companies in the treatment group do not capture phenomena related to IPO's I ran the regressions on the subset of companies in the treatment group for which I could find stock prices in the Commercial and Financial Chronicle at least back to 1919. The coefficients on q were roughly equal to the ones reported for all companies in the treatment group.

	(1)	(2)	(3)	(4)
	All	Control	Tr. -pre 26	Tr. - post 26
q-FE	0.084	0.073	0.335	0.094
	(0.047)	(0.084)	(0.161)	(0.109)
q-FD	0.112	0.1	0.322	0.077
	(0.050)	(0.133)	(0.152)	(0.064)
q-Med.	0.06	0.094	0.076	0.03
	(0.015)	(0.060)	(0.122)	(0.119)
Observations	359	142	132	60

Table 2.3: Results of regressions of investment on beginning of period q. Time and individual fixed effects are included but not reported. The first line contains the results of the fixed effects regression, whereas the second line eliminates fixed effects by first differencing. The last line is a median regression with fixed effects. The colums correspond to the subgroups. The first group includes all companies, the second only companies in the control group. The third and fourth columns report results for the treatment group pre 1926 and post 1926. Standard errors for the fixed effects and the first differences are computed with a robust covariance matrix allowing for clustering by company. For the median regression standard errors, a bootstrap procedure is used.

	(1)	(2)	(3)
	Pre-1926	24-25	27-28
FE	0.097	0.094	-0.018
	(0.047)	(0.051)	(0.034)
FD	0.157	0.06	0.023
	(0.070)	(0.042)	(0.045)
Med.	0.004	0	-0.016
	(0.024)	(0.020)	(0.022)
Observations	359	359	359

Table 2.4: Results of regressions of investment on beginning of period q and various interaction terms. Time and individual fixed effects are included but not reported. The first column reports results on an interaction dummy that is equal to q if the company is in the treatment group and the year of observation is prior to 1926. The second and third colums are defined similarly. The first line reports results of the fixed effects regression, whereas the second line eliminates fixed effects by first differencing. The last line is a median regression with fixed effects. Standard errors for the fixed effects and the first differences are computed with a robust covariance matrix allowing for clustering by company. For the median regression standard errors, a bootstrap procedure is used.
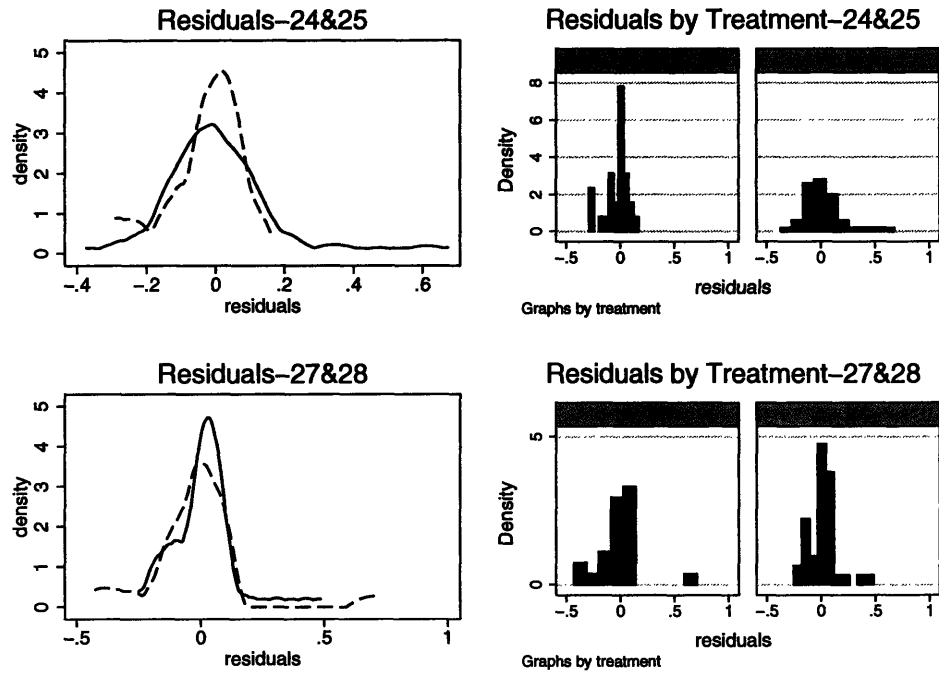
Figure 2-5: This picture presents the residuals of the (fixed effects) investment on beginning of period q regressions for the years 24 and 25 (top panel) and 27 and 28 (bottom panel). The left figure on each panel is the (kernel-smoothed) density of the residuals whereas the right figure is a histogram (10 bins) of the residuals. The solid line in the left figures corresponds to the residuals for the treatment group whereas the dashed line depicts residuals for the control group. Similarly a 0 in the right figures denotes residuals in the control group and a 1 denotes residuals in the treatment group.

of fundamental q from analysts' forecasts on company profitability. Bond and Cummins (2001) propose such a method based on I/B/E/S forecasts. Similar data are unfortunately not available for the 1920's. Accordingly, I will use a "brute" force approach to create fundamental q from reported profits.

In particular I use the methodology in Abel and Blanchard (1986) to determine fundamental q for each company. I run a 2x2 first order VAR of company profits and q on lagged company profits and q for the entire sample, assuming that the coefficients are the same for all the companies in the sample.[52] I then use the estimated coefficient matrix along with a linear approximation to the infinite horizon expression for marginal q to construct a new measure of fundamental q[53]. I used a separate discount factor for each company. To determine the weighted cost of capital for each company (from the perspective of a long termist investor) I used the CAPM in conjunction with the betas estimated in section 2.4.3 and then created a weighted cost of capital by using an interest rate of 4% for the debt of the company an interest rate of 7% for prefered stock and the remaining share of the capital structure I weighted at the cost of equity implied by the CAPM assuming a market wide expected return for common stock of 10%. Depreciation was taken to be 9%.[54]

Roughly speaking this new measure of q is meant to operationalize the notion that fundamental q is the expected sum of discounted marginal profits which are (roughly) equal to $\pi_s = \frac{\Pi_s}{K_s}$ for linear homogenous technologies. In order to create expectations for

---

[52]Icapture the presence of individual heterogeneity by including a fixed effect in each regression of the VAR. This creates a difficult estimation problem, known in the literature as the dynamic panel data problem. The problem arises because the time series dimension is very short in order to invoke standard asymptotic theory. Thus the estimates of the intercepts will be biased. I estimated the coefficients of the VAR with both standard fixed effects and the Arellano and Bond methodology. Even though the coefficients produced by the VAR were somewhat different, in both cases they led to the same conclusions about the role of fundamental "q". In this section I concentrate on the results for the fixed effects regression.

[53]For details of this procedure see Abel and Blanchard (1986)

[54]I also used a flat discount factor of 0.84 for all companies and varied the required return on the market between 7 and 12%. The results were almost identical to the ones reported here for variable discount factors suggesting that the results are not very sensitive to the specific assumptions one makes about returns etc.

the future profit rates, one uses a predictive VAR approach. Then the dynamics of the process are used to create "long run" expectations.

An obvious concern with such a procedure is its accuracy. In particular one could be worried that the estimate of fundamental q obtained in this way would be contaminated by severe measurement error which might make it very difficult to test any hypothesis of interest. A check for this is provided by running a simple regression of first differences in market based q on first differences of fundamental q for the different subgroups of companies. At the very least, one would expect actual q and the constructed measure of fundamental q to co-move closely for companies in the control group. Similarly one would expect the two measures to show disparities for companies in the treatment group prior to 1926. The results of this regression are given in columns (1)-(2) of table 2.5. Roughly half of the variation in q ($R^2 = 0.43$) is captured by the constructed measure of fundamental q for companies in the control group. The performance of this regression for companies in the treatment group prior to 1926 is -as expected- worse ($R^2 = 0.08$), suggesting that q is driven mostly by non-fundamental sources.

The next 4 columns of table 2.5 present horse races between fundamental "q" and market based q. Column (3) presents results for the treatment group prior to 1926. Time effects are included, but not reported. Individual fixed effects are eliminated by estimating all equations in first differences. The estimate for market based q is practically the same as that in Table 2.3, and the estimates on fundamental q are statistically insignificant. Column (4) runs the same regression with lagged profit rates instead of the constructed measure of fundamental q. The motivation for this regression is the following: if one assumes that profit rates follow a first order AR(1), then fundamental q would be just a scalar multiple of the lagged profit rate. Under H_1 this should be the only significant variable. Once again column (4) shows that H_0 cannot be rejected for companies in the treatment group whereas H_1 can. In fact, if one dropped time fixed effects (a Wald test confirms that they are jointly insignificant), then the coefficient on market based q becomes highly significant, whereas the coefficient on fundamental

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dq-Tr.-pre 26	Dq-Con.	Inv-Tr.-pre 26		Inv-Tr.-pre 26		Inv.-All
$\overline{Dq}_f$	0.39	2.54	0.184		0.13		0.221
	(0.24)	(0.57)	(0.185)		(0.17)		(0.14)
Dq			0.302	0.31	0.457	0.467	
			(0.171)	(0.170)	(0.172)	(0.174)	
Lag Dprofs				0.085		0.060	
				(0.089)		(0.080)	
Inter25*Treat							-.167
							(.128)
Inter25*Treat*Dq							.247
							(.117)
Observations	79	112	79	79	79	79	273
R-squared	0.08	0.43	0.29	0.29	0.25	0.25	0.15

Table 2.5: This table presents results on the relationship between "fundamental" q (q^F), market based q, and investment. The first two colums present results of a regression of first differences in actual "q" on first differences in fundamental "q". Column (1) presents these results for the treatment group prior to 1926 whereas column (2) presents these results for companies belonging to the control group. The next four columns present regressions of investment on actual, fundamental q, and the (lagged) profit rate. Columns (3)-(4) present these results for the treatment group prior to 1926 and columns (5)-(6) present the same results if one drops the (jointly insignificant) time effects. Column (7) introduces interactions between the 1925 time effect and treatment (Inter25*Treat) and interactions between the Treatment, the 1925 dummy, and beginning of period q (Inter25*Treat*Dq). The F-test that these variables are jointly 0 rejects at the 0.022 level. Robust standard errors are reported.

q and lagged profits remain insignificant. These results are reported in columns (5) and (6). Column (7) runs a regression with fundamental q, individual fixed effects and time effects for all observations and includes an interaction between a treatment dummy, the 1925 date effect, and market based q. Under H_1 this coefficient should be insignificant. However, the coefficient on the interaction is significant, suggesting that H_1 can be rejected.

2.4.4 Euler Equations

I conclude with some Euler tests. This is an alternative robustness check, with the advantage of not requiring an estimate of fundamental q. The test in this section can be motivated by the discussion in section 2.3.2. In particular I will focus on testing the overidentifying restrictions embodied in the Euler relations discussed in 2.3.2. One disadvantage of this test is that its power is likely to be very small. The reason is intuitive. This test can only detect violations of the overidentification restrictions if predictability is strong, adjustment costs are small and the rest of the errors in the investment equation (sometimes called adjustment cost shocks) are relatively unimportant. To increase the power of the test, I will accordingly focus only on investment behavior of treatment companies around 1926[55].

Table 2.6 presents results on various Euler relationships. Columns (1) and (2) estimate simple Euler relations of the form

$$q_{i,t} = E\left[e^{-(r+\delta)}\left(\pi_{i,t} + q_{i,t+1}\right)|\mathcal{F}_t\right] \tag{2.26}$$

for companies in the treatment and the control group respectively. Instruments include (beginning of period) and lagged q, along with lagged and twice lagged investment to capital ratio and profit rate.[56] Even though the point estimates are similar and economically plausible, the test of overidentifying restrictions cannot reject for the case of control companies whereas it can reject for companies in the treatment group, suggest-

---

[55]The increase in the power of the test comes from the fact that investment by a long-termist manager should not have reacted to the large fluctuations in the price during that period. However, if investment is short termist then one should be able to reject H_1 more easily precisely because of the large fluctuations in the price around this period.

[56]To account for risk premia, I also regressed $\frac{\pi_{i,t}+q_{i,t+1}}{q_{i,t}}$ on the a beta estimated separately for each company (on post 1926 CRSP data) and a constant in a Fama-Macbeth fashion. Then I included variables like q that were in the information set of the agents at time t and checked if they are jointly signifant in the usual Fama and French (1992) fashion. Variables at time t turned out to be significant for companies in the treatment group even after adjusting for a company specific beta.

ing the presence of predictability for companies in the treatment group.[57] Column (3) presents results from estimating the adjustment cost parameter from the Euler equation

$$E \left[\frac{I_{i,t}}{K_{i,t-1}} - e^{-(r+\delta)} \frac{I_{i,t+1}}{K_{i,t}} - \left(\alpha_i (1 - e^{-(r+\delta)}) + \zeta_t - e^{-(r+\delta)} \zeta_{t+1} + \frac{1}{\chi} \pi_{i,t} \right) | \mathcal{F}_{t-1} \right] = 0$$
(2.27)

on companies in the control group in two steps. First I estimate $e^{-(r+\delta)}$ from equation (2.26). Then I substitute into (2.27) and take first differences to eliminate individual fixed effects. I use twice and three times lagged q, investment/capital ratios and profit rates as instruments and adjust the standard errors for the first step estimation error. This should be viewed as yet another robustness check of the results presented in section 2.4.3. The point estimate obtained is slightly larger than the ones obtained in table 2.3 and so are the standard errors, reflecting the fact that instrumental variables are used instead of OLS to estimate this parameter. More importantly, the overidentifying restrictions cannot be rejected. These results are to be expected. If companies in the control group are not overpriced then the regressions in 2.4.3 and (2.27) present just two alternative ways of estimating the adjustment cost parameters as shown in Chirinko (1993). Interestingly, if mispricing exists, does not affect investment and the investment decision is made in a rational way, then (2.27) should continue to hold. In the appendix I show how to construct a test based on this observation. I construct a new variable y as a linear combination of differences in investment and differences in the profit rate for companies in the treatment group as follows:

$$y \overset{d}{=} \Delta \left(\frac{I_{i,1926}}{K_{i,1925}} \right) - e^{(r+\delta)} \Delta \left(\frac{I_{i,1925}}{K_{i,1924}} \right) + \frac{1}{\chi} e^{(r+\delta)} (\Delta \pi_{i,1925})$$

This linear combination depends on parameters that can be consistently estimated from

---

[57]It is unlikely that the rejection is driven by other sources of misspecification (e.g. non-linear homogenous technologies, misspecification of π_t etc) because in that case the test would reject for both the control and the treatment group.

control group observations using (2.26) and from simple regressions of investment on q like the ones performed in section 2.4.3. If (2.26) holds for companies in the treatment group, β should be 0 in the following regression of y on q_{1924} and a constant under H_1 :

$$y \stackrel{d}{=} \Delta \left(\frac{I_{i,1926}}{K_{i,1925}} \right) - e^{(r+\delta)} \left(\frac{I_{i,1925}}{K_{i,1924}} \right) + \frac{1}{\chi} e^{(r+\delta)} (\Delta \pi_{i,1925}) = \beta q_{i,1924} + \zeta + \varepsilon_{i,1926} - \varepsilon_{i,1925}$$

(2.28)

Column (4) presents results on the parameter β in (2.28) estimated on companies in the treatment group. Standard errors are adjusted for two step estimation. The test rejects H_1 since β is significant.

In conclusion, no matter how one runs the test of H_0 vs. H_1 there seems to be evidence that the companies in the treatment group did react to market based q. This seems to be true despite the fact that a significant fraction of the variation in market based q seems to have been driven by non-fundamental sources.

2.5 Conclusion

This paper addressed the question of whether investment should be expected to react to "speculative" components in stock prices. The answer obtained in the theoretical section of the paper is affirmative. In the presence of short sales restrictions and heterogenous beliefs, investors can gain by either holding the asset and reaping its dividends, or by reselling it. From an individual point of view, both are sources of value. If one further assumes that the purpose of a company is to maximize shareholder value, then the conclusion that investment will react to both fundamental and speculative sources of value follows. One can however think of situations where this short-term reasoning is no longer optimal. Indeed, investors with holding horizons that are sufficiently long might choose to disregard speculation altogether.

64

	(1)	(2)	(3)	(4)
	q+profits	q+profits	Euler Est.	Euler-Test 1926
Lagged q	1.19	1.162		
	(0.065)	(0.049)		
Profits			0.192	
			(0.600)	
q-Lag 2				-0.779
				(0.299)
Observations	184	125	82	22
Overid P-val:	0.01	0.43	0.71	
R-squared				0.27

Table 2.6: Euler Equation Tests and Tests of Overidentifying Restrictions. The first and second columns test the overidentifying restrictions embodied in (2.26) for the treatment and the control group respectively. The instruments are lagged and twice lagged q, profit rates and investment. The third column estimates the adjustment cost parameter for companies in the control group using (2.27). An efficient GMM procedure is used with a robust covariance matrix. The fourth column contains estimates for the parameter β in (2.96). Standard errors for this regression are computed with a robust covariance matrix and are adjusted for first step estimation error as described in the appendix.

This raises the empirical question: Which theory is supported by the data? The theoretical framework developed allowed a discussion of empirical tests in a unified framework. More importantly, it provided more concrete predictions about sources of predictability, excess volatility, and their strength depending on the dispersion of beliefs. These implications were tested in the framework of an episode in the 1920's. At that time, a number of companies were introduced into a market for lending stock (the so called "loan crowd"). The main finding of this paper is that the buildup of speculative components was followed by company investment as well.

One could argue that the incident studied is isolated. However, in many respects one can find many parallels between the '90's and the '20's. Large technological progress was followed by widely varying views on the growth potential of various sectors. The information superhighway was to the '90's what the automobile was to the '20's. The radio and the new advertising and distribution channels (shopping through catalogues -

the birth of large retail stores) were in many respect analogous to on-line shopping in the '90's. These technological innovations in production and distribution fueled speculation in the stock market and reduced the hurdle rates for investment in both historical periods.

One direction that was left unexplored in this paper concerns active financial policy. In the model of this paper I did not introduce any frictions or financing constraints, that would lead to a role for active financial policy. However, as discussed in Stein (1996) the presence of financing constraints can provide a further argument why investment and q would be tightly linked even if decisionmakers are long-termist. It would be interesting to study the behaviour of long-termist decisionmakers in this intertemporal model under the assumption that they have to rely on equity to finance investment. It is likely that in such a setup one would be able to derive additional relationships between equity issuance, investment and returns that would allow one to disentangle whether investment reacts to q because of short termism or because of an active financing channel. I pursue this line in current research.

2.6 Appendix: Proofs

2.6.1 Proofs for section 2.3.1

The essential difficulty in solving the filtering problem in section 2.3.1 consists in dealing with the non-linearity introduced by (2.7). If one were to replace (2.7) by:

$$df_t = -\lambda(f_t - \bar{f})dt + \sigma dZ_t^f$$

then one could replicate the arguments in Scheinkman and Xiong (2003) to show that the posterior mean process is given by:

$$
\begin{aligned}
d\widehat{f}_t^A &= -\lambda\left(\widehat{f}_t^A - \bar{f}\right)dt + \frac{\phi\sigma_s\sigma + \gamma}{\sigma_s^2}\left(ds^A - \widehat{f}_t^A dt\right) \\
&\quad + \frac{\gamma}{\sigma_s^2}\left(ds^B - \widehat{f}_t^A dt\right) + \frac{\gamma}{\sigma_D^2}\left(dD - \widehat{f}_t^A dt\right)
\end{aligned}
\tag{2.29}
$$

for agent A and similarly

$$
\begin{aligned}
d\widehat{f}_t^B &= -\lambda\left(\widehat{f}_t^B - \bar{f}\right)dt + \frac{\phi\sigma_s\sigma + \gamma}{\sigma_s^2}\left(ds^B - \widehat{f}_t^B dt\right) \\
&\quad + \frac{\gamma}{\sigma_s^2}\left(ds^A - \widehat{f}_t^B dt\right) + \frac{\gamma}{\sigma_D^2}\left(dD - \widehat{f}_t^B dt\right)
\end{aligned}
\tag{2.30}
$$

for agent B where γ is given as:

$$
\bar{\gamma} = \frac{\sqrt{\left(\lambda + \phi\frac{\sigma}{\sigma_s}\right)^2 + (1 - \phi^2)\left(2\frac{\sigma^2}{\sigma_s^2} + \frac{\sigma^2}{\sigma_D^2}\right)} - \left(\lambda + \phi\frac{\sigma}{\sigma_s}\right)}{\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}}
\tag{2.31}
$$

Then the arguments in Scheinkman and Xiong (2003) can be used to arrive at the dynamics

of the disagreement process g_t^A (respectively g_t^B):

$$dg_t^A = -\rho g_t^A dt + \sigma_g dW_g^A \tag{2.32}$$

where ρ, σ_g are given by:

$$\rho = \sqrt{\left(\lambda + \phi \frac{\sigma}{\sigma_s}\right)^2 + (1 - \phi^2)\sigma^2 \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)}$$
$$\sigma_g = \sqrt{2}\phi\sigma$$

Moreover it is easy to show that dW_g^A is orthogonal to the innovations in $d\widehat{f_t^A}$ and dW_g^B is orthogonal to the innovations in $d\widehat{f_t^B}$. The reader is referred to Scheinkman and Xiong (2003) for details.

If one wants to account for the fact that the volatility in (2.7) is non-constant, an approximate way to proceed is by means of the extended Kalman filter, which is proposed in Jazwinski (1970)[58]. This filter can be constructed by using a time varying γ (i.e. depending on the path of $\widehat{f_t^i}$) instead of the constant γ in formula (2.31):

$$\frac{d\gamma_t^i}{dt} = -2\left(\lambda + \phi\frac{\sigma}{\sigma_s}\frac{\widehat{f_t^i}}{\overline{f}}\right)\gamma_t^i + (1 - \phi^2)\sigma^2\frac{\widehat{f_t^i}}{\overline{f}} - \left(\gamma_t^i\right)^2\left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right), \quad i \in \{A, B\} \tag{2.33}$$

It is easy to verify that substituing $\widehat{f_t^i} = \overline{f}$ and requring $\frac{d\gamma_t^i}{dt} = 0$ one can recover equation (2.31). In principle one could solve γ_t^i explicitly for a given path of $\widehat{f_t^i}$. Agent $A's$ beliefs about the mean of f would then be characterized (approximately) by the two-dimensional system

---

[58]Unfortuanately, this filter does not make a claim to approximate the optimal non-linear filter, even though in applications it seems to have quite reasonable properties. Various sources discuss the properties and the efficiency of this filter for "small" noise.

(2.33), (2.29). For small λ, small σ and large $\frac{\sigma}{\sigma_s}, \frac{\sigma}{\sigma_D}$, γ_t^i will be given approximately by:

$$\gamma_t^i = \sqrt{\frac{\widehat{f_t^i}}{\overline{f}}} \frac{\sqrt{\left(\phi\frac{\sigma}{\sigma_s}\right)^2 + (1 - \phi^2)\left(2\frac{\sigma^2}{\sigma_s^2} + \frac{\sigma^2}{\sigma_D^2}\right)} - \phi\frac{\sigma}{\sigma_s}}{\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}} \tag{2.34}$$

To see why, rewrite equation (2.31) to get:

$$
\begin{aligned}
d\gamma_t^i &= \left(-2\left(\lambda + \phi\frac{\sigma}{\sigma_s}\frac{\widehat{f_t^i}}{\overline{f}}\right)\gamma_t^i + (1 - \phi^2)\sigma^2\frac{\widehat{f_t^i}}{\overline{f}} - \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)[\gamma_t^i - \overline{\gamma} + \overline{\gamma}]^2\right)dt = \\
&= \left(-2\left(\lambda + \phi\frac{\sigma}{\sigma_s}\frac{\widehat{f_t^i}}{\overline{f}}\right)\gamma_t^i + (1 - \phi^2)\left(\frac{\sigma_f^2}{\overline{f}}\right)\widehat{f_t} - \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)\left((\gamma_t^i - \overline{\gamma})^2 + \overline{\gamma}^2 + 2\overline{\gamma}(\gamma_t^i - \overline{\gamma})\right)\right)dt = \\
&= \left(-2\left(\lambda + \phi\frac{\sigma}{\sigma_s}\frac{\widehat{f_t^i}}{\overline{f}} + \overline{\gamma}\left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)\right)\gamma_t^i + (1 - \phi^2)\left(\frac{\sigma^2}{\overline{f}}\right)\widehat{f_t} - \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)\left((\gamma_t^i - \overline{\gamma})^2 - \overline{\gamma}^2\right)\right)dt
\end{aligned}
$$

If one approximates $\frac{\widehat{f_t^i}}{\overline{f}} \sim 1$ then the "solution" to this ODE is given by:

$$\gamma_t = \gamma_0 e^{-2wt} + \int_0^t e^{2w(\xi - t)}\left[(1 - \phi^2)\left(\frac{\sigma^2}{\overline{f}}\right)\widehat{f_\xi} - \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)\left((\gamma_\xi - \overline{\gamma})^2 - \overline{\gamma}^2\right)\right]d\xi \tag{2.35}$$

where:

$$w = \lambda + \phi\frac{\sigma}{\sigma_s} + \overline{\gamma}\left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)$$

w is the factor by which past γ_t are weighted. For large t one can ignore the first term in (2.35). Moreover, if w is large, then one can basically ignore the effect of past γ and approximate the above integral (as $t \to \infty$) by

$$\gamma_t = \frac{\left(\left(\frac{\sigma^2}{\overline{f}}\right)(1 - \phi^2)\widehat{f_t} + \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)\overline{\gamma}^2\right) - \left(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}\right)(\gamma_t - \overline{\gamma})^2}{2w}$$

Solving this quadratic equation and setting $\lambda = 0$ one gets (2.34). With this simplification the dimensionality of the problem can be reduced since now γ^i depends only on $\widehat{f_t}$. Replacing (2.34)

69

into (2.29) in the place of γ leads to the approximate belief processes:

$$d\widehat{f}_t^A = -\lambda\left(\widehat{f}_t^A - \bar{f}\right)dt + \tag{2.36}$$

$$\sqrt{\frac{\widehat{f}_t^A}{\bar{f}}}\left[\frac{\phi\sigma_s\sigma + \widetilde{\gamma}}{\sigma_s^2}\left(ds^A - \widehat{f}_t^A dt\right) + \frac{\widetilde{\gamma}}{\sigma_s^2}\left(ds^B - \widehat{f}_t^A dt\right) + \frac{\widetilde{\gamma}}{\sigma_D^2}\left(dD - \widehat{f}_t^A dt\right)\right] \tag{2.37}$$

$$d\widehat{f}_t^B = -\lambda\left(\widehat{f}_t^B - \bar{f}\right)dt + \tag{2.38}$$

$$\sqrt{\frac{\widehat{f}_t^B}{\bar{f}}}\left[\frac{\phi\sigma_s\sigma + \widetilde{\gamma}}{\sigma_s^2}\left(ds^A - \widehat{f}_t^A dt\right) + \frac{\widetilde{\gamma}}{\sigma_s^2}\left(ds^B - \widehat{f}_t^A dt\right) + \frac{\widetilde{\gamma}}{\sigma_D^2}\left(dD - \widehat{f}_t^A dt\right)\right] \tag{2.39}$$

where

$$\widetilde{\gamma} = \frac{\sqrt{\left(\phi\frac{\sigma}{\sigma_s}\right)^2 + (1 - \phi^2)\left(2\frac{\sigma^2}{\sigma_s^2} + \frac{\sigma^2}{\sigma_D^2}\right)} - \phi\frac{\sigma}{\sigma_s}}{\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2}}$$

Since $\frac{ds^A - \widehat{f}_t^A dt}{\sigma_s}, \frac{ds^B - \widehat{f}_t^A dt}{\sigma_s}, \frac{dD - \widehat{f}_t^A dt}{\sigma_D}$ are (standard) Brownian motions in the mind of agents of type A, it will be convenient to define the "total volatility" of $\widehat{f}_t^A$ (or $\widehat{f}_t^B$) by

$$\sigma_f = \sqrt{\left(\frac{\phi\sigma_s\sigma + \widetilde{\gamma}}{\sigma_s}\right)^2 + \left(\frac{\widetilde{\gamma}}{\sigma_s}\right)^2 + \left(\frac{\widetilde{\gamma}}{\sigma_D}\right)^2}$$

which leads to formulas (2.9) and (2.8). Moreover, as long as $\sqrt{\frac{\widehat{f}_t^A}{\bar{f}}}, \sqrt{\frac{\widehat{f}_t^B}{\bar{f}}}$ do not differ significantly from 1 (i.e. the volatility in $\widehat{f}_t^A, \widehat{f}_t^B$ is relatively small) then (2.32) will continue to be a reasonable approximation to the disagreement process.[59] Figure (2-6) demonstrates the performance of these approximations for the quantitative calibration in section 2.3.2. The top left figure compares the solution to (2.33) (obtained by an Euler Scheme) to (2.34). There are two observations about the figures. First the two volatilities comove quite closely and

---

[59]One could derive an alternative approximation to this disagreement process by subtracting df_t^A from df_t^B and then approximating all terms to the first order. Such an approximation would yield something close to the OU process used here for reasonably small ϕ. For simplicity I chose the approximate OU process described in the beginning of this section to be able to compare the results to Scheinkman and Xiong (2003).

second the posterior standard deviation (captured by γ_t^i) does not vary too much. These two observations help understand the next three panels. The top right panel is depicting the exact solution to the extended Kalman filter obtained by solving the two dimensional system (2.29) and (2.33) and the approximate filter obtained by using (2.36) instead. The two processes basically cannot be disentangled from each other, since they practically coincide. The bottom left panel depicts the performance of the extended Kalman Filter against the actual process f_t. It is easy to see that the extended Kalman Filter performs well in "recovering" the path of f_t. Finally, the bottom left panel depicts the difference in beliefs between agents A and B obtained from the approximate equation (2.32). Once again, the approximation is sufficiently good that one cannot disentangle the two processes, since they are practically identical. From these simulations it can be reasonably claimed that the approximation used is sufficiently accurate for all practical purposes.

2.6.2 Proofs for section 2.3.2

Proof. Proposition (2.1) I use a standard verification argument to verify that (2.11) provides the solution to (2.10). One can start by conjecturing a solution of the form:

$$P\left(\widehat{f}_t^A, K_t\right) = q^F\left(\widehat{f}_t^A\right) K_t + u^F\left(\widehat{f}_t^A\right) \tag{2.40}$$

Substituting this conjecture into the Hamilton Jacobi Bellman equation:

$$\max_{i_t} \left[\frac{1}{2}\sigma_f^2\frac{\widehat{f}_t^A}{\overline{f}}P_{ff} - \lambda(\widehat{f}_t^A - \overline{f})P_f + P_K\left(-\delta K + i_t\right) - rP + \widehat{f}_t^A K_t - pi_t - \frac{\chi}{2}(i_t^2)\right] = 0 \tag{2.41}$$

one arrives at the conclusion that (2.40) satisfies (2.41) if and only if the functions $q^F\left(\widehat{f}_t^A\right)$ and $u^F\left(\widehat{f}_t^A\right)$ solve the ordinary differential equations:

$$\frac{1}{2}\sigma_f^2\frac{\widehat{f}_t^A}{\overline{f}}q_{ff}^F - \lambda(\widehat{f}_t^A - \overline{f})q_f^F - (r + \delta)q^F + \widehat{f}_t^A = 0 \tag{2.42}$$

$$\frac{1}{2}\sigma_f^2\frac{\widehat{f}_t^A}{\overline{f}}u_{ff}^F - \lambda(\widehat{f}_t^A - \overline{f})u_f^F - ru^F + \frac{\left(q^F - p\right)^2}{2\chi} = 0 \tag{2.43}$$

71
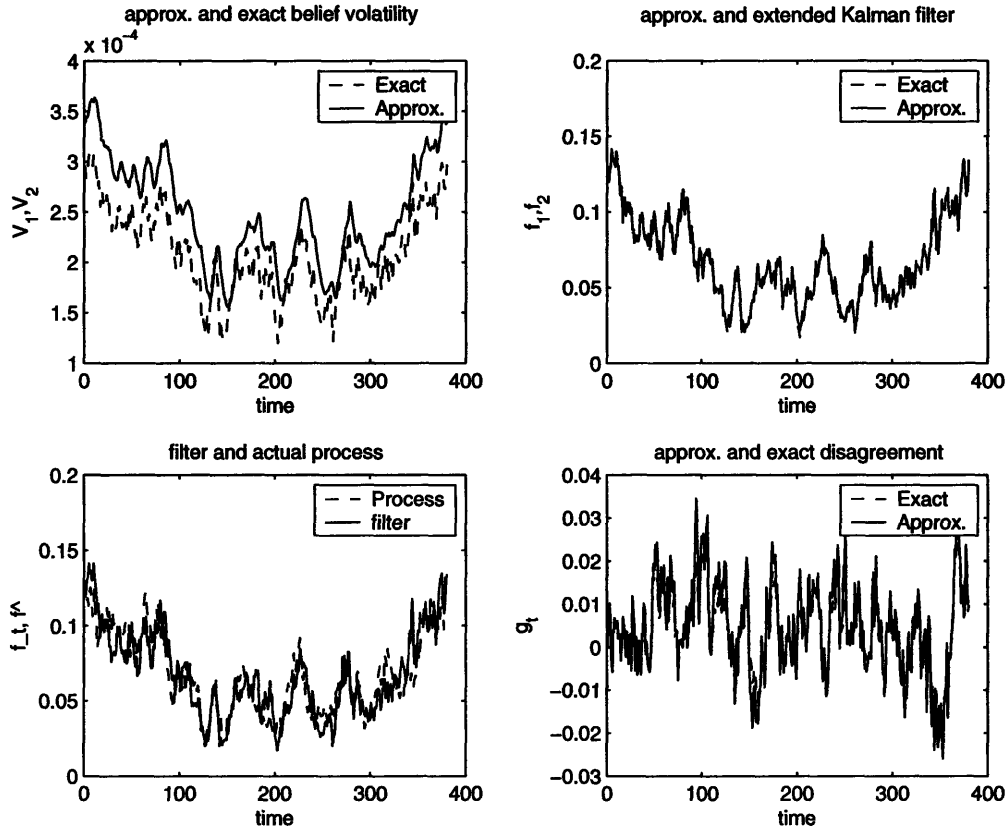
Figure 2-6: Simulation of a typical path of the model. The top left panel depicts the behaviour of the exact (under the extended Kalman Filter) and the approximate volatility of the posterior beliefs. The top right panel depicts the exact conditional mean process (under the extended Kalman Filter) and the approximation to the exact solution. The bottom left picture depicts the true f and the posterior mean $\widehat{f}$ as obtained by the extended Kalman Filter. Finally, the bottom right panel depicts the exact disagreement process (assuming both agents use the extended Kalman Filter) and the approximation proposed. The parameters for this example are the same as the ones used in Figure 2, namely: $\delta = 0.1$, $r = 0.05$, $\lambda = 0.1$, $\overline{f} = r + \delta$, $\sigma = 0.25\overline{f}$, $\sigma_D = 0.5\sigma$, $\sigma_s = \sigma$

The solution to equation (2.42) can easily be determined as[60]:

$$q^F\left(\widehat{f}_t^A\right) = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} \qquad (2.44)$$

whereas the solution to $u_1\left(\widehat{f}_t^A\right)$ is given as

$$z_1(\widehat{f}_t^A) = C_1\left(\widehat{f}_t^A - \overline{f}\right)^2 + C_2\left(\widehat{f}_t^A - \overline{f}\right) + C_3$$

with:

$$C_1 = \frac{1}{\chi}\frac{1}{r+2\lambda}\frac{1}{2}\left(\frac{1}{r+\delta+\lambda}\right)^2 \qquad (2.45)$$

$$C_2 = \frac{1}{\chi}\frac{1}{(r+\lambda)}\frac{1}{(r+\delta+\lambda)}\left[\left(\frac{\overline{f}(1-\widetilde{p})}{r+\delta}\right) + \frac{\sigma_f^2}{2}\frac{1}{\overline{f}}\frac{1}{r+2\lambda}\frac{1}{r+\delta+\lambda}\right] \qquad (2.46)$$

$$C_3 = \frac{1}{r}\left[\frac{1}{2\chi}\left(\frac{\overline{f}(1-\widetilde{p})}{r+\delta}\right)^2 + \sigma_f^2 C_1\right] \qquad (2.47)$$

where $\widetilde{p} = p\frac{\overline{f}}{r+\delta}$. The derivative of P_t w.r.t $\widehat{f}_t^A$ is given by:

$$\frac{1}{r+\delta+\lambda}K_t + 2C_1\left(\widehat{f}_t^A - \overline{f}\right) + C_2$$

As long as $2C_1\left(\widehat{f}_t^A - \overline{f}\right) + C_2 > 0$, $P_f > 0$. Since $\widehat{f}_t^A$ will always be positive it remains to check that:

$$-2C_1\overline{f} + C_2 > 0$$

or

$$\widetilde{p} < 1 - \frac{r+\delta}{r+\delta+\lambda}\frac{1}{r+2\lambda}\left(r+\lambda - \frac{\sigma_f^2}{2\overline{f}^2}\right) \overset{d}{=} \overline{p}$$

As might be expected, for the special case where $\widetilde{p} = 0$ the above equation is always satisfied

■

---

[60]In this paper only particular solutions of ODE's will be considered. Economically this means that "rational bubbles" will not be allowed, i.e. terms that grow unboundedly in expectation at the riskless rate. See Abel and Eberly (1997) on this point. In contrast only "resale premia" will be analyzed, that are determined in the next section.

2.6.3 Proofs for section 2.3.2

Two preliminary results will help in the construction of an explicit solution.

Lemma 2.1 *Consider the linear second order ordinary differential equation (ODE):*

$$\frac{\sigma_g^2}{2} y'' - \rho x y' - (r + \delta) y = 0 \tag{2.48}$$

Then there are two linearly independent solutions to this ODE and are given by

$$y_1(x) = \begin{cases} U\left(\frac{r+\delta}{2\rho}, \frac{1}{2}, \frac{\rho}{\sigma_g^2} x^2\right) & \text{if } x \le 0 \\ \frac{2\pi}{\Gamma\left(\frac{1}{2}+\frac{r+\delta}{2\rho}\right)\Gamma\left(\frac{1}{2}\right)} M\left(\frac{r+\delta}{2\rho}, \frac{1}{2}, \frac{\rho}{\sigma_g^2} x^2\right) - U\left(\frac{r+\delta}{2\rho}, \frac{1}{2}, \frac{\rho}{\sigma_g^2} x^2\right) & \text{if } x > 0 \end{cases}$$

$$y_2(x) = y_1(-x)$$

where $U()$ and $M()$ are Kummer's M and U functions[61]. $y_1(x)$ is positive, increasing and satisfies $\lim_{x \to -\infty} y_1(x) = 0, \lim_{x \to +\infty} y_1(x) = \infty$. Accordingly, $y_2(x)$ is positive, decreasing and satisfies: $\lim_{x \to -\infty} y_1(x) = \infty, \lim_{x \to +\infty} y_1(x) = 0$. Moreover any positive solution that satisfies equation (2.48) and $\lim_{x \to -\infty} y(x) = 0$ is given by: $\beta y_1(x)$ where $\beta > 0$ an arbitrary constant. Similarly any solution to (2.48) that is positive and satisfies: $\lim_{x \to \infty} y(x) = 0$ is given by $\beta y_2(x)$ where $\beta > 0$ is an arbitrary constant.

Proof. *Lemma (2.1)* The proof is essentially the same as the proof of proposition 2 in Scheinkman and Xiong (2003) and therefore large portions are omitted. If $v(z)$ is a solution to:

$$z v''(z) + \left(\frac{1}{2} - z\right) v'(z) - \frac{r+\delta}{2\rho} v(z) = 0 \tag{2.49}$$

---

[61] *These functions are described in Abramowitz and Stegun (1965) p.504.*

then $y(x) = v\left(\frac{\rho}{\sigma_g^2}x^2\right)$ satisfies (2.48). The general solution to equation (2.49) is given by[62]:

$$v(z) = \alpha M\left(\frac{r+\delta}{2\rho}, \frac{1}{2}, z\right) + \beta U\left(\frac{r+\delta}{2\rho}, \frac{1}{2}, z\right)$$

where the functions $M()$ and $U()$ are given in terms of their power series expansion in 13.1.2. and 13.1.3. of Abramowitz and Stegun (1964). The properties $y_1 > 0, y_{1x} > 0$, $\lim_{x \to -\infty} y_1(x) = 0, \lim_{x \to +\infty} y_1(x) = \infty$ can be established as in Scheinkman and Xiong (2003). It remains to show that the Wronskian of the two solutions $(y_1 y_2' - y_1' y_2)$ is different from 0 everywhere. This is immediate since $y_1(x), y_2(x) > 0$ and $y_1'(x) > 0, y_2'(x) < 0$. ∎

Lemma 2.2 *Consider the linear (inhomogenous) second order ODE:*

$$\frac{\sigma_g^2}{2}y'' - \rho x y' - (r+\delta)y = -f(x) \tag{2.50}$$

Then the general solution to (2.50) is given as:

$$y(x) = \left[\int_x^{+\infty}\left(\frac{\frac{2}{\sigma^2}f(z)y_2(z)}{y_1'(z)y_2(z) - y_1(z)y_2'(z)}\right)dz + C_1\right]y_1(x) +$$
$$+ \left[\int_{-\infty}^x\left(\frac{\frac{2}{\sigma^2}f(z)y_1(z)}{y_1'(z)y_2(z) - y_1(z)y_2'(z)}\right)dz + C_2\right]y_2(x)$$

provided that the above integrals exist. Moreover the derivative $y'(x)$ is given as:

$$y'(x) = \left[\int_x^{+\infty}\left(\frac{\frac{2}{\sigma^2}f(z)y_2(z)}{y_1'(z)y_2(z) - y_1(z)y_2'(z)}\right)dz + C_1\right]y_1'(x) +$$
$$+ \left[\int_{-\infty}^x\left(\frac{\frac{2}{\sigma^2}f(z)y_1(z)}{y_1'(z)y_2(z) - y_1(z)y_2'(z)}\right)dz + C_2\right]y_2'(x)$$

---

[62]See. Abramowitz and Stegun (1965) p. 504

Proof. *Lemma* (2.2)The proof is a basic variations of parameters argument and is omitted. For details see e.g. Section 9.3. in Rainville, Bedient and Bedient (1997). ∎

By setting $C_1 = C_2 = 0$ in the above Lemma one gets the so called particular solution, which will depend on σ, ρ, r and the specific functional form of $f(x)$. I will denote the solution $y(x)$ to this equation as:

$$G(f(x); \sigma_g, \rho, r + \delta)$$

and proceed in steps to provide a proof to proposition 2.2. The first step is to make a guess on the form of optimal investment that is verified later. In particular suppose that the firm's investment policy is given by:

Conjecture 1 *The optimal investment policy in equilibrium is given as:*

$$i_t = \frac{1}{\chi} \left(\frac{\overline{f}(1 - \widetilde{p})}{r + \delta} + \frac{\widehat{f}_t^A - \overline{f}}{r + \delta + \lambda} + 1\{g_t^A > 0\} \frac{g_t^A}{r + \delta + \lambda} + \beta y_1(-|g_t^A|) \right) \tag{2.51}$$

where β is a constant that can be determined as:

$$\beta = \frac{1}{2(r + \delta + \lambda)y_1'(0)} \tag{2.52}$$

and y_1 is the function described in the Lemma 2.1.

The next step will be to determine the equilibrium prices, stopping times for agents etc. conditional on the investment policy described. To do this it is easiest to compute the "infinite" horizon value of the company to an investor of type A conditional on the policy (2.51). One can focus without loss of generality on the determination of the reservation price for agent A since the problem for agent B is symmetric. Formally, the goal will be to determine the functional:

$$V(K_t, \widehat{f}_t^A, g_t^A) = E_t^A \left[\int_t^\infty e^{-r(s-t)} \left(\widehat{f}_s^A K_s - p i_s - \frac{\chi}{2}(i_s^2) \right) ds \right] \tag{2.53}$$

This function captures the value of the asset to an "infinite" horizon investor of type A who takes the conjectured investment policy (2.51) as given.

Proposition 2.5 *The solution to (2.53) is given by:*

$$V\left(K_t, \widehat{f}_t^A, g_t^A\right) = \left[\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda}\right] K_t +$$

$$+ \left(C_1\left(\widehat{f}_t^A - \overline{f}\right)^2 + C_2\left(\widehat{f}_t^A - \overline{f}\right) + C_3\right)$$

$$+ u(g_t^A)$$

where $u(g_t^A) < 0$ and C_1, C_2, C_3 are the same constants as in Proposition 2.1. Moreover $u_g < 0$.

Proof. *Proposition (2.5)* According to the Feynman Kac Theorem the solution $V\left(K_t, \widehat{f}_t^A, g_t^A\right)$

to (2.53) must satisfy the partial differential equation:

$$\mathcal{A}V + \widehat{f}_t^A K_t - i_t\left(p + \frac{\chi}{2}i_t\right) = 0 \tag{2.54}$$

where $\mathcal{A}$ is the infinitesimal operator given by:

$$\mathcal{A}V = \frac{\sigma_f^2}{2}\frac{\widehat{f}_t^A}{\overline{f}}V_{ff} + \frac{\sigma_g^2}{2}V_{gg} - \lambda(\widehat{f}_t^A - \overline{f})V_f - \rho g_t^A V_g + V_k\left(-\delta K_t + i_t\right) - rV$$

Conjecturing a solution of the form:

$$V = h\left(\widehat{f}_t^A\right)K_t + z\left(\widehat{f}_t^A, g_t^A\right)$$

and substituting this conjecture back into (2.54) one can determine conditions that $h()$ and $z()$ have to satisfy in order to satisfy (2.54). $h()$ has to satisfy:

$$\frac{\sigma_f^2}{2}\frac{\widehat{f}_t^A}{\overline{f}}h_{ff} + \frac{\sigma_g^2}{2}h_{gg} - \lambda(f - \overline{f})h_f - \rho g_t^A h_g - (r+\delta)h + \widehat{f}_t^A = 0$$

A particular solution is given by[63]:

$$h\left(\widehat{f}_t^A\right) = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda}$$

while $z\left(\widehat{f}_t^A, g_t^A\right)$ solves the partial differential equation:

$$\frac{\sigma_f^2}{2}\frac{\widehat{f}_t^A}{\overline{f}}z_{ff} + \frac{\sigma_g^2}{2}z_{gg} - \lambda(f - \overline{f})z_f - \rho g_t^A z_g - rz + h(f,g)i_t - i_t\left(p + \frac{\chi}{2}i_t\right) = 0 \qquad (2.55)$$

It is easy to show that:

$$h(f,g)i_t - i_t\left(p + \frac{\chi}{2}i_t\right) = \frac{1}{2\chi}\left(\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} - p\right)^2 - \frac{1}{2\chi}\left(\widetilde{b}(g_t^A)\right)^2 \qquad (2.56)$$

and

$$\widetilde{b}(g_t^A) = \beta y_1(-|g_t^A|) + 1\{g_t^A > 0\}\frac{g_t^A}{r+\delta+\lambda}$$

(2.56) allows one to derive an explicit solution for (2.55) by solving two ordinary differential equations $z_1(\widehat{f}_t^A), u(g_t^A)$ that satisfy:

$$\frac{\sigma_f^2}{2}\frac{\widehat{f}_t^A}{\overline{f}}z_{1ff} - \lambda(\widehat{f}_t^A - \overline{f})z_{1f} - rz_1 + \frac{1}{2\chi}\left(\frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} - p\right)^2 = 0 \qquad (2.57)$$

$$\frac{\sigma_g^2}{2}u_{gg} - \rho g u_g - ru - \frac{1}{2\chi}\left(\widetilde{b}(g_t^A)\right)^2 = 0 \qquad (2.58)$$

$z_1(\widehat{f}_t^A)$ solves the exact same ODE as $u^F\left(\widehat{f}_t^A\right)$ in Proposition (2.1) and thus it will be the case that:

$$z_1(\widehat{f}_t^A) = C_1\left(\widehat{f}_t^A - \overline{f}\right)^2 + C_2\left(\widehat{f}_t^A - \overline{f}\right) + C_3$$

for the same constants as in Proposition (2.1) . Finally, one can use the results in Lemma

---

[63]Obviously there are other solutions that "explode" at the rate r but we will only be interested in bounded solutions in this paper.

