Essays on the Effect of Risk on Wealth-Enhancing Investment Decisions

A Thesis Submitted to The University of Manchester for the Degree of
Doctor of Philosophy
in the Faculty of Humanities

Year of Submission
2015

David Chivers
School of Social Sciences
Economics
## Contents

Abstract ................................................................. 6  
Declaration .............................................................. 7  
Copyright Statement .................................................. 7  
Acknowledgement ....................................................... 8  

1 Fearing the Worst: The Importance of Uncertainty for Inequality 12  
1.1 Introduction ....................................................... 12  
1.2 Background Literature .......................................... 14  
1.3 The Model .......................................................... 21  
   1.3.1 Preferences and Technologies ............................. 21  
   1.3.2 The Decision Problem of Agents ......................... 25  
1.4 Equilibrium Outcomes ........................................... 25  
1.5 Distribution and Inequality .................................... 27  
1.6 Further Discussion .............................................. 29  
   1.6.1 Some Related Research ................................... 29  
      1.6.1.1 Inequality and Financial Markets ............... 29  
      1.6.1.2 Uncertainty and Macroeconomic Performance .... 34  
   1.6.2 Some Extensions ........................................... 35  
      1.6.2.1 Generalising Aspirations ......................... 35  
      1.6.2.2 Introducing Wealth ............................... 36  
      1.6.2.3 Incorporating Bankruptcy ....................... 37  
1.7 Conclusion ....................................................... 41  

2 Macroeconomic Effects of Mandated Health Benefits 50  
2.1 Introduction ....................................................... 50  
2.2 Facts and Policies .............................................. 54  
   2.2.1 U.S. Health Care .......................................... 54  
   2.2.2 Policies .................................................... 57  
2.3 The Model: Economic Environment ............................. 58  
   2.3.1 Preferences, endowments and technology .............. 59  
   2.3.2 Firm’s problem .......................................... 64  
   2.3.3 EHI and talent misallocation ........................... 65
2.4 Optimal behaviour and equilibrium

2.4.1 Firm manager

2.4.1.1 Remark on random matching

2.4.1.2 Capital

2.4.2 Workers

2.4.3 Government

2.4.4 The household’s problem

2.4.5 Health insurance

2.4.6 Steady state equilibrium

2.4.7 Analysis of competitive equilibrium

2.5 Calibration

2.6 Quantitative Analysis

2.6.1 Baseline Economy

2.6.2 Policy designs

2.6.2.1 Expansion of EHI

2.6.2.2 No EHI: Private indemnity contract

2.6.2.3 Health insurance exchange

2.6.2.4 Universal health insurance

2.6.3 Size distribution

2.6.4 Welfare analysis: conditional change

2.6.5 Welfare and risk aversion: stationary distribution

2.6.6 Policy Summary

2.7 Conclusion

Summary and Conclusions

Final Word Count: 25,782
## List of Figures

1.1 Human Capital Distribution ........................................... 49
2.1 Health expenditure in OECD countries, 1970-2012 ................. 54
2.2 EHI offer rate by establishment size, MES data .................. 55
2.3 Talent misallocation .................................................... 64
2.4 Cutoff for ability \( x \) .................................................. 76
2.5 Conditional welfare change, EHI expansion ....................... 85
2.6 Conditional welfare change, HIX ..................................... 86
2.7 Conditional welfare change, Universal Health Insurance .......... 87
2.8 Conditional welfare change, no EHI (private insurance only) .... 88
2.9 Welfare in simulations, \( \rho = 1.5 \) ................................. 89
2.10 Welfare in simulations, \( \rho = 3 \) ................................. 89
List of Tables

2.1 Income levels and insurance premium subsidies under ACA . . . . . . 58
2.2 Parameter values, baseline economy . . . . . . . . . . . . . . . . . . 79
2.3 Benchmark . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
2.4 Aggregate variables . . . . . . . . . . . . . . . . . . . . . . . . . . 83
2.5 Policy experiments . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
2.6 Aggregate variables, $\rho = 3$ . . . . . . . . . . . . . . . . . . . . . 91
Abstract

This thesis provides two general equilibrium models to analyse the macroeconomic effect of risk on wealth-enhancing investment decisions.

In chapter 1, we present an overlapping generations model in which aspirational agents face uncertainty about the returns to human capital. Investment in human capital requires external funding, implying a probability of bankruptcy that is greater the lower the human capital endowment of an agent. We show that agents with sufficiently low human capital endowments may experience such a strong influence of loss aversion that they abstain from human capital investment. We further show how this behaviour may be transmitted through successive generations to cause initial inequalities to persist. These results do not rely on any credit market imperfections.

In chapter 2 we note that most working-age Americans obtain health insurance coverage through the workplace. U.S. law requires employers that offer health plans to use a price common to all in the group. However, the value of health insurance to risk-averse agents varies with their idiosyncratic health risk. Hence, linking employment and health insurance creates a wedge between the marginal cost and benefit of insurance. Since health risk can be sizable and health insurance is part of total employee compensation, the wedge can affect firm and employee decisions. We study the impact of this wedge on occupational choice, productivity and welfare in a general equilibrium model with agents who are endowed with idiosyncratic health risk and heterogeneous managerial ability. Agents choose whether to be a worker or an entrepreneur. We find that the wedge distorts occupational choice by causing two types of misallocations. Some highly skilled individuals with adverse health shocks leave entrepreneurship while individuals with intermediate skills but favourable health shocks opt to manage firms. Four policies are analysed: expansion of employer-based health insurance; private insurance; health insurance exchanges; and universal health coverage. Factor prices are determined endogenously and programs are financed by lump sum taxes. We assess the quantitative effects of the policies on firm size, productivity, GDP, and earnings. Welfare effects may be positive or negative, vary significantly with an individual’s position in the asset and ability distributions, and are sensitive to changes in risk aversion.
Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

1. i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

2. ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

3. iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

4. iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy

see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectualproperty.pdf

in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations

see http://www.manchester.ac.uk/library/aboutus/regulations

and in The University’s policy on presentation of Theses.
Acknowledgement

Firstly, I must thank my supervisors, Keith Blackburn and Anne Villamil. Keith and I have spent many meetings together in various establishments around Manchester. From the smallest pub in Manchester to meeting with eminent economists at the Midland hotel. The incredible variety of these places perhaps best highlights our mutual interest in inequality (and whisky). Anne and I probably could not have undertaken research together if it wasn’t for Skype. After working with Anne for a period of time I felt my body clock has firmly adjusted to Chicago time. On that note, I would also like to thank Zhigang Feng, whom I have only met via Skype and not yet had the pleasure of meeting him in person.

I knew I wanted to undertake a PhD in economics from the early age of 16. I have no doubt that this was due to my first two teachers in economics, Tass Sgouros and Mike Frost. Their enthusiasm for the subject gave me an excellent grounding in economics which I feel has helped my research immensely.

I would also like to thank all the lecturers at the University of Manchester whom I have been taught and guided by. Specifically, Horst Zank for excellent discussions about economics. Ralf Becker, for all the academic advice, references and counselling. My lasting memory of him will be taking a meeting, pen in his left hand and rocking a pram in his right. Martyn Andrews, whose loud profanities after receiving an error message in Stata will surely echo in the Arthur Lewis building for an eternity. And last of all, Kyriakos Neanidis for failing me in a course - I haven’t forgotten.

I would like to thank my friends, especially my PhD colleagues, for all their support and wise words of advice. One of the PhD students, Serena Masino, taught me how to write English properly - I like to think we managed to limp over the PhD finishing line together. Jaqueson Galimberti for always being there to answer my questions. After four years of sitting next to him I don’t think he ever managed to grasp my accent (and I have never managed to grasp his research). Tommy Allan for making me arrive 10 minutes early to every lecture so we could get a good seat.

Finally, I must thank my family. I think they more than anyone know how great an achievement this is for me. Specifically, my Grandfather Peter, who seems to be the only one of my family (or for that matter, anybody) interested enough to listen to me talk about my thesis.
Introduction

Understanding the effect of risk on wealth-enhancing investment decisions - such as entrepreneurship or human capital - is important from both a development and welfare perspective. Entrepreneurship and human capital are inextricably linked with innovation, and as a consequence, the long-run growth of an economy. Excessive risk can potentially prevent an individual from choosing to invest in a project due to an individual’s preference to avoid uncertainty. Furthermore, how a society chooses to insure against certain risks - such as health cost shocks - may have an effect on the individual’s investment decision. We seek to explore these issues in two separate essays, respectively.

In chapter 1 we construct an overlapping generations model in order to demonstrate how an individual’s loss averse preferences can result in multiple equilibria without the need for capital market imperfections. Individuals face a decision of whether to invest in a risky wealth-enhancing project. We find that an increase in uncertainty or in the preference to avoid losses increases the critical value of wealth needed for individuals to invest in these projects. Hence, if initial wealth is distributed heterogeneously, initial inequality may persist and increase.

A capital market imperfection occurs when an agent is refused credit - or faces higher borrowing costs - due to insufficient collateral in the form of initial wealth. This imperfection subsequently denies the agent access to invest in wealth-enhancing projects, hence, when initial wealth is distributed heterogeneously, inequality may persist. Despite an abundance of literature utilising imperfect capital markets to explain why agents underinvest in wealth-enhancing projects, a crucial element fails to be accounted for in the agent’s decision to invest: risk. Recent empirical research has focused on the level of uncertainty being key to the savings decision making behaviour of the individual: as the level of risk increases, agents desire to insure themselves against future uncertainty increases. Consequently, individuals create a buffer-stock level of savings. As this behaviour increases as wealth decreases we suggest that when wealth-enhancing projects are risky, relatively poorer agents are less likely to invest.
The concept of loss aversion - where individual’s aversion to uncertainty increases with the size of the stake ventured - can enlighten this type of savings behaviour. It follows that aversion to losses may prevent relatively poorer agents investing in risky projects as they are comparatively less able to buffer themselves against negative shocks. Ultimately, it will be shown that loss averse preferences alone can explain the level of inequality without the need for an imperfect capital market assumption.

In chapter 2 we present an occupational choice model where agents are heterogeneously endowed with managerial talent and experience health cost shocks. Each worker is paid a wage package, which includes both a monetary wage and health insurance. Larger firms are able to offer better wage packages due to economies-of-scale in the purchasing of health insurance. The results suggest that employer-based health insurance may result in a misallocation of talent. As a result, some of the most talented individuals will be deterred from becoming an entrepreneur as a result of employer-based health insurance. We also examine counterfactual policy experiments as well as look at the potential impact of the Affordable Care Act (ACA) passed in 2010 (referred to commonly as Obamacare).