2.2 to construct the solution to (2.58). It is given by:

$$u(g) = G\left(-\frac{1}{2\chi}\left(\widetilde{b}(g_t^A)\right)^2; \sigma_g, \rho, r\right) < 0$$

To show that $u_g < 0$, observe that $\widetilde{b}(g^A)$ is a strictly increasing, positive and continuously differentiable function of g_t^A, so that:

$$\left[-\frac{1}{2\chi}\left(\widetilde{b}(g_t^A)\right)^2\right]' = -\frac{1}{\chi}\widetilde{b}(g_t^A)\widetilde{b}_g < 0$$

Differentiating (2.58) w.r.t. g_t^A gives:

$$\frac{\sigma_g^2}{2}u_{ggg} - \rho g u_{gg} - (r+\rho)u_g - \frac{1}{\chi}\widetilde{b}(g_t^A)\widetilde{b}_g = 0$$

Defining $u_g = z^d$ one can rewrite this equation as:

$$\frac{\sigma_g^2}{2}z_{gg}^d - \rho g z_g^d - (r+\rho)z^d - \frac{1}{\chi}\widetilde{b}(g_t^A)\widetilde{b}_g = 0$$

which has the particular solution:

$$G\left(-\frac{1}{\chi}\widetilde{b}(g_t^A)\widetilde{b}_g; \sigma_g, \rho, r+\rho\right)$$

which is unambiguously negative. This formal analysis can be made rigorous by invoking a set of results known as Malliavin Calculus (see e.g. Fournie et. al. (1999)). [64] ∎

With an expression for the value of the asset to an agent who does not intend to resell it ever in the future, one can proceed to guess an equilibrium pricing function and an optimal stopping policy. An informed "guess" is that the optimal stopping policy is of a particularly

---

[64]Details are available upon request

simple form: Agent A should sell once $\widehat{f}_t^A < \widehat{f}_t^B$ and agent B should sell once $\widehat{f}_t^B < \widehat{f}_t^A$. This is the case because there are no transactions costs in this model. Accordingly, each agent sells the asset once she stops having the most optimistic beliefs in the market. In particular one can re-express the reservation price for agent A as long as $g_t^A < 0$ (so that it is agent A who is the highest bidder) as

$$P(K_t, \widehat{f}_t^A, g_t^A) = V\left(K_t, \widehat{f}_t^A, g_t^A\right) + s(K_t, \widehat{f}_t^A, g_t^A) \tag{2.59}$$

Similarly, the reservation price for agent B (as long as $g_t^A > 0$ or equivalently $g_t^B < 0$) is given by symmetry as:

$$P(K_t, \widehat{f}_t^B, g_t^B) = V\left(K_t, \widehat{f}_t^B, g_t^B\right) + s(K_t, \widehat{f}_t^B, g_t^B) \tag{2.60}$$

A further conjecture that will be verified shortly is that for $g_t^A < 0$, s is given as:

$$s(K, f, g) = \beta y_1(g)K + n(g)\left(\widehat{f}_t^A - \overline{f}\right) + v(g) \tag{2.61}$$

for some functions n and v. In other words the reservation price for each agent is just the infinite horizon valuation of the dividends plus a speculative component. Using (2.59) one can get the following result:

Lemma 2.3 *If the reservation price function for agent B is given by (2.60) then the reservation price for agent A is given by:*

$$
\begin{aligned}
P(K_t, \widehat{f}_t^A, g_t^A) &= V\left(K_t, \widehat{f}_t^A, g_t^A\right) + \tag{2.62} \\
&\quad + \sup_\tau Ee^{-r\tau}\left[V\left(K_\tau, \widehat{f}_\tau^B, g_\tau^B\right) - V\left(K_\tau, \widehat{f}_\tau^A, g_\tau^A\right) + s(K_\tau, \widehat{f}_\tau^B, g_\tau^B)\right] \tag{2.63} \\
&= V\left(K_t, \widehat{f}_t^A, g_t^A\right) + \tag{2.64} \\
&\quad + \sup_\tau Ee^{-r\tau}\left[\left(\frac{g_\tau^A}{r+\delta+\lambda} + \beta y_1(-g_\tau^A)\right) K_\tau + w(\widehat{f}_\tau^A, g_\tau^A)\right] \tag{2.65}
\end{aligned}
$$

80

for $w(\widehat{f}_\tau^A, g_\tau^A)$ given by:

$$w(\widehat{f}_\tau^A, g_\tau^A) = \left[C_2 + n(-g_\tau^A)\right] g_\tau^A + C_1 \left(g_\tau^A\right)^2 + u(-g_\tau^A) - u(g_\tau^A) + \left[n(-g_\tau^A) + g_\tau^A 2C_1\right]\left(\widehat{f}_\tau^A - \overline{f}\right) + v(-g_\tau^A)$$

Proof. Lemma (2.3). The argument is essentially identical to the one given in Scheinkman and Xiong (2003). I give it here for completeness. Using (2.59) one can get:

$$P(K_t, \widehat{f}_t^A, g_t^A) =$$

$$
\begin{aligned}
&= \sup_\tau E\left[\int_0^\tau e^{-rt}\left(\widehat{f}_t^A K_t - p i_t - \tfrac{\chi}{2} i_t^2\right)dt + e^{-r\tau}P(K_\tau, \widehat{f}_\tau^B, g_\tau^B)\right] = \\
&= \sup_\tau E\left[\begin{array}{c}\int_0^\tau e^{-rt}\left(\widehat{f}_t^A K_t - p i_t - \tfrac{\chi}{2} i_t^2\right)dt + \\ + e^{-r\tau}\left[V\left(K_\tau, \widehat{f}_\tau^B, g_\tau^B\right) + \beta y_1(g_\tau^B)K_\tau + n(g_\tau^B)\left(\widehat{f}_\tau^B - \overline{f}\right) + v(g_\tau^B)\right]\end{array}\right] = \\
&= \sup_\tau E\left[\begin{array}{c}\int_0^\infty e^{-rt}\left(f K_t - p i_t - \tfrac{\chi}{2} i_t^2\right)dt + \\ + e^{-r\tau}\left[V\left(K_\tau, \widehat{f}_\tau^B, g_\tau^B\right) - V\left(K_\tau, \widehat{f}_\tau^A, g_\tau^A\right) + \beta y_1(g_\tau^B)K_\tau + n(g_\tau^B)\left(\widehat{f}_\tau^B - \overline{f}\right) + v(g_\tau^B)\right]\end{array}\right] = \\
&= V\left(K_t, \widehat{f}_t^A, g_t^A\right) + \sup_\tau E e^{-r\tau}\left[\begin{array}{c}\left(\frac{g_\tau^A}{r+\delta+\lambda} + \beta y_1(g_\tau^B)\right)K_\tau + C_2 g_\tau^A + u(g_\tau^B) - u(g_\tau^A) + \\ + C_1\left[\left(\widehat{f}_\tau^B - \overline{f}\right)^2 - \left(\widehat{f}_\tau^A - \overline{f}\right)^2\right] + n(g_\tau^B)\left(\widehat{f}_\tau^B - \overline{f}\right) + v(g_\tau^B)\end{array}\right] = \\
&= V^A + \sup_\tau E e^{-r\tau}\left[\begin{array}{c}\left(\frac{g_\tau^A}{r+\delta+\lambda} + \beta y_1(-g_\tau^A)\right)K_\tau + \left[C_2 + n(-g_\tau^A)\right]g_\tau^A + C_1\left(g_\tau^A\right)^2 + \\ + u(-g_\tau^A) - u(g_\tau^A) + \left[n(-g_\tau^A) + g_\tau^A 2C_1\right]\left(\widehat{f}_\tau^A - \overline{f}\right) + v(-g_\tau^A)\end{array}\right]
\end{aligned}
$$

where the last line follows from the identities:

$$
\begin{aligned}
\widehat{f}_\tau^B &= \widehat{f}_\tau^A + g_\tau^A \\
g_\tau^B &= -g_\tau^A
\end{aligned}
$$

Defining the function $w(\widehat{f}_\tau^A, g_\tau)$ as:

$$
\begin{aligned}
w(\widehat{f}_\tau^A, g_\tau) &= \left[C_2 + n(-g_\tau^A)\right]g_\tau^A + C_1\left(g_\tau^A\right)^2 + u(-g_\tau^A) - u(g_\tau^A) + \\
&\quad + \left[n(-g_\tau^A) + g_\tau^A 2C_1\right]\left(\widehat{f}_\tau^A - \overline{f}\right) + v(-g_\tau^A)
\end{aligned}
$$

concludes the proof. ∎

In light of Lemma (2.3) it remains to establish that (2.61) is right, i.e. that there exist appropriate functions $n(\cdot)$, $v(\cdot)$, and an appropriate constant β such that:

$$\beta y_1(g_t^A)K_t + n(g_t^A)\left(\widehat{f}_t^A - \overline{f}\right) + v(g_t^A) = \sup_\tau Ee^{-r\tau}\left[\left(\frac{g_\tau^A}{r+\delta+\lambda} + \beta y_1(-g_\tau^A)\right)K_\tau + w(\widehat{f}_\tau^A, g_\tau^A)\right]$$

(2.66)

In other words it remains to establish the existence of functions $n(g_t^A), v(g_t^A)$ and a constant β so that the Value function of the optimal stopping problem on the right hand side has the form on the left hand side inside the continuation region, i.e. inside the region where agent A finds it optimal to hold the asset. The right hand side problem is a three dimensional optimal stopping problem (in $K_t, \widehat{f}_t^A, g_t^A$) and in general there is no method to solve such problems analytically. This is in contrast to one dimensional optimal stopping problems where continuity along with smooth pasting is enough to determine the stopping region and the associated value function in most cases. Fortunately, the simple form of the conjectured continuation region allows one to solve this problem as is demonstrated in the next proposition:

Proposition 2.6 *There exist functions $n(g_t^A), v(g_t^A)$ and a constant β such that the function:*

$$s(K_t, \widehat{f}_t^A, g_t^A) = \begin{cases} \beta y_1(g_t^A)K_t + n(g_t^A)\left(\widehat{f}_t^A - \overline{f}\right) + v(g_t^A) & \text{if } g_t^A < 0 \\ \left(\frac{g_\tau^A}{r+\delta+\lambda} + \beta y_1(-g_\tau^A)\right)K_\tau + w(\widehat{f}_\tau^A, g_\tau^A) & \text{if } g_t^A \geq 0 \end{cases}$$

(2.67)

satisfies

$$\mathcal{A}V = 0 \; \text{if } g_t^A < 0$$

is twice continuously differentiable in the region $g_t^A > 0$ and in the region $g_t^A < 0$ and once cont. differentiable everywhere. The constant β is given by:

$$\beta = \frac{1}{2(r+\delta+\lambda)}\frac{1}{y_1'(0)}$$

and the functions $n(\cdot)$ and $v(\cdot)$ are given in the proof.

Proof. Proposition (2.6) The first step is to construct the Value function under the assumption that both the conjecture for the optimal stopping region and the equilibrium investment strategy is correct. Since for $g_t^A < 0$ the conjectured optimal strategy is to hold the asset, one can formulate a necessary condition for the value function s of the optimal stopping problem on the right hand side of (2.66). Namely, it has to be the case that inside this region:

$$\mathcal{A}s = 0$$

or:

$$0 = \frac{\sigma_f^2}{2}\frac{\widehat{f}_t^A}{\overline{f}}s_{ff} + \frac{\sigma_g^2}{2}s_{gg} - \lambda(\widehat{f}_t^A - \overline{f})s_f - \rho g_t^A s_g + \tag{2.68}$$

$$+ s_K\left[-\delta K + \frac{1}{\chi}\left(\frac{\overline{f}(1-\widetilde{p})}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + \beta y_1(g_t^A)\right)\right] - rs \tag{2.69}$$

An informed guess is that this PDE has a solution of the form:

$$\beta y_1(g_t^A)K_t + \zeta(\widehat{f}_t^A, g_t^A)$$

Plugging this into (??) one gets the set of equations:

$$0 = \frac{\sigma_g^2}{2}y_{1gg} - \rho g_t^A y_{1g} - (r+\delta)y_1 \tag{2.70}$$

$$0 = \frac{\sigma_f^2}{2}\frac{\widehat{f}^A}{\overline{f}}\zeta_{ff} + \frac{\sigma_g^2}{2}\zeta_{gg} - \lambda(\widehat{f}_t^A - \overline{f})\zeta_f - \rho g_t^A \zeta_g + \tag{2.71}$$

$$+ \beta y_1(g_t^A)\frac{1}{\chi}\left(\frac{\overline{f}(1-\widetilde{p})}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + \beta y_1(g_t^A)\right) - r\zeta \tag{2.72}$$

It is immediate that the function $y_1(g_t^A)$ constructed in Lemma 2.1 satisfies (2.70) by con-

struction[65]. One can determine a solution to equation (2.71) by postulating that the solution u is given by:

$$\zeta(\widehat{f}_t^A, g_t^A) = v(g_t^A) + n(g_t^A)\left(\widehat{f}_t^A - \overline{f}\right)$$

and upon substituting this conjecture into (2.71) it is easy to see that $v(g_t^A)$ and $n(g_t^A)$ have to satisfy the two ordinary differential equations:

$$\frac{\sigma_g^2}{2}v_{gg} - \rho g_t^A v_g - rv + \frac{1}{\chi}\left(\beta y_1(g_t^A)\left(\frac{\overline{f}}{r+\delta} + \beta y_1(g_t^A)\right)\right) = 0 \qquad (2.73)$$

$$\frac{\sigma_g^2}{2}n_{gg} - \rho g_t^A n_g - (r+\lambda)n(g_t^A) + \frac{1}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right)y_1(g_t^A) = 0 \qquad (2.74)$$

In the $g_t^A < 0$ region, the general solution to (2.73) and (2.74) is given by:

$$v(g_t^A) = c_1 y_1^{(r)}(g_t^A) + c_2 y_2^{(r)}(g_t^A) + v_P(g_t^A)$$

$$n(g_t^A) = \widetilde{c}_1 y_1^{(r+\lambda)}(g_t^A) + \widetilde{c}_2 y_2^{(r+\lambda)}(g_t^A) + n_P(g_t^A)$$

where $v_P(g_t^A), n_P(g_t^A)$ are the particular solutions to the above equations obtained by Lemma 2.2:

$$v_P(g_t^A) = G\left[\frac{1}{\chi}\left(\beta y_1(-|g_t^A|)\left(\frac{\overline{f}(1-\widehat{p})}{r+\delta} + \beta y_1(-|g_t^A|)\right)\right); \sigma_g, \rho, r\right]$$

$$n_P(g_t^A) = G\left[\frac{1}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right)y_1(-|g_t^A|); \sigma_g, \rho, r+\lambda\right]$$

and $y_1^{(x)}(g_t^A), y_2^{(x)}(g_t^A)$ are defined in an identical way to $y_1(g_t^A)$ and $y_2(g_t^A)$ of Lemma 2.1 with the only exception that $r + \delta$ is replaced by x. It is also clear that since $y_1(-|g_t^A|)$ is a bounded function, the above integrals are finite. Moreover, it is easy to check that the

---

[65]Moreover it is the only solution that vanishes as $g_t^A \to -\infty$

84

particular solutions to the above equations satisfy $v'_P(0) = 0$ and $n'_P(0) = 0$.[66] Finally, to keep only solutions that do not explode as $g_t^A \to -\infty$ one can set $c_2 = \tilde{c}_2 = 0$.

Observe that the conjectured Value function is of the form posited in the left hand side of equation (2.66). To conclude with the construction of a candidate pricing function, it remains to determine the constants in such a way that the resulting value function for the optimal stopping problem (2.66) is both continuous and cont. differentiable everywhere. For $g_t^A > 0$ the conjecture is that agent A resells to agent B, so that the value function for this case is given by the value of "immediate exercise" i.e.

$$s(K_t, \widehat{f}_t^A, g_t^A) = \left(\frac{g_t^A}{r + \delta + \lambda} + \beta y_1(-g_t^A) \right) K_t + w(\widehat{f}_t^A, g_t^A) \text{ if } g_t^A > 0$$

In each of the two regions ($g_t^A < 0, g_t^A > 0$) the function V is twice cont. differentiable, accordingly continuity and differentiability only needs to be enforced at $g_t^A = 0$. The left limit of V at $g_t^A = 0$ is given by:

$$\beta y_1(0) K_t + v(0) + n(0) \left(\widehat{f}_t^A - \overline{f} \right)$$

whereas the right limit is obtained by evaluating $\left(\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right) K_t + w(\widehat{f}_t^A, g_t^A)$ around $g_t^A = 0$. This yields after obvious simplifications:

$$\beta y_1(0) K_t + v(0) + n(0) \left(\widehat{f}_t^A - \overline{f} \right)$$

---

[66]Since they are symmetric around 0 and continuously differentiable.

so that continuity is immediately satisfied. Differentiability requires that:

$$
\begin{aligned}
\beta y_1'(0) &= \frac{1}{r+\delta+\lambda} - \beta y_1'(0) \\
\widetilde{c}_1 y_1^{(r+\lambda)'}(0) &= -\widetilde{c}_1 y_1^{(r+\lambda)'}(0) + 2C_1 \\
2c_1 y_1^{(r)'}(0) &= -2u'(0) + C_2 + n(0)
\end{aligned}
$$

which implies that:

$$
\begin{aligned}
\beta &= \frac{1}{2(r+\delta+\lambda)} \frac{1}{y_1'(0)} \\
\widetilde{c}_1 &= \frac{C_1}{y_1^{(r+\lambda)'}(0)} \\
c_1 &= \frac{-2u'(0) + C_2 + n(0)}{2y_1^{(r)'}(0)}
\end{aligned}
$$

∎

In order to be able to invoke a verification Theorem for optimal stopping one needs additionally the following two results:

Proposition 2.7 *The function s constructed in Proposition (2.6) satisfies:*

$$
s(\widehat{f}_t^A, g_t^A, K) =
$$

$$
\left(\frac{g_t^A}{r+\delta+\lambda} K_t + 2C_1 g_t^A (\widehat{f}_t^A - \overline{f}) + C_1 \left(g_t^A\right)^2 + C_2 g_t^A + u(-g_t^A) - u(g_t^A) \right) + s(\widehat{f}_t^A + g_t^A, -g_t^A, K_t) =
$$

$$
= V\left(K_t, \widehat{f}_t^B, g_t^B\right) - V\left(K_t, \widehat{f}_t^A, g_t^A\right) + s(K_t, \widehat{f}_t^B, g_t^B)
$$

Proof. Proposition (2.7) The definition of s in Proposition (2.6) allows one to compute $s(\widehat{f}_t^A + g_t^A, -g_t^A, K_t)$ as:

$$
s(\widehat{f}_t^A + g_t^A, -g_t^A, K_t) = \begin{cases} \beta y_1(-g_t^A)K_t + n(-g_t^A)\left(\widehat{f}_t^A + g_t^A - \overline{f}\right) + v(-g_t^A) & \text{if } g_t^A > 0 \\ \left(\frac{-g_t^A}{r+\delta+\lambda} + \beta y_1(g_t^A)\right) K_t + w(\widehat{f}_t^A + g_t^A, -g_t^A) & \text{if } g_t^A \leq 0 \end{cases}
$$

86

and since

$$w(\widehat{f}_t^A + g_t^A, -g_t^A) = -\left[C_2 + n(g_t^A)\right] g_t^A + C_1 \left(g_t^A\right)^2 + u(g_t^A) - u(-g_t^A) +$$
$$+ \left[n(g_t^A) - g_t^A 2C_1\right] \left(\widehat{f}_t^A + g_t^A - \overline{f}\right) + v(g_t^A)$$

one gets after a number of simplifications irrespective of whether $\left(g_t^A > 0\right)$ or $\left(g_t^A \leq 0\right)$

$$s(\widehat{f}_t^A, g_t^A, K_t) =$$

$$\left(\frac{g_t^A}{r+\delta+\lambda} K_t + 2C_1 g_t^A (\widehat{f}_t^A - \overline{f}) + C_1 \left(g_t^A\right)^2 + C_2 g_t^A + u(-g_t^A) - u(g_t^A)\right) + s(\widehat{f}_t^A + g_t^A, -g_t^A, K_t)$$

∎

The single most important step towards verifying the results is to verify that:

$$\mathcal{A}s \leq 0$$

in the region $(g_t \geq 0)$. In particular, the following result is true:

Proposition 2.8 *Suppose*

(A) $\widetilde{p} < \frac{\lambda}{r+\delta+\lambda}$[67]

(B) $\rho - 3\lambda - 2r > 0$

Then

$$\mathcal{A}s \leq 0$$

for $g_t^A \geq 0$.

---

[67] *It is trivial to show that under this condition $K_t \geq 0$ and moreover $\widetilde{p} < \overline{p}$*

Proof. Proposition (2.8) For $g_t^A \geq 0$ the operator A is given for any function $\tilde{V}$ as:

$$\mathcal{A}\tilde{V} = \frac{\sigma_f^2}{2} \frac{\hat{f}_t^A}{\overline{f}} \tilde{V}_{ff} + \frac{\sigma_g^2}{2} \tilde{V}_{gg} - \lambda(\hat{f}_t^A - \overline{f})\tilde{V}_f - \rho g_t^A \tilde{V}_g +$$

$$+\tilde{V}_k \left(-\delta K_t + \frac{1}{\chi} \left(\frac{\overline{f}(1-\widehat{p})}{r+\delta} + \frac{\hat{f}_t^A + g_t^A - \overline{f}}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right) \right) - r\tilde{V}$$

The conjectured s is given as:

$$s\left(K_t, \hat{f}_t^A, g_t^A\right) = \left(\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right) K_t + w(\hat{f}_t^A, g_t^A)$$

where (from Lemma 2.3) $w(\hat{f}_t^A, g_t^A)$ is given as:

$$w(\hat{f}_t^A, g_t) = \left[C_2 + n(-g_t^A) \right] g_t^A + C_1 \left(g_t^A \right)^2 + u(-g_t^A) - u(g_t^A) +$$

$$+ \left[n(-g_t^A) + g_t^A 2 C_1 \right] \left(\hat{f}_t^A - \overline{f} \right) + v(-g_t^A)$$

It will be easiest to apply the operator on each term separately:

$$A \left[\left(\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right) K_t \right] =$$

$$= \left(-\frac{(\rho+r+\delta)}{r+\delta+\lambda} g_t^A \right) K_t$$

$$+ \frac{1}{\chi} \left(\frac{g_t^A}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right) \left(\frac{\overline{f}(1-\widehat{p})}{r+\delta} + \frac{\hat{f}_t^A + g_t^A - \overline{f}}{r+\delta+\lambda} + \beta y_1(-g_t^A) \right)$$

88

$$\mathcal{A}\left[\left[C_2 + n(-g_t^A)\right]g_t^A + C_1\left(g_t^A\right)^2\right]$$

$$= -(\rho+r)g_t^A C_2 + C_1\sigma_g^2 - C_1(2\rho+r)\left(g_t^A\right)^2 +$$

$$+g\left(\frac{\sigma_g^2}{2}n_{gg}(-g_t^A) - \rho(-g_t^A)n_g(-g_t^A) - rn(-g_t^A)\right) \qquad (2.75)$$

$$-\rho g_t^A n(-g_t^A) - \sigma_g^2 n_g(-g_t^A)$$

$$\mathcal{A}\left[\left[n(-g_t^A) + g_t^A 2C_1\right]\left(\widehat{f}_t^A - \overline{f}\right)\right] =$$

$$= \left(\frac{\sigma_g^2}{2}n_{gg}(-g_t^A) - \rho g_t^A\left(-n_g(-g_t^A) + 2C_1\right) - (r+\lambda)\left[n(-g_t^A) + g_t^A 2C_1\right]\right)\cdot$$

$$\cdot\left(\widehat{f}_t^A - \overline{f}\right)$$

$$= \left(\frac{\sigma_g^2}{2}n_{gg}(-g_t^A) - \rho(-g_t^A)\left(n_g(-g_t^A)\right) - (r+\lambda)n(-g_t^A)\right)\left(\widehat{f}_t^A - \overline{f}\right)$$

$$-(\rho+r+\lambda)2C_1 g_t^A\left(\widehat{f}_t^A - \overline{f}\right)$$

Using the definitions of $u(\cdot)$ and $v(\cdot)$ from Propositions 2.6 and 2.5 the above expressions become:

$$\mathcal{A}\left[u(-g_t^A) - u(g_t^A)\right] = \left(\frac{\sigma_g^2}{2}u_{gg}(-g_t^A) - \rho(-g_t^A)u_g(-g_t^A) - ru(-g_t^A)\right) - \left(\frac{\sigma_g^2}{2}u_{gg} - \rho g_t^A u_g - ru\right) =$$

$$= -\frac{1}{2\chi}\left(\frac{2\beta y_1(-g_t^A)g_t^A}{r+\delta+\lambda} + \frac{\left(g_t^A\right)^2}{(r+\delta+\lambda)^2}\right)$$

$$\mathcal{A}\left[v(-g_t^A)\right] = \frac{\sigma_g^2}{2}v_{gg}(-g_t^A) - \rho((-g_t^A)v_g(-g_t^A)) - rv(-g_t^A)$$

$$= -\frac{1}{\chi}\left(\beta y_1(-g_t^A)\left(\frac{\overline{f}(1-\widehat{p})}{r+\delta} + \beta y_1(-g_t^A)\right)\right)$$

Collecting terms and using from Proposition 2.6 the fact that:

$$\frac{\sigma_g^2}{2} n_{gg} - \rho g_t^A n_g - (r+\lambda) n(g_t^A) = -\frac{1}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right) y_1(g_t^A)$$

and substituting it follows that:

$$\mathcal{A}\left[\left[C_2 + n(-g_t^A)\right] g_t^A + C_1 \left(g_t^A\right)^2\right] = -(\rho+r)g_t^A C_2 + C_1 \sigma_g^2 - C_1(2\rho+r)\left(g_t^A\right)^2 +$$
$$+g\left(-\frac{1}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right) y_1(-g_t^A) + \lambda n(-g_t^A)\right)$$
$$-\rho g_t^A n(-g_t^A) - \sigma_g^2 n_g(-g_t^A)$$

$$\mathcal{A}\left[\left[n(-g_t^A) + g_t^A 2C_1\right]\left(\widehat{f}_t^A - \overline{f}\right)\right] =$$
$$= \left(-\frac{1}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right) y_1(-g_t^A)\right)\left(\widehat{f}_t^A - \overline{f}\right)$$
$$- (\rho+r+\lambda)\, 2C_1 g_t^A \left(\widehat{f}_t^A - \overline{f}\right)$$

Collecting terms, simplifying and using the definitions of C_1, C_2, C_3 in Proposition 2.5 one can conclude that:

$$\mathcal{A}V = \left(-\frac{(\rho+r+\delta)}{r+\delta+\lambda} g_t^A\right) K_t +$$
$$-\frac{1}{\chi}(\rho-\lambda) g_t^A \left(2C_1 \left(\widehat{f}_t^A - \overline{f}\right) + \left[C_2 + \frac{1}{\rho-\lambda}\frac{1}{\chi}\frac{\sigma_f^2}{2}\frac{1}{\overline{f}}\frac{1}{r+2\lambda}\left(\frac{1}{r+\delta+\lambda}\right)^2\right]\right)$$
$$-\frac{1}{2\chi}(\rho-\lambda)\frac{1}{r+2\lambda}\left(\frac{\left(g_t^A\right)^2}{(r+\delta+\lambda)^2}\right)$$
$$-(\rho-\lambda) g_t^A n(-g_t^A) + \sigma_g^2 \left[C_1 - n_g(-g_t^A)\right] \qquad (2.76)$$

The first term is unambiguously negative if $K_t > 0$, the second term is negative since $(\rho - \lambda)$

is positive[68] and we have assumed $p < \widetilde{p}$ so that $2C_1 \left(\widehat{f}_t^A - \overline{f} \right) + C_2 > 0$. The third term will be negative since $\rho - \lambda > 0$. The terms in the fourth line requires some further analysis. In particular, notice that

$$\left(C_1 - n_g(-g_t^A) \right) \sigma_g^2 =$$
$$= \sigma_g^2 \left(C_1 - \frac{C_1 y_1^{(r+\lambda)\prime}(-g_t^A)}{y_1^{(r+\lambda)\prime}(0)} - n_P g(-g_t^A) \right) =$$
$$= \sigma_g^2 C_1 \left(1 - \frac{\widetilde{y_1}'(-g_t^A)}{\widetilde{y_1}'(0)} \right) - \sigma_g^2 n_P g(-g_t^A)$$

Thus one can rewrite (2.76) as:

$$-(\rho - \lambda) g_t^A n(-g_t^A) + \sigma_g^2 C_1 \left(1 - \frac{\widetilde{y_1}'(-g_t^A)}{\widetilde{y_1}'(0)} \right) - \sigma_g^2 n_P g(-g_t^A) \qquad (2.77)$$

As expected, (2.77) at $g_t^A = 0$ is 0 by smooth pasting. Thus it is sufficient to show that (2.77) is declining for $g_t^A > 0$. To do this, differentiate once w.r.t to g_t^A, to get:

$$-(\rho - \lambda) n(-g_t^A) + (\rho - \lambda) g_t^A n_g(-g_t^A) + \sigma_g^2 n_{gg}(-g_t^A) \qquad (2.78)$$

Now one can use the definition of $n(g)$ to get:

$$\sigma_g^2 n_{gg}(-g_t^A) = -\frac{2}{\chi} \left(\frac{\beta}{r + \delta + \lambda} \right) y_1(-g_t^A) + 2\rho(-g_t^A) n_g(-g_t^A) + 2(r + \lambda) n(-g_t^A)$$

---

[68] by the definition of ρ in 2.3.1

91

and substitute into (2.78) to arrive at:

$$-(\rho + \lambda)g_t^A n_g(-g_t^A)$$

$$-\frac{2}{\chi}\left(\frac{\beta}{r+\delta+\lambda}\right)y_1(-g_t^A)$$

$$-[\rho - 3\lambda - 2r]\,n(-g_t^A)$$

which will be smaller than 0 under assumption (B), since both $n(\cdot)$ and $n_g(\cdot)$ are greater than 0. ∎

Combining the properties established above one arrives at the equilibrium pricing function of Proposition 2.2. By Proposition 2.5 and the results in Proposition 2.8 it is straightforward to verify that $P(\widehat{f}_t^A, g_t^A, K_t)$ satisfies the following properties:

$$\mathcal{L}P \;=\; 0 \text{ if } g_t^A < 0$$

$$\mathcal{L}P \;\leq\; 0 \text{ if } g_t^A \geq 0$$

where:

$$LP = \max_i\left(\mathcal{A}P + fK - i\left(p + \frac{\chi}{2}i\right)\right) \leq 0$$

and

$$AP = \frac{1}{2}\frac{\sigma_f^2}{\overline{f}}\widehat{f}_t^A P_{ff} - \lambda(\widehat{f}_t^A - \overline{f})P_f + \frac{1}{2}\sigma_g^2 P_{gg} - \rho g P_g + P_K(-\delta K_t + i_t) - rP$$

Moreover $P(\widehat{f}_t^A, g_t^A, K_t)$ is C^1 everywhere and C^2 except at $g_t^A = 0$. Consider now any policy i_t and a stopping time τ. Then Ito's Lemma implies:

$$e^{-r\tau}P_\tau = P_0 + \int_0^\tau e^{-rt}APdt + \int_0^\tau dM_t$$

where dM_t is a (local) martingale. Since $P_t \geq 0$, one can conclude that $E\left(\int_0^\tau dM_t\right) \leq 0$ and thus:

$$E\left(e^{-r\tau}P_\tau\right) \leq P_0 + E\left[\int_0^\tau \mathcal{A}P + \widehat{f}_t^A K_t - i_t\left(p + \frac{\chi}{2}i_t\right)dt\right] - E\left[\int_0^\tau \widehat{f}_t^A K_t - i_t\left(p + \frac{\chi}{2}i_t\right)dt\right]$$

Then the following set of inequalities follows

$$P_0 \geq P_0 + E\left[\int_0^\tau \mathcal{L}Pdt\right] \geq P_0 + E\left[\int_0^\tau \mathcal{A}P + \widehat{f}_t^A K_t - i_t\left(p + \frac{\chi}{2}i_t\right)dt\right]$$
$$\geq E\left(e^{-r\tau}P_\tau\right) + E\left[\int_0^\tau \widehat{f}_t^A K_t - i_t\left(p + \frac{\chi}{2}i_t\right)dt\right]$$

Thus there is no set of investment policy / stopping policies that can yield more than P_0. Moreover, the conjectured investment and stopping policies turn the above inequalities into equalities. Thus, the conjectured equilibrium prices and policies form an equilibrium.

2.6.4 Proofs for section 2.3.2

Proof. *(Proposition 2.3)* Ito's Lemma for continuously differentiable functions implies that $b(g_t^A)$ satisfies:
$$e^{-(r+\delta)\Delta}b_{t+\Delta} = b_t + \int_t^{t+\Delta} e^{-(r+\delta)(s-t)}(\mathcal{A}b_s)\,ds + M_{t+\Delta} \qquad (2.79)$$

where M_{t+T} is a martingale difference satisfying:

$$E\left[M_{t+\Delta}|\mathcal{F}_t\right] = 0$$

and $\mathcal{A}b_s$ is given as:

$$\frac{\sigma_g^2}{2}b_g - \rho g b_g - (r+\delta)b$$

It is easy to show by using the definition of $y_1(\cdot)$ that:

$$\mathcal{A}b_s = -\frac{(r+\delta+\rho)}{r+\delta+\lambda} g_t^A 1\{g_t^A > 0\} < 0$$

Combining this with the fact that:

$$q_t = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f}_t^A - \overline{f}}{r+\delta+\lambda} + b\left(g_t^A\right)$$

leads to the first assertion. The last assertions can be proved by using the results in Fournie et. al (1999). In particular:

$$\frac{\partial Z}{\partial g} = E\left[\int_t^{t+\Delta} e^{-(r+\delta+\rho)(s-t)} 1\{g_s^A > 0\} ds\right]$$

which is clearly positive. A similar method can be used to show that $\frac{\partial Z}{\partial \sigma_g} > 0$. ∎

2.6.5 Proofs for section 2.3.2

Proof. The proof is a straightforward application of Ito's Lemma to

$$P_t = V_t + s_t$$

taking into account equation (2.76). To establish the rest of the result one can focus only on:

$$\widetilde{\Xi} = CE^A\left[\int_t^{t+\Delta} e^{-r(s-t)} g_s^A \left(\widehat{f}_s^A - \overline{f}\right) 1\{g_s^A > 0\} ds | \mathcal{F}_t\right]$$

and pass the expectation inside the integral and use the independence of $\widehat{f}_s^A$ and g_s^A to get that:

$$\widetilde{\Xi} = C\left[\int_t^{t+\Delta} e^{-r(s-t)} \left(E^A\left(\widehat{f}_s^A - \overline{f}\right) | \mathcal{F}_t\right) \left(E^A g_s^A 1\{g_s^A > 0\} | \mathcal{F}_t\right) ds\right]$$

and since

$$E^A\left(\hat{f}_s^A - \bar{f}|\mathcal{F}_t\right) = \left(\hat{f}_t^A - \bar{f}\right)e^{-\lambda s}$$

one gets:

$$\widetilde{\Xi} = \left(\hat{f}_t^A - \bar{f}\right)C\left[\int_t^{t+\Delta} e^{-(r+\lambda)(s-t)}\left(E^A g_s^A 1\{g_s^A > 0\}|\mathcal{F}_t\right)ds\right]$$

so that:

$$\widetilde{\Xi}_{\hat{f}_t^A} = C\left[\int_t^{t+\Delta} e^{-(r+\lambda)(s-t)}\left(E^A g_s^A 1\{g_s^A > 0\}|\mathcal{F}_t\right)ds\right] > 0$$

This term is of the same form as the term obtained in Proposition 2.3 and the rest of the results can be proved in an identical manner. ∎

2.6.6   Proofs for section 2.3.2

In order to give a proof of (2.24) it is useful to start by modelling the evolution of the capital stock for discrete time intervals:

$$K_T = K_t e^{-\delta(T-t)} + \int_t^T e^{-\delta(T-s)}i_s ds + \varepsilon_{it}$$

where ε_{it} captures adjustment cost shocks[69] and satisfies a strict exogeneity condition:

$$E\left(\varepsilon_{it}|q_{t=0..T}\right) = 0$$

so that:

$$E\left(K_T - K_t e^{-\delta(T-t)}|\mathcal{F}_t\right) = E\left(\int_t^T e^{-\delta(T-s)} i_s ds|\mathcal{F}_t\right) = \qquad (2.80)$$

$$= \frac{1}{\chi} E\left(\int_t^T e^{-\delta(T-s)} q_s ds|\mathcal{F}_t\right) + \widetilde{C} \qquad (2.81)$$

where $\widetilde{C} = -\frac{p \int_t^T e^{-\delta(T-s)} ds}{\chi}$ .Now if $q_t = \frac{\overline{f}}{r+\delta} + \frac{\widehat{f^A} - \overline{f}}{r+\delta+\lambda}$ (i.e. marginal q is equal to the long run fundamental notion of q) it is easy to demonstrate that

$$E\left(\int_0^T e^{-\delta(T-s)} q_s ds|\mathcal{F}_t\right) = B_1 + B_2 q_t \qquad (2.82)$$

where $B_2 \to \int_t^T e^{-\delta(T-s)} ds$ as $\lambda \to 0$ and $B_2 \to 0$ as $\lambda \to \infty$. This is in essence the Barnett and Sakellaris (1999) critique. Normalizing $T - t = 1$, it will be the case that $B_2 < 1$ and thus the estimate that will be obtained in a regression of i_t on *beginning of period* marginal "q"

---

[69]I haven't modelled adjustment costs and time variation in capital prices explicitly. Such a modification is easy to do. One just assumes that the adjustment cost technology is given by:

$$\frac{\chi}{2}\left(i_t + n_t\right)^2$$

where n_t is some stochastic process. Similarly one can introduce variability in prices by modifying the dividend stream to:

$$dD_s - \frac{\chi}{2}\left(i_t + n_t\right)^2 - p_t i_t$$

If adjustment costs are independent of the capital stock all of these modifications affect the rents to the adjustment technology only.

will produce a downwards biased estimate of $\frac{1}{\chi}$ in the sense that OLS will consistently estimate $B_2\frac{1}{\chi}$. If depreciation (δ) is small and fundamentals persist for a while, (i.e λ is small) then B_2 will be not much different than 1.[70]

Under the assumption that investment only reacts to long run fundamental q one can rewrite (2.81) as:

$$E\left(I_t|\mathcal{F}_{t-1}\right) = \frac{1}{\chi}Q_{t-1}^F + \widetilde{C} \tag{2.83}$$

where -in order to simplify notation- I have defined:

$$I_t = K_t - K_{t-1}e^{-\delta}$$

and

$$Q_{t-1}^F = E\left(\int_{t-1}^t e^{-\delta(t-s)}q_s^F ds|\mathcal{F}_{t-1}\right)$$

As demonstrated before, Q_{t-1}^F can be expressed as $B_1 + B_2 q_{t-1}$ under the assumptions of the model. Therefore one can rewrite (2.83) as

$$E\left(I_t|\mathcal{F}_{t-1}\right) = \frac{1}{\chi}B_1 q_{t-1}^F + \widetilde{C} \tag{2.84}$$

---

[70]In the presence of predictability this issue becomes even more involved because equation (2.19) has to be augmented by terms involving g_t^A which is correlated with beginning of period q_t. One can base a test on this fact by testing the orthogonality between beginning of period q_t and the error in the regression of investment on beginning of period q_t as proposed by Chirinko and Schaller (1996) equation (15). However, this is a test of whether bubbles exist, not whether they influence investment. It also appears to be less powerful than a test based on equation (2.19). The reason is that predictability in the Chirinko and Schaller (1996) test is multiplied by $\frac{1}{\chi}$ and the error term consists of the (possibly biased expectations error) and the adjustment cost shock. These two facts might make it difficult to observe predictability even if it exists. A practical way to obtain consistent estimates of χ is to approximate $\int_t^T e^{-\delta(T-s)}q_s ds$ by a weighted average of beginning and end of period q_t , project this quantity on beginning of period quantities and then use the predicted values in the regression. In other words to estimate two stage least squares. I used such an approach too in the empirical section of the paper and the results were unaltered.

Leading this once, one gets:

$$E\left(I_{t+1}|\mathcal{F}_t\right) = \frac{1}{\chi}B_1 q_t^F + \widetilde{C} \tag{2.85}$$

However notice that q_t^F and q_{t-1}^F are related by:

$$q_{t-1}^F = E\left[\int_{t-1}^t e^{-(r+\delta)(s-(t-1))}\widehat{f}_s ds + e^{-(r+\delta)}q_t^F|\mathcal{F}_t\right] \tag{2.86}$$

so that (2.84),(2.85), and (2.86) are related by:

$$E\left[I_t - e^{-(r+\delta)}I_{t+1} - \frac{1}{\chi}B_1\int_{t-1}^t e^{-(r+\delta)(s-(t-1))}\widehat{f}_s ds + \widetilde{C}(1 - e^{-(r+\delta)})|\mathcal{F}_{t-1}\right] = 0$$

For small r, δ this Euler Equation can be well approximated in terms of the observed profit rate:

$$E\left[I_t - e^{-(r+\delta)}I_{t+1} - \frac{1}{\chi}B_1\pi_t + \widetilde{C}(1 - e^{-(r+\delta)})|\mathcal{F}_{t-1}\right] = 0 \tag{2.87}$$

where the profit rate is given as:

$$\pi_t = \int_{t-1}^t \frac{dD_s}{K_s} = \int_{t-1}^t f_s ds + \sigma_D \int_{t-1}^t dZ_s^D$$

The results in the text follow once one defines : $C = \widetilde{C}(1 - e^{-(r+\delta)})$ and $\frac{1}{\chi}B_1 = \frac{1}{\widetilde{\chi}}$

Generalizing to arbitrary linear homogenous adjustment cost technologies

To generalize the results to arbitrary linear homogenous adjustment cost technologies and an arbitrary number of investor groups it will be most useful for expositional reasons to consider a discrete time setup and focus on a quadratic adjustment cost function for simplicity. Moreover I will assume 0 depreciation and a price of investment of 1. Once again the equilibrium price,

98

investment and selling times will have to satisfy:

$$V_t \stackrel{d}{=} D_t + P_t = \max_{j \in J} \sup_{\tau, i_s} E^j \left(\sum_{s=0}^{\tau} d^s D_s + d^\tau V_\tau \right) \tag{2.88}$$

where $j \in J$ is indexing the various groups of investors who have heterogenous beliefs, $d = \frac{1}{1+r}$, D_t is defined as

$$D_t = \left[f_t - \frac{i_{t+1}}{K_t} - \frac{\chi}{2} \left(\frac{i_{t+1}}{K_t} \right)^2 \right] K_t$$

and:

$$K_{t+1} = K_t + i_{t+1}$$

I will also assume that investment is determined at the beginning of the period. The basic idea is to show the following:

Lemma 2.4 *A set of prices, investment policies and stopping policies satisfies (2.88) if and only if it satisfies:*

$$V_t = D_t + P_t = \max_{j} \sup_{i_{t+1}} E^j \left(\left[f_t - \frac{i_{t+1}}{K_t} - \frac{\chi}{2} \left(\frac{i_{t+1}}{K_t} \right)^2 \right] K_t + dV_{t+1} \right) \tag{2.89}$$

Proof. Lemma (2.4). The proof is a generalization of the result shown in Harrison and Kreps (1978) and is available upon request. ∎

With Lemma (2.4) the rest of the steps follow essentially standard arguments. One can show that marginal q is equal to average q, where q is now given by the recursion:

$$q_t = E^{j_t^*} d \left(f_{t+1} + \frac{\chi}{2} \left(\frac{i_{t+2}}{K_{t+1}} \right)^2 + q_{t+1} \right)$$

, j_t^* is given as

$$j_t^* = \arg\max_j \sup_{i_t} E^j \left(\left[f_t - \frac{i_{t+1}}{K_t} - \frac{\chi}{2} \left(\frac{i_{t+1}}{K_t} \right)^2 \right] K_t + dV_{t+1} \right)$$

and optimal investment is:

$$\frac{i_{t+1}}{K_t} = \frac{q_t - 1}{\chi}$$

From here it is not difficult to derive that marginal and average q are equal by standard arguments (see e.g Chirinko (1993)). Suppose that now one were to define:

$$\Pi_t = f_t + \frac{\chi}{2} \left(\frac{i_{t+1}}{K_t} \right)^2 = \frac{\partial D_t}{\partial K_t}$$

If one assumes the presence of a rational agent A in the model it is immediate that under her beliefs it will no longer be the case that:

$$E^A \left[q_t - d \left[\Pi_{t+1} + q_{t+1} \right] | \mathcal{F}_t \right] = 0 \tag{2.90}$$

nor that:

$$E^A \left[\frac{i_{t+1}}{K_t} - d \left(\frac{i_{t+2}}{K_{t+1}} + \frac{1}{\chi} \Pi_t \right) + C | \mathcal{F}_t \right] = 0 \tag{2.91}$$

However, if investment is determined by a long termist rational investor (2.91) will hold even if (2.90) fails for the reasons explained in the text.

2.6.7 Proofs for section 2.4.4

For this section I use the standard assumption in the literature that adjustment costs are linear homogenous in capital, time is discrete and the adjustment cost technology contains both time

and individual fixed effects. Then by steps similar to 2.3.2 one can derive that:

$$E\left[\frac{I_{i,t}}{K_{i,t-1}} - e^{-(r+\delta)}\frac{I_{i,t+1}}{K_{i,t}} - \left(\alpha_i(1 - e^{-(r+\delta)}) + \zeta_t - e^{-(r+\delta)}\zeta_{t+1} + \frac{1}{\chi}\pi_{i,t}\right)|\mathcal{F}_{t-1}\right] = 0 \quad (2.92)$$

which can be rewritten as:

$$\frac{I_{i,t+1}}{K_{i,t}} = e^{(r+\delta)}\frac{I_{i,t}}{K_{i,t-1}} - \widetilde{\alpha}_i(e^{(r+\delta)} - 1) - \zeta_t e^{(r+\delta)} + \zeta_{t+1} - \frac{1}{\chi}e^{(r+\delta)}\pi_{i,t} + \varepsilon_{it}$$

where

$$E\left[\varepsilon_{it}|\mathcal{F}_{t-1}\right] = 0 \quad (2.93)$$

In particular:

$$E\left[\varepsilon_{it}|q_{1..t-1}\right] = 0$$

I am going to formulate the test in first differences in order to eliminate fixed effects, so that:

$$\Delta\left(\frac{I_{i,t+1}}{K_{i,t}}\right) = e^{(r+\delta)}\Delta\left(\frac{I_{i,t}}{K_{i,t-1}}\right) - \zeta - e^{(r+\delta)}\kappa(\Delta\pi_{i,t}) + \varepsilon_{it+1} - \varepsilon_{it} \quad (2.94)$$

where I have defined for convenience:

$$\kappa = \frac{1}{\chi}$$
$$\zeta = \Delta\zeta_{t+1} - \Delta\zeta_t e^{(r+\delta)}$$

The unknown parameters are $e^{(r+\delta)}$, ζ and κ. If one knew these parameters then one could determine $\varepsilon_{it+1} - \varepsilon_{it}$. To utilize the entire sample I estimate both κ and $e^{-(r+\delta)}$ from the system

101

of Euler relations:

$$q_{i,t} = e^{-(r+\delta)} E\left[\pi_{i,t} + q_{i,t+1} | \mathcal{F}_t\right] \qquad (2.95)$$

$$\frac{I_{i,t}}{K_{i,t-1}} = \alpha_i + \zeta_t + \kappa q_{i,t-1} + \varepsilon_{it}$$

estimated on all data in the control group.[71] Then one can substitute the estimated parameters into (2.94) to get:

$$\Delta\left(\frac{I_{i,1926}}{K_{i,1925}}\right) = \widehat{e^{(r+\delta)}}\Delta\left(\frac{I_{i,1925}}{K_{i,1924}}\right) - \zeta - \widehat{\kappa e^{(r+\delta)}}(\Delta\pi_{i,1925}) + \varepsilon_{i1926} - \varepsilon_{i1925}$$

To test the hypothesis of interest one can modify this equation to:

$$y \stackrel{d}{=} \Delta\left(\frac{I_{i,1926}}{K_{i,1925}}\right) - \widehat{e^{(r+\delta)}}\Delta\left(\frac{I_{i,1925}}{K_{i,1924}}\right) + \widehat{\kappa}e^{(r+\delta)}(\Delta\pi_{i,1925}) = \beta q_{i,1924} + \zeta + \varepsilon_{i1926} - \varepsilon_{i1925} \quad (2.96)$$

According to H_1, β should be 0. Moreover, q_{1924} should be orthogonal to the errors, so that an OLS regression of y on $q_{i,1924}$ will produce a consistent estimate of β. Of course standard errors need to be adjusted for the first step error. I undertake this adjustment by using the results in Newey and McFadden (2000) section 6.

---

[71]I use only the control group to simplify the computation of standard errors for the two step estimator of β.

Chapter 3

Saving and Investing for Early Retirement: A Theoretical Analysis (joint with E. Farhi)

3.1 Introduction

> Two years ago, when the stock market was soaring, 401(k)'s were swelling
> and (..) early retirement seemed an attainable goal. All you had to do was
> invest that big job-hopping pay increase in a market that produced double-
> digit gains like clockwork, and you could start taking leisurely strolls down
> easy street at the ripe old age of, say, 55. (Business Week 31 Dec 2001)

The dramatic rise of the stock market between 1995 and 2000 significantly increased the proportion of workers opting for early retirement.[1] The above quote from Business Week demonstrates the reasoning behind the decision to retire early: a booming stock market raises the amount of funds available for retirement and allows a larger fraction of

---

[1] See Gustman and Steinmeier (2002)

the population to exit the workforce prematurely. As a matter of fact, retirement savings seem to be one of the primary motivations behind investing in the stock market for most individuals.

Despite the growing relevance of early retirement, the existing literature has not studied the interactions between retirement, portfolio and consumption choice explicitly. In this paper we develop a theoretical model to address these interactions in a utility maximizing framework. We assume that agents are faced with a constant investment opportunity set and a constant wage rate while they are at work. Their utility exhibits constant relative risk aversion and is nonseparable in leisure and consumption. The major point of departure from preexisting literature is that we model the labor supply choice as an optimal stopping problem: an individual can work for a fixed amount of time and earn a constant wage but is free to exit the workforce (forever) at any time she chooses. In other words, we assume that workers can work either full time or be retired. Individuals are faced accordingly with three decisions: 1) How much to consume 2) How to invest the savings and 3) When to retire. The incentive to retire comes from a jump in their utility function, once they stop working, due to an increase in leisure. If retired, they cannot return to the workforce. We also consider two extensions of the basic framework. In the first extension we disallow the agent to choose retirement past a prespecified deadline. In a second extension we disallow her to borrow against the NPV of her human capital.