Individuals facing the choice of whether or not to start a firm will be concerned with the possible spread of returns from entrepreneurship. However, by choosing to become an entrepreneur, rather than working for a firm, exposes the individual to other idiosyncratic risks - such as healthcare costs - that would otherwise be covered by an employer. The question for the individual does not simply become what is the risk of entrepreneurship? Rather, what is the associated risks of not being employed by a firm?

These negative health shocks have so far been overlooked in the occupational choice literature which tend to focus on uncertainty of the project itself when considering the effect of risk on entrepreneurship. It follows that within economies where Employer-based Health Insurance (EHI) makes up a significant proportion of the economy’s healthcare system - such as the U.S. economy - the cost and quality of health insurance will be an important factor in the entrepreneurial decision.

The rest of the thesis is organised as follows. We present the chapters summarised above with the relevant appendices attached to the end of each chapter. Finally, a conclusion is offered to summarise the key findings of the thesis.
Chapter 1

Fearing the Worst: The Importance of Uncertainty for Inequality

1.1 Introduction

Contemporary theories of income distribution are typically based on an appeal to some form of market imperfection which creates different incentives and opportunities for different individuals. Most prominent in this research are models that rely on the imperfect functioning of capital markets for one reason or another. The key implication of these models is that agents with insufficient wealth to serve as collateral for loans may be deterred or prevented from accessing profitable investment opportunities because of high costs of borrowing or rationed availability of credit. Moreover, any initial differences in individual wealth may turn out to be persistent (permanent) fixtures such that the limiting distribution of wealth is characterised by the same agent heterogeneity and income inequality as exists to begin with. From a macroeconomic perspective, the models also yield further insights by revealing how distributional outcomes can influence aggregate economic performance in terms of growth and development.

This paper does not seek to undermine the potential significance of (financial) market imperfections in determining the relative fortunes of individuals. Rather, its aim is to highlight another factor for consideration, one that is possibly of equal importance but that has hitherto been largely (and rather surprisingly) neglected
by researchers. This is the role of risk in individual decision making when the outcomes of decisions are uncertain.¹

There are good reasons for thinking why aspects of risk and uncertainty may be important for issues of distribution. Not least of these is the precautionary motive for savings which suggests that agents’ desire to insure themselves against uncertainty leads them to create a buffer-stock level of savings that increases with the degree of uncertainty. If one thinks of this motive as being stronger for less wealthy agents (as suggested by empirical observation), then one begins to realise why poorer members of the population may be less likely to undertake wealth-enhancing ventures when such ventures are relatively risky. In addition, by entertaining the notion of intergenerational linkages, one may also start to contemplate the possibility of history-dependent behaviour and, with this, the prospect of persistent inequality.

Our basic objective in this paper is to explore the idea that, in an uncertain environment, distributional outcomes may have as much to do with the structure of preferences as they have with the functioning of markets. We approach this by appealing to some recent advancements in decision theory. Specifically, we call upon the hypothesis of aspiration-induced loss aversion as a means of enlightening the type of behaviour that we envisage. This hypothesis is based on the notion that individuals, faced with some risky prospect, have concern over attaining (or not attaining) a certain level of wealth to which they aspire. Any outcome above (below) this aspiration level is regarded as a success (failure). The result of this is a value function that reflects individuals’ weighted preferences over the overall probability of success and/or the overall probability of failure. These preferences can significantly influence the evaluation of a prospect and are absent from standard expected utility theory.

The concept of an aspiration level bears an obvious similarity to the concept of a reference point used in prospect theory. An important difference, however, is the way that outcomes are defined: models based on reference points take outcomes to be changes in wealth, whilst models based on aspiration levels take outcomes to be final states of wealth. As mentioned above, the preference for final state wealth to

¹For the purposes of this paper, we use the terms risk and uncertainty interchangeably, and do not distinguish between the two.
be above some aspiration level may have a significant influence on decision making. For example, individuals with relatively low levels of wealth to begin with will be relatively more wary about taking on risks as they are relatively less able to buffer themselves against bad outcomes. This provides the basic intuition underlying our analysis which shows how a sufficiently strong aversion to falling short of aspirations may induce the least wealthy agents to forego potentially profitable, but prohibitively risky, investment opportunities. Embedding these microfoundations in an overlapping generations framework, we also demonstrate how initial inequalities may persist as a long-run feature of distributional outcomes. The striking aspect of these results is that they are realised within the context of a financial environment in which borrowing and lending opportunities are unconstrained by any frictions. Rather than being the product of credit market imperfections, they are driven more fundamentally by the deep structure of preferences governing attitudes towards risk. Significantly, the behaviour produced by these preferences is practically identical to the behaviour produced by financial constraints. We are unaware of any other analysis to offer a similar perspective and to establish similar insights.

The remainder of the paper is organised as follows. In Section 1.2 we provide a brief overview of the literature that forms the background to our investigation. In Section 1.3 we present the model that we use to conduct our analysis. In Section 1.4 we deduce the equilibrium behaviour of agents. In Section 1.5 we establish our main results. In Section 1.6 we comment on these results within the context of other research, as well as outlining some extensions of our analysis. In Section 1.7 we make a few concluding remarks.

1.2 Background Literature

The link between income distribution and the functioning of financial markets is formally articulated in a number of analyses that form a well-established and influential body of research (e.g., Aghion and Bolton 1997; Banerjee and Newman 1993; Blackburn and Bose 2001; Galor and Zeira 1993; Piketty 1997).² Broadly

²This research developed alongside other work on income distribution that signalled a general revival of interest in the subject. Amongst this work are models of redistribution based on
speaking, this research seeks to examine the extent to which capital market imperfections of one form or another (such as asymmetric information or weak contract enforcement) can cause initial income inequalities to persist over time. The basic idea is illustrated by considering an environment in which agents face a choice between two types of production, or investment, activity: the first is a low-cost, but low-yielding, venture (e.g., subsistence production), whilst the second is a high-cost, but high-yielding, endeavour (e.g., human capital investment). Agents obtain funding for the latter by using their own endowments of wealth and by acquiring loans from financial intermediaries if necessary. Because of capital market imperfections, the terms and conditions of loan contracts depend on agents’ wealth status: poorer agents face higher costs of borrowing and/or higher requirements for collateral. The upshot is that there is a critical level of wealth below which agents are forced to undertake the low-yielding activity, whilst above which agents enjoy access to the high-yielding venture. In a dynamic setting this division of the population may endure through successive, interconnected generations of agents if other circumstances prevail, such as indivisibilities in investment. If so, then the limiting distribution of wealth is characterised by two steady states as initial inequalities persist to produce a polarisation between the rich and the poor. Evidently, since these cohorts engage in different activities with different productivities, the extent of inequality has implications for macroeconomic performance in terms of aggregate output and possibly growth.

To many observers, theories of income distribution based on capital market imperfections are compelling, not least because of the apparent pervasiveness of such frictions. Yet there is another feature of economies that is equally, if not more, pervasive and that may be worth just as much consideration: this is the existence of uncertainty, about which relatively little has been written in connection with income distribution. This is somewhat surprising, given the major role that risk can play in savings and investment decisions, the outcomes of which, being realised in the future, are typically fraught with uncertainty. For example, Krebs (2003) argues that investment in human capital is particularly susceptible to idiosyncratic, political motives (e.g., Alesina and Rodrik 1994; Persson and Tabellini 1994; Perotti 1993) and models of inequality based on neighbourhood effects (e.g., Benabou 1992; Durlauf 1993; Fernandez and Rogerson 1996).
uninsurable labour risk (due to the unpredictably of employment opportunities and job-searching time), whilst Grossman (2008) lists a variety of other reasons why such investment is risky, not least of which is individuals’ uncertainty about the distribution of post-educational earnings (because of changes in technology and demand conditions that may occur during the period of education). Another prime example of risky investment is the acquisition of equities, on which some notable observations have been made, such as the tendency of individuals to hold a smaller proportion of risky assets in their portfolios the greater is the degree of their income uncertainty and the lower is the level of their wealth (e.g., Guiso et al. 1996; Aiyagar 1994). From a distributional perspective, one’s attention is particularly drawn to the last observation since it suggests a connection between wealth accumulation and the amount of risk that individuals are willing to bear. For this reason, it is important to understand how attitudes towards risk might influence decisions that govern the relative fortunes of individuals who do not share the same wealth status.

As discussed by Guiso and Paiella (2008), there is a prevailing consensus that individuals’ aversion towards risk is decreasing in wealth. Various forms of utility function have been proposed to take account of this, though many of these are used for the purposes of tractability, rather than for their plausibility (e.g., Carroll and Kimball 1996). The idea, itself, accords with one’s intuition that, in the words of Rabin (2000), “a dollar that helps us avoid poverty is more valuable than a dollar that helps us become very rich”. Nevertheless, as pointed out by the same author, and emphasised further in Rabin and Thaler (2001), there remains a problem: under standard expected utility theory, risk averse behaviour for small stake gambles implies improbably high risk aversion for large stake gambles such that the marginal utility of wealth must decrease at an astronomical rate. Even if one ignores this, it is questionable whether the modelling of risk aversion under conventional expected utility theory is capable, by itself, of explaining why relatively poor agents choose not to pursue potentially wealth-enhancing opportunities.

One way of moving forward from the above is to think beyond the standard paradigm of expected utility by exploring other concepts in decision theory. Our attention is particularly drawn to the concept of loss aversion which entertains the idea that individuals have a stronger preference for avoiding losses than for ac-
quiring gains. The concept was first introduced by Kahneman and Tversky (1979) in their pioneering work on prospect theory.³ This theory makes two assertions about loss aversion: the first is that agents derive less utility from undertaking symmetric bets than they do from accepting the expected outcome with certainty; the second is that the extent of aversion to such bets increases with the size of the stake. The type of utility function that captures these features is one that exhibits reference-dependent asymmetry: relative to some benchmark outcome, losses are weighted more heavily than gains (i.e., the utility function is steeper for bad states than for good states). Thus, what matters to an individual when faced with some gamble is not the total amount of income that he ends up with, but rather the amount of income relative to some reference level, deviations from which are evaluated differently depending on whether they are positive or negative. Diagrammatically, the utility function is generally drawn S-shaped with a kink at the reference point, where it changes from being convex (in the domain of losses) to concave (in the domain of gains).⁴

The concept of loss aversion is particularly prominent in the field of behavioural economics and finance, where it has been subject to much investigation by decision theorists (e.g., Bleichrodt et al. 2009; Schmidt and Zank 2005) and applied by others to explain apparent anomalies and paradoxes, such as the endowment effect (e.g., Thaler 1980), the status quo bias (e.g., Samuelson and Zeckhauser 1988) and the equity premium puzzle (e.g., Benartzi and Thaler 1995). Its application in macroeconomics remains much more limited, and we are unaware of its use in any macro-type (dynamic general equilibrium) model of inequality and income distribution. Our aim in this paper is to develop such a model.⁵

Loss aversion draws attention to the importance of downside risk in shaping individuals’ preferences. In models where this is motivated by reference points, outcomes are defined in terms of changes in wealth and the extent of loss aversion is reflected in the shape of the utility function independently of probabilities. As

³This has become the dominant descriptive theory of decision making under uncertainty. It is part of the broader literature on non-expected utility, a comprehensive review of which can be found in Starmer (2000).