The major results that we obtain can be summarized as follows:

First, we show that the agent will enter retirement when a certain wealth threshold is reached. In this sense, wealth plays a dual role in our model. Not only does it determine the resources available for future consumption, it also controls the "distance" to retirement.

Second, the option to retire early strengthens the incentives to save compared to the case where early retirement is not allowed. The reason is that saving not only increases consumption in the future but also brings retirement "closer". Moreover, this incentive is wealth dependent: as the individual approaches the critical threshold, the "option" value

104

of retiring early becomes progressively more important and the saving motive becomes stronger.

Third, the marginal propensity to consume (mpc) out of wealth declines as wealth increases and early retirement becomes more likely. The intuition is simple: an increase in wealth will bring retirement closer and as a result will decrease the remaining time of the individual in the workforce. Similarly, a decline in wealth will postpone retirement. Thus, changes in wealth are somewhat counterbalanced by the behavior of the remaining NPV of income and thus the effect of a marginal change in wealth on consumption becomes attenuated. Once again this attenuation is strongest for rich individuals who are closer to their goal of early retirement.

Fourth, the optimal portfolio is tilted more towards stocks compared to the case where early retirement is not allowed. An adverse shock in the stock market will be absorbed by postponing the retirement time. Thus, the individual is more inclined to take risks, because she can always postpone her retirement time instead of cutting back her consumption if the stock market drops. Moreover, in order to bring retirement closer, the most effective way is to invest the extra savings in the stock market instead of the bond market.

Most of these results can be encapsulated in option pricing terminology: the ability to time retirement is like an American option that is most valuable when the likelihood of exercising it is high. This in turn will be the case when wealth is high. If we allow for a mandatory retirement deadline, then the value of this option also depends on the time to its expiration.

This paper is related to a number of papers in the literature, which is nicely surveyed in Ameriks and Zeldes (2001). The paper closest to ours is Bodie, Merton, and Samuelson (1992) (henceforth BMS). The major difference between BMS and this paper is the different assumption about the ability of agents to adjust their labor supply. In BMS labor can be adjusted in a continuous fashion. However, there seems to be a significant amount of evidence that labor supply is to a large extent indivisible. Most workers work

105

either full time or not at all, when they are retired. As BMS claim in the conclusion of their paper

> Obviously, the opportunity to vary continuously one's labor without cost
> is a far cry from the workings of actual labor markets. A more realistic
> model would allow limited flexibility in varying labor and leisure. One current
> research objective is to analyze the retirement problem as an optimal stopping
> problem and to evaluate the accompanying portfolio effects.

There are at least two major directions in which our results differ from BMS. First, we show that the optimal stopping decision introduces an option-type element in the decision of the individual, which is entirely absent if labor is adjusted continuously. Second, the horizon and wealth effects on portfolio and consumption choice in our paper are fundamentally different than in BMS. For instance, the holdings of stock in BMS are a *constant multiple* of the sum of (financial) wealth and human capital. This multiple is not constant in our setup, but instead depends on wealth[2]. Third, in our setup we can calibrate the parameters of the model to observed retirement decisions. In the BMS framework calibration to microeconomic data is harder, since individuals do not seem to adjust their labor supply continuously.

Moreover, the model as presented here exhibits a closer resemblance to some actual retirement systems observed in the US and other countries and can provide a "first best" benchmark against which one could measure the effectiveness of defined contribution plans, 401(k)'s etc. and their implications for optimal retirement and savings decisions. We elaborate on the necessary modifications in the conclusions to this paper. At a larger scale, one can use the insights of this model to understand why countries with larger flexibility in terms of the retirement decisions (like the US) can be expected to display increased market participation compared to countries with less flexibility (like Europe). In conclusion, the fact that labor supply flexibility is modeled in a more realistic way

---

[2] If we impose a retirement deadline, this multiple also depends on the distance to this deadline.

allows a closer mapping of the results of the model to real world institutions than is allowed by a model that exhibits continuous choice between labor and leisure.

Technically, the paper uses methods proposed by Karatzas and Wang (2000) for solving optimal consumption problems with discretionary stopping. The extension that we consider in sections 5 uses some ideas proposed in Carr, Jarrow and Myerni (1992) and Barone-Adesi and Whaley (1987), while section 6 extends the framework in He and Pages (1993) to allow for early retirement.

The role of labor supply flexibility is considered in Basak (1999) in a general equilibrium model with continuous labor/leisure choice. It is conceivable that the results presented in this paper could form the basis for a general equilibrium extension. As a matter of fact, it is well known in the macroeconomics literature that allowing indivisible labor is quite important if one is to explain the volatility of employment relative to wages. See e.g. Hansen (1985) and Rogerson (1988).

The model is also related to a strand of the literature that studies retirement decisions. A partial listing would include Stock and Wise (1990), Rust (1994), Laezar (1986), Rust and Phelan (1997), Diamond and Hausman (1984). In our setup it is possible to study retirement explicitly. We believe that the closed form solutions that we provide for the optimal retirement time could be useful in structural estimations. Furthermore, the model allows one to calibrate the parameters to commonly estimated hazard rate functions for retirement.

Some results of this paper share some similarities with results obtained in the literature on consumption and savings in incomplete markets. A highly partial listing would include Viceira (2001), Chan and Viceira (2000), Koo (1998), Caroll and Kimball (1996) on the role of incomplete markets and He and Pages (1993) and El Karoui and Jeanblanc Pique (1998) on issues related to the inability of individuals to borrow against the NPV of their future income. This literature produces some insights on why consumption as a function of wealth should be concave and has some implications on life cycle portfolio choice. However, the intuitions are quite different from the ones analyzed in this paper.

107

In the present paper the results are driven by an option component in agent's choices that is related to their ability to adjust their time of retirement. In the incomplete markets literature results are driven by agents' inability to effectively smooth their consumption due to missing markets.[3]

As this paper was completed we became aware of independent work by Dybvig and Liu (2003), who study a very similar model to the one in section 6 of this paper. Dybvig and Liu (2003) are using this model to compare institutional frameworks with voluntary vs. mandatory retirement decisions. In this paper we are interested in determining the effects of the option to retire early on portfolios and consumption decisions. Thus, our paper is suited to study the current institutional framework while the framework in Dybvig and Liu (2003) can be used to analyze the implications of alternative institutional frameworks.

The structure of the paper is as follows: Section 2 contains the model setup. Section 3 describes the solution methodology. Section 4 describes the analytical results if one places no retirement deadline. Section 5 contains an extension to the case where retirement cannot take place past a deadline. Section 6 discusses extensions to the case where a borrowing constraint is imposed and section 7 concludes. The technical details and all proofs can be found in the Appendix.

3.2 Model Setup

3.2.1 Investment Opportunity Set

The consumer can invest in the money market, where she receives a fixed strictly positive interest rate $r > 0$. Formally, the "price" of an asset invested in the money market evolves

---

[3] Chan and Viceira (2000) combines intuitions of both literatures. However, they assume labor/ leisure choices that can be adjusted continuously.

as

$$\frac{dP_0}{P_0} = rdt$$

We place no limits on the positions that can be taken in the money market. In addition the consumer can invest in a risky security with a price per share that evolves as

$$\frac{dP_1}{P_1} = \mu dt + \sigma dB_t$$

where $\mu > r$ and $\sigma > 0$ are given constants and B is a one-dimensional Brownian motion on a complete probability space (Ω, F, P).[4] We define the discount process as

$$\gamma(t) = \frac{1}{P_0(t)} = e^{-rt}$$

and the likelihood ratio process

$$Z^*(t) = \exp\left\{ - \int_0^t \kappa dB_s - \frac{1}{2}\kappa^2 t \right\}, \quad Z^*(0) = 1$$

where κ is the Sharpe ratio

$$\kappa = \frac{\mu - r}{\sigma}$$

We finally define the state price density process (or stochastic discount factor) as

$$H(t) = \gamma(t)Z^*(t), \quad H(0) = 1$$

It is a standard result, that these assumptions imply a dynamically complete market[5]. Thus, the price of a contingent claim paying a continuous dividend stream D_t, as implied by no-arbitrage is

$$\frac{E\left[\int_t^\infty H_s D_s ds \right]}{H_t}$$

---

[4]We shall denote by $F = \{F_t\}$ the P-augmentation of the filtration generated by B.
[5]See e.g. Karatzas and Shreve (1998) Chapter 1

109

3.2.2 Portfolio and Wealth Processes

An agent chooses a portfolio process π_t and a consumption process $c_t > 0$, which are progressively measurable and satisfy the standard integrability conditions given in Karatzas and Shreve (1998) sections 1 and 3. She also receives a constant income stream y_0 as long as she works and no income stream once in retirement. Retirement is an irreversible decision. Until section 3.5 we will assume that an agent can retire at any time that she chooses.

The agent is endowed with an amount of wealth $W_0 \geq -\frac{y_0}{r}$. The portfolio process π_t is the amount invested in the risky asset (the "stock market") at time t. The rest, $W_t - \pi_t$, is invested in the money market. Short selling and borrowing are both allowed.[6] As long as the agent is working, the wealth process evolves as

$$dW_t = \pi_t \left\{ \mu dt + \sigma dB_t \right\} + \left\{ W_t - \pi_t \right\} r dt - (c_t - y_0) dt \qquad (3.1)$$

Applying Ito's Lemma to the product of $H(t)$ and $W(t)$, integrating and taking expectations we get for any stochastic time τ that is finite almost surely[7]

$$E\left(H(\tau)W(\tau) + \int_0^\tau H(s)\left[c(s) - y_0\right] ds \right) \leq W_0 \qquad (3.2)$$

---

[6]Until section 3.6 of the paper we will place no extra restrictions on the (financial) wealth process. In section 3.6 we will investigate additionally the implications of the restriction $W_t \geq 0$.

[7]In detail, we apply Ito's Lemma to $H(t)W(t)$ to get:

$$H(t)W(t) + \int_0^t H(s)c(s)ds = W_0 + \int_0^t H(s)\left[\sigma\pi - \kappa W_s\right]dB_s$$

If we impose the condition $W(t) \geq -\frac{y_0}{r}$ along with non-negativity of consumption, we see that $H(t)W(t) + \int_0^t H(s)c(s)ds$ is a local martingale bounded from below and hence a supermartingale. By the optional Sampling Theorem we get:

$$E\left(H(t)W(t) + \int_0^\tau H(s)c(s)ds \right) \leq W_0$$

for any stopping time τ that is finite almost surely.

This is the well known result that one can reduce a dynamic budget constraint of the type (3.1) to a single intertemporal budget constraint of the type (3.2). If the agent is retired the above two equations continue to hold with $y_0 = 0$.

3.2.3 Leisure, Income and the Optimization Problem

To obtain closed form solutions, we will be assuming that the consumer has a utility function of the form

$$U(l_t, c_t) = \frac{1}{a} \frac{\left(l_t^{1-a} c_t^a\right)^{1-\gamma^*}}{1-\gamma^*}, \quad \gamma^* > 0 \tag{3.3}$$

where c_t is per period consumption, l_t is leisure and $0 < \alpha < 1$. We assume that the consumer is endowed with $\bar{l}$ units of leisure. l_t can only take two values l_1 or $\bar{l}$. If the consumer is working, then $l_t = l_1$, while during retirement $l_t = \bar{l}$. We will assume that the wage rate w is constant, so that the income stream is $y_0 = w(\bar{l} - l_1) > 0$. We will normalize $l_1 = 1$. Observe also that this utility is general enough so as to allow consumption and leisure to be either complements ($\gamma^* < 1$) or substitutes ($\gamma^* > 1$).The consumer maximizes expected utility

$$\max_{c_t, \pi_t, \tau} E\left[\int_0^\tau e^{-\beta t} U(l_1, c_t) dt + e^{-\beta \tau} \int_\tau^\infty e^{-\beta(t-\tau)} U(\bar{l}, c_t) dt\right] \tag{3.4}$$

where $\beta > 0$ is the agent's discount factor.[8] The easiest way to proceed is to start backwards by solving the problem

$$U_2(W_\tau) = \max_{c_t, \pi_t} E\left[\int_\tau^\infty e^{-\beta(t-\tau)} U(\bar{l}, c_t) dt\right]$$

$U_2(W_\tau)$ is the Value function once the consumer decides to retire and W_τ is the wealth

---

[8]By standard arguments the constant discount factor β could also incorporate a constant hazard rate of death λ.

at retirement. By the principle of dynamic programming, (3.4) can be rewritten as

$$\max_{c_t, W_\tau, \tau} E \left[\int_0^\tau e^{-\beta t} U(l_1, c_t) dt + e^{-\beta \tau} U_2(W_\tau) \right] \qquad (3.5)$$

It will be convenient to define the parameter γ as

$$\gamma = 1 - a(1 - \gamma^*)$$

so that we can re-express the per-period utility function as

$$U(l, c) = l^{(1-a)(1-\gamma^*)} \frac{c^{1-\gamma}}{1-\gamma}$$

It is straightforward to show under these assumptions[9], that once in retirement the Value function becomes

$$U_2(W_\tau) = \left(\bar{l}^{1-a} \right)^{1-\gamma^*} \left(\frac{1}{\theta} \right)^\gamma \frac{W_\tau^{1-\gamma}}{1-\gamma}$$

where

$$\theta = \frac{\gamma - 1}{\gamma} (r + \frac{\kappa^2}{2\gamma}) + \frac{\beta}{\gamma}$$

We will assume throughout that $\theta > 0$[10] in order to guarantee that the Value function is well defined and we will also assume that $\beta - r < \frac{\kappa^2}{2}$, in order to guarantee that retirement takes place with probability 1. It will be convenient to redefine the continuation Value function as

$$U_2(W_\tau) = K \frac{W_\tau^{1-\gamma}}{1-\gamma}$$

where

$$K = \left(\bar{l}^{1-a} \right)^{1-\gamma^*} \left(\frac{1}{\theta} \right)^\gamma$$

---

[9]See e.g. Karatzas and Shreve (1998), Chapter 3.
[10]Observe that this is guaranteed if $\gamma > 1$

112

Since $\bar{l} > l_1 = 1$ we have that

$$K^{\frac{1}{\gamma}} > \frac{1}{\theta} \text{ if } \gamma < 1 \tag{3.6}$$

$$K^{\frac{1}{\gamma}} < \frac{1}{\theta} \text{ if } \gamma > 1 \tag{3.7}$$

3.2.4 Continuous labor/leisure choice

In order to compare the results obtained in this paper to the results in BMS, we list the solution to the intertemporal consumption/portfolio/leisure choice problem, when an agent can vary labor and leisure continuously. We observe first that the solutions presented here and in the rest of the paper will depend on 2 state variables: a) the wealth at time t and b) on the distance to mandatory retirement $(T - t)$, if we impose a mandatory retirement date. Both in this section and throughout the rest of the paper we will set $t = 0$ without any loss in generality. Accordingly, we will report c_0, π_0 as the "solutions" to the problem, which will be shorthand for $c(W_0, T), \pi(W_0, T)$. However, it is important to note that under this convention T refers to the remaining periods to retirement, not the actual time of retirement.

In particular, we will consider the following 4 cases:

1. The agent never retires, and her leisure choice is fixed at $l_t = l_1 = 1$ throughout, so that her income is given by $w(\bar{l} - l_1) = y_0$.

2. The agent retires in T periods and accordingly $l_t = \bar{l}, y_t = 0$ after T periods. While working her leisure choice is fixed at $l_t = l_1 = 1$, so that her income is given by $w(\bar{l} - l_1) = y_0$.

3. The agent never retires, and her leisure choice is determined optimally on a continuum at each point in time, so that $l_t + h_t = \bar{l}$ where h_t are the hours devoted to work and the instantaneous income is wh_t.

4. The agent retires in T periods and accordingly $l_t = \bar{l}, y_t = 0$ after T periods. Her

leisure choice -while working- is determined on a continuum at each point in time, so that $l_t + h_t = \bar{l}$ where h_t are the hours devoted to work and the instantaneous income is wh_t.

By the methods in BMS it is not difficult to show that the optimal portfolio and consumption process for these cases are given by

1.

$$\pi_0 = \pi(W_0) = \frac{1}{\gamma}\frac{\kappa}{\sigma}\left(W_0 + \frac{y_0}{r}\right) \tag{3.8}$$

$$c_0 = c(W_0) = \theta\left(W_0 + \frac{y_0}{r}\right) \tag{3.9}$$

2. For $T > 0$:

$$\pi_0 = \pi(W_0, T) = \frac{1}{\gamma}\frac{\kappa}{\sigma}\left(W_0 + \frac{y_0(1 - e^{-rT})}{r}\right) \tag{3.10}$$

$$c_0 = c(W_0, T) = \frac{\theta}{\left[\left(K^{\frac{1}{\gamma}}\theta - 1\right)e^{-\theta T} + 1\right]}\left(W_0 + \frac{y_0(1 - e^{-rT})}{r}\right) \tag{3.11}$$

3.

$$\pi_0 = \pi(W_0) = \frac{1}{\gamma^*}\frac{\kappa}{\sigma}\left(W_0 + \frac{y_0}{r}\frac{\bar{l}}{(\bar{l} - l_1)}\right) \tag{3.12}$$

$$c_0 = c(W_0) = C\left(W_0 + \frac{y_0}{r}\frac{\bar{l}}{(\bar{l} - l_1)}\right) \tag{3.13}$$

4. For $T > 0$:

$$\pi_0 = \pi(W_0, T) = \frac{1}{\gamma^*}\frac{\kappa}{\sigma}\left(W_0 + \frac{y_0(1 - e^{-rT})}{r}\frac{\bar{l}}{(\bar{l} - l_1)}\right) \tag{3.14}$$

$$c_0 = c(W_0, T) = f_2(T)\left(W_0 + \frac{y_0(1 - e^{-rT})}{r}\frac{\bar{l}}{(\bar{l} - l_1)}\right) \tag{3.15}$$

114

where C is an appropriate constant and $f_2(T)$ an appropriate function of T, *but not of W_0.*

There are at least three observations about all the solutions obtained. First, the marginal propensity to consume (mpc) out of wealth ($\frac{\partial c_0}{\partial W_0}$) is constant (as in cases 1 and 3) or depends on T only (as in cases 2 and 4). Moreover, the holdings of stocks (π_0) satisfy $\frac{\partial \pi_0}{\partial W_0} = const.$ in all cases. For the cases 1 and 3, π_0 is independent of T. For the cases 2,4 the dependence of π_0 on T is captured solely by the term e^{-rT} in equations (3.10) and (3.14). No other horizon effects (dependence of the solution of T) or wealth effects (dependence on W_0) are present in the optimal portfolio choice of individuals. We will refer to these equations throughout for comparison.

3.3   Solution

We now return to the setup of section 3.2.3 and present the solution of the joint retirement /consumption and portfolio choice problem, when retirement is an irreversible discrete decision. By the notational convention adopted in the previous section, we can set without loss of generality $t = 0$ and report the optimal policies as a function of W_0.

Proposition 3.1 *Using the constants*

$$\gamma_2 = \frac{1 - 2\frac{\beta - r}{\kappa^2} - \sqrt{(1 - 2\frac{\beta - r}{\kappa^2})^2 + 8\frac{\beta}{\kappa^2}}}{2} < 0$$

$$\lambda = \left(\frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)} \frac{y_0}{r} \right)^{-\gamma}$$

$$C_2 = \frac{\left[\frac{\gamma}{1-\gamma}\frac{(\gamma_2 - 1)}{\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right)} - 1\right]\frac{y_0}{r}}{\lambda^{\gamma_2 - 1}} > 0$$

and assuming that[11]

$$\frac{r}{\theta} \frac{\left(\frac{1-\gamma}{\gamma} + \gamma_2\right)}{\gamma_2 - 1} < 1$$

the optimal policy triplet $< \widehat{c}_t, \widehat{W_{\widehat{\tau}}}, \widehat{\tau} > $ is

a) If $W_0 < \overline{W} = \dfrac{(\gamma_2 - 1)K^{\frac{1}{\gamma}}\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\dfrac{y_0}{r}$

$$\widehat{c}_t = (\lambda^* e^{\beta t} H(t))^{-\frac{1}{\gamma}} 1\{0 \leq t < \widehat{\tau}\}$$

$$\widehat{W}_{\widehat{\tau}} = \overline{W}$$

$$\widehat{\tau} = \inf\{t : \lambda^* e^{\beta t} H(t) = \underline{\lambda}\}$$

where λ^* solves the equation

$$\gamma_2 C_2 (\lambda^*)^{\gamma_2 - 1} - \frac{1}{\theta}(\lambda^*)^{-\frac{1}{\gamma}} + \frac{y_0}{r} + W_0 = 0 \tag{3.16}$$

and accordingly is a function of W_0. The optimal consumption and the optimal port-folio as a function of W_0 are given by

$$c_0 = c(W_0) = (\lambda^*(W_0))^{-\frac{1}{\gamma}}$$

$$\pi_0 = \pi(W_0) = \frac{\kappa}{\sigma}\left(\gamma_2(\gamma_2 - 1)C_2\lambda^*(W_0)^{\gamma_2 - 1} + \frac{1}{\gamma}\frac{1}{\theta}\lambda^*(W_0)^{-\frac{1}{\gamma}}\right)$$

where the notation $\lambda^*(W_0)$ is used to make the dependence of λ^* on W_0 explicit.

b) If $W_0 \geq \overline{W} = \dfrac{(\gamma_2 - 1)K^{\frac{1}{\gamma}}\theta}{\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\dfrac{y_0}{r}$ the optimal solution is to enter retirement immediately ($\widehat{\tau} = 0$) and the optimal consumption /portfolio policy is given as in Karatzas and Shreve (1998) section 3.

---

[11] *In the appendix we show that this condition is redundant in many special cases since it is implied by $\theta > 0$. We believe that $\theta > 0$ implies this inequality more generally, but haven't been able to prove it in full generality apart from some special cases given in the appendix.*

Up to the constant λ^* which is defined implicitly as the solution to (3.16) all other quantities are given explicitly. Accordingly, for a given level of wealth W_0 at time 0 it is possible to compute the associated λ^* and this in turn allows computation of optimal consumption at time 0 $(c_0 = (\lambda^*)^{-\frac{1}{\gamma}})$ and the optimal portfolio π_0. The next sections examine the properties of this solution in detail.

3.4 Properties of the solution

3.4.1 Wealth at Retirement

The wealth at retirement is given by Proposition 3.1 as

$$\overline{W} = \widehat{W}_{\widehat{\tau}} = \frac{(\gamma_2 - 1)K^{\frac{1}{\gamma}}\theta}{\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\frac{y_0}{r}$$

As Proposition 3.1 asserts, for wealth levels higher than that, it is optimal to enter retirement, whereas for wealth levels lower than that it is optimal to remain in the workforce. Not surprisingly $\overline{W}$ is strictly positive, i.e. a consumer will never go into retirement with negative wealth since there is no more income to support consumption after retirement. Also note -as might be expected- that $\overline{W}$ is increasing in y_0. This is intuitive: The incentive to go on working is coming from the additional income and if income is low, the incentive to retire early is very high. Also, it is easy to show that the incentive to retire early is increasing in $\bar{l}$, the extra leisure in retirement.

3.4.2 Optimal Consumption

We concentrate on a consumer with wealth lower than $\overline{W}$, so that she has an incentive to continue working. Optimal consumption prior to retirement is given by Proposition 3.1 as

$$c_0 = (\lambda^*)^{-\frac{1}{\gamma}}$$

where λ^* solves (3.16)

$$\gamma_2 C_2 (\lambda^*)^{\gamma_2-1} - \frac{1}{\theta} (\lambda^*)^{-\frac{1}{\gamma}} + \frac{y_0}{r} + W_0 = 0 \qquad (3.17)$$

In the appendix we show that $\theta > 0$ implies

$$1 - \gamma_2 > \frac{1}{\gamma} \qquad (3.18)$$

It is now useful to rewrite (3.17) as

$$-\gamma_2 C_2 c_0^{\gamma(1-\gamma_2)} + \frac{1}{\theta} c_0 = W_0 + \frac{y_0}{r} \qquad (3.19)$$

A first observation is that this equation is the standard relation between consumption and the sum of financial and non-financial wealth in Merton (1971) type setups with income, except for the term $-\gamma_2 C_2 c_0^{\gamma(1-\gamma_2)}$. The difference here is that the individual has an added incentive to save (since $\gamma_2 < 0$, $C_2 > 0$) for a given level of wealth, since she wants to attain early retirement. Even though we cannot provide an explicit solution to this equation we can still calculate the marginal propensity to consume out of wealth and its derivative by using the implicit function theorem. We first define the marginal propensity to consume as

$$mpc = \frac{\partial c_0}{\partial W_0}$$

We differentiate both sides of equation (3.19) w.r.t. W_0, to get

$$\left(-\gamma\gamma_2(1 - \gamma_2)C_2 c_0^{\gamma(1-\gamma_2)-1} + \frac{1}{\theta} \right) mpc = 1 \qquad (3.20)$$

One first observation from this equation is that mpc is strictly below θ since $\gamma_2 < 0$, $C_2 > 0$. Compared to the infinite horizon problem (where one stays in the workforce forever) the marginal propensity to consume out of wealth is strictly lower due to the option value embodied in (3.17). One can also study the dependence of the mpc on

118

wealth. Differentiating once more and using equation (3.18) gives

$$mpc' = -mpc^3\left(-\gamma\gamma_2(1-\gamma_2)(\gamma(1-\gamma_2)-1)C_2c_0^{\gamma(1-\gamma_2)-2}\right) < 0$$

In other words, the marginal propensity to consume out of wealth is a decreasing function of wealth and accordingly the consumption function is concave. The marginal propensity to consume is declining in wealth because an increase in W_0 brings retirement closer and thus reduces the net present value of remaining income. This in turn happens because the consumer follows a "threshold" type policy for her retirement: If wealth is high, the time to retirement is "close" and thus the increase in W_0 is counterbalanced by the equivalent decrease in the net present value of remaining income.

Reversing signs in the above argument, it is also true that the effect of a drop in wealth on consumption will be mitigated by an increase in the net present value of remaining income. Alternatively speaking, a negative shock to wealth will only partially affect consumption. A component of the drop will just postpone plans for early retirement and this will in turn increase the net present value of income to be received in the future.

These results show one important direction in which the present model sheds some new insights into the relationship between retirement, consumption and portfolio choice: in the framework of BMS a utility function of the type (3.3) along with a continuous choice of leisure and a constant wage rate would have produced a marginal propensity to consume that is independent of wealth. (One can verify this by inspection of the solutions in subsection 3.2.4). However, with endogenous retirement, wealth has a dual role. First, as in all consumption and portfolio problems, it controls the amount of resources that are available for future consumption. Second, it controls the distance to the threshold at which retirement is optimal. This second channel is behind the behavior of the mpc analyzed above.[12]

---

[12] It is also the key factor behind the behavior of optimal portfolios that will be analyzed in a subsequent subsection

3.4.3 Optimal Retirement time

The optimal retirement time in this model can be derived easily by well known formulas about the first hitting time of a Brownian Motion. To accomplish this note that Z_t defined as $Z_t = \lambda^* e^{\beta t} H_t = \lambda^* e^{(\beta - r)t - \frac{1}{2}\kappa^2 t - \kappa B_t}$ follows the process

$$\frac{dZ_t}{Z_t} = (\beta - r)\, dt - \kappa dB_t \quad Z_0 = \lambda^*$$

and an easy application of Ito's Lemma applied to $z_t = \log(Z_t)$ yields

$$dz_t = \left(\beta - r - \frac{1}{2}\kappa^2\right) dt - \kappa dB_t, \quad z_0 = \log(\lambda^*). \tag{3.21}$$

Also recall (from Proposition 3.1) that the optimal stopping time is given by the first time the process Z_t reaches $\underline{\lambda}$, where $\underline{\lambda}$ is given by

$$\underline{\lambda} = \left(\frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)} \frac{y_0}{r} \right)^{-\gamma}$$

Expressing the above quantities in logarithms we conclude that the time to retirement is given by the first hitting time of z_t to the (lower) Barrier

$$\underline{z} \hat{=} \log\left(\underline{\lambda}\right)$$

Since we have assumed that $\beta - r < \frac{\kappa^2}{2}$, the drift term in (3.21) is strictly negative and thus the stopping time is finite almost surely. The distribution of the event that retirement has not occurred until a given time T is given by the distribution of the

running minimum of a Brownian motion with drift[13]

$$P\left\{\min(z_t, t \in [0,T]) \geq \underline{z}\right\} \;=\; N\left(\frac{-(\underline{z} - \lambda^*) + (\beta - r - \frac{1}{2}\kappa^2)T}{\kappa\sqrt{T}}\right)$$

$$-e^{\left(\frac{2(\beta-r)}{\kappa^2} - 1\right)(\underline{z} - \lambda^*)} N\left(\frac{(\underline{z} - \lambda^*) + (\beta - r - \frac{1}{2}\kappa^2)T}{\kappa\sqrt{T}}\right)$$

where $N()$ is the cumulative normal distribution.

As $W \to \overline{W}$ it is easy to show that $log(\lambda^*)$ approaches $\underline{z}$ and thus the above expression becomes essentially 0, as would be expected.

3.4.4 Optimal Wealth Process and Portfolio

Combining the optimal solution of Proposition 3.1 with the intertemporal budget constraint (3.2), we arrive at the optimal wealth process

$$W_t = \frac{E\left(\int_t^{\widehat{\tau}} H_s \left(\lambda^* e^{\beta s} H_s\right)^{-\frac{1}{\gamma}} ds\right)}{H_t} - y_0 \frac{E\left[\int_t^{\widehat{\tau}} H_s\right]}{H_t} + \frac{E\left[H_{\widehat{\tau}}\overline{W}\right]}{H_t}$$

It consists of three components: The NPV (net present value) of the optimal consumption process until the retirement time $\widehat{\tau}$, the (negative of) the remaining NPV of income to be received until time $\widehat{\tau}$ and the NPV of the amount of wealth that the consumer will have at time $\widehat{\tau}$.

In other words one can think of the wealth process as a contingent claim that pays

$$\left(\lambda^* e^{\beta t} H_t\right)^{-\frac{1}{\gamma}}, \quad \text{for } 0 \leq t < \widehat{\tau}$$

$$\overline{W} \text{ at } \widehat{\tau}$$

and is short a claim that looks like a "barrier" bond, i.e. a claim that pays a constant amount y_0 until $\widehat{\tau}$ and 0 afterwards. Recall that the barrier $\underline{z} \hat{=} \log(\underline{\lambda})$ is more likely to

---

[13]Musiela and Rutkowski (1998), p.470

get hit as the stock market is rising. As one can imagine accordingly, positive shocks to the stock market reduce the value of such a barrier bond since they bring it closer to "expiration". In other words this claim has a negative "delta" in the language of contingent claim pricing. Since it enters into the wealth process with a negative sign this claim increases the incentive to take risk in the stock market. This is another manifestation of the fact that the "option value" of work increases the willingness to increase one's exposure to the stock market.

By Proposition 3.1, the optimal portfolio is given by

$$\pi_0 = \frac{\kappa}{\sigma} \left(\gamma_2(\gamma_2 - 1)C_2 \, (\lambda^*)^{\gamma_2-1} + \frac{1}{\gamma}\frac{1}{\theta} \, (\lambda^*)^{-\frac{1}{\gamma}} \right)$$

But since

$$\gamma_2 C_2 \, (\lambda^*)^{\gamma_2-1} + \frac{y_0}{r} + W_0 = \frac{1}{\theta} \, (\lambda^*)^{-\frac{1}{\gamma}}$$

from (3.16), we obtain that

$$\pi_0 = \frac{\kappa}{\sigma}\frac{1}{\gamma} \left(W_0 + \frac{y_0}{r} \right) + \frac{\kappa}{\sigma}\gamma_2 C_2 \, (\lambda^*)^{\gamma_2-1} \left((\gamma_2 - 1) + \frac{1}{\gamma} \right)$$

The first term is equal to a standard Merton type portfolio for an infinite horizon problem as in equation (3.8). The second term is positive. To see this, notice that

$$(\gamma_2 - 1) + \frac{1}{\gamma} < 0$$

by equation (3.18)and the result follows upon observing that $\gamma_2 < 0, C_2 > 0$. Moreover, one can observe that

$$C_2 \, (\lambda^*)^{\gamma_2-1} = \frac{y_0}{r} \left[\frac{\gamma}{1-\gamma} \frac{(\gamma_2 - 1)}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma} \right)} - 1 \right] \left(\frac{\lambda^*}{\underline{\lambda}} \right)^{\gamma_2-1}$$

Thus, as $\lambda^* \to \infty$ the importance of this term disappears, whereas as $\lambda^* \to \underline{\lambda}$ this

term approaches its maximal value. It is easiest to interpret this result by observing that a) λ^* is a decreasing function of wealth (W_0) and b) $\underline{\lambda}$ is the lowest value that λ^* can attain before the agent goes into retirement[14]. In words, when an agent is very poor, the relevance of early retirement is small and thus the portfolio chosen resembles a simple Merton type portfolio. By contrast as wealth increases, so does the likelihood of early retirement and this term becomes increasingly important. The option value of work increases the incentive to take risk compared to the benchmark of an infinite horizon Merton portfolio.

It is interesting to compare the above result with the results in BMS. As is shown in formulas (3.8) through (3.14) that replicate results in BMS, the amount allocated to stocks as a fraction of total resources (financial wealth + human capital) is a constant. In our framework this fraction depends on wealth, since wealth controls both the resources available for future consumption and the likelihood of attaining early retirement. Not only does the possibility of early retirement increase the incentive to save more, it also increases the incentive of the agent to invest in the stock market because this is the most effective way to obtain this goal.

3.4.5 The correlation between consumption and the stock market

In this subsection we examine the correlation between consumption and the stock market. The results of this section demonstrate a duality between the results in section 3.4.2 and 3.4.4 and show that the correlation between the stock market and consumption is constant for agents prior to retirement, despite the fact that the mpc is non-constant. The reason is quite intuitive and can be seen by examining formula (16) in Basak (1999) which continues to be true in our setup

---

[14] Actually, it is not difficult to show by the results in the appendix that $\underline{\lambda}$ is the solution to equation (3.16) if $W_0 = \overline{W}$.

$$\mu - r = -\frac{cU_{cc}}{U_c} cov\left(\frac{dP_1}{P_1}, \frac{dc}{c}\right) + \frac{U_{ch}}{U_c} cov\left(\frac{dP_1}{P_1}, dl\right) \qquad (3.22)$$

U_{ch} is the cross partial of U w.r.t the hours worked and dl is the variation in leisure. In our setup $dl = 0$ prior to retirement. Moreover, $\mu, r, \frac{cU_{cc}}{U_c}$ are constants in our framework. Accordingly, $cov\left(\frac{dP_1}{P_1}, \frac{dc}{c}\right)$ is a constant too. It is important to note that this result was obtained solely by the fact that $dl = 0$ along with the assumption of a constant investment opportunity set and CRRA utilities. In other words the consumption-CAPM holds in this framework prior to retirement.

This fact might seem surprising in light of the results we obtained for the marginal propensity to consume out of wealth. One might expect that a declining mpc would be sufficient to produce a low correlation between the stock market and consumption. The resolution of the puzzle is that a decrease in mpc in this model is accompanied by an equivalent increase in the exposure to the stock market through a portfolio that is more heavily tilted towards stocks. In other words, even though consumption becomes less responsive to shocks in the wealth process, at the same time the shocks to the wealth process become more volatile because of a riskier portfolio.[15]

An important remark is that the above discussion relies heavily on partial equilibrium. To see if labor supply flexibility can indeed explain the observed smoothness of aggregate consumption and accordingly a large equity premium one would have to study a general equilibrium version of this model (as Basak (1999) does for continuous choice of labor/leisure). In that case a fraction of the population would be entering retirement at each instant and would be experiencing consumption changes (because they receive an increased endowment of leisure). So, at the aggregate the simple consumption CAPM

$$\mu - r = -\frac{cU_{cc}}{U_c} cov\left(\frac{dP_1}{P_1}, \frac{dc}{c}\right)$$

---

[15] This can be formally shown by applying Ito's Lemma to $c(W_0)$ together with the dynamics of the wealth process (3.1) and the optimal portfolio $\pi(W_0)$ to arrive at (3.22)

would no longer hold. It can be reasonably conjectured that in this framework the behavior of the interest rate and the equity premium would be very different than in Basak (1999). Even in the base case of CRRA utilities and multiplicative technology shocks, the equity premium and the interest rate would exhibit interesting dynamics. However, this issue is beyond the scope of the present paper.

3.4.6 The marginal propensity to consume: A simple quantitative exercise

We conclude this section with a calibration exercise. Our aim is to show how the mpc varies for various levels of wealth, if an individual is saving for early retirement. We provide more quantitative results in the next two sections, when we provide an approximate solution to the problem with a deadline for retirement and consider borrowing constraints.

We picture the situation of an individual in her 30's and choose the parameter K that captures the "eagerness" to retire so that retirement will take place with a probability of more than 60% by the time she is 65[16]. For the parameter γ of risk aversion we choose 3, while the parameters affecting the investment opportunity set are chosen as follows: $\mu = 0.07, \beta = 0.03, r = 0.03, \sigma = 0.2$.

We assume that the consumer starts with an initial wealth $W_0 = 2$ and normalize $y_0 = 1$, so that all magnitudes are in units of yearly income. With these assumptions we can plot the distribution of the retirement time which is given in Figure 3-1.

The marginal propensity to consume out of wealth is plotted in Figure 3-2. A consumer begins with a marginal propensity to consume of roughly 0.029 and this number declines to 0.019 for consumers close to retirement. As we show in a subsequent section these drops are more significant once we introduce borrowing constraints.

---

[16]For the parametric assumptions that we make, it turns out that a parameter choice of $K = \left(\frac{0.4}{\theta}\right)^\gamma$ produces such a distribution of the retirement time.
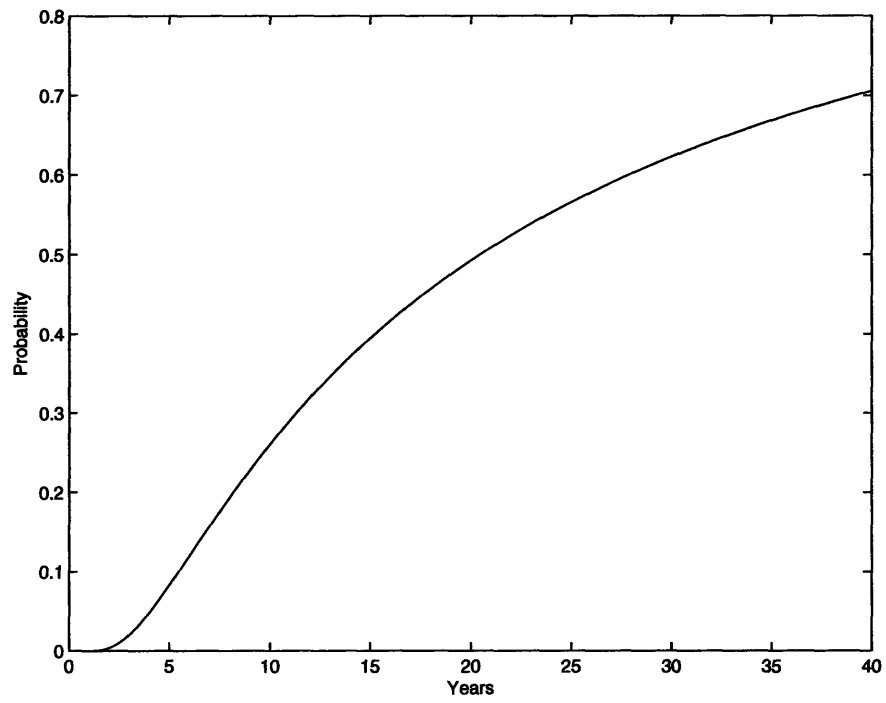
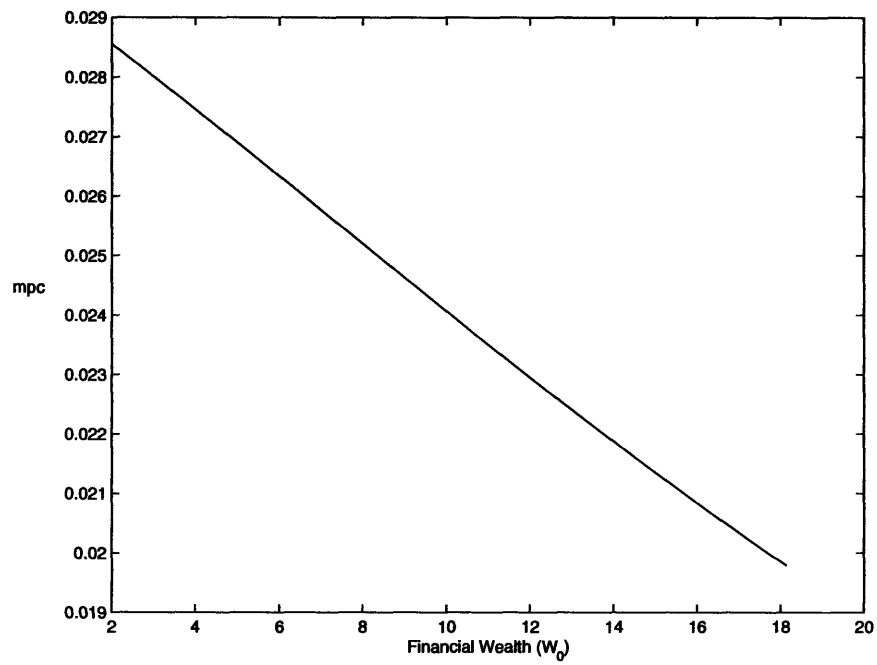Figure 3-1: Cumulative distribution of the time to retirement.

Figure 3-2: Marginal propensity to consume out of wealth. The leftmost value of wealth corresponds to $W_0 = 2$ while the rightmost value corresponds to the wealth level at which the consumer will enter retirement.

These figures suggest that a drop in the wealth of a rich individual (high W_0) will only partially affect her consumption, while it will have stronger effects on a poorer individual (low W_0). The reason is that in our framework a drop in wealth postpones retirement and thus is balanced out by an increase in the NPV of remaining income.

3.5   Retirement before a deadline

None of the claims made so far relied on restricting the time of retirement to lie in a particular interval. The exposition concentrated on the infinite horizon case because this allowed for explicit solutions to the associated optimal stopping problem. In order to analyze the case where retirement cannot take place after a prespecified deadline, one should use some approximate method to solve the associated "finite horizon"[17] optimal stopping problem. Formally, the only modification that we introduce compared to section 3.2 is that equation (3.5) becomes

$$\max_{c_t, W_\tau, \tau} E\left[\int_0^{\tau \wedge T} e^{-\beta t} U(l_1, c_t) dt + e^{-\beta \tau} U_2(W_\tau)\right]$$

where T are the remaining periods to mandatory retirement. In the appendix we first show the following analog of Proposition 3.1 that obtains a "solution" to the finite horizon problem in terms of a functional equation by using ideas proposed in Carr, Jarrow, and Myerni (1992)

Proposition 3.2 *Let T be the remaining periods to mandatory retirement. Consider the*

---

[17] An important remark on terminology: The term "finite horizon" refers to the fact that the optimal stopping region becomes a function of the deadline to mandatory retirement. The individual continues to be infinitely lived.

strictly convex function $\widetilde{V}(\lambda, T)$

$$\widetilde{V}(\lambda, T) = \lambda \frac{y_0(1 - e^{-rT})}{r} + \frac{\gamma}{1-\gamma} \lambda^{\frac{\gamma-1}{\gamma}} \left(\frac{1 - e^{-\theta T}}{\theta} + K^{\frac{1}{\gamma}} e^{-\theta T} \right)$$

$$+ \frac{\gamma}{1-\gamma}(K^{\frac{1}{\gamma}}\theta - 1)\lambda^{\frac{\gamma-1}{\gamma}} \int_0^T e^{-\theta t} N(d_{1t})\, dt$$

$$- y_0 \lambda \int_0^T e^{-rt} N(d_{2t})\, dt$$

with

$$d_{1t} = -\left(\frac{\log(\lambda) + \left(\beta - r - \frac{\kappa^2}{2}\right)t - \log(\lambda_t)}{\kappa\sqrt{t}} + \kappa^2 \left(\frac{\gamma-1}{\gamma}\right)^2 t \right)$$

$$d_{2t} = -\left(\frac{\log(\lambda) + \left(\beta - r - \frac{\kappa^2}{2}\right)t - \log(\lambda_t)}{\kappa\sqrt{t}} + \kappa^2 t \right)$$

if $\gamma > 1$ and

$$d_{1t} = \left(\frac{\log(\lambda) + \left(\beta - r - \frac{\kappa^2}{2}\right)t - \log(\lambda_t)}{\kappa\sqrt{t}} + \kappa^2 \left(\frac{\gamma-1}{\gamma}\right)^2 t \right)$$

$$d_{2t} = \left(\frac{\log(\lambda) + \left(\beta - r - \frac{\kappa^2}{2}\right)t - \log(\lambda_t)}{\kappa\sqrt{t}} + \kappa^2 t \right)$$

if $\gamma < 1$. Consider also the solution to the functional equation obtained by

$$\widetilde{V}(\lambda_t, T - t) = \frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} \lambda_t^{\frac{\gamma-1}{\gamma}}, \quad \text{for all } 0 \leq t \leq T \tag{3.23}$$

Finally, determine λ^* as the solution to

$$\widetilde{V}'(\lambda^*, T) = -W_0 \tag{3.24}$$

Then the optimal consumption, wealth at retirement and portfolio are determined as

129

a) If $W_0 < \overline{W_0} = K^{\frac{1}{\gamma}} \underline{\lambda_0}^{-\frac{1}{\gamma}}$

$$\hat{c}_t = (\lambda^* e^{\beta t} H(t))^{-\frac{1}{\gamma}} 1\{t < \hat{\tau}\}$$
$$\widehat{W_{\hat{\tau}}} = \overline{W_{\hat{\tau}}}$$
$$\hat{\tau} = \inf\{t : \lambda^* e^{\beta t} H(t) = \underline{\lambda_t}\}$$

The optimal consumption and portfolio processes as functions of W_0, T are given by

$$c_0 = c(W_0, T) = (\lambda^*)^{-\frac{1}{\gamma}}$$
$$\pi_0 = \pi(W_0, T) = -\frac{\kappa}{\sigma} \frac{\lambda^*}{\lambda^*_{W_0}}$$

b) If $W_0 \geq \overline{W_0} = K^{\frac{1}{\gamma}} \underline{\lambda_0}^{-\frac{1}{\gamma}}$ the optimal solution is to enter retirement immediately ($\hat{\tau} = 0$) and the optimal consumption /portfolio policy is given as in Karatzas and Shreve (1998) section 3.

Unfortunately, this solution to the problem is not very operational without some tractable way to solve equation (3.23). For this particular problem it turned out that an analytical approximation along the lines of Barone Adesi and Whaley (1987) seems to be performing very well in terms of actually computing the solution to (3.23) and all the associated quantities in closed form. Moreover, this approximation allows us to compare the solutions obtained in the previous section with the ones obtained here in a very direct way. The appendix establishes the following

Proposition 3.3 *Define*

$$\widetilde{V}^E(\lambda, T) = \frac{\gamma}{1-\gamma} \lambda^{\frac{\gamma-1}{\gamma}} \frac{1}{\theta} \left[\left(K^{\frac{1}{\gamma}} \theta - 1 \right) e^{-\theta T} + 1 \right] + \lambda y_0 \frac{1 - e^{-rT}}{r}$$

130

Then $\widetilde{V}(\lambda; T)$ in Proposition 3.2 is approximately given by

$$
\begin{cases}
C_{2T}\lambda^{\gamma_{2T}} + \widetilde{V}^E(\lambda, T), & \text{if } \lambda > \underline{\lambda}_0 \\[2mm]
\left(\dfrac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} (\lambda)^{\frac{\gamma-1}{\gamma}} \right) & \text{if } \lambda \leq \underline{\lambda}_0
\end{cases}
$$

where

$$
\underline{\lambda}_0 = \left(\frac{(\gamma_2 - 1)\theta_T}{\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta_T - 1\right)} \frac{y_0(1 - e^{-rT})}{r} \right)^{-\gamma}
\tag{3.25}
$$

θ_T is given by

$$
\theta_T = \frac{\theta}{\left[\left(K^{\frac{1}{\gamma}}\theta - 1\right) e^{-\theta T} + 1 \right]}
$$

C_{2T} is given by

$$
C_{2T} = \frac{\left[\frac{\gamma}{1-\gamma} \frac{(\gamma_2-1)}{\left(1+\gamma_2\frac{\gamma}{1-\gamma}\right)} - 1 \right] \frac{y_0(1-e^{-rT})}{r}}{\underline{\lambda}_0^{\gamma_2-1}}
\tag{3.26}
$$

and γ_{2T} is given by

$$
\gamma_{2T} = \frac{1 - 2\frac{\beta-r}{\kappa^2} - \sqrt{(1 - 2\frac{\beta-r}{\kappa^2})^2 + 8\frac{\beta}{(1-e^{-\beta T})\kappa^2}}}{2}
$$

This approximate $\widetilde{V}(\lambda, T)$ is continuously differentiable everywhere and $\widetilde{V}_\lambda(\lambda, T)$ maps $(0, \infty)$ into $\left(-\infty, \frac{y_0}{r}\left(1 - e^{-rT}\right)\right)$. For this approximation equation (3.24) becomes

$$
\gamma_{2T}C_{2T}(\lambda^*)^{\gamma_{2T}-1} - \frac{1}{\theta_T}(\lambda^*)^{-\frac{1}{\gamma}} + \frac{y_0(1 - e^{-rT})}{r} + W_0 = 0
\tag{3.27}
$$

Figure 3-3 gives a visual impression of the accuracy of this approximation. In particular a Binomial Tree along the lines of Cox, Ross, and Rubinstein (1978) was used in order to obtain $\underline{\lambda}_t$ numerically and the resulting solution was compared to the approximate analytical solution. The analytical approximation captures the qualitative properties of the solution to the functional equation (3.23) with good accuracy. The most important ad-

131

vantage of this approximation, is that it leads to very tractable solutions for all quantities involved. This can be seen most easily by observing that equation (3.27) is practically identical to equation (3.16). The only difference is that all the constants now depend on T. As a result all of the analysis in section 3.4 can be replicated easily for the case where a deadline for retirement is present. Economically, the only new dimension introduced is that all quantities depend explicitly on the distance to mandatory retirement, and thus one can study interaction effects between wealth and the age of an investor[18]. Here we will focus only on the implications of the model for portfolio choice.