⁴A selection of alternative functional forms can be found in Maggi (2004).

⁵Two recent macroeconomic applications of prospect theory are presented by Foellmi et al. (2011) and Rosenblatt-Wisch (2008). The focus of each these is on the aggregate implications of loss aversion in stochastic growth models.
regards the latter aspect, it has been argued on the basis of experimental evidence that a major concern for individuals in evaluating risky prospects is the overall chances of success and failure. For example, Edwards (1954) shows that individuals prefer low probabilities of large losses to high probabilities of small losses, whilst Langer and Weber (2001) and Payne (2005) reveal that individuals pay special consideration to the probabilities of winning and losing as a whole.

To take account of the above, we turn to another, more recent, concept in decision theory - namely, aspiration levels. The basic idea of this is that individuals evaluate risky prospects according to their weighted preferences over the overall probabilities of success and failure, where success and failure are defined with respect to some aspirational outcome (e.g., Diecidue and Van de Ven, 2008). Individuals whose initial wealth is above (below) their aspiration level may be thought of as seeking to maintain (improve) their status, implying the possibility of risk averse (risk loving) behaviour in the proximity of the threshold. The existence of aspiration levels has been detected in several empirical studies (e.g., Holthausen 1981; Mezias 1988) and it has been argued that individuals may use them as a means of simplifying complex decision problems (e.g., Langer and Weber 2001; Mezias et al. 2002). Like reference points, aspiration levels may be defined with respect to different target outcomes, such as maintaining initial wealth status, staying above the poverty line and avoiding situations of bankruptcy. Failure to meet such targets is assumed to result in a direct disutility (psychic) cost independently of any monetary cost. Whilst reference points and aspiration levels are closely related, there are important differences between them: the latter is a probabilistic (rather than purely behavioural) concept that takes outcomes to be final states of (rather than changes in) wealth, and that gives rise to a utility function which is discontinuous (rather than kinked) at the threshold point. The last of these features has the effect of reinforcing loss aversion.6

Part of the motivation for our analysis derives from ideas relating to precau-

---

6For a broad comparison of aspiration level models and prospect theory, see Lopes and Oden (1999). In a wider context, we note that aspiration-based preferences bear some similarity to, but are distinct from, maxi-min-type preferences. Unlike the former, the latter do not generate discontinuities and do not involve reference dependence. Compared to loss aversion, maxi-min preferences may similarly give rise to kinks, but these kinks would be observed in the indifference curves, rather than utility. The extent to which our analysis (based on aspiration theory) is transferable to other behavioural models of uncertainty is an interesting issue for further research.
tionary savings behaviour, as discussed by Carroll (2001). The basic reason for such behaviour is that, when confronted by uninsurable risk, individuals save as a means of buffering themselves against future bad shocks. Suppose that agents have some target level of consumption. A negative income shock is of greater concern to poor agents than rich agents since the former stand a higher chance of failing to reach their consumption target. Less wealthy agents are therefore inclined to have a larger “buffer stock” of savings on which they can draw.\(^7\) To the extent that these savings are less productive than other, more risky, ways of disposing of income, initial inequalities may be reinforced. By way of further illustration, consider the following example which resonates more closely with our previous discussion. Imagine two agents who are faced with a gamble (a risky investment project, perhaps) that offers an equal chance of either winning or losing the same amount of money. Suppose that these agents are identical except in terms of their initial levels of wealth which lie above their common aspiration level (a poverty line, for example). In the worst case scenario, the poorer agent falls below the aspiration level, whilst the richer agent remains above it. Thus, although the rewards are comparatively higher for the former, the stake that is being risked is comparatively higher for him as well because of the higher probability of not meeting his aspirations. Fearing the worst, this may induce the poorer agent to abstain from the gamble (forego investment) altogether. Similar to before, what this example illustrates is a tendency for less wealthy individuals to be less inclined to pursue risky, but potentially profitable, opportunities because the amount that they may lose is prohibitively costly for them (i.e., failing to achieve their aspirations counts for more than increasing their wealth). In this way, initial inequalities can cause diverse behaviour (gambling or not gambling, investing or not investing) which strengthens those inequalities and shapes distributional outcomes.

In the analysis that follows we seek to articulate the above ideas more rigorously in a simple theoretical model. The model describes an overlapping generations economy in which aspirational agents produce output using human capital, an ini-

\(^7\)Note that this argument does not necessarily mean that poor agents save more in total than rich agents. The argument refers specifically to precautionary savings, the motive for which intensifies as wealth declines because of the greater inability to buffer one’s consumption against bad shocks (e.g., Carroll 2001). This is not inconsistent with total savings being an increasing function of wealth.
tial distribution of which accounts for agent heterogeneity whilst lineage transfers of which account for intergenerational linkages. An agent accumulates human capital for himself by drawing on the human capital inherited from his parent and by undertaking his own self-improvements of knowledge and ability. The agent can maximise his human capital accumulation by making some physical investment of resources, for which he requires a loan from financial intermediaries. This is risky because the returns to human capital are uncertain and may not be high enough for an agent to achieve his aspirations after settling loan repayments. Such an outcome is more likely to occur the less is the amount of inherited human capital. Against this background, we show how agents with sufficiently low endowments of human capital may have such a strong aversion to losses that they choose not to risk borrowing for human capital investment. We further show how this behaviour can be transmitted through successive generations to cause initial inequalities to persist. As indicated earlier, the striking aspect of these results is that they do not rely on any credit market imperfections, but rather are derived within the context of a financial environment that allows borrowing and lending to take place without impediment. In accordance with Carroll (2001), there are some individuals who do not borrow, not because they are unable to do so, but because they choose not to do so. These are very different explanations, but another significant insight of our analysis is that it may be very difficult to distinguish between them since the behavioural outcomes in each case are virtually the same.

A recent related analysis to ours is that of Genicot and Ray (2014) who address the interesting matter of how aspirations are shaped and formed. In particular, the authors construct a model in which aspirations are co-determined endogenously with the distribution of income. The basic idea is that individuals form their aspirations with reference to the social environment (e.g., wealth distribution) which, in turn, evolves according to the development of aspirations. It is shown how persistent inequality may result from this. Whilst touching on similar issues, the focus of our own analysis is quite different. Like most of the literature, we do not seek to delve deeply into questions of what motivates and determines individual aspirations. Rather, taking aspirations as given, our concern is to examine how the uncertain prospect of attaining them might influence distributional outcomes through a differential impact of loss aversion on individuals’ behaviour. We
similarly show how persistent inequality may arise from this.

1.3 The Model

We consider a small open economy in which there is a constant population of mortal, reproductive agents measuring a size of unit mass. Each agent lives for two periods and belongs to a dynastic family of overlapping generations connected through transfers of human capital. Each agent has one parent and one child, inheriting capabilities from the former and imparting capabilities to the latter. Each agent is a potential investor in human capital when young, and a producer and consumer of output when old. We proceed with our formal description of the economy with reference to the circumstances facing agents of generation \( t \).

1.3.1 Preferences and Technologies

All agents have identical preferences defined over old-age consumption, or income, \( x_{t+1} \), from which they derive a lifetime utility of \( u_t = u(x_{t+1}) \). Under standard expected utility theory, an agent’s objective is to maximise \( E(u_t) \). Our departure from this involves a non-expected utility approach based on aspiration level theory. In general, this theory posits an objective function that depends not only on the expected utility of a prospect, but also on the overall probability of success and/or the overall probability of failure in attaining some target outcome. Denoting these probabilities by \( P^s \) and \( P^f \), respectively, one imagines individuals as maximising a value function of the form \( V_t = E(u_t) + \mu P^s - \lambda P^f \), where \( \mu \) and \( \lambda \) represent weighting parameters that measure the importance of success and failure.

An agent aspires to achieving some target, or reference, level of income, \( x^* > 0 \). The agent succeeds or fails in realising his aspirations according to whether \( x_{t+1} \geq x^* \) or \( x_{t+1} < x^* \). By way of simplification, we assume that agents care only about the prospect of failure (\( \mu = 0 \)), the overall probability of which is now denoted by \( P^f = P(x_{t+1} < x^*) \). For further convenience, we also assume that utility from consumption is linear, \( u(x_{t+1}) = x_{t+1} - x^* \). These features are inconsequential for our main results, as we demonstrate subsequently under various extensions of the model. As matters stand at present, the actual payoff to an agent is understood
to be either $x_{t+1} - x^*$ if $x_{t+1} \geq x^*$, or $x_{t+1} - x^* - \lambda$ if $x_{t+1} < x^*$, implying the sort of discontinuity which can account for loss aversion. The expected payoff to an agent is then given by the value function

$$V_t = E(x_{t+1} - x^*) - \lambda P(x_{t+1} < x^*).$$

(1.1)

In the first period of life an agent makes a decision about his investment, $i_t$, in human capital.\(^8\) In the spirit of other analyses (e.g., Banerjee and Newman 1993; Galor and Zeira 1993), we assume that there is a fixed cost of investment, $k > 0$, such that agents are faced with the binary choice of either $i_t = 0$ or $i_t = k$. Since all agents are endowed with zero resources to begin with, any of them who chooses the latter option must finance his investment by borrowing. The concept of human capital in our model need not be restricted to including just knowledge and skills, but may be thought of more broadly as encompassing other personal attributes (most notably, health) that enhance productive efficiency. Whatever the interpretation, and whatever choice is made, an agent accumulates human capital, $h_{t+1}$, in a way that depends on the human capital inherited from his parent, $h_t$, augmented by some additional, but uncertain, component, $\gamma_{t+1}$. Specifically,

$$h_{t+1} = \begin{cases} \beta h_t + b(1 + \gamma_{t+1}) & \text{if } i_t = 0, \\ \beta h_t + B(1 + \gamma_{t+1}) & \text{if } i_t = k, \end{cases}$$

(1.2)

where $\beta \in (0, 1)$ and $B > b > 0$.\(^9\)

The term $\gamma_{t+1}$ in (1.2) is a bounded random variable with known probability distribution, but unknown realised value at the time that agents make decisions. This variable may be thought of as capturing various personal characteristics (e.g., innate ability, health status and all-round functionality) that are randomly bestowed on agents and that agents become aware of during the course of human

---

\(^8\)This is an investment of physical resources, rather than time or effort. The latter may be treated as being already subsumed into the behaviour of agents, who may be thought of as devoting a fixed amount of time or effort to human capital production when they are young.