By equation (3.10) the portfolio of an agent with mandatory retirement in T periods and no option of retirement earlier than that is

$$\pi_0^{mand.} = \pi^{mand.}(W_0, T) = \frac{1}{\sigma}\frac{\kappa}{\gamma}\left(W_0 + y_0\frac{1 - e^{-rT}}{r}\right)$$

A constant fraction $\left(\frac{1}{\sigma}\frac{\kappa}{\gamma}\right)$ of the net present value of resources available to the individual $\left(W_0 + y_0\frac{1-e^{-rT}}{r}\right)$ is invested in the stock market irrespective of her age. In this sense the portfolio exhibits no horizon effects or wealth effects beyond the ones present in the term $\left(W_0 + y_0\frac{1-e^{-rT}}{r}\right)$: if one were to divide the stock holdings $\pi_0^{mand.}$ by the sum of financial and non-financial wealth then the result would be a constant.

However, in the presence of early retirement this result is no longer true. Even after normalizing the portfolio by $\left(W_0 + y_0\frac{1-e^{-rT}}{r}\right)$ the fraction invested exhibits time and wealth effects. One can apply the same reasoning as in section 3.4.4 to arrive at the

---

[18]As implied by the distance to mandatory retirement.

optimal holdings of stock π_0 in the presence of optimal early retirement

$$\pi_0 = \frac{\kappa}{\sigma}\frac{1}{\gamma}\left(W_0 + y_0\frac{1-e^{-rT}}{r}\right) + \frac{\kappa}{\sigma}\gamma_{2T}C_{2T}\left(\lambda^*\right)^{\gamma_{2T}-1}\left((\gamma_{2T}-1)+\frac{1}{\gamma}\right) = \qquad (3.28)$$

$$= \frac{\kappa}{\sigma}\frac{1}{\gamma}\left(W_0 + y_0\frac{1-e^{-rT}}{r}\right)$$

$$+ \frac{\kappa}{\sigma}\frac{y_0\left(1-e^{-rT}\right)}{r}\left(\frac{\lambda^*}{\lambda_0}\right)^{\gamma_{2T}-1}\gamma_{2T}\left(\frac{1}{\gamma}+(\gamma_{2T}-1)\right)\left[\frac{\gamma}{1-\gamma}\frac{(\gamma_{2T}-1)}{\left(1+\gamma_{2T}\frac{\gamma}{1-\gamma}\right)} - 1\right]$$

where λ^* is given by (3.27). Once again, next to the usual Merton type portfolio in the presence of mandatory retirement in T periods, one also obtains a portfolio that is related to both time and wealth.

Figure 3-4 plots the term:

$$\widetilde{\pi}_0 = \frac{\pi_0}{W_0 + \frac{y_0(1-e^{-rT})}{r}}$$

for various levels of wealth and periods (T) until the retirement deadline is reached. Equation (3.10) implies that this term would be constant, if no possibility of early retirement was present.

There are two patterns that emerge from Figure 3-4. First, $\widetilde{\pi}_0$ is decreasing rapidly as the relevance of early retirement becomes less important, i.e. as one approaches the retirement deadline. Moreover, the optimal portfolio is increasing in the resources available to the individual. A richer person (high W_0) invests a larger percent of her total resources in the stock market.

These findings allow an interpretation in terms of option pricing intuitions. Early retirement in this model is like an option. And as most options, its value is larger a) the more likely it is that it will be exercised (in/out of the money) and b) the more time is left until its expiration. Accordingly, the importance of the second term in (3.28) is decreasing as T and/or W_0 is small. The fact that $\widetilde{\pi}_0$ depends on both wealth and the horizon makes it important to qualify one of the conclusions that is usually reached
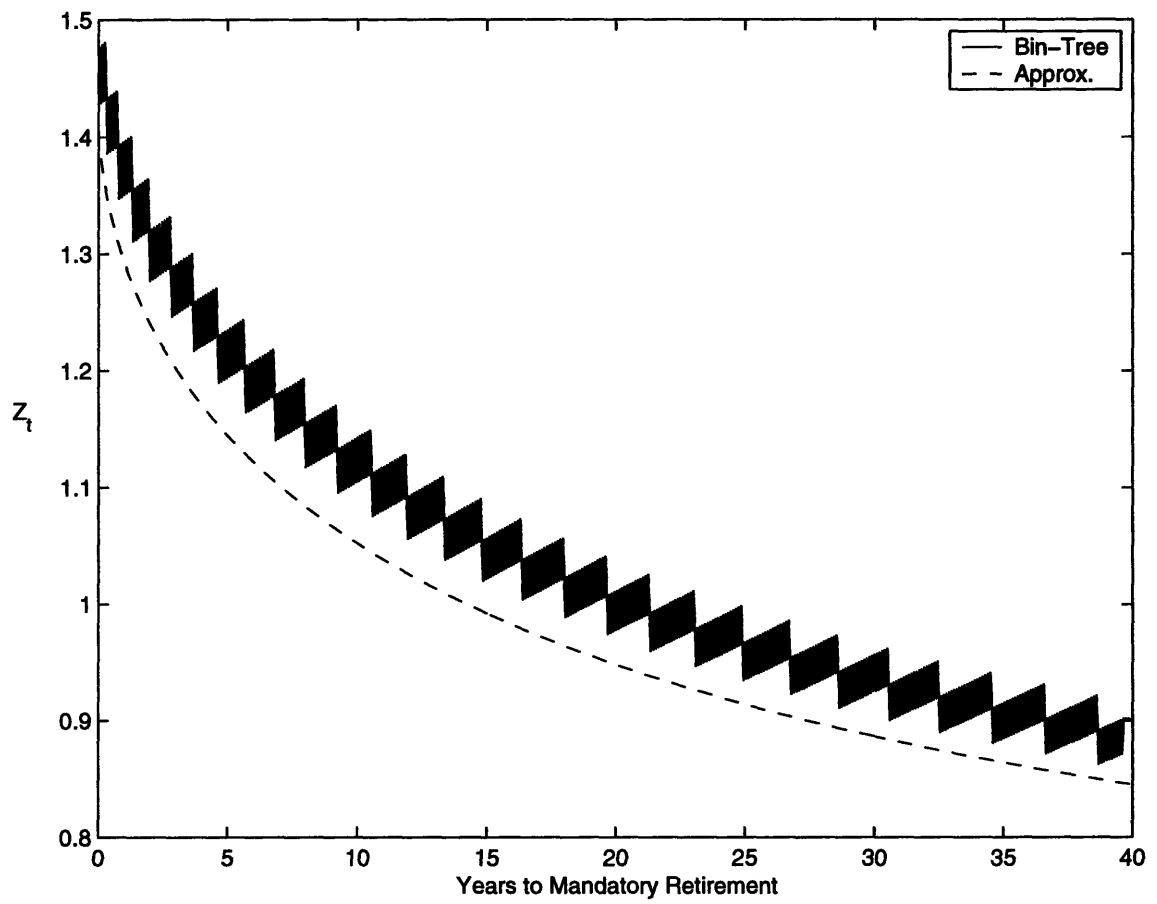
133

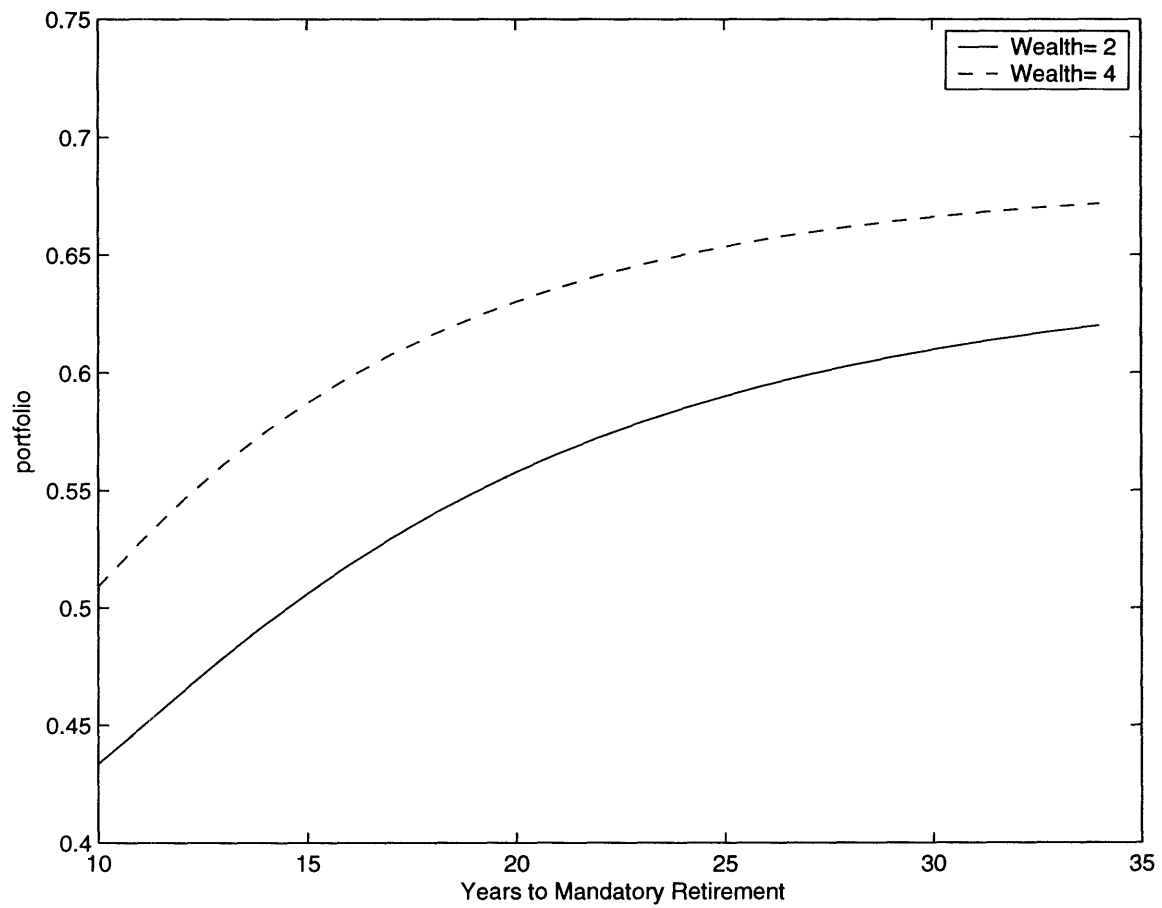Figure 3-3: Optimal exercise boundary for a binomial tree approximation and the analytical solution proposed.

Figure 3-4: Optimal portfolio $(\tilde{\pi}_0)$ as a function of time to the retirement deadline for two levels of wealth.

135

in models with continuous work/leisure choice. It is undoubtedly true that keeping W_0 and y_0 constant, a decrease in T will make an individual invest less in the stock market. However, older *but not yet retired* individuals typically have more financial wealth and as a result will have an increased incentive to invest in the stock market as a result of that. Accordingly, it could well be that a 50-year old holds more stock (both in absolute terms and relative to the sum of financial and non-financial wealth) than a 30-year old in the present model despite the fact that the one is closer to retiring than the other. By contrast in the framework of BMS the portfolio as a fraction of financial and non-financial wealth exhibits neither horizon nor wealth effects since it is a constant. One unambiguous result however, is that retirees will be holding less stock as a fraction of their total wealth than non-retirees. This is in direct analogy to BMS: taking away one margin on which individuals can adjust will necessarily lead to more risk averse behavior.

3.6 Borrowing constraint

So far, we have been assuming that the agent was able to borrow against the value of her future labor income. In this section we impose the extra restriction that it is impossible for the agent to borrow against the value of future income. Formally, we add the requirement that $W_t \geq 0$, for all $t > 0$. To preserve tractability, we assume in this section that the agent is able to go into retirement at any time that she chooses, i.e. we do not impose any deadline. This makes the problem stationary and as a result the optimal consumption and portfolio policies will be given by functions of W_0 alone. Moreover, we can set $t = 0$ without loss of generality since the optimal policies will not depend on time.

Post-retirement, the borrowing constraint is never binding because the agent receives no income and has constant relative risk aversion. This implies that once the agent is retired, her consumption, her portfolio, and her value function are the same with or without borrowing constraints. In particular, if the agent enters retirement at time τ

with wealth W_τ, her expected utility is still $U_2(W_\tau)$.

The problem of the agent is now

$$\max_{c_t, W_\tau, \tau} E\left[\int_0^\tau e^{-\beta t}U(l_1, c_t)dt + e^{-\beta\tau}U_2(W_\tau)\right] \tag{3.29}$$

subject to the borrowing constraint

$$W_t \geq 0, \forall t \geq 0, \tag{3.30}$$

and the budget constraint

$$dW_t = \pi_t\{\mu dt + \sigma dB_t\} + \{W_t - \pi_t\}rdt - (c_t - y_0 1\{t < \tau\})\,dt. \tag{3.31}$$

By arguments similar to the ones in section 3.2 we can rewrite these two constraints as

$$E\left[\int_0^\tau H_s c_s ds + H_\tau W_\tau\right] \leq E\left[\int_0^\tau H_s y_0 ds\right] + W_0,$$

$$\frac{E\left[\int_t^\tau H_s c_s ds + H_\tau W_\tau\right]}{H_t} \geq \frac{E\left[\int_t^\tau H_s y_0 ds\right]}{H_t}, \forall t \geq 0.$$

In other words, in addition to the usual budget constraint, we place the requirement that the net present value of future expenditures is always no more than the net present value of remaining income.

We present the solution in the following proposition

Proposition 3.4 *Under technical conditions (3.69) and (3.70) in the appendix[19], there exist appropriate constants $C_1, C_2, Z_L, Z_H, \gamma_1, \gamma_2$ (also given in the appendix) and a positive decreasing process X_t^* with $X_0^* = 1$ so that the optimal policy triple $< \widehat{c_t}, \widehat{W_{\widehat{\tau}}}, \widehat{\tau} >$ is*

---

[19] *We conjecture that these conditions can be shown to hold quite generally, but haven't been able to prove this. Verifying them numerically in any specific application is however straightforward.*

137

a) If $W_0 < \overline{\overline{W}} = K^{\frac{1}{\gamma}} Z_L^{-\frac{1}{\gamma}}$

$$\widehat{c}_t = (\lambda^* e^{\beta t} X_t^* H(t))^{-\frac{1}{\gamma}} 1\{t < \widehat{\tau}\}$$

$$\widehat{W}_{\widehat{\tau}} = \overline{\overline{W}}$$

$$\widehat{\tau} = \inf\{t : \lambda^* X_t^* e^{\beta t} H(t) = Z_L\}$$

and λ^* *is given by*

$$\gamma_1 C_1 (\lambda^*)^{\gamma_1 - 1} + \gamma_2 C_2 (\lambda^*)^{\gamma_2 - 1} - \frac{1}{\theta} (\lambda^*)^{-\frac{1}{\gamma}} + \frac{y_0}{r} + W_0 = 0 \qquad (3.32)$$

Using the notation $\lambda^*(W_0)$ *to make the dependence of* λ^* *on* W_0 *explicit, the optimal consumption and portfolio policy is given by*

$$c_0 = c(W_0) = (\lambda^*(W_0))^{-\frac{1}{\gamma}}$$

$$\pi_0 = \pi(W_0) = -\frac{\kappa}{\sigma} \frac{\lambda^*(W_0)}{\lambda^*_{W_0}(W_0)}$$

where $\lambda^*_{W_0}(W_0)$ *denotes the first derivative of* $\lambda^*(W_0)$ *with respect to* W_0.

b) If $W_0 \geq \overline{\overline{W}} = K^{\frac{1}{\gamma}} Z_L^{-\frac{1}{\gamma}}$ *the optimal solution is to enter retirement immediately* $(\widehat{\tau} = 0)$ *and the optimal consumption policy is given as in the standard Merton (1971) infinite horizon problem.*

In the remainder of this section we compare the results we obtained in section 3.4 to the optimal policies resulting from Proposition 3.4.

A simple intuitive argument shows that (compared to section 3.4) wealth at retirement is smaller with borrowing constraints than without: $\overline{\overline{W}} < \overline{W}$. ($\overline{W}$ is the threshold at which an individual facing no borrowing constraints goes into retirement). The reasoning is the following: Given a level of wealth $\overline{W}$, the agent will achieve the same utility $U_2(\overline{W})$ if she goes into retirement at t whether she faces borrowing constraints or not. But the expected utility the agent will achieve by postponing her retirement decision to $t + dt$ is

strictly lower if she faces borrowing constraints (the inequality is strict because there is a non-zero probability that the constraint will bind between t and $t + dt$). As a result, the value of waiting is strictly lower with borrowing constraints than without, i.e. $\overline{\overline{W}} < \overline{W}$.[20]

Figure 3-5 compares the marginal propensity to consume for the same parameter values as the ones used in Section 3.4. The figure depicts the mpc for 3 scenarios: The first scenario is one where an individual chooses her retirement time optimally on an infinite horizon without borrowing constraints. This corresponds to the case analyzed in section 3.4. The second case is one where an agent faces borrowing constraints but can never enter retirement. This corresponds to the case analyzed in He and Pages (1993). The third scenario is the one analyzed in this section, namely the situation of an individual facing both an optimal retirement decision and borrowing constraints. The wealth level is varied between 0.1 and the level at which a (borrowing unconstrained) individual enters retirement.

The basic result of figure 3-5 is that borrowing constraints amplify the effects of early retirement significantly. The behavior of the mpc for the third scenario is always between the first and the second scenarios. For large levels of wealth the presence of borrowing constraints is immaterial, and it is only the optimal stopping aspect of the problem that drives the behavior of the mpc. Thus, for large levels of wealth the behavior of the mpc is similar to the case where there are no borrowing constraints. The opposite is true for low levels of wealth. Then, the main force driving the mpc is the presence of borrowing constraints and the possibility of early retirement is immaterial.

The optimal portfolio can be computed by steps similar to section 3.4. The implicit function theorem gives

$$\frac{\lambda^*}{\lambda^*_{W_0}} = -\left(\gamma_1(\gamma_1 - 1)C_1 \, (\lambda^*)^{\gamma_1 - 1} + \gamma_2(\gamma_2 - 1)C_2 \, (\lambda^*)^{\gamma_2 - 1} + \frac{1}{\gamma}\frac{1}{\theta} \, (\lambda^*)^{-\frac{1}{\gamma}} \right) \qquad (3.33)$$

---

[20]However, this effect seems to be relatively small quantitatively. For example, with $y_0 = 1, \gamma = 3$, $\mu = 0.07, \beta = 0.03, r = 0.03, \sigma = 0.2$ and $K = \left(\frac{0.4}{\theta}\right)^{\gamma}$, we get $\overline{W} = 18.13$ and $\overline{\overline{W}} = 17.18$.
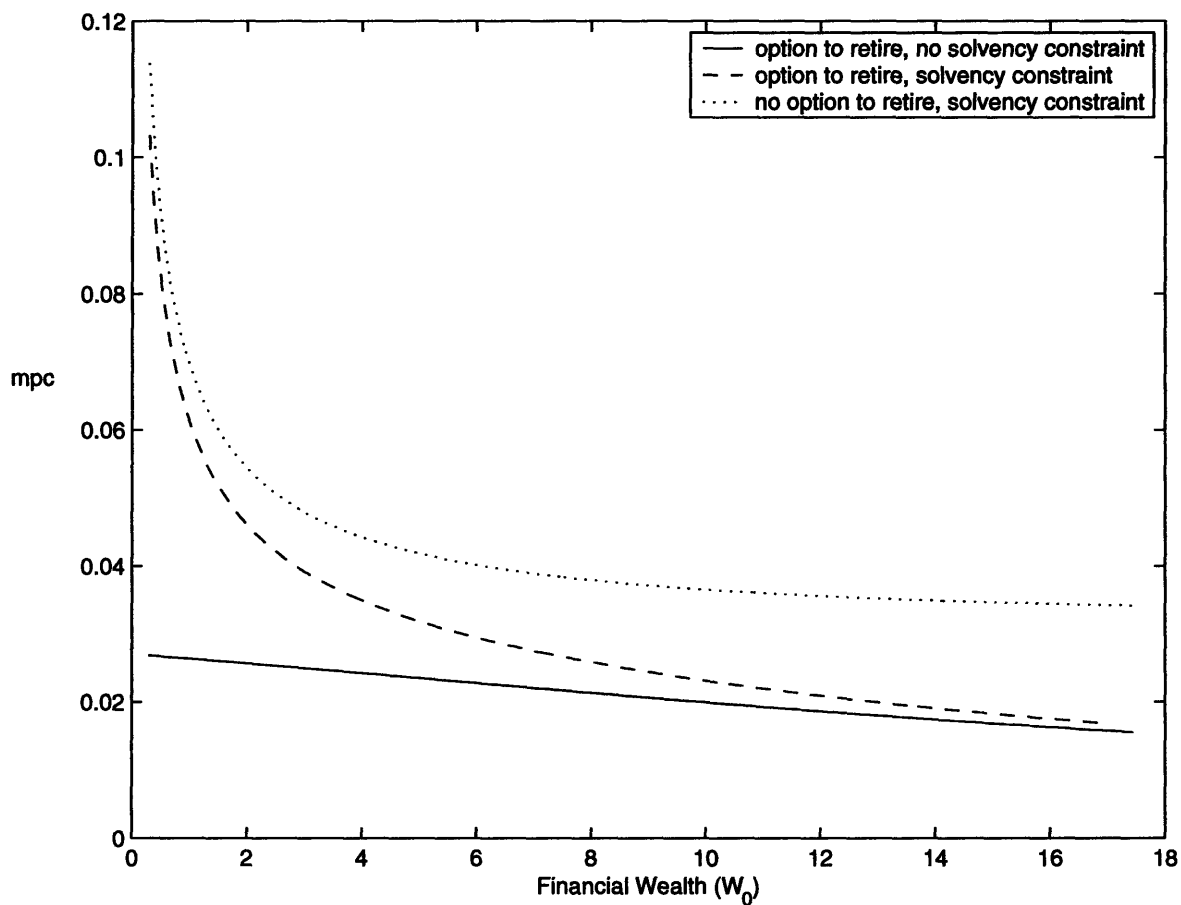
Figure 3-5: Marginal propensity to consume out of wealth. The leftmost value of wealth corresponds to $W_0 = 0.1$ while the rightmost value corresponds to the wealth level at which the consumer will enter retirement: 18.13 for the case without borrowing constraint and 17.19 for the case with borrowing constraints.

and by steps similar to section 3.4 we obtain the optimal portfolio as[21]

$$\pi_0 = \frac{\kappa}{\sigma}\frac{1}{\gamma}\left(W_0 + \frac{y_0}{r}\right) + \frac{\kappa}{\sigma}\gamma_1 C_1 (\lambda^*)^{\gamma_1 - 1}\left((\gamma_1 - 1) + \frac{1}{\gamma}\right) + \frac{\kappa}{\sigma}\gamma_2 C_2 (\lambda^*)^{\gamma_2 - 1}\left((\gamma_2 - 1) + \frac{1}{\gamma}\right).$$

The first term is equal to a standard Merton type portfolio for an infinite horizon problem. One can show that the second term is negative and decreasing in λ^* (increasing in wealth), while the third term is positive and decreasing in λ^* (increasing in wealth). π_0 is therefore decreasing in λ^* (increasing in wealth) as in section 3.4.

Figure 3-6 compares optimal portfolios for various cases. Depending on whether early retirement is allowed or not and whether borrowing constraints are imposed or not we get 4 cases. One observation is that the optimal portfolio in the presence of early retirement is more tilted towards stocks whether we impose borrowing constraints or not. Similarly, the optimal holdings of stock are smaller when one imposes borrowing constraints (whether one assumes early retirement or not). In this sense, the presence of borrowing constraints partially mitigates the incentive to invest in stock. In the presence of borrowing constraints the nominal holdings of stock for low levels of wealth approach 0.[22]

3.7 Conclusion

In this paper we proposed a simple partial equilibrium model of consumer behavior that allows for the joint determination of optimal consumption, portfolio and the retirement time of a consumer. Essentially closed form solutions were obtained for virtually all quantities of interest. The results can be summarized as follows: The ability to time one's retirement introduces an option type character to the optimal retirement decision.

---

[21] It is important to remember here that the term portfolio refers to the total holdings of stocks (in dollars).

[22] However, if one computed a portfolio as a function of wealth then that ratio would be infinite at 0 as shown by He and Pages (1993). We therefore find it more informative to speak of nominal holdings of stock.
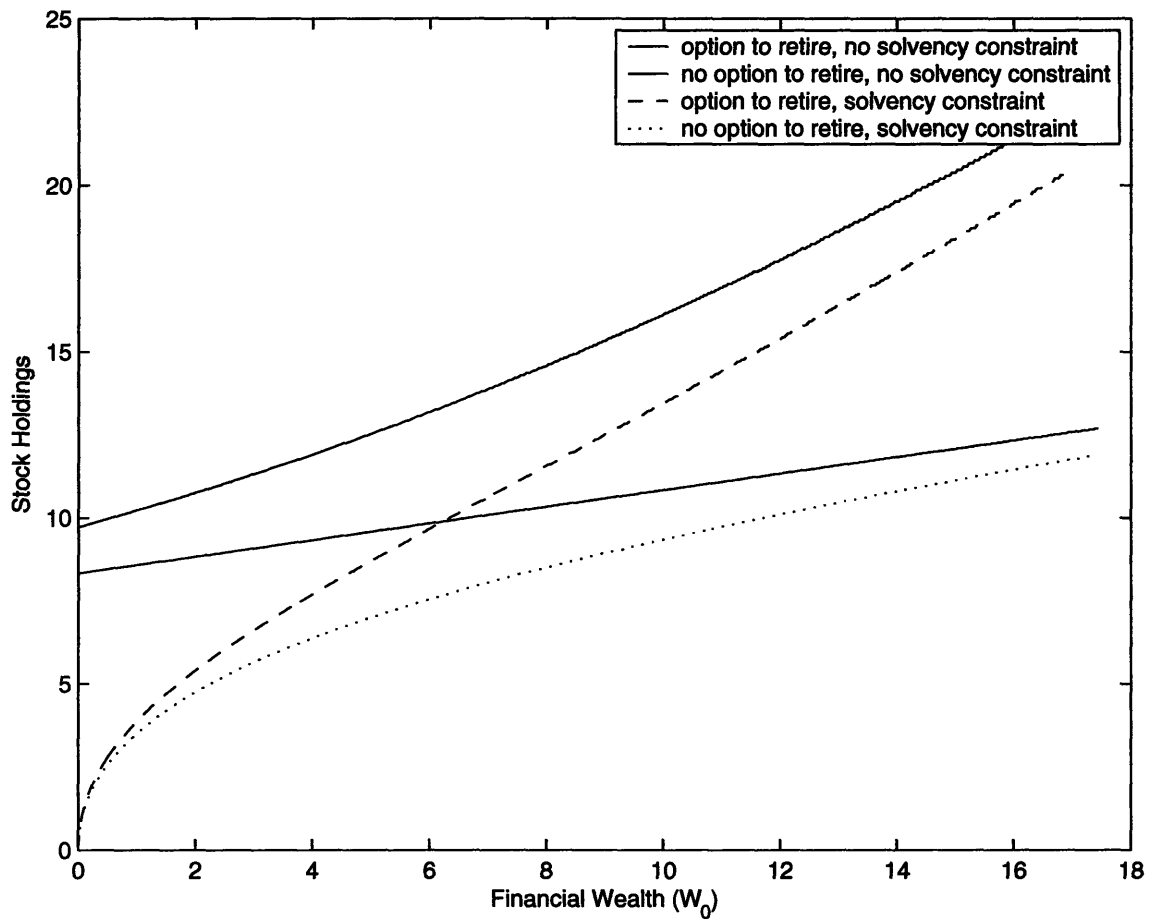
Figure 3-6: Nominal stock holdings for various combinations of constraints

This option is most relevant for individuals with a high likelihood of early retirement and affects both their incentives to consume out of current wealth and their investment decisions. Quite generally, the presence of the option value to retire will lead to portfolios that are more exposed to stock market risk. The marginal propensity to consume out of wealth will be lower as one approaches early retirement, reflecting the increased incentives to reinvest gains in the stock market in order to bring retirement "closer". The likelihood of attaining early retirement in turn is more relevant for individuals who are young and/or wealthy.

An important practical implication of the present model is that the relationship between stockholdings and age is likely to be more complicated than what is suggested in BMS. This can be most easily seen by dividing the stockholdings by total wealth (financial wealth + the net present value of future income) in order to control for the effects discussed in BMS. The resulting fraction in our model is not constant as in BMS, but has clear option pricing properties and depends on both wealth and the horizon. Even though the fraction of total wealth invested in the stock market decreases as an individual ages *for a given level of (financial) wealth*, this fraction increases as wealth increases *for a given time* to retirement. In reality middle aged individuals are likely to be richer than 25-year olds and as a result it is very likely that they will be the major holders of stock even after normalizing their stockholdings by either financial or total wealth. This might help explain the hump shaped holdings of stock as a function of age which is found in certain studies.[23]

There are many interesting extensions to this model that one could consider. A first important extension would be to include features that are realistically present in actual 401(k) type plans like tax deferral, employee matching contributions and tax provisions related to withdrawals. Then the solutions developed in this model could be used to determine the optimal saving, retirement and portfolio decisions of consumers that are contemplating retirement and taking tax considerations into account.

---

[23]See e.g. Ameriks and Zeldes (2001)

A second extension of the model would be to implement it econometrically in order to disentangle the fraction of early retirement that could be explained by variations in stock market returns. In a more stylized model Gustman and Steinmeier (2002) find some evidence that the recent extraordinary behaviour of the stock market was key in driving early retirement.

A third extension would be to consider the international evidence on stock market participation and try to link it to the flexibility of retirement systems in various countries. The model suggests that increased labor supply flexibility might be key in trying to understand why consumers participate less in the stock market in countries with less flexible retirement systems. Naturally, such an analysis would make it interesting to study the general equilibrium consequences of retirement systems as in Basak (1999). It is very likely that -even though at the individual level the consumption CAPM holds- at an aggregate level the resulting "representative agent" would behave like an agent that chooses her leisure decision on a continuum. Thus the consumption CAPM would not hold at the aggregate level and one could investigate cases where this would help in resolving certain asset pricing puzzles.

We leave these issues for future research.

3.8 Appendix

3.8.1 Proofs for section 3.3

The goal of this section is the proof of Proposition 3.1. We start with some useful definitions that are standard in the convex duality approach[24]

For a concave, strictly increasing and cont. differentiable function $U : (0, \infty) \to R$ satisfying

$$U'(0^+) = \lim_{x \downarrow 0} U'(x) = \infty \text{ and } U'(\infty) = \lim_{x \to \infty} U'(x) = 0 \qquad (3.34)$$

we can define the inverse $I()$ of $U'()$. $I()$ maps $(0, \infty) \to (0, \infty)$ and satisfies

$$I(0^+) = \infty, I(\infty) = 0$$

A very convenient concept is that of a Legendre Fenchel transform $(\widetilde{U})$ of a concave function $U : (0, \infty) \to R$

$$\widetilde{U}(y) = \max_{x>0} [U(x) - xy] = U(I(y)) - yI(y), \quad 0 < y < \infty \qquad (3.35)$$

It is easy to verify that $\widetilde{U}()$ is strictly decreasing and convex and satisfies

$$
\begin{aligned}
\widetilde{U}'(y) &= -I(y), \quad 0 < y < \infty \\
U(x) &= \min_{y>0} \left[\widetilde{U}(y) + xy \right] = \widetilde{U}(U'(x)) + xU'(x), \quad 0 < x < \infty
\end{aligned}
\qquad (3.36)
$$

The inequality

$$U(I(y)) \geq U(x) + y [I(y) - x]$$

---

[24]This section is based on Karatzas and Wang (2000). For a more explicit presentation see also Karatzas and Shreve (1998)

follows from (3.35).

With these definitions we can proceed to extend the duality approach proposed by Karatzas and Wang (2000) to address portfolio problems with discretionary stopping to a setting with income.

We start by fixing a stopping time τ and defining

$$J(W; \pi_s, c_s, \tau) = E\left[\int_0^\tau e^{-\beta t} U_1(c_t) dt + e^{-\beta \tau} U_2(W_\tau)\right]$$

where we simplify notation by defining $U_1(c_t) = U(l_1, c_t)$. We obtain the following set of inequalities for any admissible pair (c_s, π_s) and any positive number λ

$$
\begin{aligned}
J(W; \pi_s, c_s, \tau) &= E\left[\int_0^\tau e^{-\beta t} U_1(c_t) dt + e^{-\beta \tau} U_2(W_\tau)\right] \\
&\leq E\left[\int_0^\tau e^{-\beta t} \widetilde{U_1}(\lambda e^{\beta t} H(t)) dt + e^{-\beta \tau} \widetilde{U}_2(\lambda e^{\beta \tau} H(\tau))\right] \\
&\quad + \lambda E\left[H(\tau) W_\tau + \int_0^\tau H(t) c(t) dt\right] \\
&\leq E\left[\int_0^\tau e^{-\beta t} \widetilde{U_1}(\lambda e^{\beta t} H(t)) dt + e^{-\beta \tau} \widetilde{U}_2(\lambda e^{\beta \tau} H(\tau))\right] \\
&\quad + \lambda \left(W_0 + E\left[\int_0^\tau H(t) y_0 dt\right]\right)
\end{aligned}
$$

with equality if and only if

$$W_\tau = I_2\left(\lambda e^{\beta \tau} H_\tau\right) \text{ and } c(t) = I_1\left(\lambda e^{\beta t} H_t\right), \quad \text{for all } 0 \leq t \leq \tau \qquad (3.37)$$

and

$$E\left[H(\tau) W_\tau + \int_0^\tau H(t) c(t) dt\right] = W_0 + E\left[\int_0^\tau H(t) y_0 dt\right]$$

146

These inequalities are standard in the convex duality approach to solve optimal portfolio problems. The first inequality follows from the definition of $\widetilde{U_1}, \widetilde{U_2}$ as given in (3.35) and the second line from the intertemporal budget constraint (3.2).

Observing that the above set of inequalities holds for all feasible policies (π_s, c_s, τ) and all $\lambda > 0$, we have that

$$V(W_0) \leq \sup_\tau \inf_{\lambda > 0} \left[\widetilde{J}(\lambda; \tau) + \lambda W_0 \right] \tag{3.38}$$

where $V(W_0)$ is the value function of the original problem and $\widetilde{J}(\lambda; \tau)$ is given by

$$\widetilde{J}(\lambda; \tau) = E\left[\int_0^\tau \left[e^{-\beta t}\widetilde{U_1}(\lambda e^{\beta t}H(t)) + \lambda H(t)y_0 \right] dt + e^{-\beta \tau}\widetilde{U_2}(\lambda e^{\beta \tau}H(\tau)) \right]$$

Intuitively, the constant $\lambda > 0$ plays the role of a Lagrange Multiplier on the intertemporal budget constraint (3.2).[25]

An interesting observation is that since the inequalities become equalities for the policy (3.37) it follows that we could solve the consumption-portfolio easily if we fixed an arbitrary stopping time. Indeed the entire consumption path and the wealth at τ are known up to the constant $\lambda > 0$ that can be determined in such a way that the intertemporal budget constraint holds with equality. Following arguments similar to the ones in section 6 of Karatzas and Wang (2000) one can show that for any given τ

$$\chi_\tau(\lambda) \widehat{=} E\left[\int_0^\tau H(t)\left(I_1(\lambda e^{\beta t}H(t)) - y_0 \right) dt + H(\tau)I_2(\lambda e^{\beta \tau}H_\tau) \right], \quad \lambda \in (0, \infty) \tag{3.39}$$

is a continuous strictly decreasing mapping of $(0, \infty)$ to $(-E\left[\int_0^\tau H(t)y_0 dt \right], \infty)$ with a

---

[25] An important difference to Karatzas and Wang (2000) is that the income process appears inside $\widetilde{J}$. This is to be expected since the optimal stopping time for the income process affects the net present value of income to be received.

continuous strictly decreasing inverse so that in particular there exists some $\lambda^* = \widetilde{y}(W_0)$ s.t.

$$E\left[\int_0^\tau H(t)\left[I_1(\lambda^* e^{\beta t} H(t)) - y_0\right] dt + H(\tau)I_2(\lambda^* e^{\beta \tau} H_\tau)\right] = W_0$$

The existence of a portfolio financing a claim with payoffs

$$W_\tau = I_2(\lambda^* e^{\beta \tau} H_\tau) \tag{3.40}$$

$$c_t = I_1(\lambda^* e^{\beta t} H(t))1\{t < \tau\} \tag{3.41}$$

can be established by the martingale representation Theorem and is omitted. We summarize in the following result

Lemma 3.1 *For any τ that is finite almost surely we have*

$$V_\tau(W_0) = \inf_{\lambda > 0}\left[\widetilde{J}(\lambda; \tau) + \lambda W_0\right] = \widetilde{J}(\lambda^*; \tau) + \lambda^* W_0$$

where

$$\lambda^* = \widetilde{y}(W_0)$$

and the supremum is attained by (3.40),(3.41). Moreover

$$V(W_0) = \sup_\tau V_\tau(W_0) = \sup_\tau \inf_{\lambda > 0}\left[\widetilde{J}(\lambda; \tau) + \lambda W_0\right] = \sup_\tau\left[\widetilde{J}(\lambda^*; \tau) + \lambda^* W_0\right]$$

This result shows that once the optimal stopping time has been determined then the determination of the optimal consumption and portfolio strategies are easy to obtain. Karatzas and Wang (2000) show that one can reduce the entire joint portfolio-consumption-stopping problem into a pure optimal stopping problem by investigating cases in which the following inequality becomes an equality

$$V(W_0) = \sup_\tau \inf_{\lambda > 0}\left[\widetilde{J}(\lambda; \tau) + \lambda W_0\right] \leq \inf_{\lambda > 0} \sup_\tau\left[\widetilde{J}(\lambda; \tau) + \lambda W_0\right] = \inf_\lambda\left[\widetilde{V}(\lambda) + \lambda W_0\right] \tag{3.42}$$

where

$$\widetilde{V}(\lambda) = \sup_\tau \widetilde{J}(\lambda;\tau) = \sup_\tau E\left[\int_0^\tau \left[e^{-\beta t}\widetilde{U_1}(\lambda e^{\beta t}H(t)) + \lambda H(t)y_0\right]dt + e^{-\beta\tau}\widetilde{U}_2(\lambda e^{\beta\tau}H(\tau))\right]$$

(3.43)

The inequality (3.42) follows from a standard result in convex duality[26].

The interesting fact about (3.43) is that it is a standard optimal stopping problem, for which one can apply well known results. In particular, the parametric assumptions that we made in section 3.2.3 allows us to solve this optimal stopping problem explicitly. We do this in the following Lemma.

Lemma 3.2 *Assume that*

$$\frac{r}{\theta}\frac{\left(\frac{1-\gamma}{\gamma}+\gamma_2\right)}{\gamma_2-1} < 1$$

(3.44)

where

$$\gamma_2 = \frac{1 - 2\frac{\beta-r}{\kappa^2} - \sqrt{(1-2\frac{\beta-r}{\kappa^2})^2 + 8\frac{\beta}{\kappa^2}}}{2} < 0$$

The function $\widetilde{V}(\lambda)$ is strictly convex. We can obtain an explicit solution to the optimal stopping problem of (3.43). Defining

$$Z_t = \lambda e^{\beta t}H_t$$

we have that $\widetilde{V}(\lambda)$ is given by

$$C_2\lambda^{\gamma_2} - \frac{\gamma}{\gamma-1}\frac{1}{\theta}\lambda^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}\lambda, \quad if \ \lambda \ > \ \underline{\lambda}$$

$$\left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(\lambda)^{\frac{\gamma-1}{\gamma}}\right) \quad if \ \lambda \ \leq \ \underline{\lambda}$$

---

[26]See e.g. Rockafellar (1997)

149

where

$$\underline{\lambda} = \left(\frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)} \frac{y_0}{r} \right)^{-\gamma}$$

C_2 *is given by*

$$\frac{\left[\frac{\gamma}{1-\gamma}\frac{(\gamma_2-1)}{\left(1+\gamma_2\frac{\gamma}{1-\gamma}\right)} - 1\right]\frac{y_0}{r}}{\underline{\lambda}^{\gamma_2-1}}$$

The optimal stopping strategy is to stop the first time the process Z_t reaches $\underline{\lambda}$. $\widetilde{V}(\lambda)$ is continuously differentiable everywhere and $\widetilde{V}'(\lambda)$ maps $(0, \infty)$ into $\left(-\infty, \frac{y_0}{r}\right)$.

Proof. *(Lemma 3.2).* Take $0 < \lambda_1 < \lambda_2 < \infty$, $s \in (0,1)$ and $\lambda_3 = s\lambda_1 + (1-s)\lambda_2$. Denote by τ_1^* the optimal stopping time when $\lambda = \lambda_1$ and similarly for τ_2^*, τ_3^*. We then have

$$
\begin{aligned}
\widetilde{V}(\lambda_3) &= \widetilde{J}(\lambda_3; \tau_3^*) < \\
&\quad s\widetilde{J}(\lambda_1; \tau_3^*) + (1-s)\widetilde{J}(\lambda_2; \tau_3^*) \\
&\le s\widetilde{V}(\lambda_1) + (1-s)\widetilde{V}(\lambda_2)
\end{aligned}
$$

The first (strict) inequality follows from the (strict) convexity of $\widetilde{U_1}, \widetilde{U_2}$ whereas the second from the definition of $\widetilde{V}(\lambda)$ (Equation 3.43.)We proceed by calculating the solution to the optimal stopping problem. It is easy to show that

$$
\begin{aligned}
\widetilde{U}_1(\lambda) &= \max_c \frac{c^{1-\gamma}}{1-\gamma} - \lambda c = \\
&= \frac{\gamma}{1-\gamma}\lambda^{\frac{\gamma-1}{\gamma}}
\end{aligned}
$$

and

$$
\begin{aligned}
\widetilde{U}_2(\lambda) &= \max_X \left(K\frac{X^{1-\gamma}}{1-\gamma} - \lambda X\right) = \\
&= K^{\frac{1}{\gamma}}\frac{\gamma}{1-\gamma}\lambda^{\frac{\gamma-1}{\gamma}}
\end{aligned}
$$

150

so that the expression (3.43) becomes

$$\sup_{\tau \in S} E \left[\int_0^\tau e^{-\beta t} \frac{\gamma}{1-\gamma} \left(\lambda e^{\beta t} H_t \right)^{\frac{\gamma-1}{\gamma}} dt + e^{-\beta \tau} \left(\frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} \left(\lambda e^{\beta t} H_t \right)^{\frac{\gamma-1}{\gamma}} \right) + \lambda y_0 \int_0^\tau H_t dt \right]$$

Consider the process

$$Z_t = \lambda e^{\beta t} H_t = \lambda H_0 e^{(\beta - r - \frac{1}{2}\kappa^2)t - \kappa B_t}$$

so that

$$\frac{dZ_t}{Z_t} = (\beta - r) \, dt - \kappa dB_t, \quad Z_0 = \lambda$$

With this new notation we can rewrite the above optimal stopping problem as

$$\sup_{\tau \in S} E \left[\int_0^\tau e^{-\beta t} \left(\frac{\gamma}{1-\gamma} (Z_t)^{\frac{\gamma-1}{\gamma}} + y_0 Z_t \right) dt + e^{-r\tau} \left(\frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} (Z_t)^{\frac{\gamma-1}{\gamma}} \right) \right]$$

To solve the pure optimal stopping problem we proceed as in Oksendal (1998) p. 213 and we define

$$f(Z,t) = e^{-\beta t} \left(\frac{\gamma}{1-\gamma} Z^{\frac{\gamma-1}{\gamma}} + y_0 Z \right)$$

$$g(Z,t) = e^{-\beta t} \left(\frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} (Z_t)^{\frac{\gamma-1}{\gamma}} \right)$$

$f(Z,t)$ represents the per-period payoff before stopping and $g(Z)$ represents the payoff upon stopping. Then the infinitesimal Operator acting on $G(Z,t,w) = g(Z,t) + w$ gives

$$\begin{aligned}
A_Z^G &= \frac{\partial G}{\partial t} + (\beta - r) Z \frac{\partial G}{\partial Z} + \frac{1}{2} \kappa^2 Z^2 \frac{\partial^2 G}{\partial Z^2} + e^{-\beta t} \left(\frac{\gamma}{1-\gamma} Z^{\frac{\gamma-1}{\gamma}} + y_0 Z \right) \frac{\partial G}{\partial w} = \\
&= -\frac{\gamma}{1-\gamma} e^{-\beta t} Z \left[(Z)^{-\frac{1}{\gamma}} \left[K^{\frac{1}{\gamma}} \theta - 1 \right] - \frac{1-\gamma}{\gamma} y_0 \right]
\end{aligned} \tag{3.45}$$

As is shown in Oksendal (1998), it can never be optimal to stop the process while A_Z^G is positive. Accordingly the continuation region will always be contained in the set $(Z, t : A_Z^G > 0)$ For $\gamma < 1$

and the fact that $K^{\frac{1}{\gamma}}\theta - 1 > 0$ (by Equation 3.6) the operator A_Z^G will be positive whenever

$$(Z_t)^{-\frac{1}{\gamma}}\left[K^{\frac{1}{\gamma}}\theta - 1\right] \;<\; \frac{1-\gamma}{\gamma}y_0$$

$$(Z_t)^{-\frac{1}{\gamma}} \;<\; \frac{\frac{1-\gamma}{\gamma}y_0}{K^{\frac{1}{\gamma}}\theta - 1}$$

$$Z_t \;>\; \left(\frac{\frac{1-\gamma}{\gamma}y_0}{K^{\frac{1}{\gamma}}\theta - 1}\right)^{-\gamma}$$

For $\gamma > 1$ and the fact that $K^{\frac{1}{\gamma}}\theta - 1 < 0$ we obtain exactly the same inequality. This shows that a reasonable guess for the continuation region is

$$\underline{Z} \le Z_t < \infty$$

for some $\underline{Z}$ satisfying

$$\underline{Z} \le \left(\frac{\frac{1-\gamma}{\gamma}y_0}{K^{\frac{1}{\gamma}}\theta - 1}\right)^{-\gamma}$$

To determine the solution, we apply the standard methodology of smooth pasting, i.e. we search for $\phi(Z_t)$ satisfying the following properties

$$-\beta\phi + (\beta - r)Z\frac{\partial\phi}{\partial Z} + \frac{1}{2}\frac{\partial^2\phi}{\partial Z^2}Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0Z\right) \;=\; 0 \quad \text{on } U \qquad (3.46)$$

$$\phi(Z_t) \;\ge\; \left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(Z_t)^{\frac{\gamma-1}{\gamma}}\right) \qquad (3.47)$$

$$-\beta\phi + (\beta - r)Z\frac{\partial\phi}{\partial Z} + \frac{1}{2}\frac{\partial^2\phi}{\partial Z^2}Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0Z\right) \;\le\; 0 \quad \text{on } R\backslash U \qquad (3.48)$$

$$\phi(Z_t) \text{ is } C^1,\ C^2 \text{ on } R \qquad (3.49)$$

where U is the continuation region and D is the exercise boundary. The general solution to (3.46) is given by

$$\phi(Z) = C_1 Z^{\gamma_1} + C_2 Z^{\gamma_2} - \frac{\gamma}{\gamma - 1}Z^{\frac{\gamma-1}{\gamma}}\frac{1}{\theta} + \frac{y_0}{r}Z$$

where

$$\gamma_{1,2} = \frac{1 - 2\frac{\beta - r}{\kappa^2} \pm \sqrt{(1 - 2\frac{\beta - r}{\kappa^2})^2 + 8\frac{\beta}{\kappa^2}}}{2}$$

152

It is straightforward to verify that

$$\gamma_1 > 0, \gamma_2 < 0, \ \gamma_1 + \gamma_2 = 1 - 2\frac{\beta - r}{\kappa^2}$$

Since the continuation region is of the form $\underline{Z} < Z < \infty$ we require

$$C_1 = 0$$

and thus we are left with determining the optimal exercise point and the constant C_2 and $\underline{Z}$. We can do that by invoking (3.49) to get the set of conditions

$$C_2 \underline{Z}^{\gamma_2} - \frac{\gamma}{\gamma - 1}\frac{1}{\theta}\underline{Z}^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}\underline{Z} = \frac{\gamma}{1 - \gamma}K^{\frac{1}{\gamma}}(\underline{Z})^{\frac{\gamma-1}{\gamma}} \tag{3.50}$$

$$\gamma_2 C_2 \underline{Z}^{\gamma_2 - 1} - \frac{1}{\theta}\underline{Z}^{-\frac{1}{\gamma}} + \frac{y_0}{r} = -K^{\frac{1}{\gamma}}\underline{Z}^{-\frac{1}{\gamma}} \tag{3.51}$$

Notice that we can rewrite the above as

$$C_2 \underline{Z}^{\gamma_2 - 1} - \frac{\gamma}{\gamma - 1}\frac{1}{\theta}\underline{Z}^{-\frac{1}{\gamma}} + \frac{y_0}{r} = \frac{\gamma}{1 - \gamma}K^{\frac{1}{\gamma}}(\underline{Z})^{-\frac{1}{\gamma}}$$

$$\gamma_2 C_2 \underline{Z}^{\gamma_2 - 1} - \frac{1}{\theta}\underline{Z}^{-\frac{1}{\gamma}} + \frac{y_0}{r} = -K^{\frac{1}{\gamma}}\underline{Z}^{-\frac{1}{\gamma}}$$

Solving for $\underline{Z}$ leads to

$$\underline{Z}^{-\frac{1}{\gamma}} = \frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\frac{y_0}{r}$$

Observe that $\underline{Z}^{-\frac{1}{\gamma}} > 0$ for $\gamma > 1$ since $\gamma_2 < 0$ and $K^{\frac{1}{\gamma}}\theta - 1 < 0$. For $\gamma < 1$ we can show that $\left(1 + \gamma_2\frac{\gamma}{1-\gamma}\right) < 0$ and accordingly $\underline{Z}^{-\frac{1}{\gamma}} > 0$ as follows: $1 + \gamma_2\frac{\gamma}{1-\gamma} = \frac{1-\gamma+\gamma\gamma_2}{1-\gamma}$, so that it is equivalent to show that $1 - \gamma + \gamma\gamma_2 < 0$. Using the fact that $\theta > 0$, we see that $\beta > \frac{1-\gamma}{\gamma}(\frac{\kappa^2}{2\gamma} - (\beta - r))$. This in turn implies that

$$1 - \gamma + \gamma\gamma_2 \le 1 - \gamma(\frac{1}{2} + \frac{\beta - r}{\kappa^2}) - |1 - \gamma(\frac{1}{2} + \frac{\beta - r}{\kappa^2})| \le 0.$$

This verifies that $\underline{Z} > 0$ for $\gamma < 1$ too. Similarly,

$$C_2 = \frac{\left[\frac{\gamma}{1-\gamma}\frac{(\gamma_2-1)}{\left(1+\gamma_2\frac{\gamma}{1-\gamma}\right)} - 1\right]\frac{y_0}{r}}{\underline{Z}^{\gamma_2-1}} > 0$$

since

$$\frac{\gamma}{1-\gamma}\frac{(\gamma_2-1)}{\left(1+\gamma_2\frac{\gamma}{1-\gamma}\right)} - 1 = \frac{-\frac{1}{1-\gamma}}{\left(1+\gamma_2\frac{\gamma}{1-\gamma}\right)} > 0$$

The previous considerations allow us to guess that the solution to the optimal stopping problem under consideration is given by

$$C_2 Z^{\gamma_2} - \frac{\gamma}{\gamma-1}\frac{1}{\theta}Z^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}Z, \quad \text{if } Z > \underline{Z} \tag{3.52}$$

$$\left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(Z_t)^{\frac{\gamma-1}{\gamma}}\right) \quad \text{if } Z \leq \underline{Z} \tag{3.53}$$

We proceed to verify that this is indeed the optimal stopping time by considering the rest of the conditions (namely (3.48) and (3.47)). To verify (3.47) we need to show that

$$C_2 Z^{\gamma_2} - \frac{\gamma}{\gamma-1}\frac{1}{\theta}Z^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}Z \geq \left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}Z^{\frac{\gamma-1}{\gamma}}\right)$$

for $Z \geq \underline{Z}$. We do this by considering the difference

$$T(Z) = C_2 Z^{\gamma_2} + \frac{\gamma}{\gamma-1}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)Z^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}Z$$

It is clear that $T(Z)$ satisfies: $T(\underline{Z}) = 0$ and $T'(\underline{Z}) = 0$ by construction. The claim that $T(Z) \geq 0$ for $Z > \underline{Z}$ will be proved if we can show that $T'(Z) \geq 0$ for $Z > \underline{Z}$. $T'(Z)$ is given by

$$T'(Z) = C_2\gamma_2 Z^{\gamma_2-1} + Z^{-\frac{1}{\gamma}}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right) + \frac{y_0}{r}$$

or

$$T'(Z) = C_2\gamma_2 \left(\frac{Z}{\underline{Z}}\right)^{\gamma_2-1} \underline{Z}^{\gamma_2-1} + \left(\frac{Z}{\underline{Z}}\right)^{-\frac{1}{\gamma}} \underline{Z}^{-\frac{1}{\gamma}} \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right) + \frac{y_0}{r}$$

Observe that for $Z > \underline{Z}$ we have that

$$\left(\frac{Z}{\underline{Z}}\right)^{\gamma_2-1} < \left(\frac{Z}{\underline{Z}}\right)^{-\frac{1}{\gamma}}$$

and accordingly

$$C_2\gamma_2 \left(\frac{Z}{\underline{Z}}\right)^{\gamma_2-1} \underline{Z}^{\gamma_2-1} > C_2\gamma_2 \left(\frac{Z}{\underline{Z}}\right)^{-\frac{1}{\gamma}} \underline{Z}^{\gamma_2-1}$$

since $\gamma_2 < 0$, $(\gamma_2 - 1) < -\frac{1}{\gamma}$, so that

$$T'(Z) \geq \left(\frac{Z}{\underline{Z}}\right)^{-\frac{1}{\gamma}} \left[C_2\gamma_2\underline{Z}^{\gamma_2-1} + \underline{Z}^{-\frac{1}{\gamma}}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)\right] + \frac{y_0}{r} \tag{3.54}$$

But

$$C_2\gamma_2\underline{Z}^{\gamma_2-1} + \underline{Z}^{-\frac{1}{\gamma}}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right) = -\frac{y_0}{r}$$

by the fact that $T'(\underline{Z}) = 0$. Accordingly (3.54) becomes

$$T'(Z) \geq \left[1 - \left(\frac{Z}{\underline{Z}}\right)^{-\frac{1}{\gamma}}\right] \frac{y_0}{r} > 0$$

for $Z > \underline{Z}$. This verifies that $T(Z) > 0$ for $Z > \underline{Z}$.

We are left with checking that (3.48) holds. This will be true if

$$\underline{Z} < \left(\frac{\frac{1-\gamma}{\gamma}y_0}{K^{\frac{1}{\gamma}}\theta - 1}\right)^{-\gamma} \tag{3.55}$$

since this will guarantee that

$$-r\phi + \phi'(\beta - r) + \phi''Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0Z\right) \leq 0 \quad \text{on } R\backslash U$$

by the observations we made about the sign of A_G^Z in (3.45).[27] To check (3.55) we need to

---

[27]Observe that for $Z < \underline{Z}$ the function under consideration becomes: $\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(Z_t)^{\frac{\gamma-1}{\gamma}}$

show that

$$\underline{Z}^{-\frac{1}{\gamma}} = \frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\frac{y_0}{r} > \frac{\frac{1-\gamma}{\gamma}y_0}{K^{\frac{1}{\gamma}}\theta - 1}$$

We need to distinguish cases. For $\gamma < 1$, $\left(K^{\frac{1}{\gamma}}\theta - 1\right) > 0$ and thus the above inequality can be rewritten as

$$\frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)r} > \frac{1 - \gamma}{\gamma}$$

or equivalently

$$\frac{r}{\theta}\frac{\left(\frac{1-\gamma}{\gamma} + \gamma_2\right)}{\gamma_2 - 1} < 1 \tag{3.56}$$

Observe that we arrive at the same inequality for $\gamma > 1$ too with identical steps. This is exactly assumption (3.44). The previous reasoning allows us to obtain the function $\widetilde{V}(\lambda)$ by observing that $Z(0) = \lambda e^{\beta 0}H_0 = \lambda$, so that

$$\widetilde{V}(\lambda) = C_2\lambda^{\gamma_2} - \frac{\gamma}{\gamma - 1}\frac{1}{\theta}\lambda^{\frac{\gamma-1}{\gamma}} + \frac{y_0}{r}\lambda, \quad \text{if } \lambda > \underline{\lambda}$$

$$\widetilde{V}(\lambda) = \left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(\lambda)^{\frac{\gamma-1}{\gamma}}\right) \quad \text{if } \lambda \leq \underline{\lambda}$$

where

$$\underline{\lambda} = \underline{Z} = \left(\frac{(\gamma_2 - 1)\theta}{\left(1 + \gamma_2 \frac{\gamma}{1-\gamma}\right)\left(K^{\frac{1}{\gamma}}\theta - 1\right)}\frac{y_0}{r}\right)^{-\gamma}$$

The function $\widetilde{V}(\lambda)$ is continuously differentiable everywhere and convex. Accordingly, we can calculate the derivative

$$\widetilde{V}'(\lambda) = \gamma_2 C_2\lambda^{\gamma_2-1} - \frac{1}{\theta}\lambda^{-\frac{1}{\gamma}} + \frac{y_0}{r}, \quad \text{if } \lambda > \underline{\lambda}$$

$$\widetilde{V}'(\lambda) = \left(-K^{\frac{1}{\gamma}}\lambda^{-\frac{1}{\gamma}}\right) \quad \text{if } 0 < \lambda \leq \underline{\lambda}$$

The range of $\widetilde{V}'(\lambda)$ for positive λ is $(-\infty, \frac{y_0}{r})$ implying that the equation

$$\widetilde{V}'(\lambda) = -W_0$$

156

will always have a solution as long as $W_0 \in (-\frac{y_0}{r}, \infty)$, since $\widetilde{V}'(\lambda)$ is an increasing continuous function. ∎

Remark 3.1 *Notice that in contrast to Karatzas and Wang (2000) the function $\widetilde{V}(\lambda)$ does not have to be decreasing due to the presence of income. As a matter of fact we show that $\widetilde{V}'(\lambda)$ takes values in $(-\infty, \frac{y_0}{r})$.*

Remark 3.2 *Assumption 3.44 can be shown to be always satisfied as long as $\theta > 0$ in two special cases: i) if $\gamma > 1$ and $\beta \geq r$ or ii) if $\beta = r$. We conjecture that $\theta > 0$ is sufficient for assumption 3.44 more generally but we haven't been able to prove it algebraically.*

Proof. *Remark (3.2)*To see this, observe that we can rewrite (3.44) as

$$r + (\theta - r)\gamma \; > \; (\theta - r)\gamma_2\gamma,$$
$$r \; > \; (r - \theta)\gamma(1 - \gamma_2).$$

This inequality is clearly satisfied if $\theta \geq r$, that is if

$$\frac{\gamma - 1}{\gamma}\frac{\kappa^2}{2\gamma} + \frac{\beta - r}{\gamma} \geq 0.$$

Observe that this last equation is verified if $\gamma > 1$ and $\beta \geq r$.To show ii) assume now that $\beta = r$ and that γ is arbitrary as long as $\theta > 0$. Multiplying both numerator and denominator of (3.44) by γ_2 and using the fact that$\gamma_2(\gamma_2 - 1) = \frac{2r}{\kappa^2}$ we reduce to showing that

$$\frac{\kappa^2}{2\theta}\left[\frac{\gamma_2}{\gamma} + \frac{2r}{\kappa^2}\right] < 1$$

Now using the definition of $\theta = r - \frac{1}{2}\frac{1-\gamma}{\gamma}\frac{\kappa^2}{\gamma}$ and the fact $\gamma_1 + \gamma_2 = 1$ we reduce the above problem to checking that the ratio

$$\frac{r - \frac{1}{2}\frac{\kappa^2}{\gamma}(\gamma_1 - 1)}{r - \frac{1}{2}\frac{1-\gamma}{\gamma}\frac{\kappa^2}{\gamma}} < 1$$

which will be trivially the case if $\gamma > 1$. For $\gamma < 1$ this will be true if

$$\gamma_1 - 1 > \frac{1 - \gamma}{\gamma}$$

but this is immediate since

$$\gamma_1 > \frac{1}{\gamma}$$

■

The following result shows the connection between $\widetilde{V}'(\lambda)$ and $\chi(\lambda; \widehat{\tau}_\lambda)$ of Equation (3.39). $\widehat{\tau}_\lambda$ is the optimal stopping rule associated with λ and is given in Lemma 3.2.

Lemma 3.3 $\widetilde{V}'(\lambda)$ *and* $\chi(\lambda; \widehat{\tau}_\lambda)$ *are related by*

$$\widetilde{V}'(\lambda) = -\chi(\lambda; \widehat{\tau}_\lambda)$$

Proof. *(Lemma 3.3).*The proof is very similar to Karatzas and Wang (2000). The convexity of $\widetilde{U}_j, j = 1, 2$ implies that

$$\widetilde{U}_j'(y)(x - y) \le \widetilde{U}_j(x) - \widetilde{U}_j(y) \le \widetilde{U}_j'(x)(x - y)$$

so that (for $|h| < \lambda$) we get

$$\widetilde{V}(\lambda + h) - \widetilde{V}(\lambda) = \widetilde{V}(\lambda + h) - \widetilde{J}(\lambda; \widehat{\tau}_\lambda) \ge \widetilde{J}(\lambda + h; \widehat{\tau}_\lambda) - \widetilde{J}(\lambda; \widehat{\tau}_\lambda) \ge$$
$$\ge hE\left[\int_0^{\widehat{\tau}_\lambda} \left[H(t)\widetilde{U}_1'(\lambda e^{\beta t} H(t)) + H(t)y_0\right] dt + H(\widehat{\tau}_\lambda)\widetilde{U}_2(\lambda e^{\beta \widehat{\tau}_\lambda} H(\widehat{\tau}_\lambda))\right] = -\chi(\lambda; \widehat{\tau}_\lambda)$$

where we have made use of the definition of $\chi(\lambda; \widehat{\tau}_\lambda)$ in Equation (3.39) and Equation (3.36). The result follows after taking limits as $h \to 0$. ■

This Lemma shows that the derivative of $\widetilde{V}(\lambda)$ informs us of the amount of initial wealth

that would be needed in order to sustain a stream of income and retirement wealth of

$$\hat{c}_t = I_1(\lambda e^{\beta t} H(t)) 1\{t < \hat{\tau_\lambda}\}$$

$$\widehat{W_{\hat{\tau_\lambda}}} = I_2(\lambda e^{\beta \hat{\tau_\lambda}} H(\hat{\tau_\lambda}))$$

To prove Proposition 3.1 we use this observation in order to replace the inequality in (3.42) with an equality sign and thus compute the Value function of the problem of interest.

Proof. *(Proposition 3.1)* We will verify that the triplet

$$\hat{c}_t = (\lambda^* e^{\beta t} H(t))^{-\frac{1}{\gamma}} 1\{0 \le t < \hat{\tau}\}$$

$$\hat{\tau} = \inf\{t : \lambda^* e^{\beta t} H(t) = \underline{\lambda}\}$$

$$\widehat{W_{\hat{\tau}}} = I_2(\lambda^* e^{\beta \hat{\tau_\lambda}} H(\hat{\tau_\lambda})) = I_2(\underline{\lambda}) = \overline{W}$$

where λ^* is given by (3.16) is an optimal policy. We start by showing that this policy is feasible. To see this, consider the function $\widetilde{V}(\lambda)$ as obtained in Lemma 3.2. Since $\widetilde{V}(\lambda)$ is strictly convex, and $\widetilde{V'}(\lambda)$ maps $(0, \infty)$ to $(-\frac{y_0}{r}, \infty)$ we know that there exists a unique $\lambda^* > 0$ s.t.

$$\widetilde{V}(\lambda^*) + \lambda^* W_0 = \inf_{\lambda > 0}\left[\widetilde{V}(\lambda) + \lambda W_0\right]$$

which can be rewritten as

$$\widetilde{V}(\lambda) + \lambda W_0 \ge \widetilde{V}(\lambda^*) + \lambda^* W_0 \quad \forall \lambda > 0$$

Moreover, λ^* as obtained in equation (3.16) minimizes $\left[\widetilde{V}(\lambda) + \lambda W_0\right]$ over all $\lambda > 0$, since

$$\widetilde{V'}(\lambda) = -W_0$$

By Lemma 3.3

$$W_0 = -\widetilde{V'}(\lambda^*) = -\chi(\lambda^*; \hat{\tau}_{\lambda^*}) = E\left[\int_0^{\hat{\tau}_{\lambda^*}} \left[H(t)(\lambda^* e^{\beta t} H(t))^{-\frac{1}{\gamma}}) - H(t)y_0\right] dt + H(\hat{\tau}_{\lambda^*})\overline{W}\right]$$

159

so that we can create portfolios that can finance the consumption stream

$$\widehat{c}_t = I_1(\lambda^* e^{\beta t} H(t)) 1\{t < \widehat{\tau}_{\lambda^*}\}$$

and the retirement Wealth

$$\widehat{W}_{\widehat{\tau}_{\lambda^*}} = I_2(\lambda^* e^{\beta \widehat{\tau}_{\lambda^*}} H(\widehat{\tau}_{\lambda^*})) = \overline{W}$$

We now verify optimality of this policy as follows

$$
\begin{aligned}
V(W_0) \;\geq\; & E\left[\int_0^{\widehat{\tau}_{\lambda^*}} e^{-\beta t} U_1(I_1(\lambda^* e^{\beta t} H(t))) dt + e^{-\beta \widehat{\tau}_{\lambda^*}} U_2\left(I_2(\lambda^* e^{\beta \widehat{\tau}_{\lambda^*}} H(\widehat{\tau}_{\lambda^*}))\right)\right] = \\
= \; & E\left[\int_0^{\widehat{\tau}_{\lambda^*}} e^{-\beta t} \widetilde{U}_1(I_1(\lambda^* e^{\beta t} H(t))) dt + e^{-\beta \widehat{\tau}_{\lambda^*}} \widetilde{U}_2\left(I_2(\lambda^* e^{\beta \widehat{\tau}_{\lambda^*}} H(\widehat{\tau}_{\lambda^*}))\right)\right] + \\
& \lambda^* E\left[H(\widehat{\tau}_{\lambda^*}) \widehat{W}_{\widehat{\tau}_{\lambda^*}} + \int_0^{\widehat{\tau}_{\lambda^*}} H(t) \widehat{c}_t dt\right] \\
= \; & E\left[\int_0^{\widehat{\tau}_{\lambda^*}} \left[e^{-\beta t} \widetilde{U}_1(I_1(\lambda^* e^{\beta t} H(t))) + \lambda^* H(t) y_0\right] dt + e^{-\beta \widehat{\tau}_{\lambda^*}} \widetilde{U}_2\left(I_2(\lambda^* e^{\beta \widehat{\tau}_{\lambda^*}} H(\widehat{\tau}_{\lambda^*}))\right)\right] + \lambda^* W_0 = \\
= \; & \widetilde{V}(\lambda^*) + \lambda^* W_0 = \inf_{\lambda > 0}\left[\widetilde{V}(\lambda) + \lambda W_0\right]
\end{aligned}
$$

The first equality follows from the definitions of $\widetilde{U}_1, I_1, \widetilde{U}_2, I_2$. The second equality follows from the intertemporal budget constraint and the last from the definition of $\widetilde{V}(\lambda)$. The fact that $V(W_0) \geq \inf_{\lambda > 0}\left[\widetilde{V}(\lambda) + \lambda W_0\right]$ along with (3.42) delivers the result that

$$V(W_0) = \inf_{\lambda > 0}\left[\widetilde{V}(\lambda) + \lambda W_0\right]$$

In particular the optimal policies are given by $< \widehat{c}_t, \widehat{W}_{\widehat{\tau}_{\lambda^*}}, \widehat{\tau}_{\lambda^*} >$. The final claim of the proposition concerns the optimal portfolio. To actually compute it in feedback form we make use of formula (3.8.24) in Karatzas and Shreve (1998)

$$\pi_0 = -\frac{\kappa}{\sigma} \frac{\lambda^*(W_0)}{\lambda^*_{W_0}(W_0)}$$

160

where $\lambda^*(W_0)$ solves equation (3.16). The implicit function theorem gives

$$\frac{\lambda^*}{\lambda^*_{W_0}} = -\left(\gamma_2(\gamma_2 - 1)C_2\lambda^{*\gamma_2 - 1} + \frac{1}{\gamma}\frac{1}{\theta}\lambda^{*-\frac{1}{\gamma}}\right) \tag{3.57}$$

∎

3.8.2 Proofs for section 3.5

Proposition 3.2 can be established by virtually identical steps as Proposition 3.1. The only substantial difference to section 3.3 is that now $\widetilde{V}(\lambda, T)$ solves the optimal stopping problem

$$\widetilde{V}(\lambda, T) = \sup_{\tau \leq T} E\left[\int_0^\tau \left[e^{-\beta t}\widetilde{U_1}(\lambda e^{\beta t}H(t)) + \lambda H(t)y_0\right] dt + e^{-\beta \tau}\widetilde{U}_2(\lambda e^{\beta \tau}H(\tau))\right] \tag{3.58}$$

Accordingly Proposition 3.2 will be established once we can show the following result

Lemma 3.4 *The function $\widetilde{V}(\lambda, T)$ is strictly convex in λ.. Let*

$$Z_t = \lambda e^{\beta t}H_t$$

and consider the solution $\underline{Z}_t, t \in [0, T]$ to the functional equation

$$\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}\underline{Z}_t^{\frac{\gamma-1}{\gamma}} = \underline{Z}_t\frac{y_0(1 - e^{-r(T-t)})}{r} + \frac{\gamma}{1-\gamma}\underline{Z}_t^{\frac{\gamma-1}{\gamma}}\left(\frac{1 - e^{-\theta(T-t)}}{\theta} + K^{\frac{1}{\gamma}}e^{-\theta(T-t)}\right)$$

$$+ \frac{\gamma}{1-\gamma}(K^{\frac{1}{\gamma}}\theta - 1)\underline{Z}_t^{\frac{\gamma-1}{\gamma}}\int_t^T e^{-\theta(s-t)}N\left(d_{1s-t}\right) dt$$

$$- y_0\underline{Z}_t\int_t^T e^{-rt}N\left(d_{2s-t}\right) dt$$

where

$$d_{1s-t} = -\left(\frac{\log(\underline{Z}_t) + \left(\beta - r - \frac{\kappa^2}{2}\right)(s-t) - \log(\underline{Z}_s)}{\kappa\sqrt{s-t}} + \kappa^2\left(\frac{\gamma-1}{\gamma}\right)^2(s-t)\right)$$

$$d_{2s-t} = \left(\frac{\log(\underline{Z}_t) + \left(\beta - r - \frac{\kappa^2}{2}\right)(s-t) - \log(\underline{Z}_s)}{\kappa\sqrt{s-t}} + \kappa^2(s-t)\right)$$

if $\gamma > 1$ and

$$d_{1s-t} = \left(\frac{\log(\underline{Z}_t) + \left(\beta - r - \frac{\kappa^2}{2}\right)(s-t) - \log(\underline{Z}_s)}{\kappa\sqrt{s-t}} + \kappa^2\left(\frac{\gamma-1}{\gamma}\right)^2(s-t)\right)$$

$$d_{2s-t} = \left(\frac{\log(\underline{Z}_t) + \left(\beta - r - \frac{\kappa^2}{2}\right)(s-t) - \log(\underline{Z}_s)}{\kappa\sqrt{s-t}} + \kappa^2(s-t)\right)$$

if $\gamma < 1$. The optimal stopping strategy to (3.58) is to stop the first time the process Z_t reaches $\underline{Z}_t$.

Proof. *(Lemma 3.4)* To save on notation we set without loss of generality $t = 0$. The first assertion can be established along the same lines as the proof of Lemma 3.2. The second assertion follows by an argument similar to Carr, Jarrow, and Myerni (1992) Applying Ito's Lemma to $e^{-\beta t}\widetilde{V}(Z_t, T - t)$[28] one gets

$$e^{-\beta T}\widetilde{V}(Z_T, 0) = \widetilde{V}(\lambda, T) + \int_0^T e^{-\beta t}\frac{\partial\widetilde{V}}{\partial Z}dZ_t + \tag{3.59}$$
$$+ \int_0^T \left[e^{-\beta t}\left(\frac{\partial^2\widetilde{V}}{\partial Z^2}\frac{\kappa^2}{2}Z_t^2 + \frac{\partial\widetilde{V}}{\partial Z}Z_t(\beta - r) - r\widetilde{V} + \frac{\partial\widetilde{V}}{\partial t}\right)\right]dt$$

Using the fact that inside the continuation region

$$\frac{\partial^2\widetilde{V}}{\partial Z^2}\frac{\kappa^2}{2}Z_t^2 + \frac{\partial\widetilde{V}}{\partial Z}Z_t(\beta - r) - r\widetilde{V} + \frac{\partial\widetilde{V}}{\partial t} = -\left(\frac{\gamma}{1-\gamma}Z_t^{\frac{\gamma-1}{\gamma}} + y_0 Z_t\right)$$

---

[28]Even though the second derivative of this expression can have jumps one can use a generalization of Ito's Lemma to obtain this expression that only requires the function to be in C_1.

one can rewrite (3.59) as

$$\frac{\gamma}{1-\gamma}e^{-\beta T}K^{\frac{1}{\gamma}}Z_T^{\frac{\gamma-1}{\gamma}} \;=\; \widetilde{V}(Z_T,0)=\widetilde{V}(\lambda,T)+\int_0^T e^{-\beta t}\frac{\partial \widetilde{V}}{\partial Z}dZ_t+$$

$$-\int_0^T e^{-\beta t}\left(\frac{\gamma}{1-\gamma}Z_t^{\frac{\gamma-1}{\gamma}}+y_0 Z_t\right)1\{Z_t>\underline{Z_t}\}dt+$$

$$+\frac{\gamma}{\gamma-1}K^{\frac{1}{\gamma}}\theta\int_0^T e^{-\beta t}Z_t^{\frac{\gamma-1}{\gamma}}1\{Z_t<\underline{Z_t}\}dt$$

where $\{Z_t>\underline{Z_t}\}$ is the continuation region and $\{Z_t<\underline{Z_t}\}$ is the stopping region.[29] Adding and subtracting

$$\int_0^T e^{-\beta t}\left(\frac{\gamma}{1-\gamma}Z_t^{\frac{\gamma-1}{\gamma}}+y_0 Z_t\right)1\{Z_t<\underline{Z_t}\}dt$$

and taking expectations one gets

$$\widetilde{V}(\lambda,T) \;=\; E\left(\int_0^T e^{-\beta t}\left(\frac{\gamma}{1-\gamma}Z_t^{\frac{\gamma-1}{\gamma}}+y_0 Z_t\right)dt+\frac{\gamma}{1-\gamma}e^{-\beta T}K^{\frac{1}{\gamma}}Z_T^{\frac{\gamma-1}{\gamma}}\right)$$

$$+\frac{\gamma}{1-\gamma}E\left(\int_0^T e^{-\beta t}\left(Z_t^{\frac{\gamma-1}{\gamma}}(K^{\frac{1}{\gamma}}\theta-1)+\frac{\gamma-1}{\gamma}y_0 Z_t\right)1\{Z_t<\underline{Z_t}\}dt\right)$$

The first part of this term coincides with the term that one would get in the problem where retirement would be mandatory in T periods. The second term is an "early exercise premium". It captures the extra value of being able to stop at any time prior to T. One can obtain an explicit expression for the first part of this equation

$$E\left(\int_0^T e^{-\beta t}\left(\frac{\gamma}{1-\gamma}Z_t^{\frac{\gamma-1}{\gamma}}+y_0 Z_t\right)dt+\frac{\gamma}{1-\gamma}e^{-\beta T}K^{\frac{1}{\gamma}}Z_T^{\frac{\gamma-1}{\gamma}}\right) \;=$$

$$Z_0\frac{y_0(1-e^{-rT})}{r}+$$

$$+\frac{\gamma}{1-\gamma}Z_0^{\frac{\gamma-1}{\gamma}}\left(\frac{1-e^{-\theta T}}{\theta}+K^{\frac{1}{\gamma}}e^{-\theta T}\right)$$

To determine the second part needs knowledge of the unknown $\underline{Z_t}$. However using the fact that

$$\widetilde{V}(\underline{Z_0},T)=\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}\underline{Z_0}^{\frac{\gamma-1}{\gamma}}$$

---

[29]To obtain the result that the continuation and the stopping regions take this form one can proceed as in Oksendal (1998)

allows one to obtain an integral equation for $\underline{Z_t}, t = 0..T$ as follows

$$\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}\underline{Z_t}^{\frac{\gamma-1}{\gamma}} = \underline{Z_t}\frac{y_0(1-e^{-r(T-t)})}{r} + \frac{\gamma}{1-\gamma}\underline{Z_t}^{\frac{\gamma-1}{\gamma}}\left(\frac{1-e^{-\theta(T-t)}}{\theta} + K^{\frac{1}{\gamma}}e^{-\theta(T-t)}\right)$$
$$+\frac{\gamma}{1-\gamma}E\left(\int_t^T e^{-\beta s}\left(Z_s^{\frac{\gamma-1}{\gamma}}(K^{\frac{1}{\gamma}}\theta-1)+\frac{\gamma-1}{\gamma}y_0 Z_s\right)1\{Z_s < \underline{Z_s}\}ds|Z_t = \underline{Z_t}\right)$$

Using Fubini's Theorem we can further rewrite

$$E\left(Z_s 1\{Z_s < \underline{Z_s}\}|Z_t = \underline{Z_t}\right) = \underline{Z_t}e^{(\beta-r)(s-t)}N\left(\widetilde{d_{2s-t}}\right)$$
$$E\left(Z_s^{\frac{\gamma-1}{\gamma}}1\{Z_s < \underline{Z_s}\}|Z_t = \underline{Z_t}\right) = \underline{Z_t}^{\frac{\gamma-1}{\gamma}}e^{(\beta-\theta)(s-t)}N\left(\widetilde{d_{1s-t}}\right) \quad \text{if } \gamma > 1$$

$\widetilde{d_{1s-t}}, \widetilde{d_{2s-t}}$ are given in the statement of the Lemma. ∎

The rest of the assertions of proposition 3.2 follow easily from results established in section 3.3.

We proceed with the proof of Proposition 3.3

Proof. *(Proposition 3.3)* Only a sketch is given. The idea behind the approximation is to

observe that

$$\widetilde{V}^E(\lambda;T) = E\left[\int_0^T \left[e^{-\beta t}\widetilde{U_1}(\lambda e^{\beta t}H(t)) + \lambda H(t)y_0\right]dt + e^{-\beta \tau}\widetilde{U}_2(\lambda e^{\beta \tau}H(\tau))\right] =$$
$$= \frac{\gamma}{1-\gamma}\lambda^{\frac{\gamma-1}{\gamma}}\frac{1-e^{-\theta T}}{\theta} + \lambda y_0\frac{1-e^{-rT}}{r} + \frac{\gamma}{1-\gamma}\lambda^{\frac{\gamma-1}{\gamma}}K^{\frac{1}{\gamma}}e^{-\theta T}$$

which can be shown by elementary methods. The next step is to study the difference between

$$P(\lambda;T) = \widetilde{V}(\lambda;T) - \widetilde{V}^E(\lambda;T)$$

which we will refer to as the early exercise premium. One can then show that inside the

continuation region the "early exercise premium" $P(\lambda; T)$ solves the PDE

$$-\beta P + P_Z(\beta - r) + \frac{1}{2}P_{ZZ}Z^2\kappa^2 - P_T = 0$$

By the same approximation idea as in Barone-Adesi and Whaley (1987) we will postulate a solution of the form $P = Y(T)f(Z, Y(T))$, take $Y(T) = 1 - e^{-\beta T}$ and ignore P_Y. This allows to reduce the problem to the determination of solutions of the equation

$$f_Z(\beta - r) + \frac{1}{2}Z^2\kappa^2 f_{ZZ} - \frac{\beta}{Y(T)}f = 0$$

which is a simple linear ODE. The solution is given just as in the infinite horizon case by

$$f(Z) = C_{2T}Z^{\gamma_{2T}}$$

where

$$\gamma_{2T} = \frac{1 - 2\frac{\beta - r}{\kappa^2} - \sqrt{(1 - 2\frac{\beta - r}{\kappa^2})^2 + 8\frac{\beta}{Y(T)\kappa^2}}}{2}$$

To determine the complete solution we require continuity and smooth pasting of $\widetilde{V}(\lambda; T)$ to $\frac{\gamma}{1-\gamma}\lambda^{\frac{\gamma-1}{\gamma}}K^{\frac{1}{\gamma}}$. Then by arguments identical to the infinite horizon case we get (3.25) and (3.26). The rest of the results follow easily. ∎

3.8.3 Proofs for section 3.6

In what follows we sketch how to obtain the solution to this problem and prove proposition 3.4. The basic modification of the approach used so far is that $\widetilde{V}(\lambda)$ needs to be minimized over a set of decreasing processes in a manner analogous to He and Pages (1993). The reader is refered to that paper for a number of technical details.

We start by fixing a stopping time τ and defining

$$J(W; \pi_s, c_s, \tau) = E\left[\int_0^\tau e^{-\beta t}U_1(c_t)dt + e^{-\beta\tau}U_2(W_\tau)\right]$$

165

for any admissible pair (c_s, π_s) satisfying (3.30) and (3.31). Let λX_t be a non-increasing process starting at $X_0 = 1$ with $\lambda > 0$. We obtain the following set of inequalities for any admissible pair (c_s, π_s)

$$
\begin{aligned}
J(W; \pi_s, c_s, \tau) &= E\left[\int_0^\tau e^{-\beta t} U_1(c_t)dt + e^{-\beta \tau} U_2(W_\tau)\right] \qquad (3.60) \\
&\leq E\left[\int_0^\tau e^{-\beta t} \widetilde{U_1}(\lambda X_t e^{\beta t} H(t))dt + e^{-\beta \tau} \widetilde{U}_2(\lambda X_\tau e^{\beta \tau} H(\tau))\right] \qquad (3.61) \\
&\quad + \lambda E\left[X_\tau H(\tau) W_\tau + \int_0^\tau X_t H(t)c(t)dt\right]
\end{aligned}
$$

Integrating by parts and using the fact that $X_0 = 1$, the second term of the right hand side can be rewritten as

$$
\begin{aligned}
E\left[\int_0^\tau X_t H_t c_t dt + X_\tau H_\tau W_\tau\right] &= E\left[\int_0^\tau X_t H_t(c_t - y_0)dt + X_\tau H_\tau W_\tau + \int_0^\tau X_t H_t y_0 dt\right] = \\
&\quad E\left[\int_0^\tau X_t H_t y_0 dt + H_\tau W_\tau + \int_0^\tau H_t(c_t - y_0)dt\right] \\
&\quad + E\left[\int_0^\tau H_t \frac{E_t\left[\int_t^\tau H_s(c_s - y_0)ds + H_\tau W_\tau\right]}{H_t} dX_t\right]
\end{aligned}
$$

so that we have

$$J(W; \pi_s, c_s, \tau) \leq E\left[\int_0^\tau e^{-\beta t}\widetilde{U_1}(\lambda X_t e^{\beta t}H(t))dt + e^{-\beta \tau}\widetilde{U}_2(\lambda X_\tau e^{\beta \tau}H(\tau))\right]$$

$$+\lambda E\left[\int_0^\tau X_t H_t y_0 dt + H_\tau W_\tau + \int_0^\tau H_t(c_t - y_0)dt\right]$$

$$+\lambda E\left[\int_0^\tau H_t \frac{E_t\left[\int_t^\tau H_s(c_s - y_0)ds + H_\tau W_\tau | F_t\right]}{H_t} dX_t\right]$$

$$\leq E\left[\int_0^\tau e^{-\beta t}\widetilde{U_1}(X_t e^{\beta t}H(t))dt + e^{-\beta \tau}\widetilde{U}_2(X_\tau e^{\beta \tau}H(\tau))\right]$$

$$+\lambda\left(W_0 + E\left[\int_0^\tau X_t H_t y_0 dt\right]\right)$$

where the last inequality comes from the fact that

$$\frac{E_t\left[\int_t^\tau H_s(c_s - y_0)ds + H_\tau W_\tau\right]}{H_t}dX_t \leq 0, \text{ and}$$

$$E\left[\int_0^\tau H_t(c_t - y_0)dt + H_\tau W_\tau\right] \leq W_0.$$

The equality occurs if and only if

$$W_\tau = I_2\left(e^{\beta \tau}X_\tau H_\tau\right) \text{ and } c(t) = I_1\left(e^{\beta t}X_t H_t\right), \text{ for all } 0 \leq t \leq \tau \qquad (3.62)$$

and

$$E\left[H(\tau)W_\tau + \int_0^\tau H(t)c(t)dt\right] = W_0 + E\left[\int_0^\tau H(t)y_0 dt\right]$$

and

$$\frac{E_t \left[\int_t^\tau H_s(c_s - y_0)ds + H_\tau W_\tau \right]}{H_t} dX_t = 0.$$

Evaluating the above set of inequalities at the optimal stopping time τ, and observing that it holds for all X_t decreasing, we have that

$$V(W_0) \le \sup_\tau \inf_{\{\lambda, X_t\}} \left[\widetilde{J}(\{X_t, \lambda\}; \tau) + \lambda W_0 \right] \tag{3.63}$$

where $V(W_0)$ is the value function of the original problem and $\widetilde{J}(X_t; \tau)$ is given by

$$\widetilde{J}(\{X_t\}; \tau, \lambda) = E \left[\int_0^\tau \left[e^{-\beta t} \widetilde{U_1}(\lambda X_t e^{\beta t} H(t)) + \lambda X_t H(t) y_0 \right] dt + e^{-\beta \tau} \widetilde{U}_2(\lambda X_t e^{\beta \tau} H(\tau)) \right]$$

Let

$$\widetilde{V}(\{X_t\}, \lambda) =$$

$$= \sup_\tau \widetilde{J}(\{X_t\}; \tau, \lambda) \tag{3.64}$$

$$= \sup_\tau E \left[\int_0^\tau \left[e^{-\beta t} \widetilde{U_1}(\lambda X_t e^{\beta t} H(t)) + \lambda X_t H(t) y_0 \right] dt + e^{-\beta \tau} \widetilde{U}_2(\lambda X_t e^{\beta \tau} H(\tau)) \right], \tag{3.65}$$

and

$$\widetilde{V}(\lambda) = \inf_{\{X_t\}} \widetilde{V}(\{X_t\}, \lambda) \tag{3.66}$$

and define the process Z_t:

$$Z_t = \lambda e^{\beta t} X_t H_t$$

168

We now proceed by analogy to the case without borrowing constraints. It can be shown that[30]

$$V(W_0) = \sup_{\tau} \inf_{\{\lambda, X_t\}} \left[\widetilde{J}(\{X_t\}; \tau, \lambda) + \lambda W_0 \right] = \inf_{\{\lambda, X_t\}} \sup_{\tau} \left[\widetilde{J}(\{X_t\}; \tau, \lambda) + \lambda W_0 \right] = \inf_{\lambda} \left[\widetilde{V}(\lambda) + \lambda W_0 \right].$$

$$(3.67)$$

The optimal policy functions are given by

$$W_\tau = I_2 \left(\lambda^* e^{\beta \tau} X_t^* H_\tau \right) \text{ and } c(t) = I_1 \left(\lambda^* e^{\beta t} X_t^* H_t \right), \quad \text{for all } 0 \le t \le \tau \qquad (3.68)$$

where λ^*, τ^*, X_t^* solve (3.67). To solve the infimization over the space of decreasing processes one can proceed in a fashion analogous to He Pages (1993) to construct the Value of the min-max game of equation (3.67). The following generalization of Lemma 3.2 is required

Lemma 3.5 *For appropriate constants* C_1, C_2, Z_L, Z_H *(given in the proof) define the function* $\widetilde{V}(\lambda)$ *as*

$$\widetilde{V}(\lambda) = C_1 \lambda^{\gamma_1} + C_2 \lambda^{\gamma_2} - \frac{\gamma}{\gamma - 1} \lambda^{\frac{\gamma - 1}{\gamma}} \frac{1}{\theta} + \frac{y_0 \lambda}{r} \text{ if } Z_L \le \lambda \le Z_H,$$

$$\widetilde{V}(\lambda) = \left(\frac{\gamma}{1 - \gamma} K^{\frac{1}{\gamma}} (\lambda)^{\frac{\gamma - 1}{\gamma}} \right) \text{ if } \lambda < Z_L,$$

$$\widetilde{V}(\lambda) = C_1 Z_H^{\gamma_1} + C_2 Z_H^{\gamma_2} - \frac{\gamma}{\gamma - 1} Z_H^{\frac{\gamma - 1}{\gamma}} \frac{1}{\theta} + \frac{y_0 Z_H}{r} \text{ if } \lambda > Z_H,$$

Assume moreover that

$$-\beta \widetilde{V} + (\beta - r) \lambda \frac{\partial \widetilde{V}}{\partial \lambda} + \frac{1}{2} \frac{\partial^2 \widetilde{V}}{\partial \lambda^2} \lambda^2 \kappa^2 + \left(\frac{\gamma}{1 - \gamma} \lambda^{\frac{\gamma - 1}{\gamma}} + y_0 \lambda \right) \le 0 \quad \text{for } \lambda < Z_L \qquad (3.69)$$

---

[30]This proof is available upon request. It is ommited because it effectively replicates the steps in Karatzas and Wang (2000), combined with the results in He and Pages (1993)

and

$$\widetilde{V} \geq \left(\frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} (\lambda)^{\frac{\gamma-1}{\gamma}} \right) \quad everywhere \qquad (3.70)$$

Then $\widetilde{V}(\lambda)$ provides the value to the game

$$\widetilde{V}(\lambda) = \sup_{\tau} \inf_{\{X_t\}} \left[\widetilde{J}(\{X_t\}, \tau) \right] = \inf_{\{X_t\}} \sup_{\tau} \left[\widetilde{J}(\{X_t\}, \tau) \right]$$

Define also the process

$$Z_t = \lambda e^{\beta t} H_t$$

The optimal stopping policy is to stop once Z_t crosses Z_L whereas the optimal X_t decreases once $Z_t = Z_H$.

Proof. *(Lemma 3.5)* We give a sketch. To keep the notation consistent with section 3.3 we focus at time 0 without loss of generality and use the fact that at time 0 $Z_0 = \lambda$. We will denote $Z = Z_0$ for convenience. The purpose is to determine the value $\phi(Z)$ of the game

$$\phi(Z) = \sup_{\tau} \inf_{\{X_t\}} \left[\widetilde{J}(\{X_t\}, \tau) \right] = \inf_{\{X_t\}} \sup_{\tau} \left[\widetilde{J}(\{X_t\}, \tau) \right]$$

i.e. to fix a given initial value of the multiplier $\lambda = Z = Z_0$ and determine a decreasing process X_t^* and a stopping time τ^* so that X_t^* minimizes $\widetilde{J}$ conditional on λ and τ and τ^* maximizes $\widetilde{J}$ conditional on X_t^* and λ. In this context it is not difficult to establish a verification theorem, asserting that $\phi(Z)$ is the value of the game, as long as we can find a function $\phi(Z)$ and two barriers Z_L and Z_H with $Z_L < Z_H$ satisfying

$$-\beta\phi + (\beta - r)\,Z\frac{\partial\phi}{\partial Z} + \frac{1}{2}\frac{\partial^2\phi}{\partial Z^2}Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0 Z\right) \;=\; 0 \ \text{for}\ Z\in(Z_L, Z_H) \quad (3.71)$$

$$\phi(Z) \;\geq\; \left(\frac{\gamma}{1-\gamma}K^{\frac{1}{\gamma}}(Z_t)^{\frac{\gamma-1}{\gamma}}\right) \quad (3.72)$$

$$-\beta\phi + (\beta - r)\,Z\frac{\partial\phi}{\partial Z} + \frac{1}{2}\frac{\partial^2\phi}{\partial Z^2}Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0 Z\right) \;\leq\; 0 \ \text{ for}\ Z < Z_L \quad (3.73)$$

$$\phi(Z)\ \text{is}\ C^1,\ C^2\ \text{a.e.} \quad (3.74)$$

$$\frac{\partial\phi}{\partial Z} \;\leq\; 0\ \text{everywhere} \quad (3.75)$$

$$\frac{\partial\phi}{\partial Z} \;=\; 0\ \text{for}\ Z\in(Z_H, \infty) \quad (3.76)$$

A proof of this verification Theorem can be given along the lines of Theorem 3 in He and Pages (1993) and standard arguments for optimal stopping problems along the lines of Oksendal (1998) and is available upon request. We now proceed to construct a function and two barriers that satisfy these equations. The general solution to

$$-\beta\phi + (\beta - r)\,Z\frac{\partial\phi}{\partial Z} + \frac{1}{2}\frac{\partial^2\phi}{\partial Z^2}Z^2\kappa^2 + \left(\frac{\gamma}{1-\gamma}Z^{\frac{\gamma-1}{\gamma}} + y_0 Z\right) = 0$$

is given by

$$\phi(Z) = C_1 Z^{\gamma_1} + C_2 Z^{\gamma_2} - \frac{\gamma}{\gamma-1}Z^{\frac{\gamma-1}{\gamma}}\frac{1}{\theta} + \frac{y_0 Z}{r}$$

where γ_1 and γ_2 are given by

$$\gamma_1 = \frac{1 - 2\frac{\beta-r}{\kappa^2} + \sqrt{(1 - 2\frac{\beta-r}{\kappa^2})^2 + 8\frac{\beta}{\kappa^2}}}{2}$$

and

$$\gamma_2 = \frac{1 - 2\frac{\beta-r}{\kappa^2} - \sqrt{(1 - 2\frac{\beta-r}{\kappa^2})^2 + 8\frac{\beta}{\kappa^2}}}{2}$$

It is straightforward to verify that

$$\gamma_1 > 0, \gamma_2 < 0.$$

171

To enforce the condition (3.74) we will search for Z_L, Z_H and C_1, C_2 so that

$$C_1 Z_L^{\gamma_1} + C_2 Z_L^{\gamma_2} - \frac{\gamma}{\gamma-1} Z_L^{\frac{\gamma-1}{\gamma}} \frac{1}{\theta} + \frac{y_0 Z_L}{r} = \left(\frac{\gamma}{1-\gamma} K^{\frac{1}{\gamma}} (Z_L)^{\frac{\gamma-1}{\gamma}} \right)$$

$$\gamma_1 C_1 Z_L^{\gamma_1-1} + \gamma_2 C_2 Z_L^{\gamma_2-1} - Z_L^{-\frac{1}{\gamma}} \frac{1}{\theta} + \frac{y_0}{r} = \left(-K^{\frac{1}{\gamma}} (Z_L)^{-\frac{1}{\gamma}} \right)$$

$$\gamma_1 C_1 Z_H^{\gamma_1-1} + \gamma_2 C_2 Z_H^{\gamma_2-1} - Z_H^{-\frac{1}{\gamma}} \frac{1}{\theta} + \frac{y_0}{r} = 0$$

$$-\beta \left(C_1 Z_H^{\gamma_1} + C_2 Z_H^{\gamma_2} - \frac{\gamma}{\gamma-1} Z_H^{\frac{\gamma-1}{\gamma}} \frac{1}{\theta} + \frac{y_0 Z_H}{r} \right) + \left(\frac{\gamma}{1-\gamma} Z_H^{\frac{\gamma-1}{\gamma}} + y_0 Z_H \right) = 0$$

For notational simplicity, we will define

$$A_1 = C_1 Z_L^{\gamma_1-1},$$

$$A_2 = C_2 Z_L^{\gamma_2-1},$$

$$B = (Z_L)^{-\frac{1}{\gamma}}$$

$$C = \frac{Z_H}{Z_L}$$

With this new notation the above 4x4 system becomes

$$A_1 + A_2 = -\frac{y_0}{r} - \frac{\gamma}{\gamma-1} \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta} \right) B$$

$$\gamma_1 A_1 + \gamma_2 A_2 = -\frac{y_0}{r} - \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta} \right) B$$

$$\gamma_1 A_1 C^{\gamma_1-1} + \gamma_2 A_2 C^{\gamma_2-1} = B \frac{C^{-\frac{1}{\gamma}}}{\theta} - \frac{y_0}{r}$$

$$\beta \left(A_1 C^{\gamma_1-1} + A_2 C^{\gamma_2-1} \right) = -y_0 \left(\frac{\beta}{r} - 1 \right) + \frac{\gamma}{1-\gamma} \left[1 - \frac{\beta}{\theta} \right] BC^{-\frac{1}{\gamma}}$$

The first two equations allow us to solve for A_1 and A_2 as functions of B

$$A_1 = \frac{\frac{y_0}{r}(1-\gamma_2) + \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta} \right) B \left(1 - \gamma_2 \frac{\gamma}{\gamma-1} \right)}{\gamma_2 - \gamma_1},$$

$$A_2 = \frac{-\frac{y_0}{r}(1-\gamma_1) - \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta} \right) B \left(1 - \gamma_1 \frac{\gamma}{\gamma-1} \right)}{\gamma_2 - \gamma_1}.$$

The last two equations also allow us to solve for A_1 and A_2 as functions of B and C

$$A_1 = \frac{\frac{y_0}{r}\left(1 - \gamma_2 + \frac{r\gamma_2}{\beta}\right) + \left[\gamma_2\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta}\right]B}{C^{\gamma_1 - 1}(\gamma_2 - \gamma_1)},$$

$$A_2 = \frac{-\frac{y_0}{r}\left(1 - \gamma_1 + \frac{r\gamma_1}{\beta}\right) - \left[\gamma_1\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta}\right]B}{C^{\gamma_2 - 1}(\gamma_2 - \gamma_1)}.$$

By equating the A_1 and A_2 obtained from the two subsystems, we get

$$\frac{\frac{y_0}{r}\left(1 - \gamma_2 + \frac{r\gamma_2}{\beta}\right) + \left[\gamma_2\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta}\right]B}{C^{\gamma_1 - 1}(\gamma_2 - \gamma_1)} - \frac{\frac{y_0}{r}(1 - \gamma_2) + \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)B\left(1 - \gamma_2\frac{\gamma}{\gamma-1}\right)}{\gamma_2 - \gamma_1} = 0,$$

$$\frac{\frac{y_0}{r}\left(1 - \gamma_1 + \frac{r\gamma_1}{\beta}\right) + \left[\gamma_1\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta}\right]B}{C^{\gamma_2 - 1}(\gamma_2 - \gamma_1)} - \frac{\frac{y_0}{r}(1 - \gamma_1) + \left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)B\left(1 - \gamma_1\frac{\gamma}{\gamma-1}\right)}{\gamma_2 - \gamma_1} = 0$$

.We can rewrite these two equations as

$$B = \frac{\frac{y_0}{r}\left[(1 - \gamma_2)C^{\gamma_1 - 1} - \left(1 - \gamma_2 + \frac{r\gamma_2}{\beta}\right)\right]}{\gamma_2\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta} - C^{\gamma_1 - 1}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)\left(1 - \gamma_2\frac{\gamma}{\gamma-1}\right)},$$

$$B = \frac{\frac{y_0}{r}\left[(1 - \gamma_1)C^{\gamma_2 - 1} - \left(1 - \gamma_1 + \frac{r\gamma_1}{\beta}\right)\right]}{\gamma_1\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta} - C^{\gamma_2 - 1}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)\left(1 - \gamma_1\frac{\gamma}{\gamma-1}\right)}.$$

so we finally get the following non-linear equation for C

$$\frac{\frac{y_0}{r}\left[(1 - \gamma_2)C^{\gamma_1 - 1} - \left(1 - \gamma_2 + \frac{r\gamma_2}{\beta}\right)\right]}{\gamma_2\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta} - C^{\gamma_1 - 1}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)\left(1 - \gamma_2\frac{\gamma}{\gamma-1}\right)} =$$

$$\frac{\frac{y_0}{r}\left[(1 - \gamma_1)C^{\gamma_2 - 1} - \left(1 - \gamma_1 + \frac{r\gamma_1}{\beta}\right)\right]}{\gamma_1\frac{\gamma}{\gamma-1}(\frac{1}{\theta} - \frac{1}{\beta})C^{-\frac{1}{\gamma}} - \frac{C^{-\frac{1}{\gamma}}}{\theta} - C^{\gamma_2 - 1}\left(K^{\frac{1}{\gamma}} - \frac{1}{\theta}\right)\left(1 - \gamma_1\frac{\gamma}{\gamma-1}\right)}$$

Thus we are left with determining C from this equation and then, substituting above to

obtain B, A_1 and A_2.Given A_1, A_2, B, C , we can recover Z_L, Z_H and C_1, C_2. Conditions (3.72)

and (3.73) are stated as part of the assumptions of the Lemma (compare equations (3.69) and (3.70)). Finally, condition (3.75) can be shown by elementary methods. ∎

The solution proposed has the same form as the one obtained in Section 3.3. An agent should enter retirement when her wealth is sufficiently high. This will occur when Z_t is sufficiently low, which in turn is more likely to be the case when the stock market experiences a period of good returns. Similarly, the borrowing constraints will bind once Z_t is high, which will typically be associated with a period of low performance in the stock market. The consumption process will posses a similar behavior to the one described in He and Pages (1993).