\(^9\)If the concept of human capital is confined to schooling and education, then the target outcome, $x^*$, may be thought of as reflecting an underlying target for academic achievement to which individuals aspire. For example, $x_{t+1} \geq x^*$ may correspond to the case in which a student graduates successfully, whilst $x_{t+1} < x^*$ may represent the case in which a student barely passes (or perhaps drops out of college altogether).
capital formation. For any given realisation of \( \gamma_{t+1} \), an agent ends up with more human capital if he invests resources in improving his capabilities than if he foregoes such investment. The basic role played by \( \gamma_{t+1} \) is to inject uncertainty into agents’ future incomes by creating uncertainty about their future productive efficiency. An alternative approach would be to assume that output production is, itself, stochastic (due to technology shocks), in which case a similar analysis could be conducted to obtain essentially the same results. We choose the present modelling strategy arbitrarily and for no particular reason.\(^{10}\)

For simplicity, we assume that \( \gamma_{t+1} \) is uniformly distributed over the interval \((-c, c)\) with probability density function \( f(\gamma_{t+1}) = \frac{1}{2c} \), where \( c < 1 \). The mean and variance of \( \gamma_{t+1} \) are therefore 0 and \( \frac{c^2}{3} \), respectively. In view of the latter, a measure of uncertainty in our model is provided by \( c \), an increase in which corresponds to a mean-preserving spread in the distribution of \( \gamma_{t+1} \).

In the second period of life an agent produces output, \( y_{t+1} \), using his human capital according to

\[
y_{t+1} = Ah_{t+1},
\]

where \( A > 0 \). Given this, the agent realises a final income of \( x_{t+1} \) which determines his final consumption and final utility.

An agent’s consumption depends on what action he took when young. If the agent abstained from human capital investment \((i_t = 0\) in (1.2)), then he consumes all of his realised output, \( A[\beta h_t + b(1 + \gamma_{t+1})] \). If the agent engaged in human capital investment \((i_t = k\) in (1.2)), then he consumes whatever output is left over after paying back lenders in return for his loan of \( k \). Agents borrow from competitive financial intermediaries that have access to a perfectly elastic supply of loanable funds at the world rate of interest, \( r > 0 \). As we have emphasised, there are no credit market imperfections in our model. In particular, lenders do not face any problems of asymmetric information (e.g., observing the incomes of

---

\(^{10}\)Having said this, we note that our approach is well-motivated by our earlier discussion about the various risks associated with human capital investment. The formulation in (2) can be likened to the stochastic human capital production technologies used by Grossman (2008) and Krebs (2003), to whom we referred in our discussion. The main focus of these authors is on the role of human capital risk in influencing growth, though the former establishes this role with reference to initial wealth inequalities. Our own focus is centred primarily on distribution in an environment with initial human capital inequalities.
agents) or contract enforcement (e.g., preventing agents from absconding with loans). Problems that might otherwise arise from these - such as moral hazard, costly state verification and strategic defaulting - are therefore redundant and do not play any role in our analysis. To focus attention even further, we also abstract from any risk of bankruptcy by assuming that agents are always able to repay their loans by assuming that \( A[\beta h_t + B(1 + \gamma_{t+1})] \geq (1 + r)k \) for all values of \( h \) and \( \gamma \). We return to these aspects later when we broaden the context of our analysis to study issues relating to the functioning of financial markets. For now, we note that, since competition between intermediaries drives their profits to zero, the rate of interest charged on loans to agents is simply equal to intermediaries’ own cost of borrowing, \( r \). Given this, then the size of an agent’s loan repayment is \( (1 + r)k \), which leaves \( A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + r)k \) to be consumed. In summary, we may write the consumption profile of agents under alternative choices as

\[
x_{t+1} = \begin{cases} 
A[\beta h_t + b(1 + \gamma_{t+1})] & \text{if } i_t = 0, \\
A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + r)k & \text{if } i_t = k.
\end{cases} \tag{1.4}
\]

Of ultimate concern to each agent is his expected payoff in (1.1), which incorporates his fear of failing to achieve his aspirations. We assume that there is no such anxiety if an agent foregoes human capital investment, which acts as a safety option in the sense of always yielding an income at least equal to the reference level, \( A[\beta h_t + b(1 + \gamma_{t+1})] > x^* \).\textsuperscript{11} The agent’s expected payoff in this case is simply

\[
V_{t|t=0} = A[\beta h_t + b] - x^*. \tag{1.5}
\]

Conversely, investment in human capital means that aspirations may or may not be realised depending on whether \( A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + r)k \geq x^* \) or \( A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + r)k < x^* \). Accordingly, we may deduce a critical value of \( \gamma_{t+1} - \hat{\gamma}_{t+1} \), say - such that aspirations are attained if \( \gamma_{t+1} \geq \hat{\gamma}_{t+1} \) and are unattained if \( \gamma_{t+1} < \hat{\gamma}_{t+1} \). That is,

\[
A[\beta h_t + B(1 + \hat{\gamma}_{t+1})] = (1 + r)k + x^*. \tag{1.6}
\]

\textsuperscript{11}This can be ensured by assuming that \( A[\beta h_0^L + b(1 - c)] > x^* \), where \( h_0^L \) denotes the lowest initial human capital endowment amongst agents.
Given this, we may then compute an agent’s expected payoff from human capital investment as

\[ V_t|_{it=k} = \int_{-c}^{c} \{A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + r)k\} f(\gamma_{t+1}) d\gamma_{t+1} \]  
(1.7)

\[ - \lambda \int_{-c}^{c} f(\gamma_{t+1}) d\gamma_{t+1} - x^* \]

1.3.2 The Decision Problem of Agents

The basic choice that agents confront is whether or not to invest in human capital. Obviously, an agent will choose to invest if his expected payoff from doing so is no less than his expected payoff from not doing so: that is, if \( V_t|_{it=k} \geq V_t|_{it=0} \).

Our subsequent analysis of inequality is essentially concerned with identifying circumstances under which the above condition is satisfied (or not satisfied). Casual observation at this stage suggests a potentially important role for aspirations through the second term on the right-hand-side of (1.7). This term captures an agent’s expected utility loss associated with the possibility of failing to achieve his aspirations. Such an event occurs with probability

\[ P(x_{t+1} < x^*) = P(\gamma_{t+1} < \hat{\gamma}_{t+1}) = \int_{-c}^{\hat{\gamma}_{t+1}} f(\gamma_{t+1}) d\gamma_{t+1} = \frac{\hat{\gamma}_{t+1} + c}{2c}, \]

and the intensity of aversion towards it is measured by \( \lambda \).

1.4 Equilibrium Outcomes

The key factor that dictates agents’ behaviour towards human capital investment is \( \hat{\gamma}_{t+1} \) which governs the probability of failing to achieve aspirations. Solving for this variable is our starting point for characterising the equilibrium of the model. Having done this, we then proceed to deduce other equilibrium properties that form the basis of our subsequent analysis of inequality.

From (1.6), we have

\[ \hat{\gamma}_{t+1} = \frac{x^* + (1 + r)k - A(\beta h_t + B)}{AB} \equiv \gamma(h_t). \]  
(1.8)

Evidently, \( \gamma(h_t) < 0 \) which means that the probability of failing to achieve as-
pirations is higher the lower is the amount of inherited human capital. *Ceteris paribus*, a lower human capital inheritance implies a lower production of output and a greater probability of falling short of target consumption.

Given the above, we may now deduce the equilibrium behaviour of agents. We do so by recalling the expression in (1.7) which gives an agent’s expected payoff from investing in human capital. Using (1.8), we may re-write this expression as

\[ V_{it} = A(\beta h_t + B) - (1 + r)k - \lambda \left( \frac{\gamma(h_t) + c}{2c} \right). \]  

(1.9)

As stated earlier, an agent will choose to invest in human capital if doing so yields an expected payoff that is no less than the expected payoff from not investing - that is, if \( V_{it} \geq V_{it} = 0 \). Using (1.5) and (1.9), this condition can be written as \( A(\beta h_t + B) - (1 + r)k \geq \lambda \left( \frac{\gamma(h_t) + c}{2c} \right) \). We assume that the left-hand-side of this expression (which is the expected difference in income between investing and not investing in human capital) is strictly positive in order to avoid the degenerate case in which no agent ever invests. The right-hand-side (which is the expected disutility from not achieving aspirations) is deduced to be an increasing function of \( c \), a decreasing function of \( h_t \) and an increasing function of \( \lambda \).\(^{12}\) Accordingly, the condition is less likely to be satisfied the greater is the degree of uncertainty, the lower is the amount of inherited human capital and the greater is the influence of aspirations. The crucial role played by aspirations is evident. Both a greater degree of uncertainty and a lower endowment of human capital make it more likely that agents will fail in realising their target consumption. Their aversion to this may incline them not to borrow to finance human capital investment when the probability of failure is high. Naturally, for any given probability of failure, the disincentive to borrow is stronger the greater is the influence of agents’ loss aversion. In the absence of any such influence (i.e., when \( \lambda = 0 \)) uncertainty plays no role as agents become purely risk-neutral.

\(^{12}\)Note that the effect of \( c \) is positive because \( \gamma_{t+1} < 0 \) in (8). This follows from the fact that \( A(\beta h_t + B) - (1 + r)k - x^* > 0 \) must hold if human capital investment is ever to be chosen (otherwise, the expected income from this investment would be less than target income).
1.5 Distribution and Inequality

The economy starts off with some initial distribution of human capital that accounts for agent heterogeneity and income inequality. Our principal concern is to study the role of aspiration-induced loss aversion in governing how the distribution changes over time and the extent to which initial inequalities may vanish or persist. We proceed to do this by determining the lineage dynamics for each dynasty that show the transition of human capital from one generation to the next. Then, for any given initial distribution of human capital, we may use this information to infer the dynamic processes operating at the aggregate level and thereby deduce possible long-run distribution outcomes.