The rest of the proposition follows steps similar to section 3.3.

Chapter 4

Hedging Sudden Stops and Precautionary Recessions: A quantitative framework (joint with R. Caballero)

4.1 Introduction

Most emerging economies need to borrow from abroad as they catch up with the developed world. Unfortunately, even well managed emerging economies are subject to the "sudden stop" of capital inflows. At a moment's notice, and with only limited consideration of initial external debt and conditions, these economies may be required to reverse the capital inflows that supported the preceding boom.

The deep contractions triggered by this sudden tightening of external financial constraints have great costs for these economies. Not surprisingly, local policymakers struggle to prevent such crises. During the cycle, the anti-crisis mechanism entails deep "precautionary recessions:" tight monetary and fiscal contractions at the first sign of symptoms of a potential external crunch. Over the medium run, the tools of choice are the build up of large stabilization funds and international reserves, and taxation of capital inflows. All of these *precautionary* mechanisms are extremely costly.

The case of Chile illustrates the point well. Chile's business cycle is highly correlated with the price of copper, its main export good: so much so that this price has become a signal to foreign and domestic investors, and to policymakers alike, of aggregate Chilean conditions. As a result of the many internal and external reactions to this signal, the decline in Chilean economic activity when the price of copper falls sharply is many times larger than the annuity value of the income effect of the price decline. This contrasts with the scenario in Australia, a developed economy not exposed to the possibility of a sudden stop, where similar terms-of-trade shocks are fully absorbed by the current account, with almost no impact on domestic activity and consumption.[1]

In this paper we do not attempt to explain why international financial markets treat Chile and Australia so differently, or even why the signal for the sudden stops should be so correlated with the price of copper. Instead, we take the "sudden stops" feature as a description of the environment and characterize the optimal hedging strategies under

---

[1]See , e.g., Caballero (2001).

different assumptions about imperfections in hedging markets. We also characterize the precautionary business cycle that arises when hedging opportunities are very limited.[2]

The main technical contribution of the paper is a model that is stylized enough to allow extensive analytical characterization, but is also flexible and realistic enough to generate quantitative guidance. The model has two central features: First is the *sudden stop,* which we characterize as a probabilistic event that, once triggered, requires the country to reduce the pace of external borrowing significantly. Second is a *signal.* Sudden stops have some element of predictability to them. We start by studying a simple environment where there is a perfect signal (e.g., the high yield spread, or the price of an important commodity for the country) that triggers a sudden stop once it crosses a well defined threshold. We then study the more realistic case where the threshold is blurred and a sudden stop, while increasingly likely as the signal deteriorates, may occur at any time.

Within this model, we develop two substantive themes. First we characterize *precautionary recessions*: that is, in the absence of perfect hedging, the business cycle of the economy follows the signal even if no sudden stop actually takes place. This is because as the likelihood of a sudden stop rises, the country goes into a precautionary recession. Consumption is cut to reduce the extent of the adjustment required if a sudden stop does take place.

The second and main theme is aggregate hedging strategies. An adequate hedging strategy not only reduces the extent of the crisis in the case of a sudden stop, but also reduces the need for hoarding scarce assets and for incurring sharp precautionary recessions as the signal deteriorates. We study two polar-opposite types of generic hedging instruments or strategies, as well as their intermediate cases. At one end, the hedging

---

[2]We focus on the aggregate financial problem vis-a-vis the rest of the world, in an environment where domestic policy is managed optimally and decentralization is not a source of problems. Needless to say, these assumptions seldom hold in practice. Such failures compound the problems we deal with in this paper by exacerbating the country's exposure to sudden stops. See, e.g., Caballero and Krishnamurthy (2001, 2003) and Tirole (2002) for articles dealing with decentralization problems. The literature on government's excesses is very extensive. See, e.g., Burnside et al (2003) for a recent incarnation.

contract relaxes the sudden stop constraint one-for-one. That is, each dollar of hedge can be used to relax the constraint. The other extreme is motivated by crowding out: if the resources obtained from the hedge facilitate the withdrawal of other lenders, then the main effect of the hedge is to reduce the country's debt in bad states of the world but not to provide fresh resources. By its nature, this type of hedging is an excellent substitute for precautionary recessions, but it cannot remove the sudden stop entirely.

The paper concludes with an illustration of our main results for the case of Chile. We estimate the probability of a sudden stop as a function of the price of copper and calibrate the parameters needed to obtain sharp consumption drops such as those experienced by Chile in the recent contraction of 1998/99. We then describe different hedging strategies and their impact on the volatility and levels of consumption. We discuss credit lines and their indexation to the signal as one way of reducing asymmetric information problems. For example, we argue that Chile could virtually eliminate sudden stops, precautionary recessions, and its large accumulation of precautionary assets, with a credit line that rises nonlinearly with the price of copper. The cost of this line, if fairly priced, should be around 1-2 percent of GDP. This is very little when compared with the costs of the precautionary measures currently undertaken, including large accumulation of reserves, limited short-term borrowing, and precautionary recessions.[3]

Currently, although futures markets exist for much of the income-flow effects of commodity price fluctuations, the size of the financial problem is much larger than that. The markets required for these strategies do not exist, at least in the magnitude required. In this sense, our framework also serves to highlight quantitatively the usefulness of these markets and allows us to begin gauging the potential size of the markets to be developed.[4]

---

[3]In his Nobel lecture, Robert Merton (1998) highlights the enormous savings than can be obtained by designing adequate derivatives and other contracting technologies to deal with risk management. He also argues that emerging-market economies stand to gain the most. Our findings in this paper fully support his views.

[4]See, e.g., Krugman (1988), Froot, Scharfstein, Stein (1989), Haldane (1999), Caballero (2001), for articles advocating commodity indexation of emerging markets debt. The contribution of this paper relative to that literature is to offer a quantitative framework and to link the hedging need not to commodities per se, but simply as a signal of much costlier financial constraints.

In section 2 we describe the environment and characterize the optimization problem once a sudden stop has been triggered — this provides the boundary conditions for the "precautioning phase." In section 3 we study the phase that precedes a sudden stop when the country self-insures. The main goal of this section is to characterize precautionary recessions. Section 4 describes aggregate hedging strategies under different imperfections and the degrees of crowding-out in these markets. Section 5 illustrates our results through an application to the case of Chile. Along the way, we outline an econometric approach to gauging the likelihood of a sudden stop and its correlation with an underlying signal. Section 6 concludes and is followed by an extensive technical appendix.

4.2 The Environment and the Sudden Stop Value Function

Intertemporal smoothing implies that, during the catch up process, emerging economies typically experience substantial needs for borrowing from abroad. For a variety of reasons that we do not model here, this dependence on borrowing is a source of fragility. The sudden tightening of financial constraints, or the mere anticipation of such an event, generates large drops in consumption. In this section, we formalize such an environment and characterize the optimization problem once a sudden stop has occurred. The next sections complete the description by characterizing the phase that precedes the crisis.

4.2.1 The Environment

Endowment and Preferences

We assume that the endowment grows at some constant rate, g, during $0 \leq t < \infty$. Thus, the income process $(y(t), t \geq 0)$, is described by:

$$\frac{dy(t)}{y(t)} = g\,dt, \qquad y(0) = y_0, \qquad g > 0.$$

179

Two aspects of this process are to be highlighted. First, it is deterministic. In the economies we wish to characterize, sudden stops are significantly more important than endowment shocks as triggers of deep contractions.[5] From this perspective, the main importance of endowment shocks is their "collateral" effect, and hence their potential to trigger a sudden stop. We simplify the model along this dimension and group any collateral effects contained in endowment shocks with the signal process (to be described below). Second, since $g > 0$, there is an incentive for the country to borrow early on.

Reflecting its initial net debtor position, the country starts with financial wealth, $X(0) = X_0 < 0$. However, the country's total wealth must be positive at all times:

$$X_t > -\frac{y_t}{r-g} \qquad t \geq 0,$$

where r denotes the riskless interest rate and it exceeds the rate of growth of the endowment, $r > g$.

Let c_t and c_t^* represent date t consumption and "excess" consumption, respectively, with:

$$c_t^* \equiv c_t - \kappa y_t, \qquad 0 \leq \kappa < 1.$$

The representative consumer maximizes:

$$E_t \left[\int_t^\infty u(c_s^*) e^{-\delta(s-t)} \, ds \right] \tag{4.1}$$

with

$$u(c_t^*) = \frac{c^{*1-\gamma}}{1-\gamma}, \qquad \delta > 0, \gamma > 0.$$

The parameters δ and γ are the discount rate and risk aversion coefficient, respectively. For simplicity, we assume $r = \delta$ throughout. The functional form of the utility function captures an external habit formation, with a habit level that is increasing at the rate of

---

[5]Of course, sudden stops reduce growth as well, but we capture these effects directly through the decline in consumption. In this sense, y can be thought of as potential, rather than actual, output.

the country's growth rate. This is a natural assumption in economies that exhibit strong growth, since future generations are likely to have a higher standard of living than the current ones.[6]

A frictionless benchmark

It is instructive to pause and study the solution of problem (4.1) subject to a standard intertemporal budget constraint (and absent a sudden stop constraint):

$$dX_t = (rX_t - c_t^* + y_t^*)\, dt$$

$$X_0 \text{ given}$$

with $y^* \equiv (1 - \kappa)y$.

By standard methods it can be shown that the solution to this deterministic problem is given by:

$$c_t^* = r\left(X_0 + \frac{y_0^*}{r - g}\right) \quad \text{for all } t > 0,$$

and the "total resources" of the country remain constant throughout:

$$X_t + \frac{y_t^*}{r - g} = X_0 + \frac{y_0^*}{r - g} \tag{4.2}$$

so that

$$\lim_{t \to \infty} \frac{X_t}{y_t^*} = -\frac{1}{r - g}$$

and, accordingly:

$$\lim_{t \to \infty} \frac{X_t}{y_t} = -\frac{1 - \kappa}{r - g}.$$

---

[6]Accordingly, this type of utility function allows us to impose (in a simple reduced form) a barrier on the amount of indebtedness of the country at any point in time of:

$$X_t > -\frac{(1 - \kappa)y_t}{r - g} \qquad t \geq 0.$$

and on saving, since:

$$y_t - c_t \leq (1 - \kappa)y_t \qquad t \geq 0.$$

181

Moreover, for any level of the debt-to-income ratio below its limit value, the ratio $\frac{x_t}{y_t^*}$ decreases monotonically to $-\frac{1}{r-g}$.

This case serves as a frictionless benchmark in what follows.

Signal

There is a publicly observable signal, s_t, correlated with the sudden stop and, for simplicity, uncorrelated with world endowments. This signal follows a diffusion process:

$$ds_t = \mu dt + \sigma dB_t.$$

In our basic model, s_t is a perfect signal and a crisis is triggered the first time the signal reaches a threshold, $\underline{s}$, from above. We associate with this event the stochastic time, τ:

$$\tau = \inf\{t \in (0, \infty) : s_t \leq \underline{s}\}.$$

In our second model, the signal is imperfect. In this case, a sudden stop can happen at any point in time. It is only its likelihood that is (smoothly) influenced by the signal s_t. In this case, the specification of the stochastic time, τ, depends on the realization of a stochastic jump process, the intensity of which is given by

$$\lambda_t = \exp(\alpha_0 - \alpha_1 s_t).$$

Sudden Stops

We place no limits on the country's borrowing ability up to the stochastic time τ. At this time, the country faces a "sudden stop." We do not model the informational or contractual factors behind this constraint, or the complex bargaining and restructuring process that follows once the sudden stop is triggered. Since our goal is to produce a quantitative assessment of the hedging aspects of the problem, we look for a realistic and fairly robust (across models) constraint. For this, we simply model the sudden stop as a

182

temporary and severe constraint on the rate of external borrowing. In particular, since the "natural" aggregator of a country's total wealth is the net present value of its total resources, we place the constraint on:

$$X_\tau + \frac{y_\tau^*}{r-g}.$$

We assume that at time τ, financial markets require the country to increase its total resources by $\zeta \geq e^{gT}$ within T periods. Formally:

$$\left(X_{\tau+T} + \frac{y_{\tau+T}^*}{r-g}\right) \geq \zeta \left(X_\tau + \frac{y_\tau^*}{r-g}\right). \tag{4.3}$$

It is obvious that this constraint will be always binding because as we showed in a previous subsection, at the unconstrained benchmark $\left(X_t + \frac{y_t^*}{r-g}\right)$ remains constant at all times. It is then straightforward to show that this constraint can be expressed as a constraint on the maximum allowable amount of debt/gdp at time $\tau+T$, as a function of the debt/gdp ratio at time τ. To see this, redefine

$$\zeta = e^{gT} + \phi(1 - e^{-(r-g)T}) \tag{4.4}$$

and observe that the constraint becomes:

$$X_{\tau+T} \geq X_\tau \left[e^{gT} + \phi(1 - e^{-(r-g)T})\right] + \frac{y_\tau^* \phi(1 - e^{-(r-g)T})}{r-g} \tag{4.5}$$

or

$$\frac{X_{\tau+T}}{y_{\tau+T}^*} \geq \frac{X_\tau}{y_\tau^*} \frac{\left[e^{gT} + \phi(1 - e^{-(r-g)T})\right]}{e^{gT}} + \frac{\phi(1 - e^{-(r-g)T})}{e^{gT}(r-g)}$$

Higher levels of ϕ imply higher levels of ζ and thus make the constraint tighter. It is also trivial to verify that for the special case $\phi = 0$ this constraint literally reduces to

the requirement that debt/gdp cannot grow any further between τ and $\tau + T$:[7]

$$\frac{X_{\tau+T}}{y^*_{\tau+T}} \geq \frac{X_\tau}{y^*_\tau}$$

Finally, it is interesting to note that one can think of (4.5) as a constraint on the minimum balance-of-trade surpluses the country has to accumulate over the next T periods. To see this, start with the intertemporal budget constraint:

$$X_\tau + \int_\tau^{\tau+T} e^{-r(t-\tau)} y^*_t \, dt = \int_\tau^{\tau+T} e^{-r(t-\tau)} c^*_t \, dt + e^{-rT} X_{\tau+T}$$

or:

$$X_\tau - e^{-rT} X_{\tau+T} = \int_\tau^{\tau+T} e^{-r(t-\tau)} c^*_t \, dt - \frac{y^*_\tau (1 - e^{-(r-g)T})}{r - g}$$

and replace (4.5) into it, to obtain:

$$\int_\tau^{\tau+T} e^{-r(t-\tau)} \left(c^*_t - y^*_t\right) \, dt \leq X_\tau (1 - (e^{-(r-g)T} + \phi e^{-rT}(1 - e^{-(r-g)T}))) - \phi e^{-rT} \int_\tau^{\tau+T} e^{-r(t-\tau)} y^*_t \, dt$$

or, finally:

$$\int_\tau^{\tau+T} e^{-r(t-\tau)} \left(y^*_t - c^*_t\right) \, dt \geq X_\tau ((e^{-(r-g)T} + \phi e^{-rT}(1 - e^{-(r-g)T})) - 1) + \phi e^{-rT} \int_\tau^{\tau+T} e^{-r(t-\tau)} y^*_t \, dt.$$

This gives us the constraint in terms of the balance-of-trade surpluses required from a country that is faced with a sudden stop. These required surpluses rise with the level of debt (recall that $X_\tau < 0$), ϕ, and the endowment of the country.

---

[7]Even for the $\phi = 0$ case the constraint we consider is binding at the time of the sudden stop. This is due to the fact that the debt/gdp ratio is always growing for an unconstrained country as we showed in section 2.1.1.

4.2.2 The Optimization Problem

Let us assume for now that the only financial instrument is riskless debt, so there are no hedging instruments indexed to s or τ. In this case, the country maximizes the expected utility of a representative consumer:

$$V(X_t, y_t^*) = \max_{c_t^*} \mathrm{E}\left[\int_0^\infty \frac{c_t^{*1-\gamma}}{1-\gamma} e^{-rt}\, dt\right]$$

s.t.

$$dX_t = (rX_t - c_t^* + y_t^*)\, dt$$

$$X_{\tau+T} \geq \overline{X}_{\tau+T} \equiv X_\tau\left[e^{gT} + \phi(1 - e^{-(r-g)T})\right] + \phi\frac{y_\tau^*(1 - e^{-(r-g)T})}{r-g}, \quad 0 \leq \phi < e^{rT}$$

$$\frac{dy_t^*}{y_t^*} = g\, dt$$

$$ds_t = \mu dt + \sigma dB_t$$

$$\lim_{t\to\infty} X(t)e^{-rt} = 0, \quad a.s.$$

and either

Perfect signal: $\tau = \inf\{t \in (0, \infty) : s_t < \underline{s}\}$ or

Imperfect Signal: $\tau = \inf\{t \in (0, \infty) : \int_0^\tau \lambda(s_t) dt < z\}$, $\lambda_t = e^{\alpha_0 - \alpha_1 s_t}$, $z \sim \exp(1)$

where z is independent of the standard Brownian F_t–Filtration.

4.2.3 The Sudden Stop Value Function and Amplification

We solve the optimization problem in three steps. Starting backwards, we first solve for the post-crisis period, then for the sudden-stop period, and finally for the period preceding the sudden stop. In this section we present the first two, and trivial, steps. The goal of these is to find the value function at the time of the sudden stop, $V^{SS}(X_\tau, y_\tau)$, which then can be used to find the solution of the optimization and hedging problems before the crisis takes place.

Post Sudden Stop: $t > \tau + T$

Since we made the simplifying assumption that the country suffers only one financial crisis, the maximization problem for $t > \tau + T$ is simply:

$$V(X_{\tau+T}, y^*_{\tau+T}) = \max_{c^*_t} \int_{\tau+T}^{\infty} \frac{c^{*1-\gamma}_t}{1-\gamma} e^{-r(t-(\tau+T))} \, dt \tag{4.6}$$

$$s.t.$$

$$\int_{\tau+T}^{\infty} c^*_t e^{-r(t-(\tau+T))} \, dt = X_{\tau+T} + \int_{\tau+T}^{\infty} y^*_t e^{-r(t-(\tau+T))} \, dt \tag{4.7}$$

This problem has the trivial solution:

$$c^*_t = rW^e_2 \qquad \tau + T < t < \infty \tag{4.8}$$

with

$$W^e_2 = X_{\tau+T} + \frac{y^*_{\tau+T}}{r-g}. \tag{4.9}$$

This constant can be interpreted as the "excess" wealth at date $\tau + T$ (that is, the wealth in excess of that which is needed to cover the reservation-consumption level κy_t).

Sudden Stop: $\tau \leq t \leq \tau + T$

With the continuation value function, $V(X_{\tau+T}, y^*_{\tau+T})$, from above, we can now write the maximization problem for the sudden stop phase:

$$V^{SS}(X_\tau, y^*_\tau) = \max_{c^*_t} \int_{\tau}^{\tau+T} \frac{c^{*1-\gamma}_t}{1-\gamma} e^{-r(t-\tau)} \, dt + e^{-rT} V(X_{\tau+T}, y^*_{\tau+T})$$

$$s.t.$$

$$\int_{\tau}^{\tau+T} c^*_t e^{-r(t-\tau)} \, dt = X_\tau + \int_{\tau}^{\tau+T} y^*_t e^{-r(t-\tau)} \, dt - e^{-rT} X_{\tau+T}$$

$$X_{\tau+T} \geq \overline{X}_{\tau+T}. \tag{4.10}$$

Since the country is growing and wishes to expand its consumption at a faster rate than the constraint allows, the sudden stop constraint, (4.10), is always binding (see the appendix for a formal proof):

$$X_{\tau+T} = \overline{X}_{\tau+T}. \tag{4.11}$$

Given this result, the optimization problem is straightforward. It is a deterministic consumption problem subject to a final wealth condition. A few steps of algebra show that:

$$c_t^* = \frac{r}{1 - e^{-rT}} W_1^e, \qquad \tau < t < \tau + T \tag{4.12}$$

with

$$W_1^e = X_\tau + \frac{y_\tau^*(1 - e^{-(r-g)T})}{r - g} - e^{-rT}\overline{X}_{\tau+T}.$$

It is easy to see from these expressions that as $\overline{X}_{\tau+T}$ rises, c_t^* falls. The country has to cut back consumption during the crisis in order to satisfy a tighter sudden-stop constraint.

We are now ready to determine the sudden-stop value function:

$$
\begin{aligned}
V^{SS}(X_\tau, y_\tau^*) &= \int_\tau^{\tau+T} \frac{(rW_1^e)^{1-\gamma}}{1-\gamma} e^{-r(t-\tau)} \, dt + \\
&\quad + e^{-rT} \int_{\tau+T}^\infty \frac{(rW_2^e)^{1-\gamma}}{1-\gamma} e^{-r(t-(\tau+T))} \, dt \\
&= \frac{(rW_1^e)^{1-\gamma}}{1-\gamma} \left(\frac{1 - e^{-rT}}{r} \right) + \frac{(rW_2^e)^{1-\gamma}}{1-\gamma} \left(\frac{e^{-rT}}{r} \right).
\end{aligned}
\tag{4.13}
$$

It is easy to verify that our careful choice of the constraint pays off at this stage. The value function simplifies to:

$$
\begin{aligned}
V^{SS}(X_\tau, y_\tau^*) &= K \left(\frac{1}{r} \right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \\
&= K V(X_\tau, y_\tau^*)
\end{aligned}
\tag{4.14}
$$

187

where $V(X_\tau, y_\tau^*)$ denotes the value function in the *absence* of a sudden-stop constraint and K is a constant given by:

$$K = (1 - e^{-rT})^\gamma (1 - e^{-rT}\phi(1 - e^{-(r-g)T}) - e^{-(r-g)T})^{1-\gamma} + e^{-rT} \left(\phi(1 - e^{-(r-g)T}) + e^{gT}\right)^{1-\gamma}.$$

(4.15)

That is, the dimension of the state space effectively is reduced from two, (X_τ, y_τ), to one, $\left(X_\tau + \frac{y_\tau^*}{r-g}\right)$. Moreover, *up to the constant K*, we have arrived at a value function that is identical to the one in the problem *without* sudden stops. This problem corresponds to the trivial case of an infinite horizon consumption-savings problem under certainty. Note that for $\gamma > 1$, which we assume throughout, $K > 1$ and increases with ϕ.[8]

While the particular simplicity of our formulae is due to stylized assumptions, the basic message is more general. The model also can be understood as an approximation to a potentially more complicated specification of constraints that result in higher marginal disutility of debt in the event of a crisis.

4.3 Precautionary Recessions

Let us now address our main concern and begin to characterize the country's optimization problem prior to the sudden-stop phase. Since we have assumed that there are no hedging instruments contingent on s or τ, the country's only mechanism for reducing the cost of a sudden stop is to cut consumption and borrowing before it takes place. We show that in addition to a precautionary savings result, the amount of self-insurance varies over time, because sudden stops have some elements of predictability in them. In particular,

---

[8]Observe that for $\gamma > 1$ the function $\frac{Z^{1-\gamma}}{1-\gamma}$ is negative for all $Z > 0$. Thus $K > 1$ reflects that the constrained value function is lower than the continuation value function for the unconstrained problem. The reason we need $\gamma > 1$ is that the flow aspect of the constraint implies that having a higher X_τ does not relax the constraint one for one, because the lender does not reduce its demands by the same amount during the sudden stop. When γ is below one, this effect is strong enough to discourage saving for the crisis.

when the signal of a sudden stop deteriorates, the country falls into what can be labelled as a "precautionary recession:" that is, a sharp reduction in consumption to limit the cost of the potential sudden stop. Such behavior is widely observed in emerging market economies, where private decisions and macroeconomic policy tighten on the face of external risk.

In practice, such a problem is complex for many reasons, one of the most important being the uncertainty that surrounds the factors that trigger such crises. In order to isolate the main issues, we proceed in two steps. First, we study a case where there is a perfect (stochastic) signal: A sudden stop occurs when this signal hits a minimum threshold, $\underline{s}$, for the first time. Second, we add (local) uncertainty: while sudden stops still are more likely as the signal deteriorates, they can occur at any time.

4.3.1 Perfect Signal: The Threshold Model

In the threshold model, the dynamic programming problem is:

$$V(X_t, s_t, y_t^*) = \max_{c_u^*} E\left[\int_t^\tau \frac{c_u^{*\,1-\gamma}}{1-\gamma} e^{-r(u-t)}\, du + e^{-r(\tau-t)} V^{SS}(X_\tau, y_\tau^*)|F_t\right] \qquad (4.16)$$

where

$$\tau = \inf\{(t : s \le \underline{s}) \wedge \infty\} \qquad (4.17)$$

and the evolution of (X_t, s_t, y_t) is given by:

$$dX_t = (rX_t - c_t^* + y_t^*)\, dt \qquad (4.18)$$

$$\frac{dy_t}{y_t^*} = g\, dt \qquad (4.19)$$

$$ds_t = \mu dt + \sigma dB_t. \qquad (4.20)$$

The boundaries of the value function, $V(X_t, s_t, y_t)$, can be found readily. On one end,

189

we showed in the previous section that:

$$V(X_\tau, \underline{s}, y_\tau^*) = K \left(\frac{1}{r} \right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \tag{4.21}$$

where $K > 1$ measures the intensity of the crisis and is given in (4.15).

On the other end, we show in the appendix that as s_t goes to infinity, the value function converges to the value function of the deterministic problem with no sudden stops:

$$\lim_{s \to \infty} V(X_\tau, s, y_\tau^*) = \left(\frac{1}{r} \right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g} \right)^{1-\gamma}}{1-\gamma}. \tag{4.22}$$

That is, the value function becomes independent of the signal as the crisis event becomes less and less likely.

The value function V satisfies the following Hamilton-Jacobi-Bellman (henceforth HJB) equation:

$$0 = \max_{c_t^*} \left\{ \frac{c_t^{*1-\gamma}}{1-\gamma} - V_X c_t^* \right\} - rV + V_X \left(rX + y_t^* \right) + V_{y^*} g y_t^* + V_s \mu + \frac{1}{2} V_{ss} \sigma^2 \tag{4.23}$$

subject to the boundary conditions (4.21) and (4.22).

In the appendix, we show that the solution has the form:

$$V(X_t, s_t, y_t^*) = a(s_t)^\gamma \frac{\left(X_\tau + \frac{y_t^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \tag{4.24}$$

for some twice differentiable function $a(s_t)$. The latter satisfies the boundary conditions:

$$\lim_{s \to \infty} a(s) = \left(\frac{1}{r} \right) \tag{4.25}$$

$$a(\underline{s}) = K^{1/\gamma} \left(\frac{1}{r} \right) \tag{4.26}$$

and solves an ordinary differential equation that can be obtained in three steps. First,

190

carry out the maximization in (4.23) with respect to c_t^* to get:

$$c_t^* = (V_X)^{-1/\gamma}.\tag{4.27}$$

Second, substitute this into (4.23). Third, divide the resulting expression by $\dfrac{\left(X_\tau + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$. These steps yield:

$$1 - ra + \mu a_s + \frac{1}{2}\sigma^2\left((\gamma-1)\frac{(a_s)^2}{a} + a_{ss}\right) = 0\tag{4.28}$$

which can be solved numerically subject to (4.25) and (4.26). We show in the appendix that the solution to this ODE exists and is unique. We plot this function (multiplied by r) in the left panel of Figure 4-1. The function $a(s)$ is decreasing with respect to s. As the signal deteriorates toward $\underline{s}$, this function rises rapidly, reflecting the increasing value of wealth (and marginal wealth) as the sudden stop becomes more likely.

This setup allows for an explicit characterization of the optimal consumption policy in feedback form. Given the value function, it follows from (4.27) that the optimal consumption policy takes the form:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{a(s_t)}.\tag{4.29}$$

The effect of the signal can be seen clearly in this expression. As s rises, $a(s)$ falls toward its frictionless limit. Conversely, when the signal worsens, consumption falls for any given level of income and debt. This is precisely the precautionary recession mechanism: as the crisis becomes imminent, consumption falls in order to reduce external debt and hence exposure to the sudden stop.

Applying Ito's Lemma to the right hand side of (4.29), we obtain the country's (excess) consumption process:

$$\frac{dc_t^*}{c_t^*} = \frac{1}{2}(\gamma+1)\left(\frac{\sigma a_s}{a}\right)^2 dt - \frac{a_s}{a}\sigma dB_t.\tag{4.30}$$

191

This process reinforces the above conclusion and shows that consumption has a nontrivial diffusion term. Without hedging, consumption "picks up" the volatility and fulfills the function of a hedging strategy.

Up to the functions $\frac{a_s}{a}$, which can be evaluated numerically, the rest of the terms in this expression have a straightforward interpretation. The first term is positive and captures the effect of precautionary savings, whereas the second term $\left(-\frac{a_s}{a}\sigma\right)$ is also positive and captures the sensitivity of consumption to signal news. Positive news about the signal increases consumption whereas negative news decreases it. This is the outcome of the country's attempt to accumulate resources as the sudden stop becomes imminent. As can be seen in the right panel of Figure 1, which plots $\frac{a_s}{a}$, this effect becomes more intense as the signal deteriorates. That is, the sensitivity of consumption to news about the signal increases during downturns.

4.3.2 Imperfect Signal: The jump model

Countries do not have a perfect signal for their sudden stops. Variables such as terms-of-trade and the conditions of international financial markets raise the probability of a sudden stop, but it is never as stark as in the threshold model. In this section we capture this dimension of reality by making the trigger of a sudden stop a probabilistic function of the underlying signal.

The setup and optimization problem are exactly as in (4.16)-(4.20), with the exception of the stopping time, τ, in (4.17). Now the sudden stop is a Poisson Jump-Event with intensity:

$$\lambda(s_t) = e^{\alpha_0 - \alpha_1 s_t}, \qquad \alpha_1 \geq 0.$$

Thus, we replace (4.17) for:

$$\tau = \inf\{t \in (0, \infty) : \int_0^\tau \lambda(s_t)dt < z\}, \quad \lambda(s_t) = e^{\alpha_0 - \alpha_1 s_t}, \quad z \sim \exp(1).$$

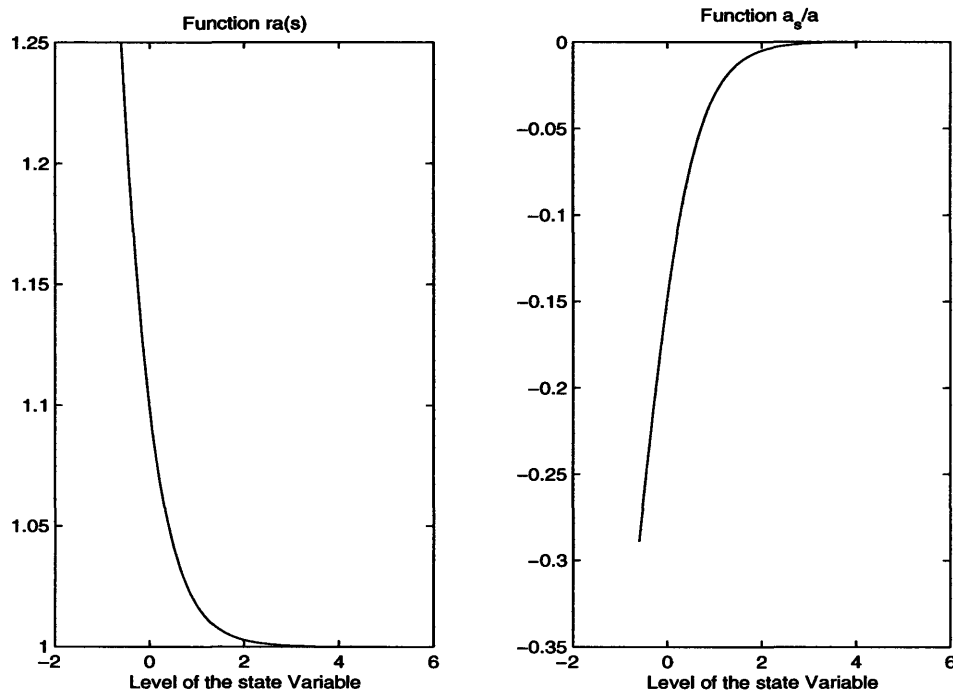Once the sudden stop takes place, events unfold exactly as in the threshold model.

192

Figure 4-1: Functions $ra(s)$ (left panel) and $\frac{a_s}{a}$ (right panel). The parameters used for this example are: $\underline{s}= -0.6, s_0 = 0, r = 0.09, T = 1, \phi = 0.35e^{rT}, \sigma = 0.23, \gamma = 7, a = 0.03, \kappa = 0.8, y_0 = 1, X_0 = -0.5.$The drift of the state variable is set to 0.

In particular, the value function still takes the form:

$$V^{SS}(X_\tau, y_\tau^*) = K \left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}.$$

The only difference is that V^{SS} now can be reached from any s for which $\lambda(s) > 0$, rather than from just $\underline{s}$. In this case the HJB equation for the value function is:

$$0 = \max_{c^*}\{\frac{c^{*1-\gamma}}{1-\gamma} - V_X c_t^*\} - rV + V_X(rX + y^*) + V_{y^*} y^* g + V_s \mu + \frac{1}{2}\sigma^2 V_{ss} + \lambda(s_t)\left[V^{SS} - V\right].$$

With essentially identical steps as in the previous subsection, we can show that our stylized framework still yields a simple solution of the form:

$$V = b(s_t)^\gamma \frac{\left(X_t + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}.$$

Precautionary Savings

Let us pause and focus on the case where sudden stops are totally unpredictable, $\alpha_1 = 0$. Since in this case $b(s)$ is no longer a function of the signal s, replacing the function V in the HJB equation yields a simple algebraic equation for b:

$$1 - rb + \frac{b}{\gamma}\lambda\left[\left(\frac{b}{\underline{b}}\right)^\gamma - 1\right] = 0.$$

where

$$\underline{b} \equiv K^{\frac{1}{\gamma}}\left(\frac{1}{r}\right).$$

It is now straightforward to obtain the (excess) consumption function from the envelope theorem:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{b}.$$

Applying simple differentiation to this expression, we obtain the country's (excess) con-

sumption process:

$$\frac{dc_t^*}{c_t^*} = \frac{1}{\gamma}\lambda\left[\left(\frac{b}{\bar{b}}\right)^\gamma - 1\right] dt$$

Relative to the threshold model, there is a new drift term in the excess consumption process. This term reflects the additional precautionary savings attributable to the local uncertainty introduced by the strictly positive probability of a sudden stop taking place in the next instant. But, because this probability is not correlated with the signal s, there are no precautionary recessions. In this case, the pattern of saving for self-insurance is not a source of business cycles.

Precautionary Recessions

Let us now return to the general case, where $\alpha_1 > 0$. After a few simplifications, substituting the function V in the HJB equation yields the following ODE:

$$1 - rb + b_s\mu + \left[(\gamma - 1)\frac{(b_s)^2}{b} + b_{ss}\right]\frac{1}{2}\sigma^2 + \frac{b}{\gamma}\lambda(s_t)\left[\left(\frac{b}{\bar{b}}\right)^\gamma - 1\right] = 0, \qquad (4.31)$$

which differs from (4.28) only in the last term.

The boundary conditions are also different. First, when s goes to infinity there is still a strictly positive probability of a crisis. Second, as conditions worsen, there is no equivalent to the threshold $\underline{s}$ where a crisis happens with probability one. Let

$$b^* = \lim_{s\to\infty} b(s).$$

$$b_* = \lim_{s\to-\infty} b(s).$$

b^* and b_* are given by:

$$b^* = \frac{1}{r}$$
$$b_* = K^{1/\gamma}\frac{1}{r}.$$

195

Again, it is straightforward to obtain the (excess) consumption function from the envelope theorem:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{b(s_t)}.$$

(4.32)

Finally, applying Ito's lemma to the right hand side of (4.32), we obtain the country's (excess) consumption process:[9]

$$\frac{dc_t^*}{c_t^*} = \left[\frac{\gamma+1}{2}\left(\frac{\sigma b_s}{b}\right)^2 + \frac{1}{\gamma}\lambda(s_t)\left[\left(\frac{b}{\bar{b}}\right)^{\gamma} - 1\right]\right]dt - \frac{b_s}{b}\sigma dB_t$$

(4.33)

This case integrates the insights of the previous models. There is ongoing precautionary savings attributable to local uncertainty, but the amount varies with the signal. For the same reasons of the previous section, $\frac{b_s}{b}$ is strictly negative, so that the conclusions of the previous section also carry through here. A deteriorating signal increases the need to accumulate resources and accordingly makes consumption respond more to a one-standard-deviation increase in the signal by a factor of $-\frac{b_s}{b}\sigma$. Finally, the drift term includes a new precautionary term $\left(\frac{1}{\gamma}\lambda(s_t)\left[\left(\frac{b}{\bar{b}}\right)^{\gamma} - 1\right]\right)$, which we also found in the case $\alpha_1 = 0$, and captures precautioning against the Poisson jump event that can occur at any time. In the application section of the paper we quantify these effects in the context of Chile.

4.4 Aggregate Hedging

Precautionary savings and recessions are very costly and imperfect self-insurance mechanisms for smoothing the impact of a sudden stop. In this section, we enlarge the options of the country and allow it to hedge using derivatives and insurance contracts. Of course, the effectiveness of the hedging strategy depends on the contracts and instruments that are available to the country, how these contracts enter into the sudden stop constraint,

---

[9]We are obviously focusing on a case where the jump has not yet taken place so that $dq = 0$, where q is the poisson event.

and how accurate the crisis-signal is.

The issue of how the fresh funds relax the country's constraint is at the core of the current debate about optimal assistance mechanisms and is not well understood. We do not try to solve this debate but rather characterize optimal hedging strategies under different scenarios. We begin by exploring two polar cases: In the first case, the hedge cannot relax the constraint directly but only improves the initial conditions of the country once it hits the constraint. This would be the case, for example, when the resources from the hedge are used entirely to pay other lenders.[10] In the second case, the country can buy hedging that directly relaxes the sudden stop constraint. That is, each dollar received from the hedge can be used to fulfill the sudden stop constraint.

In this section, we isolate the above distinction and focus on the simpler threshold model. We do this for analytical tractability, because we can use complete-markets tools in this model that allow us to obtain closed-form solutions. We use these results as an approximation benchmark for the more realistic imperfect signal case, which is the focus of the empirical section.

4.4.1 Hedging Precautionary Recessions

Assume for the moment that the country has no mechanism of injecting net resources into the sudden stop constraint. However, the country faces complete hedging markets before the sudden stop arises. In this sense, we can interpret the sudden stop as a time when all financial markets close and the country is left only with its resources at the outset of the crisis: X_{τ^-}.

An alternative interpretation is that the hedge is used (crowded out) by existing lenders and the sudden-stop constraint remains unchanged, except for the positive effect of a reduced initial debt (hence, there is a reduction in the required balance of trade surplus).

---

[10]Recall that our model has no straight default. Implicitly, however, it does allow for limited rescheduling.

Let us re-write the dynamic budget constraint for $t \geq 0$ as:

$$dX_t = (rX_t - c_t^* + y_t^*)dt + \pi_t dF_t$$

where dF_t denotes the profit/loss in the futures position and π_t is the number of future contracts.

Since under the assumption of zero risk premium and any constant convenience yield for copper, d, we have:

$$F_t = S_t e^{(r-d)(T-t)}$$

we can apply Ito's lemma to obtain:

$$dF_t = \sigma F_t dB_t.$$

Defining the portfolio process as $p_t = \pi_t F_t$, we obtain the new dynamic budget constraint:

$$dX_t = (rX_t - c_t^* + y_t^*)dt + p_t \sigma dB_t. \tag{4.34}$$

Thus we modify the dynamic programming problem in (4.16) to allow for a hedging portfolio p_t (correspondingly, we refer to this hedging portfolio as p−hedging):

$$V(X_t, s_t, y_t^*) = \max_{c_u^*, p_u} E\left[\int_t^\tau \frac{c_u^{*\,1-\gamma}}{1-\gamma} e^{-r(u-t)}\, du + e^{-r(\tau-t)} V^{SS}(X_\tau, y_\tau^*)|F_t\right] \tag{4.35}$$

where

$$\tau = \inf\{(t : s \leq \underline{s}) \wedge \infty\}$$

and the evolution of (X_t, s_t, y_t) is now given by:

$$dX_t = (rX_t - c_t^* + y_t^*)\,dt + p_t\sigma dB_t \qquad (4.36)$$

$$\frac{dy_t^*}{y_t^*} = g\,dt \qquad (4.37)$$

$$ds_t = \mu dt + \sigma dB_t. \qquad (4.38)$$

It turns out that the solution to this problem is simpler than the no-hedging problem because we can use well-known techniques from the complete markets case.[11] Following a derivation similar to the no-hedging case, one verifies that the value function of the problem in the presence of complete hedging is given by (see the appendix):

$$V(X_t, y_t^*, s_t) = a^p(s_t)^\gamma \frac{\left(X_t + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$$

with

$$a^p(s_t) = \frac{1}{r}\left(1 + e^{-\lambda_1(s-\underline{s})}(K^{\frac{1}{\gamma}} - 1)\right) \qquad (4.39)$$

$$\lambda_1 = \frac{\mu + \sqrt{\mu^2 + 2\sigma^2 r}}{\sigma^2}.$$

Correspondingly, the consumption policy in feedback form is:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{a^p(s_t)}. \qquad (4.40)$$

While it may appear from this expression that not much has changed with respect to the no-hedging case, this is not so. To see this, apply Ito's lemma to the right hand side and simplify (see the appendix), to obtain:

$$dc^* = 0.$$

---

[11]This is possible because markets are complete at all dates but τ.

That is, excess consumption is constant throughout the pre-sudden stop phase. *There are no more precautionary recessions*, and the signal does not affect consumption.

How can this be reconciled with the consumption expression in (4.40)? The answer is in the behavior of X_t. While in the no-hedging case X_t was simply made of riskless debt, it now includes a hedging portfolio component, p_t:

$$p_t = -\lambda_1 \left(\frac{e^{-\lambda_1(s-\underline{s})}(K^{\frac{1}{\gamma}} - 1)}{1 + e^{-\lambda_1(s-\underline{s})}(K^{\frac{1}{\gamma}} - 1)} \right) \left(X_t + \frac{y_t^*}{r - g} \right)$$

which once replaced in

$$dX_t = (rX_t - c_t^* + y_t^*)dt + p_t \sigma dB_t$$

implies that all the variation in precautioning is absorbed by X_t rather than by consumption.[12] The hedging portfolio, p, is always negative. Its absolute value is largest when the signal is at the trigger point $\underline{s}$ and goes to 0 when the signal goes to infinity. This means that the country is always shorting the asset that is perfectly correlated with the signal, and the amount of shorting rises as the signal worsens. The counterpart of this investment position is a reduction in the countries' external debt as the signal deteriorates.

However, this form of insurance will not entirely remove the consumption drop during the sudden stop. All the country can do is to arrive at the sudden stop with less debt, and hence to reduce the size of the trade surpluses it is required to have during the crisis. Because of this drop, the country still cuts consumption and borrowing throughout the pre-sudden stop phase for precautionary reasons. It is simply no longer state (signal)-dependent.

---

[12]The appendix provides a formal proof of this statement, which follows steps similar to those in Karatzas and Wang (2000), and return to a fuller characterization in the implementation section.

Optimal consumption before the sudden stop is constant at (see the appendix):

$$c^* = \frac{X_0 + \frac{y_0^*}{r-g}}{E\left[\int_0^\tau e^{-rt}\,dt + e^{-r\tau}\frac{K^{1/\gamma}}{r}\right]}.$$

Since $K > 1$, it clearly is lower than the level of consumption in a framework without sudden stops:

$$c^* < c^{*NSS} = r\left(X_0 + \frac{y_0^*}{r-g}\right).$$

4.4.2 Hedging Sudden Stops

Let us now assume that there is a second asset, H, with the property that it can be excluded from the sudden stop constraint and hence can be used directly to overcome the forced savings problem. This can be thought of as a credit line that does not crowd out alternative funding options (we refer to this case as h–hedging).

Formally, the problem becomes:

$$V(X_t, s_t, y_t^*) = \max_{c_u^*, p_u, \widetilde{p}_u, dN_u} E\left[\int_t^\tau \frac{c_u^{*\,1-\gamma}}{1-\gamma}e^{-r(u-t)}\,du + e^{-r(\tau-t)}V^{SS}(X_\tau, H_t, y_\tau^*)|F_t\right] \quad (4.41)$$

where

$$\tau = \inf\{(t : s \le \underline{s}) \wedge \infty\} \quad (4.42)$$

and the evolution of (X_t, H_t, s_t, y_t) is now given by:

$$
\begin{aligned}
dX_t &= (rX_t - c_t^* + y_t^*)\,dt + p_t\sigma dB_t - \xi dN_t & (4.43)\\
dH_t &= rH_t + \widetilde{p}_t\sigma dB_t + dN_t & (4.44)\\
\frac{dy_t^*}{y_t^*} &= g\,dt & (4.45)\\
ds_t &= \mu dt + \sigma dB_t. & (4.46)\\
H_t &\ge 0, \quad \xi \ge 1 & (4.47)
\end{aligned}
$$

Observe that we now have a second portfolio process $\tilde{p}_t$ and an increasing process dN_t representing additions into the second type of asset. Moreover, every addition into H_t from X_t is accompanied by a markup fee, $\xi \geq 1$. Since the income flows arrive in the form of X, the presence of a markup would seem to imply that H_t is dominated by X_t. However, this is not the case precisely because H_t has the ability to relax the constraint: $X_{\tau+T} \geq \overline{X_\tau}$. In other words we assume that at time τ the holdings of the asset H_τ will be added to $X_{\tau-}$ in order to relax the constraints as we explain below.