From our previous analysis, an agent will choose to invest in human capital accumulation if
\[ A(B - b) - (1 + r)k \geq \lambda \left[ \frac{\gamma(h^t) + c}{2c} \right]. \]
When holding with equality, this condition can be used to determine a critical inheritance of human capital \( \hat{h} \), say - which is necessary for an agent to make such a choice. That is,
\[ A(B - b) - (1 + r)k = \lambda \left[ \frac{\gamma(\hat{h}) + c}{2c} \right]. \] (1.10)

Since \( \gamma_h(h) < 0 \), this expression implies that only those agents for whom \( h_t \geq \hat{h} \) will willingly borrow to finance human capital investment; all other agents for whom \( h_t < \hat{h} \) will choose not to do this. Evidently, the precise value of the human capital threshold depends on certain key parameters. In particular, we may write \( \hat{h} = h(c, \lambda) \), where \( h_c(c, \lambda) > 0 \) and \( h_\lambda(c, \lambda) > 0 \). Thus agents face a higher threshold the greater is the degree of uncertainty and/or the stronger is the influence of aspirations. The critical role played by the latter is again self-evident: in the absence of it (\( \lambda = 0 \)), agents’ inheritance of human capital would be irrelevant as all of them would invest in human capital accumulation since \( A(B - b) - (1 + r)k > 0 \).

Given the above, together with (1.2), we may conclude that the intergenerational evolution of human capital for an individual dynasty satisfies
\[ h_{t+1} = \begin{cases} 
\beta h_t + b(1 + \gamma_{t+1}) & \text{if } h_t < \hat{h}, \\
\beta h_t + B(1 + \gamma_{t+1}) & \text{if } h_t \geq \hat{h}.
\end{cases} \] (1.11)
These lineage transition equations are portrayed in Figure 1.1. Each of them corresponds to a stable stochastic difference equation which is bounded according to the bounds on $\gamma_{t+1}$ (i.e., $\gamma_{t+1} \in (-c, c)$). The intersections with the $45^\circ$ line are given by the stationary points associated with these bounds; that is,

$$h^* = \frac{b(1 \pm c)}{1 - \beta}; \quad h^{**} = \frac{B(1 \pm c)}{1 - \beta}.$$  \hspace{1cm} (1.12)

The transition equations are drawn under the parameter restriction $b(1 + c) < (1 - \beta)\hat{h} < B(1 - c)$. The restriction rules out the possibility that all agents automatically end up either investing or not investing in human capital. The restriction also means that any lineage that chooses to invest at some point will never alter its choice subsequently. We do this for illustrative purposes and to present our analysis in line with the rest of the literature (e.g. Galor and Zeira, 1993).

The long-run distribution of human capital in our economy is straightforward to characterise. The only investors in human capital accumulation are those agents who are well-endowed with human capital to begin with; these are agents for whom $h_0 \geq \hat{h}$, implying convergence to some high steady state equilibrium. All other agents who start off with relatively low human capital endowments remain forever as non-investors; these are agents for whom $h_0 < \hat{h}$, implying convergence to some low steady state equilibrium. In terms of income distribution, there is always the possibility that some generation of investors will find themselves worse off \textit{ex post} than if they had not invested. Yet this does not affect the dynamics of human capital distribution and such agents are strictly better off \textit{ex ante} in terms of their higher expected income from investing. The same remarks can be made about the actual and expected payoffs of agents.

As indicated already, the key factor in explaining our results is the existence of loss aversion caused by aspirations. The extent to which this impacts on an agent’s behaviour depends on his likelihood of failing to realise his aspirations which, in turn, depends on his inherited human capital. The lower is this inheritance, the

\footnote{For example, if the parameter restrictions do not hold it might be the case that some offspring of individuals sufficiently close to $\hat{h}$ might invest in human capital even if their parent does not invest.}
higher is the probability of failure and the stronger is the influence of loss aversion associated with this. For agents with \( h_t < \hat{h} \), the disutility suffered from under-achievement is sufficiently high as to deter human capital investment. The number of agents for which this is true depends on both the extent of uncertainty and the strength of aspirational influence. This follows from the fact that, as noted above, \( \hat{h} \) is an increasing function of both \( c \) and \( \lambda \). Thus, a higher value of either of these induces more agents to forego human capital investment.

### 1.6 Further Discussion

The foregoing analysis establishes our main results. In what follows we make several additional observations about how these results relate to other relevant research and how they survive under some extensions of the model. As regards the former, two issues attract our attention - the role of capital market imperfections in explaining inequality and the effect of uncertainty on aggregate economic activity. As regards the latter, three modifications are considered - the generalisation of aspirational preferences, the introduction of initial wealth endowments and the incorporation of bankruptcy considerations.

#### 1.6.1 Some Related Research

**1.6.1.1 Inequality and Financial Markets**

The persistence of inequality in our model is reflected in the existence of multiple, history-dependent long-run equilibria associated with threshold effects that explain how limiting outcomes depend on initial conditions. These are the key features of models of income distribution based on credit market imperfections (e.g., Banerjee and Newman 1993; Galor and Zeira 1993). The significant and novel aspect of our analysis is that it abstracts from any such imperfections and focuses purely on individuals’ aspiration-based (loss averse) preferences as a means of accounting for inequality. With this in mind, it is interesting to note that our results are observationally equivalent to those obtained under a reformulation of our model in which the assumption of aspirations is replaced by an assumption of credit market imperfections. We demonstrate this as follows.
Suppose that aspirations are absent so that \( x^* = \lambda = 0 \). Instead, assume that borrowers have an opportunity of strategically defaulting by taking flight and avoiding loan repayments. Lenders spend resources on tracking down borrowers and retrieving some part of their incomes. This is now a model of imperfect enforcement of loan contracts, where the cost of enforcement, denoted \( e \), and the non-retrievable fraction of borrowers’ income, denoted \( 1 - \delta \in (0, 1) \), provide measures of financial market imperfection.

As we shall see, the existence of credit market frictions drives a wedge between the lending and borrowing rates of financial intermediaries. Denoting the former by \( R_{t+1} \), the size of an agent’s loan repayment is \( (1 + R_{t+1})k \), implying an income of \( A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k \) if the repayment is actually made. If not (i.e., if defaulting occurs), then the agent earns \( (1 - \delta)A[\beta h_t + B(1 + \gamma_{t+1})] \). Evidently, the agent will choose not to default provided that

\[
\delta A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k \geq (1 - \delta)A[\beta h_t + B(1 + \gamma_{t+1})].
\]

From this, we may determine a critical value of \( \gamma_{t+1} \) - denoted \( \bar{\gamma}_{t+1} \) - above (below) which loan contracts are honoured (reneged upon). That is,

\[
\delta A[\beta h_t + B(1 + \bar{\gamma}_{t+1})] = (1 + R_{t+1})k.
\]

Naturally, \( \bar{\gamma}_{t+1} \) is increasing in \( R_{t+1} \) and decreasing in \( h_t \): \textit{ceteris paribus}, the higher is the interest rate on loans, or the lower is the inherited amount of human capital, the more productive must be a borrower if he is to respect his debt obligations.

The probability that he will choose otherwise - that is, the probability of defaulting - is given by

\[
P(\gamma_{t+1} < \bar{\gamma}_{t+1}) = \int_{-c}^{\bar{\gamma}_{t+1}} f(\gamma_{t+1})d\gamma_{t+1} = \frac{\bar{\gamma}_{t+1} + c}{2c}.
\]

If defaulting does not occur (i.e., if \( \gamma_{t+1} \geq \bar{\gamma}_{t+1} \)), then intermediaries receive the full repayment of loans, \( (1 + R_{t+1})k \). If defaulting does occur (i.e., if \( \gamma_{t+1} < \bar{\gamma}_{t+1} \)), then intermediaries retrieve \( \delta A[\beta h_t + B(1 + \gamma_{t+1})] \) amount of output at a cost of \( e \). As before, competition between intermediaries drives their expected profits to zero. Since the cost of borrowing is \( (1 + r)k \), this break-even condition is given by

\[
(1 + r)k = \int_{-c}^{\bar{\gamma}_{t+1}} (1 + R_{t+1})k f(\gamma_{t+1})d\gamma_{t+1} + \int_{-c}^{\bar{\gamma}_{t+1}} \{\delta A[\beta h_t + B(1 + \gamma_{t+1})] - e\} f(\gamma_{t+1})d\gamma_{t+1}.
\]
For any given $\gamma_{t+1}$, this expression determines the contractual interest rate on loans, $R_{t+1}$. We may write the expression in a different way by combining it with (1.13) to obtain

\[
(1 + R_{t+1})k - (1 + r)k = \int_{-c}^{\gamma_{t+1}} \delta A[\beta h_t + B(1 + \gamma_{t+1})]f(\gamma_{t+1})d\gamma_{t+1} \\
- \int_{-c}^{\gamma_{t+1}} \{\delta A[\beta h_t + B(1 + \gamma_{t+1})] - e\}f(\gamma_{t+1})d\gamma_{t+1}.
\]

(1.15)

This shows the interest rate spread between lending and borrowing. The magnitude of the spread depends on how much a lender expects to lose when a borrower defaults on his loan. Essentially, (1.15) states that the contractual interest rate is set as a simple mark-up over intermediaries’ cost of borrowing, where the size of mark-up is equal to the expected net income lost due to defaulting. This mark-up rule may be simplified to

\[
(1 + R_{t+1})k = (1 + r)k + \frac{\delta AB(\gamma_{t+1} + c)^2}{4c} + e \left(\frac{\gamma_{t+1} + c}{2c}\right).
\]

(1.16)

As above, there is a positive relationship between $R_{t+1}$ and $\gamma_{t+1}$: ceteris paribus, intermediaries set a higher contractual interest rate the more likely it is that defaulting will occur.

The expressions in (1.13) and (1.16) define a simultaneous equations system in $R_{t+1}$ and $\gamma_{t+1}$. Under the parameter restriction $e \leq \delta A(\beta h_t + B) - (1+r)k \leq \delta ABc$,

\[14\]To be sure, observe from (13) that the first integral term on the right-hand-side of (15) is equal to $\int_{-c}^{\gamma+1} (1 + R_{t+1})k f(\gamma_{t+1})d\gamma_{t+1}$ which measures the expected amount of non-repayment when defaulting occurs. Conversely, the second integral term on the right-hand-side of (15) gives the expected amount of income that is seized from a defaulter, net of enforcement costs.
there exists a unique feasible solution to this system, as given by\textsuperscript{15}

\[
\gamma_{t+1} = c - \frac{e}{\delta AB} - \frac{\sqrt{4\delta ABC[\delta A(\beta h_t + B) - (1 + r)k - e] + e^2}}{\delta AB} \equiv \gamma(c, e, h_t), \quad (1.17)
\]

\[
R_{t+1} = r + \delta AB\left[\frac{(\gamma(c, e, h_t) + e)^2}{4ck} + e\left[\frac{\gamma(c, e, h_t) + e}{2c}\right]\right] \equiv R(c, e, h_t). \quad (1.18)
\]

The result in (1.17) implies that \(\gamma_c(c, e, h_t) > 0\), \(\gamma_e(c, e, h_t) > 0\) and \(\gamma_h(c, e, h_t) > 0\), and the result in (1.18) reveals similarly that \(R_c(c, e, h_t) > 0\), \(R_e(c, e, h_t) > 0\) and \(R_h(c, e, h_t) < 0\).\textsuperscript{16} In words, the greater is the degree of uncertainty, the larger is the cost of contract enforcement and the lower is the amount of inherited human capital, the higher is the probability of defaulting on loans and the higher is the contractual interest rate on loans. The effects of uncertainty are due to the fact that the loan repayment is a non-linear (specifically, concave) function of \(\gamma_{t+1}\).

To be sure, recall that the repayment is \(\delta A[\beta h_t + B(1 + \gamma_{t+1})]\) if \(\gamma_{t+1} < \gamma_{t+1}\), but \((1 + R_{t+1})k\) if \(\gamma_{t+1} \geq \gamma_{t+1}\). The expected repayment is therefore reduced by a mean-preserving spread in the distribution of \(\gamma_{t+1}\). Intermediaries compensate for this by charging a higher interest rate on loans which increases the likelihood that defaulting will occur. The effects of enforcement costs and inherited human capital operate in a similar way. An increase in \(e\) or a decrease in \(h_t\) reduces intermediaries’ expected net returns which raises the contractual interest rate and makes defaulting more likely.

Having established the above, we may now turn our attention to the equilibrium behaviour of agents. An agent’s expected payoff from investing in human capital

\textsuperscript{15}Details of the derivations can be found in an Appendix.

\textsuperscript{16}Verification of these results is again contained in the Appendix. Note that one does not need to assume that financial intermediaries are able to observe human capital directly. As shown in (17) and (18), \(\gamma_{t+1}\) and \(R_{t+1}\) are functions of \(Ah_t\), which is the output produced by parents. One needs only to assume that this is observable (in which case, of course, \(h_t\) can be trivially inferred anyway).
is

\[ V_t|_{i_t=k} = \int_{\gamma_{t+1}}^c \{ A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k\} f(\gamma_{t+1})d\gamma_{t+1} \]
\[ + \int_{-c}^{\gamma_{t+1}} (1 - \delta) A[\beta h_t + B(1 + \gamma_{t+1})] f(\gamma_{t+1}) d\gamma_{t+1}. \]  \hspace{1cm} (1.19)

Using (1.14) and (1.17), we may write this as

\[ V_t|_{i_t=k} = \int_{-c}^c A[\beta h_t + B(1 + \gamma_{t+1})] f(\gamma_{t+1}) d\gamma_{t+1} - (1 + r)k \]
\[ - e \int_{-c}^{\gamma_{t+1}} f(\gamma_{t+1}) d\gamma_{t+1} \]
\[ = A(\beta h_t + B) - (1 + r)k - e \left[ \frac{\gamma(c,e,h_t) + c}{2c} \right]. \]  \hspace{1cm} (1.20)

The condition for human capital investment to be chosen is \( V_t|_{i_t=k} \geq V_t|_{i_t=0} \) which, from (1.20) and (1.5) (with \( x^* = 0 \)), implies \( A(B - b) - (1 + r)k \geq e \left[ \frac{\gamma(c,e,h_t) + c}{2c} \right]. \)

As before, this condition is less likely to be satisfied the greater is the degree of uncertainty, \( c \), and/or the lower is the amount of inherited human capital, \( h_t \), each of which raises the right-hand-side term. Additionally, this is true for a larger cost of contract enforcement, \( e \). When holding with equality, the condition implies a critical level of \( h_t \) - denoted \( \bar{h} \) - above (below) which human capital investment is chosen (declined). That is,

\[ A(B - b) - (1 + r)k = e \left[ \frac{\gamma(c,e,\bar{h}) + c}{2c} \right]. \]  \hspace{1cm} (1.21)

In turn, this expression yields \( \bar{h} = h(c,e) \), where \( h_c(c,e) > 0 \) and \( h_e(c,e) > 0 \). Thus agents face a higher human capital threshold the greater is the degree of uncertainty (as was present before) and/or the higher is the cost of contract enforcement (as was absent before).

The observational equivalence between these results and those established earlier is self-evident (compare, in particular, (1.21) with (1.10)). Essentially, \( e \) substitutes for \( \lambda \) in a way that makes the effects of credit market imperfections very similar to the effects of aspirations (or loss aversion). This finding echoes the
sentiments of Carroll (2001) who argues that, in many instances of uncertainty, the existence of a precautionary savings motive can generate behaviour that is virtually indistinguishable from the behaviour that emerges from the existence of liquidity constraints. The same is true with respect to loss aversion and credit market frictions. The outcomes may be similar in the sense that some individuals forego borrowing opportunities, but the reasons are fundamentally different: in the case of loss aversion, there is a self-imposed reluctance towards borrowing; in the case of credit market frictions, there is an externally-imposed obstacle to borrowing.

1.6.1.2 Uncertainty and Macroeconomic Performance

The distributional effects of uncertainty in our model have implications for aggregate productive activity in the economy. This follows from the fact that the productivity of agents who invest in human capital is different from the productivity of agents who refrain from such investment. Since the number of investors and non-investors is determined by the degree of uncertainty, then so too is the total level of output. In this way, our analysis bears on other research that seeks to explore the link between uncertainty and macroeconomic performance.\(^{17}\)

Let \( g_t(h_t) \) be the probability density function of human capital at time \( t \). Suppose that human capital is initially distributed over the interval \((h, \bar{h})\) such that \( \int_{h}^{\bar{h}} g_t(h_t) dh_t = 1 \) (corresponding to the unit mass of agents). In each period there is the same population of non-investors in human capital, \( \int_{h}^{\bar{h}} g_t(h_t) dh_t \), and the same population of investors in human capital, \( \int_{h}^{\bar{h}} g_t(h_t) dh_t \). From (1.3) and (1.11), the expected output of a non-investor is \( A(\beta h_t + b) \), whilst the expected output of an investor is \( A(\beta h_t + B) \). Accordingly, the expected total (or average) level of output

\(^{17}\)For example, there is a large body of research that shows how uncertainty (or volatility) can influence long-term growth (either positively or negatively) through various factors (e.g., Aghion and Saint-Paul 1998; Blackburn and Varvarigos 2008; de Hek 1999; Jones et al. 2005; Martin and Rogers 2000). Whilst we do not consider long-term growth, our analysis identifies another factor that can create a link between uncertainty and macroeconomic performance - namely, the impact of uncertainty on distribution outcomes.
in the economy is given by

\[ Y_{t+1} = \int_{\hat{h}}^{\Gamma} A(\beta_h t + b)g_t(h_t)dh_t + \int_{\hat{h}}^{\Gamma} A(\beta_h t + B)g_t(h_t)dh_t \]

\[ = A\beta \int_{\hat{h}}^{\Gamma} h_t g_t(h_t)dh_t + A[b \int_{\hat{h}}^{\Gamma} g_t(h_t)dh_t + B \int_{\hat{h}}^{\Gamma} g_t(h_t)dh_t]. \]  

(1.22)

Recall that \( \hat{h} = h(c, \lambda) \), where \( h_c(c, \lambda) > 0 \) and \( h_{\lambda}(c, \lambda) > 0 \). Since \( B > b \), it follows that, for any given \( g_t(h_t) \), an increase in \( c \), which increases \( \hat{h} \), implies a decrease in the second right-hand-side term of (1.22), thus causing a decrease in \( Y_{t+1} \). In words, a greater degree of uncertainty is associated with a lower average level of output as fewer agents choose to invest in human capital. It may also be noted that an increase in \( \lambda \) has the same effect, meaning that a stronger influence of aspirations (or loss aversion) reduces macroeconomic performance.

1.6.2 Some Extensions

1.6.2.1 Generalising Aspirations

Our treatment of aspirations in moulding agents’ preferences led to the objective function in (1.1). Whilst this allows for any arbitrary reference level of income, \( x^* \), it is based on the assumption (made for convenience) that agents are affected by their aspirations only because of the disappointment of not achieving them, not because of the satisfaction from fulfilling them. In what follows we modify this feature.

Let \( P(x_{t+1} \geq x^*) \) denote the probability of success in attaining aspirations. An agent’s objective function is now given by

\[ V_t = E(x_{t+1} - x^*) + \mu P(x_{t+1} \geq x^*) - \lambda P(x_{t+1} < x^*). \]

(1.23)

The expression in (1.6) for determining \( \hat{\gamma}_{t+1} \) continues to apply, and the agent’s
expected payoff from human capital investment is given by

\[
V_{t|i_t=k} = \int_{-c}^{c} \left\{ A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k \right\} f(\gamma_{t+1}) d\gamma_{t+1} \\
+ \mu \int_{-c}^{c} f(\gamma_{t+1}) d\gamma_{t+1} - \lambda \int_{-c}^{\gamma_{t+1}} f(\gamma_{t+1}) d\gamma_{t+1} - x^*.
\]

(1.24)

\[
= A(\beta h_t + B) - (1 + r)k + \mu - (\mu + \lambda) \left[ \frac{\gamma(h_t) + c}{2c} \right].
\]

(1.25)

Since the expected payoff to a non-investor is \( V_{t|i_t=0} = A[\beta h_t + b] - x^* + \mu \), the condition for investment to be chosen is \( A(B - b) - (1 + r)k \geq (\mu + \lambda) \left[ \frac{\gamma(h_t) + c}{2c} \right] \), from which we may deduce a critical level of human capital, \( \hat{h} \), above which an agent invests and below which an agent does not invest: that is,

\[
A(B - b) - (1 + r)k = (\mu + \lambda) \left[ \frac{\gamma(h_t) + c}{2c} \right].
\]

(1.26)

Clearly, the results of our main analysis (which can be recovered by setting \( \mu = 0 \)) are not significantly altered by the generalisation of preferences reflected in (1.23).