To be more precise, let us re-write the sudden stop constraint as:

$$X_{\tau+T} \geq X_{\tau-} \left(\phi(1 - e^{-(r-g)T}) + e^{gT} \right) + \phi \frac{y_\tau^*}{r-g}(1 - e^{-(r-g)T}) \overset{d}{=} \overline{X_\tau^-}, \quad 0 \leq \phi < e^{rT}. \tag{4.48}$$

Notice that the constraint now reads $X_{\tau-}$ rather than X_τ. This is because the country can use its holdings of H_τ to relax the constraint directly, so that we can write:

$$X_{\tau+} = X_{\tau-} + H_\tau. \tag{4.49}$$

It follows that the constraint will be non-binding if

$$H_\tau \geq h^* \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right) \tag{4.50}$$

where

$$h^* \equiv \left(\phi(1 - e^{-(r-g)T}) + e^{gT} - 1 \right). \tag{4.51}$$

To see why this is so, note that for $H_\tau = h^* \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)$ at time τ^+ total resources become:

$$
\begin{aligned}
X_{\tau+} &= X_{\tau-} + H_\tau \\
&= X_{\tau-} \left(\phi(1 - e^{-(r-g)T}) + e^{gT} \right) + \phi \frac{y_\tau^*}{r-g}(1 - e^{-(r-g)T}) + \left(e^{gT} - 1 \right) \frac{y_\tau^*}{r-g} \\
&= \overline{X_\tau^-} + \frac{y_{\tau+T}^*}{r-g} - \frac{y_\tau^*}{r-g}.
\end{aligned}
$$

However, as we established in equation (4.2), the unconstrained solution automatically satisfies the relation:

$$X_{\tau+} + \frac{y_\tau^*}{r-g} = X_{\tau+T} + \frac{y_{\tau+T}^*}{r-g},$$

which implies that $X_{\tau+T} \geq \overline{X_\tau}^-$, i.e. the constraint is satisfied automatically for $H_\tau \geq h^* \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)$.

A modification of the martingale approach of Karatzas and Wang (2000) is particularly well suited to finding and characterizing the solution $< c_t^*, p_t, \widetilde{p}_t, X_t, H_t, dN_t >$ in this context. Applying Ito's lemma to $e^{-rt}X_t$ and $e^{-rt}H_t$, and because we have a bounded portfolio and a stopping time that is finite almost surely, for any feasible consumption-portfolio plan we get:

$$\begin{aligned}
E(e^{-r\tau}X_\tau) &= X_0 + E\left(\int_0^\tau (y_t^* - c_t^*) e^{-rt} \, dt - \xi \int_0^\tau e^{-rt} dN_t \right) \\
E(e^{-r\tau}H_\tau) &= H_0 + E\left(\int_0^\tau e^{-rt} dN_t \right).
\end{aligned}$$

Combining the above two equations we get:

$$E(e^{-r\tau}(X_\tau + \xi H_\tau)) = X_0 + \xi H_0 + E\left(\int_0^\tau (y_t^* - c_t^*) e^{-rt} \, dt \right)$$

We will assume that at time 0 the country starts with $H_0 = 0$, so that we end up with:

$$E(e^{-r\tau}(X_\tau + \xi H_\tau)) = X_0 + E\left(\int_0^\tau (y_t^* - c_t^*) e^{-rt} \, dt \right). \tag{4.52}$$

An interesting interpretation of this equation is as follows: Suppose the country is offered a contingent credit line that it can structure as it desires. I.e., it can choose H_τ "state by state." But assume too that it pays a "markup," ξ, on this credit line. Then the price of the credit line is:

$$P_{ccl} = \xi E(e^{-r\tau}H_\tau).$$

With this definition we can rewrite (4.52) as:

$$E(e^{-r\tau}X_\tau) = (X_0 - P_{ccl}) + E\left(\int_0^\tau (y_t^* - c_t^*) e^{-rt}\, dt\right)$$

which is the standard budget constraint (e.g. of Section 4.1.) after subtracting the initial cost of the contingent credit line from X_0.

From now on, it will prove useful to express the amount of the credit line as a fraction of $\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right)$ so that:

$$h_\tau = \frac{H_\tau}{\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right)}.$$

For values of $h < h^*$ the constraint will be binding and consumption will drop during the sudden stop. It is not difficult to show that in this case consumption after the sudden stop $(t > \tau + T)$ is equal to:

$$c_{\tau+T}^* = r\left(1 + h^*\right)\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right).$$

which follows from (4.4) and (4.8).

Given X_{τ^-}, this is the same quantity as in the no-hedging case because the sudden stop constraint binds as long as $h < h^*$. However, the effect of the credit line is to raise consumption during the sudden stop phase $(\tau < t < \tau + T)$ since H_τ is entirely used during this period . Consumption during the crisis is now equal to:

$$c_{\tau+}^* = \frac{r}{1 - e^{-rT}}\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right)(1 + h - e^{-rT}\phi(1 - e^{-(r-g)T}) - e^{-(r-g)T})$$

which for $h = h^*$ rises to

$$r(1 + h^*)\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right) = c_{\tau+T}^*.$$

Accordingly, the value function at time τ is:

$$V^{SS} = C \frac{\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \tag{4.53}$$

$$C = \left(\frac{1}{r}\right)^\gamma \left[(1+h^*)^{1-\gamma} 1_{\{h=h^*\}} + K_1 1_{\{h<h^*\}}\right]$$

where:[13]

$$K_1 = (1 - e^{-rT})^\gamma (1 + h - e^{-rT}(1+h^*))^{1-\gamma} + e^{-rT}(1+h^*)^{1-\gamma} < K.$$

Now the optimization problem has the same form as before except that the intertemporal constraint takes into account the fact that the consumer essentially faces two types of assets, X and H.

Adopting the Cox-Huang (1989) methodology and its application to problems involving a random stopping time in Karatzas and Wang (2000), we are able to reduce the problem to the following static problem:

$$\min_k \max_{c_t, X_\tau, H_\tau} E \left[\int_0^\tau \frac{c_t^{*1-\gamma}}{1-\gamma} e^{-rt}\, dt + e^{-r\tau} V^{SS}(X_\tau, H_\tau) - k \left(\int_0^\tau c_t e^{-rt}\, dt + e^{-r\tau}(X_\tau + \xi H_\tau) \right) \right]$$

$$s.t.\ E \left(\int_0^\tau c_t^* e^{-rt}\, dt + e^{-r\tau}(X_\tau + \xi H_\tau) \right) = X_0 + E \left(\int_0^\tau y_t^* e^{-rt}\, dt \right)$$

where V^{SS} is given by (4.53).

The presence of dynamically complete markets allows us to reduce the problem to an essentially static problem. Parties can contract on the payoffs to be transferred "state

---

[13]It might seem puzzling why $C = \left(\frac{1}{r}\right)^\gamma (1+h^*)^{1-\gamma}$ when $h = h^*$ and not $C = \left(\frac{1}{r}\right)^\gamma$. The reason is that total resources at time τ^+ are now given by $(1+h^*)\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right)$. Actually one can easily show in the framework of the threshold model that in the absence of a markup, it will be the case that $(1+h^*)\left(X_{\tau^-} + \frac{y_\tau^*}{r-g}\right) = X_\tau^{unc} + \frac{y_\tau^*}{r-g}$, where X_τ^{unc} denotes the level of X_τ in the complete absence of sudden stops.

by state." In other words, the objective inside the brackets is maximized state by state and the optimal payoff is replicated dynamically.

It is easy to show that in this framework one can derive optimal consumption to be:

$$c_t^* = k^{-\frac{1}{\gamma}}, t \in (0, \tau).$$

where k is a constant that is determined in such a way that the intertemporal budget constraint is satisfied.

The crucial step is the maximization of the problem involving the continuation value function at time τ. In the appendix we show that the solution to this problem is:

$$X_{\tau-} = \frac{k^{-\frac{1}{\gamma}}}{r(1 + h^*)} \Gamma^{-\frac{1}{\gamma}} - \frac{y_\tau^*}{r - g}$$

$$h^{opt} = h^* - (1 - e^{-rT})(1 + h^*) \left[1 - \left(\frac{\Gamma}{\xi} \right)^{\frac{1}{\gamma}} \right]$$

with

$$0 < \Gamma \equiv \xi - \frac{(\xi - 1)}{e^{-rT}(1 + h^*)} < 1.$$

Notice that as $\xi \to 1$, $h \to h^*$. As might be expected in this case, where it costs nothing to avoid the constraint, the country will optimally choose $h = h^*$. If $\xi > 1$ it will be true that $h < h^*$ and some consumption adjustment during the sudden stop will be required.

Because of the homotheticity of the problem, the optimal credit line ratio does not depend on the level of initial wealth. To complete the solution to the overall problem, let us return to the time 0 budget constraint and combine everything to solve for c^*:

$$c^* = \left[E\left[\int_0^\tau e^{-rt}\,dt \right] + \frac{E\left[e^{-r\tau}\right]}{r} \frac{(1+h^{opt}\xi)\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)} \right]^{-1} \left(X_0 + \frac{y_0^*}{r-g} \right) =$$

$$= \left[\frac{1 - e^{-\lambda(s-\underline{s})}}{r} + \frac{e^{-\lambda(s-\underline{s})}}{r} \frac{(1+h^{opt}\xi)\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)} \right]^{-1} \left(X_0 + \frac{y_0^*}{r-g} \right)$$

where as usual:

$$\lambda = \frac{\mu + \sqrt{\mu^2 + 2\sigma^2 r}}{\sigma^2} \tag{4.54}$$

The credit line reduces the magnitude of the sudden stop, which translates into higher pre-sudden stop consumption. However, since $h^{opt} < h^*$, there is still scope for hedging; p_t takes the slack and eliminates the remaining need for a precautionary recession. The economy still suffers through the sudden stop, but significantly less than without hedging.

4.4.3 Imperfect Signal

In general, crises will be correlated with a signal but not perfectly. This complicates the p-insurance case but the nature of the solution is similar to that of the threshold model. On the other hand, the fact that τ can occur for any s, considerably complicates the h-hedging case. In this section we develop the former case to contrast it with the perfect-signal scenario. We return to the $h-$hedging case in the empirical section.

The steps of the derivation are similar to those in the threshold model, and we relegate them to the appendix. The optimal (excess) consumption and portfolio policies become:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g} \right)}{b^H(s_t)} \tag{4.55}$$

$$p_t = \frac{b_s^H(s_t) \left(X_t + \frac{y_t^*}{r-g} \right)}{b^H(s_t)} \tag{4.56}$$

Unfortunately the function $b^H(s)$ can no longer be characterized in closed form, but it

can be computed numerically.

Applying Ito's lemma to the right hand side of (4.55), we obtain the process:

$$\frac{dc_t^*}{c_t^*} = \left(\frac{\lambda(s_t)}{\gamma} \left[\left(\frac{b}{b^H(s_t)} \right)^\gamma - 1 \right] \right) dt. \tag{4.57}$$

Comparing this expression to that of the unhedged case in (4.33), shows that the possibility of hedging eliminates the diffusion term. However, unlike the hedged case in the threshold model, there is still pre-sudden-stop precautionary savings and these fluctuate over time as the signal changes. This happens because the sudden-stop jump is not directly contractible.

4.5   An Illustration: The Case of Chile

In this section we illustrate our main results through an application to the case of Chile. This is a good case study since Chile is an open economy, with most of its recent business cycle attributable to capital flows' volatility. Moreover, the price of copper (its main export) is an excellent indicator of investor attitude toward Chile.

4.5.1   Calibration

While our purpose here is only illustrative, and hence our search for parameters is rather informal, we spend some time describing our estimation/calibration of the key function $\lambda(s)$.

Estimating $\lambda(s)$

Sudden stops are very severe but rare events. This makes it hard to estimate $\lambda(s)$ with any precision. However, we still highlight our procedure because it provides a good starting point for an actual implementation. Similarly, a key issue is determining the components of the signal, s. Given the limited goal of this section, we use only the logarithm of the

price of copper. In a true implementation it also would be worth including some global risk-financial indicator, such as the U.S. high yield spread or the EMBI+, and removing slow moving trends from the price of copper.[14]

With these caveats behind us, let us describe the reduced-form Markov regime switching model we use to estimate $\lambda(s)$. Unlike the standard regime-switching model, ours has time-varying transition probabilities. Our left-hand side variable is aggregate demand, Y_t, which we assume to be generated by the process:[15]

$$Y_t = \mu_0 + 1_{\{z_t=1\}}\mu_1 + \sigma_{z_t}\varepsilon, \quad \varepsilon \sim N(0,1), \mu_1 < 0.$$

The growth rate of Y_t depends on an unobserved regime z_t that takes values 0 and 1. When the value of the regime is 0, the regime is "normal" and growth is just a normal variable with mean μ_0 and variance σ_0^2. Otherwise, the regime is "abnormal" and growth has a lower mean, $\mu_0 + \mu_1$, and a variance σ_1^2. The transition matrix between the two states is assumed to be the following:

$$\Pr(z_{t+1} = 0|z_t = 0, X_{s,s\in(t,t+1)}) = \exp^{-\int_t^{t+1} \lambda(x_u)du} = p_{00} \quad \Pr(z_{t+1} = 1|z_t = 0, X_{s,s\in(t,t+1)}) = p_{01}$$

$$\Pr(z_{t+1} = 0|z_t = 1, X_{s,s\in(t,t+1)}) = p_{10} \qquad \Pr(z_{t+1} = 1|z_t = 1, X_{s,s\in(t,t+1)}) = q$$

where $p_{01} = 1 - p_{00}$, $p_{10} = 1 - q$, and X_t stands for the regressors that enter the hazard function (the logarithm of the price of copper in this application).[16] For the purpose of estimating $\lambda(s)$, we use annual GDP data starting from 1976 to 1999 and monthly

---

[14]Removing the slow moving trend not only seems to improve the fit but also is a key ingredient in designing long run insurance and hedging contracts. Few investors/insurers are likely to be willing to hold long-term risk on a variable that may turn out to be non-stationary.

[15]All aggregate quantity data are highly correlated during large crises; hence the particular series used is immaterial for our purposes.

[16]Strictly speaking the quantity $\exp^{-\int_t^{t+1} \lambda(x_u)du}$ is unobservable, since we cannot observe the continuous path of copper, only discrete points. However, we can obtain copper data at reasonable high frequencies so that we can safely ignore this issue and calculate the integral by a Rieman sum as

$$\exp^{-\int_t^{t+1} \lambda(x_u)du} \approx \exp^{-\sum \lambda(x_u)\Delta u}$$

copper data for the same period (source, International Financial Statistics).[17]

For estimation, we use a Bayesian approach that is suitable for the very few datapoints that we have. Bayesian inference seems natural in this context, since it allows exact statements that do not require asymptotic justification. To estimate the parameters of interest (namely (α_0, α_1)) we use a multimove Gibbs Sampler as described in Kim and Nelson (1998,1999) and is based on the filtering algorithm of Hamilton (1989). For details we refer the reader to these references. The basic idea is to augment the parameter set by treating the unobserved states as parameters. Then we fix a draw from the posterior distribution of $(\mu_0, \mu_1, \sigma_0, \sigma_1, q, \alpha_0, \alpha_1)$ and conditional on these parameters we draw from the posterior distribution of the states in a sigle step as described in Kim and Nelson (1999). Given the draw from the posterior distribution of the states we sample the conditional means μ_0, μ_1 using a conjugate normal / trucated normal prior which leads to a normal / trucated normal posterior.[18] Conditional on the states and the draw from (μ_0, μ_1) se sample from the posterior distribution of (σ_0, σ_1) using inverse gamma priors which lead to posteriors in the same class. Similarly, fixing the states and the draw from $(\mu_0, \mu_1, \sigma_0, \sigma_1)$ we draw q using a conjugate beta prior leading to conjugate beta posterior. The sampling of (α_0, α_1) presents the difficulty that there does not seem to be a natural conjugate prior to use and thus we use a Metropolis Hastings-Random walk-accept-reject step where we sample from a bivariate normal centered at the previous iterations' estimate as described in Robert and Casella (1999). This provides us with a new draw form $(\mu_0, \mu_1, \sigma_0, \sigma_1, q, \alpha_0, \alpha_1)$ and conditional on this new draw we can iterate the algorithm by filtering the states again, based on the new draw etc. It is then a standard result in Bayesian computation that the stationary distribution of the parameters sampled with these procedure (treating the unobserved states as parameters

---

[17]There is a caveat here. Our extended-sample starts from 1972, but all the years up to 1976 are part of a deep contraction due to political turmoil combined with extremely weak external conditions. However, since the extended-sample starts during abnormal years these do not influence the estimation of $\lambda(s)$ (which is estimated from the transitions from normal to abnormal states). It is in this sense that the sample relevant for the estimation of the latter starts in 1976. For details see below.

[18]For details see Albert and Chib (1993).

too) coincides with the posterior distribution of the parameters.[19]

A first output of this procedure is Figure 4-2 which plots the probability of being in an abnormal state. The model recognizes roughly three years out of the 24 as abnormal years: the early 1980s and the recent episode following the Asian/Russian crises. The early 1980s episode corresponds to a devastating debt-crisis, and was significantly more severe than the recent one. In fact, the recent episode appears to be a mix of a milder sudden stop and a precautionary recession.

An interesting observation about these probabilities is that they allow us to identify the abnormal regimes with great confidence. To improve the tightness of the posterior confidence intervals concerning the parameters of interest (α_0, α_1) we observe that conditional on the states the (log-) likelihood function becomes:

$$\sum_{z_i=0} [\chi_i \log(p_{00}(\alpha_0, \alpha_1)) + (1 - \chi_i) \log(p_{01}(\alpha_0, \alpha_1))]$$

where χ_i takes the value 1 if there is no transition to state 1, and takes the value 0 otherwise. It is interesting to notice that the likelihood for (α_0, α_1) depends on the data only through the filtered states. In other words all other parameters of the model including the GDP-data are relevant for our purpose of estimating (α_0, α_1) only to the extent that they influence the identification of the states. In other words if we were to condition directly on the states we would be able to get rid of the noise introduced by filtering.

Given the few data points and the quite clear identification of the abnormal states

---

[19]To reduce computational time and satisfy the technical conditions required for the applicability of the Gibbs Sampler, we used proper priors for (μ_0, μ_1) and (α_0, α_1) and improper priors for the rest of the parameters. The proper prior for (μ_0, μ_1) was Normal / Truncated normal with means (0,0) and a diagonal covariance matrix. The standard deviations where chosen to be roughly 5 times the range of the sample, so that the priors had effectively no influence on the estimation. Similarly for (α_0, α_1) we used independent normal priors centered at 0 with a standard deviation of 10. Even when we experimented with more diffuse priors the algorithm produced virtually identical results but computational time was significantly increased, because convergence was significantly slower. Most importantly though, for the results that we report in Table 1 and that we use in the calibration exercise we used completely flat priors.
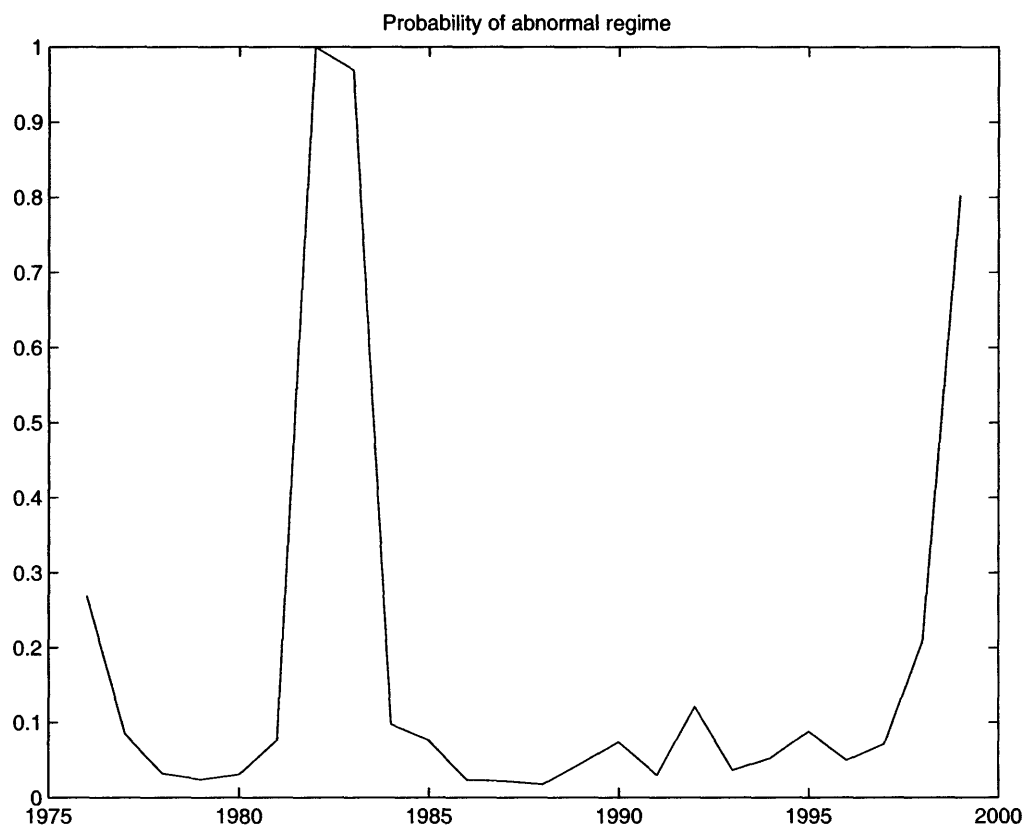
Figure 4-2: Probability of abnormal regime

Posterior Distributions							
	mean	std. Deviation	10	25	50	75	90
α_0	-2.522	1.067	-3.826	-3.027	-2.316	-1.789	-1.420
α_1	-5.170	3.387	-9.259	-6.802	-4.708	-3.040	-1.515

Table 4.1: Posterior Distributions of the Parameters (α_0, α_1) Conditional on the States that are Identified as Abnormal. 10, 25 etc. refer to the respective quantiles.

we report directly the posterior distribution of (α_0, α_1) conditional on the early 80's and 1999 being the abnormal states. we report the posterior distributions of the parameters (α_0, α_1) conditional on the states that are identified as abnormal.[20] Table 4.1 reports these results.

Correspondingly, in what follows we use $\alpha_0 = 2.5, \alpha_1 = 5.2$ as our benchmark values for the function $\lambda(s)$.

Other parameters

We calibrate ϕ and T to generate the average cumulative consumption drop caused by the sudden stops in the sample, which is about 8-10% of GDP. One such configuration is $T = 1$ and $\phi = 0.35e^{rT}$.

The parameters κ and γ are calibrated to generate reasonable amounts of steady-state debt (and hence reasonable amounts of insurance needs) together with significant precautionary fluctuations (see below, in particular, to explain much of the precautionary recession experienced by Chile at the early stages of the 1998-9 episode). For this, we set $\kappa = 0.8$ and $\gamma = 7$.

The interest rate r is set to 0.09 and the growth rate to 0.03. The latter is a constant-rate approximation to a path that grows significantly faster than that level for a few years, while the country is catching up, and then decelerates below that level forever.

---

[20] I.e., states that have a posterior probability of being abnormal above 0.75. The conditioning is done in order to increase the precision of the estimates and seems to be warranted just by a visual inspection of Figure 4-2.

We normalize initial GDP (y_0) to 1, and set the initial debt-to-GDP ratio, X_0, to 0.5.

Finally, we approximate the process for copper by a driftless Brownian motion with constant volatility σ, which we estimate using monthly data from 1972 to 1999 (source, IFS). We find a value of roughly $\sigma = 0.23$ and normalize the initial value of s to 0.

4.5.2 Aggregate Hedging

Recall that our concern in this economy is to stabilize both precautionary recessions and sudden stops. Let us start with the former.

Hedging Precautionary Recessions

Evaluating at $\frac{c_t}{y_t} \approx 1$, and $s = 0$ in the calculation of $\frac{b_s}{b}$, we can approximate the volatility of log consumption in our model before the sudden stop takes place by:

$$-(1 - \kappa)\frac{b_s}{b}\sigma \approx 0.01.$$

That is, despite the absence of income fluctuations in our model, fluctuations in precautionary behavior can account for about a fourth of consumption volatility in the regimes that are characterized as normal by our algorithm. More importantly, the contribution of precautionary fluctuations is particularly relevant near sudden stops, as shown by Figure 3.

Panel (a) in this figure shows a random realization of the path s that runs for nearly eight years before a sudden stop takes place. The main features of this path are not too different from the realization of the price of copper during the 1990s. In particular, the large rise in s near the middle of the path followed by a sharp decline toward the end of the period is reminiscent of the path of the price of copper from 1996 to 2000.

The dashed line in the bottom left panel (b) illustrates the corresponding path of (excess) consumption generated by our model without hedging. The drift in the process is due to average precautionary savings. More interestingly, one can clearly see the
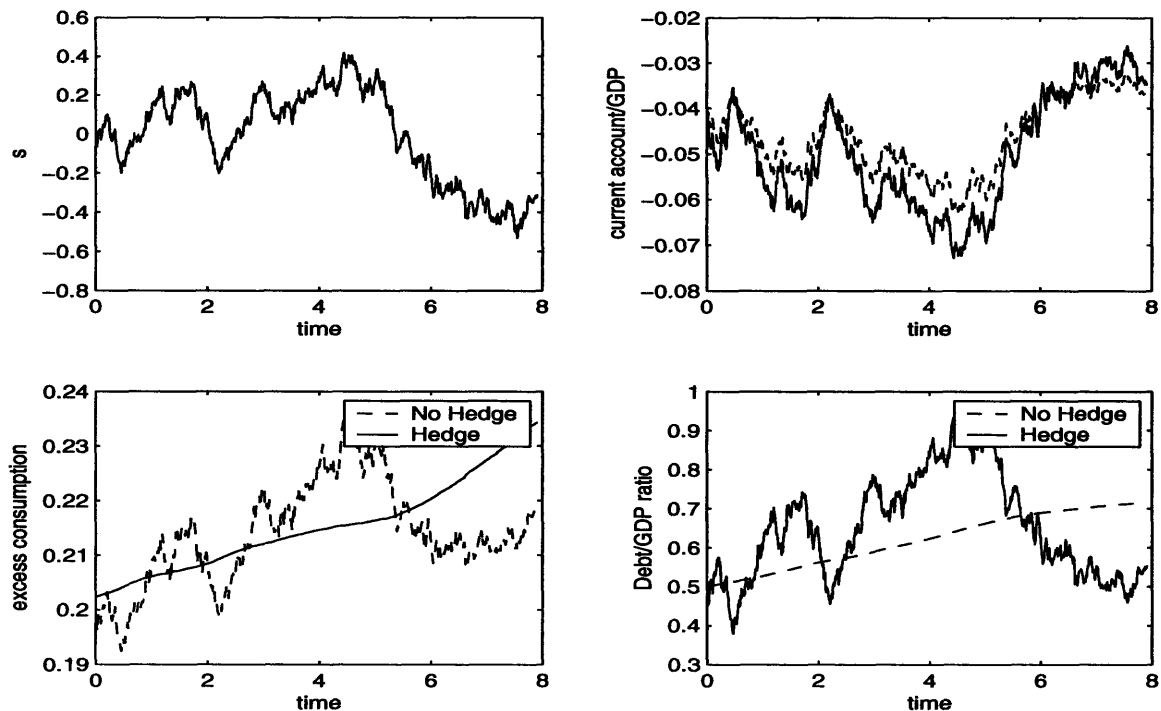
Figure 4-3: p-Hedging for the Chilean case

precautionary recessions caused by the decline in s. The dashed line in the top right panel (c) shows the impact of the latter on the current account. In particular, the sharp decline in the deficit in the current account from 6 to about 3.5 percent of GDP toward the end of the sample. This suggests that about half of the current account adjustment observed in Chile during the 1998-9 crisis can be accounted by *optimal* precautionary behavior in the absence of hedging. The rest of the adjustment could be accounted for by excess adjustment (some have argued that the central bank contracted monetary policy excessively during this episode) and by the partial sudden stop itself (the "sovereign" spread tripled during this episode).

Now, let the country hedge by shorting copper futures, which do not relax the sudden stop directly but only through their effect on X_τ (p-hedging). The solid lines in panels

(b) to (d) show the paths corresponding to the dashed paths of the unhedged economy. Panel (b) clearly shows that hedging, even of this very limited kind, virtually eliminates precautionary fluctuations. Interestingly, panel (c) shows that the insulation of consumption from precautionary cycles is not done through a smoothing of the current account, which looks virtually identical in the hedged and unhedged economies. The difference comes from the fact that the improvement in the current account as s deteriorates comes from hedge-transfers in favor of the country. The latter are reflected in a sharp decline in external debt as s worsens (panel (d)).

As a practical matter, it is important to point out that the hedge-ratios required to achieve this success are very large. The size of the implied portfolios can be calculated as (evaluating at $s = 0$):

$$\frac{b_s}{b}(X_0 + (1 - \kappa)\frac{y_t^*}{r - g}) \approx -0.56.$$

That is, the notional amount in the futures position would have to exceed 50 percent of GDP. This is very large when compared with the existing futures markets for copper, pointing to the need to develop these or related markets for sudden-stop-insurance purposes. However, it is not a disproportionate number when compared with the very large and much costlier practice of accumulating international reserves (around 30 percent of GDP).

Finally, and as indicated in section 4, this strategy — if subject to full crowding out— is not very effective in smoothing the sudden stop itself. Our estimates show that in case of a sudden stop, consumption with hedging of this sort exceeds consumption in the unhedged case only by about 2% for consumption drops of roughly 8-10%.

Hedging Sudden Stops

Let us now analyze the opposite extreme and assume that hedging directly relaxes the sudden stop constraint (h-hedging). Full insurance in this case requires a transfer in the

	(2.5,5.2)	(5,5.2)	(2.3,0)
$s = 0$	0.536	0.247	0.517
$s = 1.5\sigma$	0.323	0.139	0.523

Table 4.2: Values of $E[e^{-r\tau}]$ for Different Combinations of (α_0, α_1).

case of a sudden stop of:

$$h^* \left(X_0 + \frac{(1 - \kappa)y_0}{r - g} \right) = \left(\phi(1 - e^{-(r-g)T}) + e^{gT} - 1 \right) \left(X_0 + \frac{(1 - \kappa)y_0}{r - g} \right) \approx 0.15.$$

Importantly, note that 15 percent of GDP exceeds the 8-10 percent we calibrated the decline in consumption to be in the case of a sudden stop. This is because 15 percent covers not only the drop in consumption in case of a sudden stop but also the precautionary savings and recession that precedes the sudden stop.

In order to calculate the NPV of such a claim, we must multiply the number above by $E[e^{-r\tau}]$. Since the latter depends on the initial value of the signal s and on the estimates of α's, which are very imprecise, we report in Table 4.2 the value of $E[e^{-r\tau}]$ for different combinations of these parameters.

Using the benchmark case, we see that "fairly" priced full insurance would cost the country about 8 percent of GDP (0.15 times 0.536) if contracted when the price of copper is at "normal" levels and about 5 percent (0.15 times 0.323) when at very high levels. While these amounts are small when compared with the cost of sudden stops and precautionary recessions, they still involve amounts which are probably too large for these countries to undertake, even if such markets existed.

Fortunately, much of the advantage of insurance can be obtained with significantly

smaller amounts of insurance. In what follows, we take h as given and solve:

$$\max_{c_s^*, X_\tau} E \left[\int_0^\tau e^{-rs} \frac{c_s^{*1-\gamma}}{1-\gamma} ds + e^{-r\tau} V^{SS}(X_{\tau-}) \right]$$

s.t.

$$dX_t = (rX_t - c_t^* + y_t^*) dt$$

$$X_{0+} = X_{0-} - P_0$$

$$P_0 = E \left[e^{-r\tau} h(X_{\tau-} + \frac{y_\tau^*}{r-g}) \right].$$

In other words we assume that the country starts with an initial debt X_{0-}, and purchases insurance costing P_0 which is priced fairly and denominated as a fraction $h < h^*$ of the quantity $X_{\tau-} + \frac{y_\tau^*}{r-g}$. The setup is common knowledge to both the borrower and the lender, so the lender prices this claim understanding the subsequent optimization problem of the borrower and accordingly the resulting path of $X_{\tau-} + \frac{y_\tau^*}{r-g}$. Using this particular type of contingent claim results in a problem that we can easily solve numerically with the tools we developed in the previous sections, plus a fixed point problem for the initial price of the insurance, P_0.[21] The results are summarized in Table 4.3.

The first row shows P_0, the second c_0^*, while the third row shows the maximum value attained by excess consumption along the path before the sudden stop. The next three rows show consumption right before the sudden stop takes place, during the sudden stop, and after the crisis, respectively. The final row describes the standard deviation of $c_{\tau+}^*$ for a sample of 1000 simulations. The first three columns present values for $s_0 = 0$, while the last three columns do the same for $s_0 = 1.5\sigma$.

There are three main lessons to be learned from this table. First, clearly much of the cost of a lack of insurance is paid for with the large precautionary behavior that the economy needs to undertake without insurance. The level dimension of this effect can

---

[21] Even though the contingent claim signed by the borrower and the lender seems complicated, it isn't since $X_{\tau-} + \frac{y_\tau^*}{r-g}$ is not varying very much and can be approximated reasonably well by $X_0 + \frac{y_0^*}{r-g}$, especially for large h.

s_0:	0			1.5σ		
h/h^*:	0	0.4	0.8	0	0.4	0.8
Initial Cost	0.000	0.032	0.063	0.000	0.019	0.039
c_0^*	0.198	0.243	0.252	0.213	0.247	0.252
c_{max}^*	0.226	0.248	0.253	0.243	0.252	0.254
$c_{\tau-}^*$	0.212	0.244	0.252	0.224	0.247	0.253
Standard Deviation of $c_{max}^* - c_{\tau-}^*$	0.010	0.002	0.002	0.011	0.002	0.001
$c_{\tau+}^*$	0.125	0.175	0.229	0.131	0.177	0.230
$c_{\tau+T}^*$	0.300	0.269	0.259	0.313	0.271	0.260
Standard Deviation of $c_{\tau+}^*$	0.013	0.002	0.001	0.014	0.002	0.001

Table 4.3: h—Hedging

be seen in the value of c_0^*, while the cyclical component can be seen in the difference between $c^*{}_{max} - c_{\tau-}^*$. Second, much of the precautionary costs can be removed with very limited amounts of insurance; a value of $h = 0.4$ does nearly the same as one of 0.8. Third, the sudden stop itself is significantly harder to smooth, but it is still possible to make a difference with h significantly less than h^*.

Asymmetric Beliefs and Contingent Credit Lines

Before concluding, we discuss an important practical issue. Sudden stops are not entirely exogenous to the country's actions and there is significant asymmetric information about these actions between borrowers (the country) and lenders. Suppose then that the financial markets overstate (from the viewpoint of the country) the constant part of the hazard, α_0. In particular, we assume that the lender takes this constant to be α_0^L. The borrower on the other hand takes this number to be $\alpha_0^B \ll \alpha_0^L$. Let the expectation operators of the borrower and the lender be E^L and E^B, respectively, so that the problem

becomes:

$$\max_{c_s^*, X_\tau} E^B \left[\int_0^\tau e^{-rs} \frac{c_s^{*1-\gamma}}{1-\gamma} ds + e^{-r\tau} V^{SS}(X_{\tau-}) \right]$$

s.t.

$$dX_t = \left(rX_t - c_t^* + y_t^* \right) dt$$

$$X_{0+} = X_{0-} - P_0$$

$$P_0 = E^L \left[e^{-r\tau} h(X_{\tau-}^{opt} + \frac{y_\tau^*}{r-g}) \right].$$

In this case the country obviously will find the price of the insurance "unfairly" high. One way to reduce the extent of this problem is for the country to add another contingency to the credit line. It is clear that in this case the country would benefit from making the line contingent on the value of s. Since the country sees crises not related to s as much less likely than the markets – for example, it may be certain that it will not run up fiscal deficits — it is optimal for it not to pay for insurance in states when s is high. Let us capture this feature by assuming that the borrower and the lender enter an agreement whereby the lender agrees to pay:

$$f(s_\tau) h \left(X_{\tau-}^{opt} + \frac{y_\tau^*}{r-g} \right)$$

to the borrower and the fraction of h paid out depends on s_τ:

$$f(s_\tau) = \frac{\exp(-w(s_\tau + 2\sigma))}{1 + \exp(-w(s_\tau + 2\sigma))}$$

Obviously $0 < f(s_\tau) < 1$. Taking $\alpha_0^L = 2.5$, $\alpha_0^B = 5$ and $w = 3$, i.e. a claim that pays quite steeply when and only when s_τ is very low (2 standard deviations below 0), the results are reported in the first half of Table 4.4, and contrasted with the case without the additional contingency.

By adding the additional state contingency, a country that is confident of its "good

		uncontingent		$w=3$	
h/h^*:	0	0.4	0.8	0.4	0.8
Initial Cost	0.000	0.032	0.063	0.009	0.020
c_0^*	0.219	0.247	0.250	0.238	0.246
$c_{\tau-}^*$	0.219	0.246	0.251	0.245	0.255
$c_{\tau+}^*$	0.128	0.175	0.228	0.155	0.187
Standard Deviation of $c_{\tau+}^*$	0.009	0.002	0.001	0.016	0.026

Table 4.4: $h-$Hedging with asymmetric Beliefs

behavior" is able to lower the price of the claim without incurring a very significant rise in risk exposure.

4.6 Final Remarks

In this paper we characterized many aspects of sudden stops, precautionary recessions, and the corresponding aggregate hedging strategies. For this, we built a model simple enough to shed analytical light on some of the key issues but realistic enough to provide some quantitative guidance.

We showed that even after removing all other sources of uncertainty, the combination of infrequent sudden stops and the mere anticipation of them has the potential to explain a large share of the volatile business cycle experienced by emerging market economies. This important source of volatility could be overcome with suitable aggregate hedging strategies, but these would require developing large new financial markets.

In our application, we hinted at one aspect of the insurance arrangements that could facilitate the development of such markets by reducing the inherent asymmetric information problems. We argued for credit lines and financial instruments being contingent on indicators that are exogenous to the country. Thus, for example, Chile could remove many of the signaling problems it fears from the IMF's Contingent Credit Lines by adding a clause linking the size of the line to the price of copper. Mexico could do the same by

indexing its line to the price of oil and US GDP, and so on. Moreover, it is highly unlikely that the broad non-specialists markets could be tapped without such contingencies.

There are several aspects of an aggregate risk management strategy that we left unexplored. In particular, we did not model the maturity structure of debt and how this changes with the signal, s. Similarly, we did not discuss accumulation of international reserves, as these are dominated by our credit lines. However, building reserves financed with long-term debt may partially substitute for overpriced credit lines. In order to address these issues properly, we also need to have a better understanding of the behavior of the supply side as the signal worsens. Of course, a country would want to postpone accumulating reserves and borrowing long term until the crisis is imminent, but it is highly unlikely that the lenders will go along with this strategy. We intend to explore these dimensions in future work.

4.7 Appendix

4.7.1 Propositions and Proofs for section 2.3.

Proposition 4.1 *The optimal solution of the τ period problem is:*

$$c_t^{*opt} = (k_1)^{-\frac{1}{\gamma}}, \quad \tau < t < \tau + T \tag{4.58}$$

$$X_{\tau+T} = \max\{\overline{X}, \frac{(k_1)^{-\frac{1}{\gamma}}}{r} - \frac{y_\tau^* e^{gT}}{r-g}\} \tag{4.59}$$

and k_1 is determined as the (unique) solution to:

$$X_\tau + \frac{y_\tau^*(1 - e^{-(r-g)T})}{r-g} = \frac{(k_1)^{-\frac{1}{\gamma}}(1 - e^{-rT})}{r} + e^{-rT}X_{\tau+T} \tag{4.60}$$

Moreover the maximum in (4.59) is always given by $\overline{X}$. That is, the constraint is always binding. Combining the above statements, we have that the optimal consumption-wealth pair is given by (4.12) and (4.11).

Proof. To establish the first assertion we need to show that the proposed policy satisfies (we normalize $\tau = 0$ and thus $y_\tau e^{gT} = y_T$ without loss of generality)

$$\int_0^T e^{-ru}\frac{(c_u^{*opt})^{1-\gamma}}{1-\gamma}du + e^{-rT}V(X_T^{opt} + \frac{y_T^*}{r-g}) \geq \int_0^T e^{-ru}\frac{c_u^{*1-\gamma}}{1-\gamma}du + e^{-rT}V(X_T + \frac{y_T^*}{r-g})$$

for any other admissible policy pair c_u^*, X_T^*, for $0 < u < T$. By feasibility we have

$$X_0 + \int_0^T e^{-ru}y_u^*du \geq \int_0^T e^{-ru}c_u^*du + e^{-rT}X_T$$

$$X_T \geq \overline{X}$$

where we can focus without loss of generality on deviations that satisfy the first equation as an equality. By concavity of the utility function and the terminal Value function it follows

223

that:

$$\int_0^T e^{-ru}\frac{c_u^{*1-\gamma}}{1-\gamma}du + e^{-rT}V(X_T + \frac{y_T^*}{r-g})$$
$$- \left(\int_0^T e^{-ru}\frac{(c_u^{*opt})^{1-\gamma}}{1-\gamma}du + e^{-rT}V(X_T^{opt} + \frac{y_T^*}{r-g})\right)$$

$$= \int_0^T e^{-ru}\frac{c_u^{*1-\gamma} - (c_u^{*opt})^{1-\gamma}}{1-\gamma}du + e^{-rT}\left(V(X_T + \frac{y_T^*}{r-g}) - V(X_T^{opt} + \frac{y_T^*}{r-g})\right)$$
$$\leq \int_0^T e^{-ru}(c_u^{*opt})^{-\gamma}(c_u^* - c_u^{*opt})du + e^{-rT}V'(X_T^{opt} + \frac{y_T^*}{r-g})(X_T - X_T^{opt})$$

Now we need to distinguish between two cases: a) if $X_T^{opt} > \overline{X}$ then $V'^{opt} = k_1 = \left(c_u^{*opt}\right)^{-\gamma}$ for a constant k_1 such that the budget constraint (4.60) is satisfied and the result follows upon substituting in the last expression:

$$k_1\left(\int_0^T e^{-ru}(c_u^* - c_u^{*opt})du + e^{-rT}(X_T - X_T^{opt})\right) = 0$$

since both policies satisfy:

$$X_0 + \int_0^T e^{-ru}y_u^*du = \int_0^T e^{-ru}c_u^*du + e^{-rT}X_T$$

b)When $X_T^{opt} = \overline{X}$, it is still the case that $\left(c_u^{*opt}\right)^{-\gamma} = k_1$, and because the alternative candidate policy is feasible, it also satisfies $X_T \geq \overline{X} = X_T^{opt}$. Because of concavity of V this

also implies $V'(\overline{X} + \frac{y_T^*}{r-g}) < k_1$. These two arguments can be combined to get:

$$\int_0^T e^{-ru} \left(c_u^{*opt} \right)^{-\gamma} (c_u^* - c_u^{*opt}) du + e^{-rT} V'(\overline{X} + \frac{y_T^*}{r-g})(X_T - X_T^{opt}) =$$

$$\int_0^T e^{-ru} k_1 (c_u^* - c_u^{*opt}) du + e^{-rT} V'(\overline{X} + \frac{y_T^*}{r-g})(X_T - X_T^{opt}) <$$

$$\int_0^T e^{-ru} k_1 (c_u^* - c_u^{opt}) du + e^{-rT} k_1 (X_T - X_T^{opt}) = 0$$

This verifies the optimality of the proposed policy.

For the second assertion, that the constraint is always binding, the argument runs as follows. Suppose otherwise. Then k_1 is given as:

$$\frac{(k_1)^{-\frac{1}{\gamma}}}{r} = X_\tau + \frac{y_\tau^*}{r-g}$$

Under our (counterfactual) assumption we should have:

$$X_\tau \left(\phi(1 - e^{-(r-g)T}) + e^{gT} \right) + \phi \frac{y_\tau^*}{r-g}(1 - e^{-(r-g)T}) \leq X_\tau + \frac{y_\tau^*(1 - e^{gT})}{r-g}$$

or

$$X_\tau(1 - \phi(1 - e^{-(r-g)T}) - e^{gT}) \geq -\frac{y_\tau^*}{r-g}(1 - \phi(1 - e^{-(r-g)T}) - e^{gT})$$

or

$$X_\tau \leq -\frac{y_\tau^*}{r-g} \tag{4.61}$$

since

$$(1 - \phi(1 - e^{-(r-g)T}) - e^{gT}) < 0$$

225

But (4.61) contradicts non-negativity of consumption when combined with the transverality condition $\lim_{t\to\infty} e^{-rt} X_t = 0$. ∎

4.7.2 Propositions and Proofs for section 3.1

In this section we proof all the steps of the perfect signal case. The proofs of the imperfect signal case follow the same steps but has a few steps which are much harder to proof, hence we only validate them though our numerical simulations.

The proof consists of the following steps. First we establish rigorously the boundary conditions that the value function should satisfy. Then we prove the existence of a unique solution to the ODE presented in the text subject to the required boundary conditions. Then we verify the "tail" condition on the value function that we use in order to apply a (classical) verification Theorem along the lines of Fleming and Soner (1993) p.172.

Proposition 4.2 *The value function of the problem with sudden stops for $s > \underline{s}$ satisfies:*

$$\lim_{s\to\infty} V(X_\tau, s, y_\tau^*) = \left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \tag{4.62}$$

Proof. We focus on the $\gamma > 1$ case. The proof proceeds in two steps and does not depend on whether we allow hedging or not (i.e., whether we require that the portfolio $p = 0$ or whether p is determined as part of the optimization problem). First one obtains an upper bound on the value function of the problem with sudden stops, which in this case is naturally given by the value function of the problem in the absence of sudden stops:

$$V^{NSS} = \left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$$

To see why this is an upper bound notice that the following inequality holds for any feasible policy:

$$E\left[\int_t^\tau e^{-r(u-t)} \frac{(c_u^*)^{1-\gamma}}{1-\gamma} du + 1\{\tau < \infty\} e^{-r(\tau-t)} K \left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau + \frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma} | F_t \right]$$

226

$$\leq E\left[\int_t^\tau e^{-r(u-t)}\frac{(c_u^*)^{1-\gamma}}{1-\gamma}du + 1\{\tau<\infty\}e^{-r(\tau-t)}\left(\frac{1}{r}\right)^\gamma\frac{\left(X_\tau+\frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}|F_t\right] =$$

$$E\left[\int_t^\tau e^{-r(u-t)}\frac{(c_u^*)^{1-\gamma}}{1-\gamma}du + 1\{\tau<\infty\}e^{-r(\tau-t)}V^{NSS}(X_\tau,y_\tau^*)|F_t\right]$$

since for $\gamma>1$, $\dfrac{\left(X_\tau+\frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}<0$. This holds true for any feasible consumption /portfolio policies and thus, when one evaluates this inequality at the optimal plan c_u^{*opt},p_u^*:

$$V(X_t,s_t,y_t^*) =$$

$$E\left[\int_t^\tau e^{-r(u-t)}\frac{\left(c_u^{*opt}\right)^{1-\gamma}}{1-\gamma}du + 1\{\tau<\infty\}e^{-r(\tau-t)}K\left(\frac{1}{r}\right)^\gamma\frac{\left(X_\tau^{opt}+\frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}|F_t\right] \leq$$

$$E\left[\int_t^\tau e^{-r(u-t)}\frac{\left(c_u^{*opt}\right)^{1-\gamma}}{1-\gamma}du + 1\{\tau<\infty\}e^{-r(\tau-t)}\left(\frac{1}{r}\right)^\gamma\frac{\left(X_\tau^{opt}+\frac{y_\tau^*}{r-g}\right)^{1-\gamma}}{1-\gamma}|F_t\right]$$

$$\leq \max_{c_u^*,p_u} E\left[\int_t^\tau e^{-r(u-t)}\frac{(c_u^*)^{1-\gamma}}{1-\gamma}du + 1\{\tau<\infty\}e^{-r(\tau-t)}V(X_\tau,y_\tau^*)|F_t\right] = V^{NSS}(X_t,y_t^*) =$$

$$\left(\frac{1}{r}\right)^\gamma\frac{\left(X_t+\frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$$

The last equality follows from the principle of dynamic programming. Since this holds for all s_t it is also true as $s_t\to\infty$. This establishes the upper bound. The lower bound is established upon observing that the consumption /portfolio policy in the absence of sudden stops c^{NSS},p^{NSS} (=0) is still a feasible consumption /portfolio plan in the presence of sudden stops (since it satisfies the intertemporal budget constraint and also $X_t>-\frac{y_t^*}{r-g}$ for all t). It is also within the class of policies that mandate $p_u=0$. Accordingly we have the two inequalities:

$$V^{SS,p=0}(X_t,s_t,y_t^*) \geq E\left[\int_t^\tau e^{-r(u-t)}\frac{\left(c_u^{*NSS}\right)^{1-\gamma}}{1-\gamma}du + e^{-r(\tau-t)}1\{\tau<\infty\}K\left(\frac{1}{r}\right)^\gamma\frac{\left(X_\tau^{NSS}+\frac{y_\tau^*}{r-a}\right)^{1-\gamma}}{1-\gamma}|F_t\right]$$

$$V^{SS,\,p=p^{opt}}(X_t, s_t, y_t^*) \geq E\left[\int_t^\tau e^{-r(u-t)}\frac{\left(c_u^{*NSS}\right)^{1-\gamma}}{1-\gamma}du + e^{-r(\tau-t)}1\{\tau < \infty\}K\left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau^{NSS} + \frac{y_\tau^*}{r-a}\right)^{1-\gamma}}{1-\gamma}|F_t\right]$$

However by standard arguments it is easy to show that as $s \to \infty$, $\Pr(\tau < Q) \to 0$ for all finite Q, and thus

$$\lim_{s\to\infty} E\left[\int_t^\tau e^{-r(u-t)}\frac{\left(c_u^{*NSS}\right)^{1-\gamma}}{1-\gamma}du + e^{-r(\tau-t)}K1\{\tau < \infty\}\left(\frac{1}{r}\right)^\gamma \frac{\left(X_\tau^{NSS} + \frac{y_\tau^*}{r-a}\right)^{1-\gamma}}{1-\gamma}|F_t\right] =$$

$$= E\left[\int_t^\infty e^{-r(u-t)}\frac{\left(c_u^{*NSS}\right)^{1-\gamma}}{1-\gamma}du|F_t\right] = V^{NSS}(X_t, y_t^*) =$$

$$\left(\frac{1}{r}\right)^\gamma \frac{\left(X_t + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$$

Since the upper and the lower bound coincide to V^{NSS} as $s \to \infty$, the claim is established.