### 1.6.2.2 Introducing Wealth

Intergenerational linkages in our model take place through the serendipitous intra-family transfers of human capital. In other models these linkages are the result of altruistic bequests of wealth which may influence opportunities for borrowing in the presence of capital market imperfections. We outline how our model may be extended to include a bequest motive.

Suppose that agents derive utility from their own consumption, \( c_{t+1} \), and the bequests they leave to their offspring, \( q_{t+1} \). Ex post (i.e., after incomes have been realised), their aim is to maximise \( u(c_{t+1}, q_{t+1}) = \Phi c_{t+1}^{\phi} q_{t+1}^{1-\phi} \) (\( \phi \in (0, 1) \), \( \Phi = [\phi(1 - \phi)^{1-\phi}]^{-1} \)) subject to \( c_{t+1} + q_{t+1} = z_{t+1} \), where \( z_{t+1} \) is final income. They do this by choosing \( c_{t+1} = \phi z_{t+1} \) and \( q_{t+1} = (1 - \phi)z_{t+1} \), implying an indirect utility of \( U(z_{t+1}) = z_{t+1} \). An agent evaluates this with reference to some threshold outcome, \( z^* \), that he aspires to attain and that he fails to do so with probability \( P(z_{t+1} < z^*) \). Ex ante (i.e., before any decisions have been made), the agent’s
value function is
\[ V_t = E(z_{t+1} - z^*) - \lambda P(z_{t+1} < z^*). \] (1.27)

Suppose that parents invest bequests on behalf of their children (e.g., in a trust fund) who become entitled to their inheritance when old. This means that agents are unable to use bequests to finance human capital investment when young. The expression for final income is therefore given by \( z_{t+1} = x_{t+1} + (1 + r)q_t \), where \( x_{t+1} \) is determined according to (1.4).

Assume that an agent’s threshold income is some fixed value above his inheritance, \( z^* = (1 + r)q_t + x^* \). Thus agents aspire to be better off than they would be from relying solely on the altruism of their parents. Given this, then (1.27) can be re-written as (1.1). All of our original results can be re-established, with straightforward implications for the dynamics of bequests, \( q_{t+1} = (1 - \phi)[x_{t+1} + (1 + r)q_t] \).

### 1.6.2.3 Incorporating Bankruptcy

A notable feature maintained throughout our analysis has been the absence of any bankruptcy considerations. This has been useful in terms of focusing attention and simplifying the analysis. Nevertheless, it is instructive to see how our model of aspirations can be adapted to study issues of bankruptcy, issues that figure prominently in many areas of macroeconomics. In doing this, we are able to demonstrate further the broader context of our analysis and its potential for further application.

We allow for the possibility that agents who invest in human capital may be unable to repay their loans. Under such circumstances, agents declare bankruptcy and intermediaries seize whatever output is produced. Based on this, one might plausibly consider the avoidance of bankruptcy to be a fairly natural candidate for aspirations. Adopting this criteria, an agent succeeds or fails in achieving his aspirations according to whether \( x_{t+1} > 0 \) or \( x_{t+1} = 0 \). His value function is then given by
\[ V_t = E(x_{t+1}) - \lambda P(x_{t+1} = 0). \] (1.28)

Obviously, bankruptcy is not an issue if an agent abstains from human capital investment, in which case his consumption is \( A[\beta h_t + b(1 + \gamma_{t+1})] \). If the agent invests in human capital, then his consumption is determined as follows. The
possibility of bankruptcy means that intermediaries charge an interest rate on loans, \( R_{t+1} \), which is different from their interest cost of borrowing, \( r \). The size of an agent’s loan repayment is therefore given by \( (1 + R_{t+1})k \). The agent is able to make this repayment if \( A[\beta h_t + B(1 + \gamma_{t+1})] \geq (1 + R_{t+1})k \), in which case his consumption is \( A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k \); otherwise, if \( A[\beta h_t + B(1 + \gamma_{t+1})] < (1 + R_{t+1})k \), the agent is bankrupt and his consumption is zero as lenders appropriate whatever output is produced. These observations lead us to define a critical value of \( \gamma_{t+1} - \tilde{\gamma}_{t+1} \), say - such that bankruptcy is avoided if \( \gamma_{t+1} \geq \tilde{\gamma}_{t+1} \) and is unavoided if \( \gamma_{t+1} < \tilde{\gamma}_{t+1} \). That is,

\[
A[\beta h_t + B(1 + \gamma_{t+1})] = (1 + R_{t+1})k.
\]

(1.29)

Naturally, \( \tilde{\gamma}_{t+1} \) is increasing in \( R_{t+1} \) and decreasing in \( h_t \): *ceteris paribus*, the higher is the interest rate on loans, or the lower is the inherited amount of human capital, the more productive must be a borrower if he is to be able to make his loan repayment. The probability that he is unable to do this - that is, the probability of bankruptcy - is given by \( P(x_{t+1} = 0) = P(\gamma_{t+1} < \tilde{\gamma}_{t+1}) = \int_{-c}^{\tilde{\gamma}_{t+1}} f(\gamma_{t+1})d\gamma_{t+1} = \frac{\tilde{\gamma}_{t+1} + c}{2c} \). This is also the probability of failing to achieve aspirations.

If bankruptcy is not declared (i.e., if \( \gamma_{t+1} \geq \tilde{\gamma}_{t+1} \)), then intermediaries are paid back in full, earning a return of \( (1 + R_{t+1})k \). If bankruptcy is declared (i.e., if \( \gamma_{t+1} < \tilde{\gamma}_{t+1} \)), then agents’ proclamations are verified and intermediaries retrieve all of the output produced, \( A[\beta h_t + B(1 + \gamma_{t+1})] \). In keeping with our original analysis, we abstract from any capital market imperfections that might arise in the case of bankruptcy. In particular, we assume that intermediaries can verify bankruptcy claims at zero cost, or can directly observe the output of agents. We shall return to this later. For now, we note that the zero profit condition on intermediaries implies

\[
(1 + r)k = \int_{-c}^{\tilde{\gamma}_{t+1}} (1 + R_{t+1})kf(\gamma_{t+1})d\gamma_{t+1} + \int_{-c}^{\tilde{\gamma}_{t+1}} A[\beta h_t + B(1 + \gamma_{t+1})]f(\gamma_{t+1})d\gamma_{t+1}.
\]

(1.30)

In conjunction with (1.29), this gives the following expression for the interest rate
spread between borrowing and lending:

\[
(1 + R_{t+1})k - (1 + r)k = \int_{-c}^{\tilde{\gamma}_{t+1}} A[\beta h_t + B(1 + \tilde{\gamma}_{t+1})]f(\gamma_{t+1})d\gamma_{t+1} - \int_{-c}^{\tilde{\gamma}_{t+1}} A[\beta h_t + B(1 + \gamma_{t+1})]f(\gamma_{t+1})d\gamma_{t+1}.
\] (1.31)

The size of spread depends on how much a lender expects to lose when a borrower claims that he is bankrupt and fails to repay his loan. Similar to our model of strategic defaulting, (1.31) determines the contractual interest rate as a mark-up over intermediaries’ cost of borrowing, where the size of mark-up is equal to the expected net income lost due to bankruptcy. This mark-up rule may be simplified to

\[
(1 + R_{t+1})k = (1 + r)k + \frac{AB(\tilde{\gamma}_{t+1} + c)^2}{4c}.
\] (1.32)

As before, there is a positive relationship between \(R_{t+1}\) and \(\tilde{\gamma}_{t+1}\): ceteris paribus, intermediaries set a higher contractual interest rate the more likely it is that bankruptcy will be declared.

The expressions in (1.29) and (1.32) define a simultaneous equations system in \(R_{t+1}\) and \(\tilde{\gamma}_{t+1}\). Under the parameter restriction \(0 \leq A(\beta h_t + B) - (1 + r)k \leq ABc\), there exists a unique feasible solution to this system, as given by

\[
\tilde{\gamma}_{t+1} = c - 2\sqrt{\frac{c[A(\beta h_t + B) - (1 + r)k]}{AB}} \equiv \gamma(c, h_t),
\] (1.33)

\[
R_{t+1} = r + \frac{AB[\gamma(c, h_t) + c]^2}{4ck} \equiv R(c, h_t).
\] (1.34)

The result in (1.33) implies that \(\gamma_c(c, h_t) > 0\) and \(\gamma_h(c, h_t) < 0\), and the result in (1.34) reveals similarly that \(R_c(c, h_t) > 0\) and \(R_h(c, h_t) < 0\). Thus both the probability of bankruptcy and the rate of interest on loans are increased by an increase in the degree of uncertainty and a decrease in the amount of inherited

---

\(^{18}\)From (29), the first integral term on the right-hand-side of (31) is equal to \(\int_{-c}^{\tilde{\gamma}_{t+1}} (1 + R_{t+1})kf(\gamma_{t+1})d\gamma_{t+1}\) which is the expected amount of non-repayment when bankruptcy is declared. Conversely, the second integral term on the right-hand-side of (31) is the expected amount of income that is claimed in the case of bankruptcy.

\(^{19}\)The solution can be obtained directly from (17) and (18) by setting \(\delta = 1\) and \(e = 0\).
human capital.\footnote{These results, and the intuition for them, are the same as those obtained in our model of strategic defaulting.}

Given the above, we deduce the equilibrium behaviour of agents as follows. An agent’s expected payoff from investing in human capital is

$$V_{t|i_t=k} = \int_{\gamma_{t+1}}^{c} \{ A[\beta h_t + B(1 + \gamma_{t+1})] - (1 + R_{t+1})k \} f(\gamma_{t+1})d\gamma_{t+1}$$

$$- \lambda \int_{-c}^{\gamma_{t+1}} f(\gamma_{t+1})d\gamma_{t+1}$$

(1.35)

Using (1.30) and (1.33), we may write this as

$$V_{t|i_t=k} = \int_{-c}^{c} A[\beta h_t + B(1 + \gamma_{t+1})] f(\gamma_{t+1})d\gamma_{t+1} -(1+r)k$$

$$- \lambda \int_{-c}^{\gamma_{t+1}} f(\gamma_{t+1})d\gamma_{t+1}$$

$$= A(\beta h_t + B) - (1 + r)k - \lambda \left[ \frac{\gamma(c, h_t) + c}{2c} \right].$$

(1.36)

The condition for human capital investment to be chosen is $V_{t|i_t=k} \geq V_{t|i_t=0}$ which, from (1.36) and (1.5), implies $A(B - b) - (1 + r)k \geq \lambda \left[ \frac{\gamma(h_t, c) + c}{2c} \right]$. This is very similar to the condition obtained in our original analysis and displays the same properties. Specifically, the condition is less likely to be satisfied the greater is the degree of uncertainty, $c$, the lower is the amount of inherited human capital, $h_t$, and the greater is the influence of aspirations, $\lambda$. Likewise, the condition can be used to infer a critical level of human capital, $\bar{h}$, above (below) which human capital investment is chosen (declined). That is,

$$(B - b) - (1 + r)k = \lambda \left[ \frac{\gamma(\bar{h}, c) + c}{2c} \right].$$

(1.37)

In turn, this expression yields $\bar{h} = h(c, \lambda)$, where $h(c, \lambda) > 0$ and $h(\lambda, c, \lambda) > 0$. The intuition for these results is exactly the same as before: agents who are relatively poor in terms of their human capital endowments may choose to forego risky, but potentially profitable, human capital investment because of their relatively strong
aversion to failing in their aspirations. The present analysis provides a specific application of this, where loss aversion means bankruptcy aversion and where the aspiration to succeed means the aspiration to avoid bankruptcy.