∎

Before we can invoke a standard verification Theorem we establish the existence of a solution to the following ODE:

$$\gamma C^{\frac{\gamma-1}{\gamma}} - \gamma rC + \mu C_s + \frac{1}{2}C_{ss}\sigma^2 = 0 \tag{4.63}$$

Even though this second order non-linear ODE does not seem to have any closed form solution it is not hard to establish that it has a unique solution that satisfies the required boundary conditions. This is done in the following proposition:

Proposition 4.3 *The ODE (4.63) has a unique solution satisfying the boundary conditions (4.25), (4.26), with the change in variables $a^\gamma = C$.*

Proof. The steady state of this 2nd order nonlinear ODE can be obtained at once as:

$$C^{SS} = \left(\frac{1}{r}\right)^\gamma \tag{4.64}$$

228

The Theorem now basically follows from the stable manifold Theorem upon reformulating the system as a system of two first order ODE's, realizing that the system has one positive and one negative eigenvalue and applying the stable manifold Theorem (see e.g. Verhulst(2000) p.33.) The argument is particularly easy and thus we do not provide any details apart from the fact that the linearized system has one positive and one negative eigenvalue. Indeed, one can approximate the solution by means of a linear approximation around the steady state value to get:

$$\gamma \left(C^{SS}\right)^{\frac{\gamma-1}{\gamma}} + (\gamma - 1)\left(C^{SS}\right)^{-\frac{1}{\gamma}}(C - C^{SS}) - r\gamma C + \mu C_s + \frac{1}{2}C_{ss}\sigma^2 = 0 \qquad (4.65)$$

The two eigenvalues of the characteristic polynomial of this equation are:

$$\lambda_{1,2} = \frac{\mu \pm \sqrt{\mu^2 + 2\sigma^2 r}}{\sigma^2} \qquad (4.66)$$

Obviously one eigenvalue is positive, one is negative and this establishes the claim after a few straightforward steps and the use of the stable manifold Theorem. ∎

Remark 4.1 *A simple reformulation of the ODE (4.63), yields the function $a(s)$ that we use in the text and is obtained by defining:*

$$C = a^\gamma$$

and rewriting the ODE as:

$$\gamma a^{\gamma-1} + -\gamma r a^\gamma + \mu\gamma a^{\gamma-1}a_s + \frac{1}{2}\sigma^2\left(\gamma(\gamma-1)a^{\gamma-2}(a_s)^2 + \gamma a^{\gamma-1}a_{ss}\right) = 0$$

or:

$$1 - ra + \mu a_s + \frac{1}{2}\sigma^2\left((\gamma-1)\frac{(a_s)^2}{a} + a_{ss}\right) = 0 \qquad (4.67)$$

subject to the boundary constraints:

$$a(\underline{s}) \;=\; K^{\frac{1}{\gamma}}\left(\frac{1}{r}\right) \tag{4.68}$$

$$\lim_{s\to\infty} a(s) \;=\; \left(\frac{1}{r}\right) \tag{4.69}$$

The last step in verifying the fact that the conjectured Value function is indeed the Value function for the problem at hand is to invoke a verification Theorem along the lines of Fleming and Soner (1993) p. 172. Before we do that we prove an almost obvious Lemma that is of independent interest:

Lemma 4.1 *Consider a consumption policy that has the feedback form:*

$$c_t^* = A(s_t)\left(X_t + \frac{y_t^*}{r-g}\right), \quad K^{-1/\gamma}r \le A(s_t) \le r \;\; \forall s_t \in (-\underline{s}, +\infty)$$

Then $X_t \ge X_t^{NSS}$, where X_t^{NSS} is the asset process that results when one uses the optimal consumption policy in the absence of sudden stops:

$$c_t^{NSS} = r\left(X_t^{NSS} + \frac{y_t^*}{r-g}\right)$$

Proof. This result is trivial. One solves for the asset process that results from the two policies to find that

$$X_t^{NSS} = X_0 - \frac{y_0^*}{r-g}\left(e^{gt} - 1\right)$$

The equivalent calculation for the optimal policy c_t conditional on a path of s_t gives:

$$
\begin{aligned}
d\left(e^{-\left(rt - \int_0^t A(s_u)du\right)} X_t\right) &= -(r - A(s_t))e^{-\left(rt - \int_0^t A(s_u)du\right)} X_t + e^{-\left(rt - \int_0^t A(s_u)du\right)} dX_t = \\
&= e^{-\left(rt - \int_0^t A(s_u)du\right)}\left[y_0^* \frac{r - A(s_t) - g}{r - g} e^{gt}\right]
\end{aligned}
$$

Now integrating both sides and rearranging gives:

$$
\begin{aligned}
X_t &= X_0 + \frac{y_0^*}{r-g} e^{\left(rt - \int_0^t A(s_u)du\right)} \left(\int_0^t e^{-\left((r-g)i - \int_0^i A(s_u)du\right)} (r - A(s_i) - g)di \right) = \\
&\quad X_0 - \frac{y_0^*}{r-g} \left(e^{gt} - e^{\left(rt - \int_0^t A(s_u)du\right)} \right)
\end{aligned}
$$

The result now follows since $A(s_t) \leq r \ \forall s_t \in (\underline{s}, +\infty)$. ∎

To invoke a verification Theorem we finally need to show that:

Lemma 4.2 *Assume* $\left(\frac{1}{r}\right)^\gamma \leq C(s_t) \leq K \left(\frac{1}{r}\right)^\gamma \ \forall s_t \in (\underline{s}, +\infty)$. *Then*

$$
\lim_{t \to \infty} E \left[e^{-rt} 1\{\tau > t\} C(s) \frac{\left(X_t^{opt} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \right] = 0
$$

Proof. For any t it is the case that

$$
E \left[e^{-rt} 1\{\tau > t\} C(s) \frac{\left(X_t^{opt} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \right] \geq
$$

$$
E \left[e^{-rt} 1\{\tau > t\} K \left(\frac{1}{r}\right)^\gamma \frac{\left(X_t^{opt} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \right] \geq
$$

$$
K E \left[e^{-rt} 1\{\tau > t\} \left(\frac{1}{r}\right)^\gamma \frac{\left(X_t^{NSS} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \right] \quad \to \quad 0
$$

The first inequality follows from the assumption and the fact that $\dfrac{\left(X_t^{opt} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$ is a negative number. The second from the previous lemma and the monotonicity of the value function and the last limit follows from the fact that one can trivially show that in the standard model without sudden stops $e^{-rt} \left(\frac{1}{r}\right)^\gamma \dfrac{\left(X_t^{NSS} + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} \to 0$ ∎

The final step is to prove that any solution to $a(s)$ (respectively $C(s)$) will be bounded. This can be done in a rather straightforward manner:

Proposition 4.4 *The function $a(s) = C^{\frac{1}{\gamma}}$, is a decreasing function that stays between $\frac{1}{r}K^{1/\gamma}$ and $\frac{1}{r}$.*

Proof. We will derive the result by means of two contradictions. Suppose $a(s)$ solves (4.67) subject to the boundary conditions (4.68),(4.69). To establish the claim we just need to show that $a(s)$ is everywhere decreasing. (If it exited the "band" $[\frac{1}{r}, \frac{1}{r}K^{1/\gamma}]$ it would have to have at least one section where it would be increasing). To establish that it is decreasing, we are going to establish 2 contradictions by studying points that could be local maxima or minima. So suppose indeed that there exists one point s^* that is a local extremum. Then $a_s(s^*) = 0$ and accordingly we have 2 cases:

1) $1 - ra < 0$. Then it has to be true (in order to satisfy the ODE) that $a_{ss} > 0$, so that s^* would have to be a local minimum. But in order for $a(s)$ to satisfy the Boundary condition at infinity there would have to exist another $s^{**} > s^*$, that would have to be local maximum, which is impossible since it would still be in the region where $1 - ra < 0$. (Since -necessarily- $a(s^*) < a(s^{**})$).

2) $1 - ra > 0$. Then it has to be true that $a_{ss} < 0$, so that s^* would have to be a local maximum. But in order to satisfy the boundary condition at infinity there would have to be a local minimum s^{**} to the right of s^* satisfying $a(s^{**}) < a(s^*)$ and $a_{ss}(s^{**}) > 0$ which is impossible. ∎

Now one can apply a standard verification Theorem and verify that indeed the proposed function V is the Value function for the problem.

The following Lemma derives the properties of the consumption process:

Lemma 4.3 *The optimal consumption process satisfies:*

$$\frac{dc_t^*}{c_t^*} = \frac{1}{2}(\gamma + 1)\left(\frac{\sigma a_s}{a}\right)^2 dt - \frac{a_s}{a}\sigma dB_t$$

Proof. Straightforward application of Ito's Lemma to:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{a(s_t)}$$

232

along with equation (4.67.) ∎

4.7.3 Propositions and Proofs for section 4.1.

We first show that the excess consumption process, c_t^*, is constant. There are two ways to see this. The first one is to apply Ito's lemma to c_t^* and observe that both parts of the resulting Ito process (informally speaking the dt and the dB_t terms) are identically zero. Another more direct way is to use the martingale methodology developed by Karatzas and Wang (2000) to deal with this problem along with standard formulas for distributions of hitting times of Brownian Motion. This approach is particularly appealing in the context of this section because of the presence of complete markets (up to the point where the sudden stop takes place).

We start with the first approach. We have the following result:

Lemma 4.4 *Under perfect hedging "excess" consumption, c_t^*, is constant for $t < \tau$*

Proof. To simplify notation somewhat we define

$$a^p(s_t) = \frac{1}{r}\left(e^{-\lambda_1(s-\underline{s})}(K^{\frac{1}{\gamma}} - 1) + 1\right) \tag{4.70}$$

and observe that this function solves the ordinary differential equation:

$$\frac{1}{2}\sigma^2 a_{ss}^p + \mu a_s^p - r a^p + 1 = 0 \tag{4.71}$$

The consumption policy can then be reexpressed by means of (4.40) and (4.70) as

$$c_t^* = \frac{1}{a^p(s_t)}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)$$

and the portfolio policy as:

$$p_t = \frac{a_s}{a}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)$$

233

Applying Ito's Lemma to the right hand side of this expression gives:

$$dc^* = C_1 dt + C_2 dB_t$$

where

$$
\begin{aligned}
C_1 &= -\frac{1}{a^p}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)\left(\frac{a_s^p}{a^p}\mu - \left(\frac{a_s^p \sigma}{a^p}\right)^2 + \frac{a_{ss}^p}{a^p}\frac{\sigma^2}{2}\right) + \\
&\quad + \frac{1}{a^p}\left(rX_t - \frac{1}{a^p}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right) + (1-\kappa)y_t + g\frac{y_t(1-\kappa)}{r-g}\right) + \\
&\quad - \frac{1}{a^p}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)\left(\frac{\sigma a_s^p}{a^p}\right)^2 \\
&= -\frac{1}{(a^p)^2}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)\left(\frac{\sigma^2}{2}a_{ss}^p + a_s^p\left(r - \frac{1}{2}\sigma^2\right) - ra^p + 1\right) = 0
\end{aligned}
$$

and

$$C_2 = -\frac{1}{a^p}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)\left(\frac{\sigma a_s^p}{a^p}\right) + \frac{1}{a^p}\left(X_t + \frac{y_t(1-\kappa)}{r-g}\right)\left(\frac{\sigma a_s^p}{a^p}\right) = 0$$

∎

In other words perfect hedging leads to complete smoothing in our example, despite the possibility of a fall in consumption when the state variable crosses the critical level.

Let us prove the same result with the martingale methodology developed by Cox and Huang (1989) and Karatzas, Lechoszky and Shreve (1987), as it will facilitate the proofs later on. For our exposition we will be using the results in Karatzas and Wang (2000).

Lemma 4.5 *Without loss of generality, take $t = 0$. The optimal policy is given as:*

$$
\begin{aligned}
c_u^* &= (k^{-1/\gamma})1\{u < \tau\} \\
X_\tau &= \left(\frac{1}{r}\left(\frac{k}{K}\right)^{-1/\gamma} - \frac{y_\tau^*}{r-a}\right)
\end{aligned}
$$

where k solves the intertemporal budget constraint:

$$X_0 + \frac{y_0^*}{r-g} = k^{-1/\gamma}\left(E\int_0^\tau e^{-ru}du + e^{-r\tau}\frac{K^{1/\gamma}}{r}1\{\tau < \infty\}\right)$$

234

or

$$k = \left(\frac{X_0 + \frac{y_0^*}{r-g}}{\left(E \int_0^\tau e^{-ru} du + e^{-r\tau} \frac{K^{1/\gamma}}{r} 1\{\tau < \infty\} \right)} \right)^{-\gamma}$$

Proof. We provide only a sketch of the proof. The reader is referred to Karatzas and Wang (2000) for details.[22]

The objective is:

$$\max_{c_u^e, p_u} E \left(\int_0^\tau \frac{c_u^{*1-\gamma}}{1-\gamma} e^{-ru} du + e^{-r\tau} V(X_\tau, y_\tau^*) 1\{\tau < \infty\} \right)$$

s.t.

$$dX_t = (rX_t - c_t^* + y_t^*) dt + \sigma p_t dB_t$$

and the rest of the evolution equations remain unchanged.

We proceed to show formally that the proposed strategy is optimal borrowing from Karatzas and Wang (2000). Let us denote for an alternative admissible strategy (c^*, X_τ^*) satisfying the intertemporal budget constraint with equality the Value of the objective:

$$J(c,p) = E \left(\int_0^\tau \frac{c_u^{*1-\gamma}}{1-\gamma} e^{-ru} du + e^{-r\tau} V(X_\tau, y_\tau^*) 1\{\tau < \infty\} \right)$$

where

$$V(X_\tau, y_\tau) = K \left(\frac{1}{r} \right)^\gamma \frac{\left(X_t + \frac{y_t^*}{r-g} \right)^{1-\gamma}}{1-\gamma}$$

---

[22]Karatzas and Wang (2000) do not strictly cover the case we are considering here, since our problem is on an infinite horizon. Even though it seems easy to extend their approach to cover our case too, we refrain from doing so and just note that the techniques used for the limited hedging case at the beginning of section 4 along with the observation about the optimal policy and the differentiability of the a^H−function are enough to apply a standard verification Theorem.

Let $\widetilde{f}$ denote the Fenchel - Legendre Transform of the function f:

$$\widetilde{f}(k) = \max_x f(x) - kx$$

so that this maximization e.g. for

$$\widetilde{\frac{c_u^{*1-\gamma}}{1-\gamma}}e^{-ru} = \max_x e^{-ru}\left(\frac{c_u^{*1-\gamma}}{1-\gamma} - kc_u^*\right)$$

yields:

$$c_u^{*opt} = (k^{-1/\gamma})1\{u < \tau\}$$

and similarly for X_τ.

Then it is the case that:

$$
\begin{aligned}
J(c,p) &= E\left(\int_0^\tau \frac{c_u^{*1-\gamma}}{1-\gamma}e^{-ru}du + e^{-r\tau}V(X_\tau, y_\tau^*)1\{\tau < \infty\}\right) \leq \\
&\leq E\left(\int_0^\tau \widetilde{\frac{c_u^{*1-\gamma}}{1-\gamma}}e^{-ru}du + e^{-r\tau}\widetilde{V(X_\tau, y_\tau^*)}1\{\tau < \infty\}\right) + \\
&\quad +k\left(E\int_0^\tau e^{-ru}c_u^* du + e^{-r\tau}X_\tau 1\{\tau < \infty\}\right) \\
&\leq E\left(\int_0^\tau \widetilde{\frac{c_u^{*1-\gamma}}{1-\gamma}}e^{-ru}du + e^{-r\tau}\widetilde{V(X_\tau, y_\tau)}1\{\tau < \infty\}\right) + k\left(X_0 + \frac{y_0^*}{r-g}\right) = \\
&= E\left(\int_0^\tau \frac{\left(c_u^{*opt}\right)^{1-\gamma}}{1-\gamma}e^{-ru}du + e^{-r\tau}V(X_\tau^{opt}, y_\tau^*)1\{\tau < \infty\}\right) + \\
&\quad +k\left[\left(X_0 + \frac{y_0^*}{r-g}\right) - \left(X_0 + \frac{y_0^*}{r-g}\right)\right] \\
&= E\left(\int_0^\tau \frac{\left(c_u^{*opt}\right)^{1-\gamma}}{1-\gamma}e^{-ru}du + e^{-r\tau}V(X_\tau^{opt}, y_\tau^*)1\{\tau < \infty\}\right)
\end{aligned}
$$

This verifies optimality of the strategy. We will now use the optimal strategy to obtain the Value function using well known formulas about the first hitting times of Brownian Motions.

$$V(X_0 + \frac{y_0^*}{r-g}) =$$

$$= \left(E \int_0^\tau \frac{k^{-(1-\gamma)/\gamma}}{1-\gamma} e^{-ru} du + e^{-r\tau} \frac{k^{-(1-\gamma)/\gamma}}{1-\gamma} \left(\frac{K^{1/\gamma}}{r} \right)^\gamma \left(\frac{K^{1/\gamma}}{r} \right)^{1-\gamma} 1\{\tau < \infty\} \right) =$$

$$= \frac{\left(X_0 + \frac{y_0^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \left(E \int_0^\tau e^{-ru} du + e^{-r\tau} \frac{K^{1/\gamma}}{r} 1\{\tau < \infty\} \right)^\gamma =$$

$$= \frac{\left(X_0 + \frac{y_0^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \left(\frac{1}{r} \left(1 - e^{-\lambda_1(s-\underline{s})} \right) + e^{-\lambda_1(s-\underline{s})} \frac{K^{1/\gamma}}{r} 1\{\tau < \infty\} \right)^\gamma =$$

$$= \frac{\left(X_0 + \frac{y_0^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \left(\frac{1}{r} \right)^\gamma \left(1 - e^{-\lambda_1(s-\underline{s})} + e^{-\lambda_1(s-\underline{s})} K^{\frac{1}{\gamma}} \right)^\gamma =$$

$$= \frac{\left(X_0 + \frac{y_0^*}{r-g} \right)^{1-\gamma}}{1-\gamma} \left(\frac{1}{r} \right)^\gamma \left(e^{-\lambda_1(s-\underline{s})} (K^{\frac{1}{\gamma}} - 1) + 1 \right)^\gamma$$

The only important step in this straightforward derivation is going from the second to the third line. This involves a well known formula which can be found e.g. in Harrison (1985) p. 47 ∎

4.7.4 Proofs and Propositions for section 4.2

Proposition 4.5 *For* $1 < \xi < \xi^*$ *,where:*

$$\xi^* \stackrel{d}{=} \frac{1}{1 + e^{-rT}(1+h^*) \left[\left(1 - \frac{h^*}{(1+h^*)(1-e^{-rT})} \right)^\gamma - 1 \right]}$$

the solution to the problem (4.72) is given by:

$$c_t^* = k^{-\frac{1}{\gamma}}, t \in (0, \tau)$$

$$X_{\tau-} = \frac{k^{-\frac{1}{\gamma}}}{r(1+h^*)}\Gamma^{-\frac{1}{\gamma}} - \frac{y_\tau^*}{r-g}$$

$$h^{opt} = h^* - (1 - e^{-rT})(1+h^*)\left[1 - \left(\frac{\xi}{\Gamma}\right)^{-\frac{1}{\gamma}}\right]$$

Γ *is defined by*

$$\Gamma \stackrel{d}{=} \frac{1-\xi}{e^{-rT}(1+h^*)} + \xi$$

and h^* is given in the text.
$k^{-\frac{1}{\gamma}}$ *is given by:*

$$k^{-\frac{1}{\gamma}} = \left[E\left[\int_0^\tau e^{-ru}du\right] + \frac{E\left[e^{-r\tau}\right]}{r}\frac{(1+h^{opt}\xi)\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)}\right]^{-1}\left(X_0 + \frac{y_0^*}{r-g}\right)$$

and since:

$$E\left[e^{-r\tau}\right] = e^{-\lambda(s-\underline{s})}$$

where λ is given in (4.54), we get finally that:

$$c_t^* = k^{-\frac{1}{\gamma}} = \left[\frac{1-e^{-\lambda(s-\underline{s})}}{r} + \frac{e^{-\lambda(s-\underline{s})}}{r}\frac{(1+h^{opt}\xi)\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)}\right]^{-1}\left(X_0 + \frac{y_0^*}{r-g}\right), t \in (0, \tau)$$

Proof. Adopting the Cox-Huang (1989) methodology and its application to problems involving a random stopping time Karatzas-Wang (2000) we are able to reduce the problem to

238

the following static problem:

$$\min_{k} \max_{c_u^*, X_\tau, H_\tau} E\left[\int_0^\tau e^{-ru} \frac{c_u^{*1-\gamma}}{1-\gamma} du + e^{-r\tau} V^{SS}(X_\tau, R_\tau) - k\left(\int_0^\tau e^{-ru} c_u + e^{-r\tau}(X_\tau + \xi H_\tau) \right) \right]$$

where V^{SS} is given by (4.53). By maximizing the objective inside the integral one can derive optimal consumption to be:

$$c_t^* = k^{-\frac{1}{\gamma}}, t \in (0, \tau)$$

The crucial step is the maximization of the problem involving the continuation value function at time τ:

$$\max_{H_t, X_t} \left[V^{SS}(X_\tau, H_\tau) - k(X_\tau + \xi H_\tau) \right] \tag{4.72}$$

where V^{SS} is given above. Assuming for the moment that

$$H_\tau < \xi^* \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)$$

we can solve the problem (4.72) using (4.53), to get:

$$\left(\frac{1}{r} \right)^\gamma \left[(1 - e^{-rT})^\gamma B \left(B \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right) + H_\tau \right)^{-\gamma} + e^{-rT} (1 + h^*)^{1-\gamma} \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)^{-\gamma} \right] = k$$

$$\left(\frac{1}{r} \right)^\gamma (1 - e^{-rT})^\gamma \left(B \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right) + H_\tau \right)^{-\gamma} = k\xi$$

where:

$$B = 1 - e^{-rT}(1 + h^*)$$

Let us define

$$k_1 = k \left(\frac{1}{r} \right)^{-\gamma}$$

so that we arrive at:

$$k_1 \xi B + e^{-rT} (1+h^*)^{1-\gamma} \left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)^{-\gamma} = k_1$$

or:

$$\left(X_{\tau-} + \frac{y_\tau^*}{r-g} \right)^{-\gamma} = \frac{k_1(1-\xi B)}{e^{-rT} (1+h^*)^{1-\gamma}}$$

or

$$
\begin{aligned}
X_{\tau-} + \frac{y_\tau^*}{r-g} &= \left(\frac{k_1(1-\xi B)}{e^{-rT}(1+h^*)^{1-\gamma}} \right)^{-\frac{1}{\gamma}} = \\
&= \left(\frac{k_1(1-\xi(1-e^{-rT}(1+h^*)))}{e^{-rT}(1+h^*)(1+h^*)^{-\gamma}} \right)^{-\frac{1}{\gamma}} = \\
&= \frac{k_1^{-\frac{1}{\gamma}}}{(1+h^*)} \left[\frac{1-\xi}{e^{-rT}(1+h^*)} + \xi \right]^{-\frac{1}{\gamma}}
\end{aligned}
$$

It is an interesting observation (that we will use later) to note that since $\xi > 1, \xi B < 1$ we get:

$$\Gamma \overset{d}{=} \frac{1-\xi}{e^{-rT}(1+h^*)} + \xi \tag{4.73}$$

and we have that $0 < \Gamma < 1, \Gamma^{-\frac{1}{\gamma}} > 1$.

Now the holdings of the second (hedging) asset are determined as:

$$(1-e^{-rT})^\gamma \left(B \left(\frac{k_1(1-\xi B)}{e^{-rT}(1+h^*)^{1-\gamma}} \right)^{-\frac{1}{\gamma}} + H_\tau \right)^{-\gamma} = k_1 \xi$$

$$\left(B \left(\frac{k_1(1-\xi B)}{e^{-rT}(1+h^*)^{1-\gamma}} \right)^{-\frac{1}{\gamma}} + H_\tau \right)^{-\gamma} = \frac{k_1 \xi}{(1-e^{-rT})^\gamma}$$

240

$$H_\tau = \left(\frac{k_1\xi}{(1-e^{-rT})\gamma}\right)^{-\frac{1}{\gamma}} - B\left(\frac{k_1(1-\xi B)}{e^{-rT}(1+h^*)^{1-\gamma}}\right)^{-\frac{1}{\gamma}}$$

To get the ratio

$$h = \frac{H_\tau}{X_{\tau-} + \frac{y_\tau^*}{r-g}}$$

we combine the preceding equations to get:

$$h = (1-e^{-rT})(1+h^*)\left(\frac{\xi}{\Gamma}\right)^{-\frac{1}{\gamma}} - (1-e^{-rT}(1+h^*))$$

We first verify that indeed $h < h^*$. This is true since:

$$
\begin{aligned}
h &= (1-e^{-rT})(1+h^*)\left(\frac{\xi}{\Gamma}\right)^{-\frac{1}{\gamma}} - \left(1+h^* - e^{-rT}(1+h^*)\right) + h^* \\
&= (1-e^{-rT})(1+h^*)\left[\left(\frac{\xi}{\Gamma}\right)^{-\frac{1}{\gamma}} - 1\right] + h^* < h^*
\end{aligned}
$$

since:

$$\frac{\xi}{\Gamma} = \frac{\xi}{\frac{1-\xi}{e^{-rT}(1+h^*)} + \xi} = \frac{1}{\frac{1-\xi}{\frac{\xi}{e^{-rT}(1+h^*)}} + 1} > 1$$

(since $\xi > 1$) and thus $(1-e^{-rT})(1+h^*)\left[\left(\frac{\xi}{\Gamma}\right)^{-\frac{1}{\gamma}} - 1\right] < 0$.

Finally we want to provide conditions in order to exclude that h becomes negative and this will be the case whenever:

$$\frac{\xi}{\Gamma} > \left(1 - \frac{h^*}{(1+h^*)(1-e^{-rT})}\right)^{-\gamma}$$

or

$$\frac{1}{\xi} < e^{-rT} \left(1 + h^*\right) \left[\left(1 - \frac{h^*}{(1+h^*)(1-e^{-rT})}\right)^\gamma - 1 \right] + 1$$

or

$$\xi > \xi^* \stackrel{d}{=} \frac{1}{1 + e^{-rT}\left(1+h^*\right)\left[\left(1 - \frac{h^*}{(1+h^*)(1-e^{-rT})}\right)^\gamma - 1\right]} \tag{4.74}$$

Accordingly we only focus on cases where $\xi < \xi^*$.

As might be expected due to the homotheticity of the problem, the optimal reserve ratio does not depend on the level of initial wealth. To complete the solution to the overall problem let us return to the time 0 budget constraint and combine everything together to get:

$$
X_0 + \frac{y_0^*}{r-g} = k^{-\frac{1}{\gamma}} E\left[\int_0^T e^{-ru}du\right] + k^{-\frac{1}{\gamma}} \frac{E\left[e^{-r\tau}\right]}{r} \frac{\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)} + \\
+ h^{opt}\xi k^{-\frac{1}{\gamma}} \frac{E\left[e^{-r\tau}\right]}{r} \frac{\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)}
$$

which shows that k can be determined as:

$$k^{-\frac{1}{\gamma}} = \left[E\left[\int_0^T e^{-ru}du\right] + \frac{E\left[e^{-r\tau}\right]}{r} \frac{(1+h^{opt}\xi)\Gamma^{-\frac{1}{\gamma}}}{(1+h^*)} \right]^{-1} \left(X_0 + \frac{y_0^*}{r-g}\right)$$

and we have that $\Gamma < 1, \Gamma^{-\frac{1}{\gamma}} > 1$. The final step of the proof is to use the formula for $E\left[e^{-r\tau}\right]$ from section 4.2. ∎

Remark 4.2 *The optimal portfolios can be derived in a manner similar to section 4.2. Their magnitude is much smaller in this case.*

4.7.5 Propositions and Proofs for section 4.3

We give a sketch of the claims. Formal proofs would proceed along the lines of the respective proofs in section 3.1. We focus only on p_t insurance for simplicity. In this case the Bellman

Equation becomes:

$$0 = \max_{c^*}\{\frac{(c^*)^{1-\gamma}}{1-\gamma} - V_X c^*\} + \max_p\{\frac{1}{2}\sigma^2 p^2 V_{XX} + V_{Xs}\sigma^2 p\} -$$

$$-rV + V_X(rX + y^*) + V_y y^* g + V_s\mu + \frac{1}{2}\sigma^2 V_{ss} + \lambda(s_t)\left[\underline{C}\frac{\left(X_t + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma} - V\right]$$

Once again the optimal consumption and portfolio policies are given by (assuming a C^2- Value function):

$$c^* = V_X^{-\frac{1}{\gamma}}$$

and

$$p = -\frac{V_{Xs}}{V_{XX}}$$

Notice that the portfolio strategy is a "pure" hedging strategy, in the sense that there are no demands due to risk premia (we assumed them away by positing that the sources of risk are uncorrelated with aggregate consumption growth). In this way we are able to distill out the pure hedging component, excluding mean-variance motives. As before we conjecture a value function of the form:

$$V(X_t, y_t, s_t) = C^H(s_t)\frac{\left(X_t + \frac{y_t^*}{r-g}\right)^{1-\gamma}}{1-\gamma}$$

where $C^H(s_t)$ satisfies the same boundary conditions as in the no-hedging case (section 3.2.). Under this conjecture the optimal portfolio process becomes:

$$p = \frac{C_{s_t}^H(s_t)\left(X_t + \frac{y_t^*}{r-g}\right)}{\gamma C^H(s_t)}$$

Notice that the sign and magnitude of hedging is influenced by the derivative of $C(s_t)$

with respect to s_t. One can show that this is a decreasing function and thus not surprisingly the hedging demands are negative, i.e. they involve short sales. Plugging back into the value function and simplifying we get a second order non-linear ODE for $C(s_t)$.

$$\gamma C^{H\frac{\gamma-1}{\gamma}} - r\gamma C^H + C_s^H \mu + C_{ss}^H \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 \frac{(1-\gamma)(C_s^H)^2}{\gamma C^H} + \lambda(s_t)\left[\underline{C} - C^H\right] = 0$$

Now defining

$$C^H = (b^H)^\gamma$$

we get

$$\begin{aligned}
0 &= \gamma\left(b^H\right)^{\gamma-1} - r\gamma(b^H)^\gamma + \gamma\mu(b^H)^{\gamma-1}b_s^H + \frac{\sigma^2}{2}\gamma(b^H)^{\gamma-1}\left[(\gamma-1)\frac{(b_s^H)^2}{b^H} + b_{ss}^H\right] \\
&+ \frac{1}{2}\sigma^2 \frac{(1-\gamma)\gamma^2(b_s^H)^2\left[b^H\right]^{2\gamma-2}}{\gamma(b^H)^\gamma} + \lambda(s_t)\left[A - (b^H)^\gamma\right]
\end{aligned}$$

or after simplifying:

$$0 = 1 - rb^H + \mu b_s^H + \frac{\sigma^2}{2}b_{ss}^H + \frac{b^H}{\gamma}\lambda(s_t)\left[\frac{\underline{C}}{(b^H)^\gamma} - 1\right]$$

Notice that once again this ODE is very similar to the ODE with no hedging, namely:

$$0 = 1 - rb + b_s\mu + \left[(\gamma-1)\frac{(b_s)^2}{b} + b_{ss}\right]\frac{1}{2}\sigma^2 + \frac{b}{\gamma}\lambda(s_t)\left[\frac{\underline{C}}{b^\gamma} - 1\right]$$

up to the absence of the term $\frac{1}{2}\sigma^2(\gamma-1)\frac{(b_s)^2}{b}$. In terms of the $b-$function the optimal

244

consumption and portfolio policies become:

$$c_t^* = \frac{\left(X_t + \frac{y_t^*}{r-g}\right)}{b^H(s_t)}$$

$$p_t = \frac{b_s^H(s_t)\left(X_t + \frac{y_t^*}{r-g}\right)}{b^H(s_t)}$$

Lemma 4.6 *Under perfect hedging "excess" consumption, i.e.*

$$c_t - \kappa y_t = c_t^*., \ t < \tau$$

follows the process:

$$\frac{dc_t^*}{c_t^*} = \left(\frac{\lambda(s_t)}{\gamma}\left[\frac{C}{C(s_t)} - 1\right]\right) dt$$

Proof. The proof proceeds in a virtually identical way as in section 4.2. ∎

Bibliography

[1] Abel, Andrew B. "Optimal Investment Under Uncertainty," *American Economic Review*, *73*(1):228–33 (1983).

[2] Abel, Andrew B. and Olivier J. Blanchard. "The Present Value of Profits and Cyclical Movements in Investment," *Econometrica*, *54*(2):249–73 (1986).

[3] Abel, Andrew B. and Janice C. Eberly. "A Unified Model of Investment under Uncertainty," *American Economic Review*, *84*(5):1369–84 (1994).

[4] Abel, Andrew B. and Janice C. Eberly. "An Exact Solution for the Investment and Value of a Firm Facing Uncertainty, Adjustment Costs, and Irreversibility," *Journal of Economic Dynamics and Control*, *21*(4-5):831–52 (1997).

[5] Abel, Andrew B. and Janice C. Eberly. "Investment and q With Fixed Costs: An Empirical Analysis." Working Paper, 2002.

[6] Abramowitz, Milton and Irene Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.

[7] Albert, J. H. and Chib S. "Bayes Inference via Gibbs Sampling of Autoregressive Time Series subject to Markov mean and Variance Shifts," *J. Bus. Econ. Stat.*, *11*(1):1–15 (1993).

[8] Allen, Franklin, et al. "Finite Bubbles with Short Sale Constraints and Asymmetric Information," *Journal of Economic Theory*, *61*(2):206–29 (1993).

[9] Ameriks, John and Stephen P. Zeldes. "How Do Household Portfolio Shares Vary With Age?." Working Paper, Columbia GSB, 2001.

[10] Baker, Malcolm, et al. "When Does the Stock Market Matter? Stock Prices and the Investment of Equity Dependent Firms," *Quarterly Journal of Economics*, (3):203–218 (2003).

[11] Baker, Malcolm and Jeffrey Wurgler. "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*, *55*(5):2219–57 (2000).

[12] Barnett, Steven A. and Plutarchos Sakellaris. "A New Look at Firm Market Value, Investment, and Adjustment Costs," *Review of Economics and Statistics*, *81*(2):250–60 (1999).

[13] Barone-Adesi, Giovanni and Robert E. Whaley. "Efficient Analytic Approximation of American Option Values," *Journal of Finance*, *42*(2):301–320 (1987).

[14] Basak, Suleyman. "On the Fluctuations in Consumption and Market Returns in the Presence of Labor and Human Capital: An Equilibrium Analysis," *Journal of Economic Dynamics and Control*, *23*(7):1029–64 (1999).

[15] Berk, Jonathan B., et al. "Optimal Investment, Growth Options, and Security Returns," *Journal of Finance*, *54*(5):1553–1607 (1999).

[16] Bertsekas, Dimitri P. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[17] Blanchard, Olivier, et al. "The Stock Market, Profit, and Investment," *Quarterly Journal of Economics*, *108*(1):115–36 (1993).

[18] Blanchard, Olivier Jean. "Bubbles, Liquidity traps, and Monetary Policy.Comments on Jinushi et al, and on Bernanke.." Working Paper, M.I.T., 2000.

[19] Blanchard, Olivier Jean and Stanley Fischer. *Lectures on macroeconomics*. Cambridge, Mass. and London: MIT Press, 1989.

[20] Bodie, Zvi, et al. "Labor Supply Flexibility and Portfolio Choice in a Life Cycle Model," *Journal of Economic Dynamics and Control, 16*(3-4):427–49 (1992).

[21] Bond, Stephen and Jason Cummins. "Noisy share prices and the Q model of investment." IFS Working Papers W01/22, Institute for Fiscal Studies, 2001.

[22] Bond, Stephen R. and Jason G. Cummins. "The Stock Market and Investment in the New Economy: Some Tangible Facts and Intangible Fictions," *Brookings Papers on Economic Activity, 0*(1):61–108 (2000).

[23] Burnside, C., et al. "Prospective Deficits and the South East Asian Crisis," *Journal of Political Economy - forthcoming* (2003).

[24] Caballero, Ricardo J. *Macroeconomic Volatility in Reformed Latin America*. (first Edition), *25*. Washington, D.C.: Inter-American Development Bank, 2001.

[25] Caballero, Ricardo J. and Mohamad L. Hamour. "Speculative Growth." Working Paper, M.I.T., 2002.

[26] Caballero, Ricardo J. and Arvind Krishnamurthy. "International and Domestic Collateral Constraints in a Model of Emerging Market Crises," *Journal of Monetary Economics., 48*(3):513–548 (2001).

[27] Caballero, Ricardo J. and Arvind Krishnamurthy. "Excessive Dollar Debt: Financial Development and Underinsurance," *Journal of Finance.* (2003).

[28] Campbell, John Y. and Luis M. Viceira. *Strategic Asset Allocation*. Clarendon Lectures in Economics, Oxford University Press, 2002.

[29] Carr, Peter, et al. "Alternative Characterizations of American Put Options," *Mathematical Finance, 2*(2):87–106 (1992).

[30] Carroll, Christopher D. and Miles S. Kimball. "On the Concavity of the Consumption Function," *Econometrica, 64*(4):981–92 (1996).

[31] Chan, Yeung Lewis and Luis Viceira. "Asset Allocation with Endogenous Labor Income: The Case of Incomplete Markets." Working Paper, Harvard University, 2000.

[32] Chirinko, Robert S. "Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications," *Journal of Economic Literature, 31*(4):1875–1911 (1993).

[33] Chirinko, Robert S. and Huntley Schaller. "Bubbles, Fundamentals, and Investment: A Multiple Equation Testing Strategy," *Journal of Monetary Economics, 38*(1):47–76 (1996).

[34] Chirinko, Robert S. and Huntley Schaller. "Business Fixed Investment and "Bubbles": The Japanese Case," *American Economic Review, 91*(3):663–80 (2001).

[35] Cochrane, John H. "Production-Based Asset Pricing and the Link between Stock Returns and Economic Fluctuations," *Journal of Finance, 46*(1):209–37 (1991).

[36] Cochrane, John H. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy, 104*(3):572–621 (1996).

[37] Cox, John C. and Chi-fu Huang. "Optimal consumption and portfolio policies when asset prices follow a diffusion process," *J. Econom. Theory, 49*(1):33–83 (1989).

[38] Cox, John C., et al. "Option Pricing: A Simplified Approach," *Journal of Financial Economics, 7*(3):229–263 (1979).

[39] DeMarzo, Peter M. "An Extension of the Modigliani-Miller Theorem to Stochastic Economies with Incomplete Markets and Interdependent Securities," *Journal of Economic Theory, 45*(2):353–69 (1988).

[40] Detemple, Jerome and Shashidhar Murthy. "Equilibrium Asset Prices and No-Arbitrage with Portfolio Constraints," *Review of Financial Studies*, *10*(4):1133–74 (1997).

[41] Diamond, Peter A. and Jerry A. Hausman. "Individual Retirement and Savings Behavior," *Journal of Public Economics*, *23*(1-2):81–114 (1984).

[42] Diebold, Francis X., et al. *Regime Switching with Time-Varying Transition Probabilities in Nonstationary Time Series Analysis and Cointegration*, chapter 10, 283–302. Oxford University Press, 1994.

[43] Duffie, Darrell, et al. "Modeling Sovereign Yield Spreads: A Case Study of Russian Debt." Graduate School of Business, Stanford University, forthcoming, Journal of Finance, 2000.

[44] Dybvig, Philip H. and Hong Liu. "Voluntary or Mandatory Retirement," *Work in progress* (2003).

[45] El Karoui, Nicole and Monique Jeanblanc-Picque. "Optimization of Consumption with Labor Income," *Finance and Stochastics*, *2*(4):409–440 (1998).

[46] Fama, Eugene F. and Kenneth R. French. "The Cross-Section of Expected Stock Returns," *Journal of Finance*, *47*(2):427–65 (1992).

[47] Fama, Eugene F. and Kenneth R. French. "Size and Book-to-Market Factors in Earnings and Returns," *Journal of Finance*, *50*(1):131–55 (1995).

[48] Fischer, Stanley and Robert C. Merton. "Macroeconomics and Finance: The Role of the Stock Market," *Carnegie-Rochester Conference Series on Public Policy*, *21*:57–108 (1984).

[49] Fleming, Wendell H. and H. Mete Soner. *Controlled Markov processes and viscosity solutions*, *25*. New York: Springer-Verlag, 1993.

[50] Fournie, Eric, et al. "Applications of Malliavin Calculus to Monte Carlo Methods in Finance," *Finance and Stochastics*, *3*(4):391–412 (1999).

[51] Froot, Kenneth A., et al. "LDC Debt: Forgiveness, Indexation, and Investment Incentives," *Journal of Finance*, *44*(5):1335–50 (1989).

[52] Gilchrist, Simon, et al. "Do Stock Price Bubbles Influence Corporate Investment." Working Paper, 2002.

[53] Griliches, Zvi and Jerry A. Hausman. "Errors in Variables in Panel Data," *Journal of Econometrics*, *31*(1):93–118 (1986).

[54] Grossman, Sanford J. and Oliver D. Hart. "A Theory of Competitive Equilibrium in Stock Market Economies," *Econometrica*, *47*(2):293–329 (1979).

[55] Grossman, Sanford J. and Joseph E. Stiglitz. "On Value Maximization and Alternative Objectives of the Firm," *Journal of Finance*, *32*(2):389–402 (1977).

[56] Gustman, Alan L. and Thomas L. Steinmeier. "A Structural Retirement Model," *Econometrica*, *54*(3):555–84 (1986).

[57] Gustman, Alan L. and Thomas L. Steinmeier. "Retirement and the Stock Market Bubble," *NBER Working Paper*, *9404* (2002).

[58] Haldane, Andy. "Private Sector Involvement in Financial Crisis: Analytics and Public Policy Approaches," *Financial Stability Review*, *7* (1999).

[59] Hamilton, James D. "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, *57*(2):357–384 (1989).

[60] Hansen, Gary. "Indivisible labor and the business cycle," *Journal of Monetary Economics*, *16*(3):309–327 (1985).

[61] Harrison, J. Michael. *Brownian motion and stochastic flow systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, New York: John Wiley & Sons Inc., 1985.

[62] Harrison, J. Michael and David M. Kreps. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations," *Quarterly Journal of Economics*, *92*(2):323–36 (1978).

[63] Hart, Oliver D. "On the Optimality of Equilibrium when the Market Structure is Incomplete," *Journal of Economic Theory*, *11*(3):418–43 (1975).

[64] Hayashi, Fumio. "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, *50*(1):213–24 (1982).

[65] He, Hua and Henri F. Pages. "Labor Income, Borrowing Constraints, and Equilibrium Asset Prices," *Economic Theory*, *3*(4):663–696 (1993).

[66] Hellwig, Martin F. "Bankruptcy, Limited Liability, and the Modigliani-Miller Theorem," *American Economic Review*, *71*(1):155–70 (1981).

[67] Hong, Harrison and Jeremy Stein. "Differences of Opinion, Short Sales Constraints and Market Crashes," *Review of Financial Studies (forthcoming)* (2002).

[68] Jazwinski, Andrew H. *Stochastic Processes And Filtering Theory*. Academic Press, New York and London, 1970.

[69] Karatzas, Ioannis, et al. "Optimal portfolio and consumption decisions for a "small investor" on a finite horizon," *SIAM J. Control Optim.*, *25*(6):1557–1586 (1987).

[70] Karatzas, Ioannis and Steven E. Shreve. *Brownian motion and stochastic calculus* (2 Edition), *113*. New York: Springer-Verlag, 1991.

[71] Karatzas, Ioannis and Steven E. Shreve. *Methods of mathematical finance*, *39*. New York: Springer-Verlag, 1998.

[72] Karatzas, Ioannis and Hui Wang. "Utility maximization with discretionary stopping," *SIAM J. Control Optim.*, *39*(1):306–329 (2000).

[73] Kim, C. J. and C. R. Nelson. "Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching," *Rev. Econ. Statist.*, *80*(1):188–201 (1998).

[74] Kim, Chang-Jin and Charles R. Nelson. *State-space models with regime switching Classical and Gibbs-sampling approaches with applications.* Cambridge MA and London: MIT Press, 1999.

[75] Koo, Hyeng Keun. "Consumption and Portfolio Selection with Labor Income: A Continuous Time Approach," *Mathematical Finance*, *8*(1):49–65 (1998).

[76] Kothari, S.P. and Jay Shanken. "Book-To-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis," *Journal of Financial Economics*, 169–203 (1997).

[77] Lamont, Owen A. "Investment Plans and Stock Returns," *Journal of Finance*, *55*(6):2719–45 (2000).

[78] Lamont, Owen A. and Charles M. Jones. "Short Sales Constraints and Stock Returns," *Journal of Financial Economics (forthcoming)* (2002).

[79] Lazear, Edward P. *Retirement from the Labor Force*, 305–55. Handbook of Labor Economics, vol. I eds. O. Ashenfelter and R. Layard, Elsveier, 1986.

[80] Lindenberg, Eric B. and Stephen A. Ross. "Tobin's q Ratio and Industrial Organization," *Journal of Business*, *54*(1):1–32 (1981).

[81] Magill, Michael and Martine Quinzii. *Theory of incomplete markets Volume 1.* Cambridge and London: MIT Press, 1996.

[82] Meeker, Edward J. *Short Selling*. Harper and Brothers Publishers, New York and London, 1932.

[83] Merton, Robert C. "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, *3*(4):373–413 (1971).

[84] Morck, Randall, et al. "The Stock Market and Investment: Is the Market a Sideshow?," *Brookings Papers on Economic Activity*, *0*(2):157–202 (1990).

[85] Morris, Stephen. "Speculative Investor Behavior and Learning," *Quarterly Journal of Economics*, *111*(4):1111–33 (1996).

[86] Musiela, Marek and Marek Rutkowski. *Martingale Methods in Financial Modelling* (Fffirst Edition), *36*. Applications of Mathematics. New York: Springer-Verlag, 1998.

[87] Naik, Vasanttilak. "Asset Prices in Dynamic Production Economies with Time-Varying Risk," *Review of Financial Studies*, *7*(4):781–801 (1994).

[88] Newey, Whitney K. and Daniel McFadden. "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics, North Holland*, *4*

[89] Nicholas, Tom. "Why Schumpeter Was Right: Innovation, Market Power, and Creative Destruction in 1920s America," *The Journal of Economic History*, *63*(4) (2003).

[90] Øksendal, Bernt. *Stochastic differential equations* (5 Edition). Berlin: Springer-Verlag, 1998. An introduction with applications.

[91] Olivier, Jacques. "Growth-Enhancing Bubbles," *International Economic Review*, *41*(1):133–51 (2000).

[92] Paul, Krugman. "Financing vs. Forgiving a Debt Overhang," *Journal of Development Economics*, *29*:253–68 (1988).

[93] Polk, Christopher and Paola Sapienza. "The Real Effects of Investor Sentiment." Working Paper, Kellog School of Management, 2002.

[94] Rainville, Earl D., et al. *Elementary Differential Equations*. Prentice Hall, Eighth Edition, 1997.

[95] Robert, Christian P. and George Casella. *Monte Carlo statistical methods*. New York: Springer-Verlag, 1999.

[96] Rockafellar, R. Tyrrell. *Convex analysis*. Princeton Landmarks in Mathematics, Princeton, NJ: Princeton University Press, 1997.

[97] Rogerson, Richard. "Indivisible labor, lotteries and equilibrium," *Journal of Monetary Economics*, *21*(1):3–16 (1988).

[98] Rust, John. *Structural Estimation of Markov Decision Processes*, 3081–3143. Handbook of Econometrics, vol. 4 eds. Engle R.F and McFadden D.L,, Amsterdam: North-Holland, 1994.

[99] Rust, John and Christopher Phelan. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, *65*(4):781–831 (1997).

[100] Scheinkman, Jose and Wei Xiong. "Overconfidence and Speculative Bubbles," *Journal of Political Economy (forthcoming)* (2003).

[101] Shleifer, Andrei. *Inefficient markets: An introduction to behavioral finance*. Oxford and New York: Oxford University Press, 2000.

[102] Stein, Jeremy C. "Rational Capital Budgeting in an Irrational World," *Journal of Business*, *69*(4):429–55 (1996).

[103] Stock, James H. and David A. Wise. "Pensions, the Option Value of Work, and Retirement," *Econometrica*, *58*(5):1151–80 (1990).

[104] Tirole, Jean. "Inefficient Foreign Borrowing: A Dual-and-Common Agency Perspective." Mimeo, Toulouse, 2002.

[105] Verhulst, Ferdinand. *Nonlinear differential equations and dynamical systems* (Second Edition). Universitext, Berlin: Springer-Verlag, 1996.

[106] Viceira, Luis. "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income," *Journal of Finance*, *56*(2):433–470 (2001).