A final observation relates to our earlier demonstration of how behaviour under aspirational influences can look almost identical to behaviour under capital market imperfections. The same result can be established in the present context and with a different type of imperfection to the one considered previously. Assuming that aspirations are absent, suppose that borrowers and lenders face an ex post informational asymmetry such that only the former can directly observe the realisation of $\gamma_{t+1}$, whereas the latter must incur expenditures to make this observation. Such expenditures are incurred if a borrower claims bankruptcy since the claim needs to be verified in order to prevent a borrower from falsely declaring that he is unable to repay his loan. This is now a model of costly state verification, where the cost of verification provides a measure of credit market imperfections (e.g., Diamond 1984; Gale and Hellwig 1985; Townsend 1979). The formal structure of the model is readily obtained from our earlier model of strategic defaulting by setting $\delta = 1$ and reinterpreting $e$ as a verification (rather than enforcement) cost. The end result is the expression in (1.26) which again bears a striking resemblance to (1.37).

1.7 Conclusion

This paper has sought to make a theoretical contribution to the literature on inequality and income distribution. Its approach has been to focus on the structure of preferences, rather than the functioning of markets, as a way of explaining the diverse behaviour and diverse fortunes of individuals who face uncertainty. This offers a new perspective on why some individuals do not pursue potentially wealth-enhancing opportunities: the reason is not that they are prevented from doing so, but rather that it is not in their interests to do so. For such agents, the loss that may be incurred on a risky venture is simply too great to make the venture attractive, even if it offers the prospect of high rewards. For other agents, the same loss may be of much less concern so that the venture is taken on. These cohorts of individuals are the less wealthy and the more wealthy members of the
population. The former may stand to gain relatively more if an investment goes well, but they also risk losing relatively more if the investment goes wrong. Under the influence of aspiration-induced loss aversion, it is the disliking of losses, rather than the liking of gains, that matters most to individuals, which is why the less wealthy of them may abstain from opportunities that could make them better off.

Our analysis raises important questions about the appropriateness and effectiveness of strategies aimed at redistribution. If inequality was the result of market imperfections, then the natural prescription for reducing inequality would be the attenuation of these imperfections. In the case of financial markets, for example, improvements in monitoring and enforcement would presumably make borrowing more accessible to greater numbers of individuals who might otherwise be denied loans. But if inequality was rooted in the deep structure of preferences, it is much less obvious what options are available and feasible. And if this source of inequality was mistaken for another (a possibility that we alluded to), then one may end up with well-meaning strategies that are ineffectual (and perhaps even worse if they are costly to implement). One possible approach suggested specifically by our analysis is the enhancement of human capital accumulation (e.g., through better quality public education and health programmes) that may push individuals over the human capital threshold by making their aspirations more attainable. In a broader context, to the extent that the poor wealth status of some individuals may make them unable (because of imperfections) or unwilling (because of preferences) to try to improve their status, lump-sum redistributions from the rich to the poor may offer the most straightforward means of reducing inequality. Further exploration of these issues is an avenue worth pursuing.
References


44


Appendix

The results in (1.17) and (1.18) are derived as follows. Combining (1.13) and (1.16) yields the quadratic equation

\[
0 = \delta AB \gamma_{t+1}^2 - 2(\delta AB c - e)\gamma_{t+1} - \{4c[\delta A(\beta h_t + B) - (1 + r)k] - \delta AB c^2 - 2ec\}. \tag{A1}
\]

Hence

\[
\gamma_{t+1} = c - \frac{e}{\delta AB} \pm \frac{\sqrt{4\delta AB c[\delta A(\beta h_t + B) - (1 + r)k - e] + e^2}}{\delta AB}. \tag{A2}
\]

A sufficient condition for ruling out complex roots is \(A(\beta h_t + B) \geq (1 + r)k + e\). Given this, together with the fact that \(\gamma_{t+1} \leq c\), the only possible solution to (A2) is when \(\sqrt{\cdot}\) enters negatively, as shown in (1.17). The restriction \(\delta A(\beta h_t + B) \leq (1 + r)k + \delta AB c\) ensures that \(\gamma_{t+1} \geq -c\) as well. Having obtained (1.17), the result in (1.18) is obtained by appropriate substitution in (1.16).

The properties of the functions \(\gamma(c, e, h_t)\) and \(R(c, e, h_t)\) are established as follows. From (1.17) and (1.18), one finds that

\[
\gamma_c(\cdot) = 1 - \frac{2[\delta A(\beta h_t + B) - (1 + r)k - e]}{\sqrt{4\delta AB c[\delta A(\beta h_t + B) - (1 + r)k - e] + e^2}}, \tag{A3}
\]

\[
\gamma_e(\cdot) = -\frac{1}{\delta AB} \left[1 - \frac{(2\delta AB c - e)}{\sqrt{4\delta AB c[\delta A(\beta h_t + B) - (1 + r)k - e] + e^2}}\right] \tag{A4}
\]

\[
\gamma_h(\cdot) = -\frac{2\delta A\beta c}{\sqrt{4\delta AB c[\delta A(\beta h_t + B) - (1 + r)k - e] + e^2}}, \tag{A5}
\]

\[
R_c(\cdot) = \frac{\gamma_c(\cdot) + c}{e^2} \left\{\frac{\delta AB[(c\gamma_c(\cdot) + c) + (c\gamma_c(\cdot) - c)]}{4k} + \frac{e(c\gamma_c(\cdot) - c)}{2}\right\}, \tag{A6}
\]

\[
R_e(\cdot) = \frac{\delta AB(\gamma(\cdot) + c)\gamma_e(\cdot)}{2ck} + \frac{e\gamma_e(\cdot) + \gamma(\cdot) + c}{2c}, \tag{A7}
\]

\[
R_h(\cdot) = \frac{\delta AB[\gamma(\cdot) + c]\gamma_h(\cdot)}{2ck} + \frac{e\gamma_h(\cdot)}{2c}. \tag{A8}
\]

Under the above parameter restrictions, it is deduced that \(\gamma_c(\cdot) > 0, \gamma_e(\cdot) > 0, \gamma_h(\cdot) < 0, R_c(\cdot) > 0, R_e(\cdot) > 0\) and \(R_h(\cdot) < 0\).
Figure 1.1: Human Capital Distribution
Chapter 2

Macroeconomic Effects of Mandated Health Benefits

2.1 Introduction

Rising and significant health care expenditures challenge many countries. In the United States, health care accounted for nearly 18 percent of GDP in 2012, which exceeds the OECD average of about 9 percent. In 2010 the U.S. passed the most significant regulatory overhaul of its health care system since the creation of the Medicare and Medicaid programs in 1965. The 2010 reform is called the Patient Protection and Affordable Care Act (ACA). Because 90 percent of working age people in the U.S. who have health insurance obtain coverage from their employer, a report by Council of Economic Advisers (2009) argued that the current health care system imposes a heavy “tax” (health care cost) on small businesses and that the ACA would reduce this burden on small firms and encourage entrepreneurial activity. Small firms are important in the U.S. because they employ at least half of workers, and businesses with fewer than 20 employees account for a significant fraction of net employment growth (e.g., 25% from 1992 to 2005).

This paper focuses on the links among healthcare policy, small firms and the

---

1Medicare and Medicaid were the first U.S. public health insurance programs. Medicare provides federal health insurance for individuals at least age 65 or disabled, that paid into the system. Medicaid covers low income groups designated by statute such as children or pregnant women.
macroeconomy. We address two questions. How does a country’s health care policy influence an individual’s choice to become an entrepreneur or worker? How does health care policy affect GDP, firm size, wages, interest rates and welfare? In order to answer these questions, we take four alternative health care policies as given and investigate the policies’ implications for occupational choice and macroeconomic performance. We construct an occupational choice model in which individuals differ in ability, assets, and health shocks. The government maintains a balanced budget and uses lump sum taxes to pay for the benefits it provides. We then use the model to assess quantitatively the macroeconomic effects of key aspects of health care policies.

The main finding is that U.S. healthcare policies can distort occupational choice. The frictions are generated by two distinctive features: First, most working-age Americans obtain health insurance coverage through the workplace, employment-based health insurance (EHI). Second, U.S. law requires employers that offer health plans to use a price common to all employees. The value of health insurance to risk-averse agents varies with their idiosyncratic health risk. In a market without frictions, compensation reflects individual ability and exogenous shocks, and marginal utility is equalized across workers. We show that the link between employment and health insurance creates a wedge between the marginal cost and benefit of insurance. Since health risk can be sizable and insurance is part of total employee compensation, this wedge can distort firm and employee decisions. Our goal is to identify precisely this distortion and assess the quantitative impact of four policies: expansion of EHI; private insurance only; health insurance exchanges; and universal health coverage. To our knowledge, such healthcare policy induced distortions have not been examined previously.

To accomplish our goals, we construct a general equilibrium model of occupational choice with heterogeneous agents and a credit market. Individuals are risk averse, live for many periods, and choose to operate a firm and employ others or become a worker. Wages are determined endogenously and healthcare policy is given. There are three sources of heterogeneity. The first is the standard Lu-

\footnote{The Employee Retirement Income Security Act of 1974 (ERISA), amended by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), requires employers to offer health plans at a common price.}
cas (1978) “span of control” talent to manage a firm. The second is a health shock. The third is worker productivity. As in our paper, Jeske and Kitao (2009), Fang and Gavazza (2011), Aizawa and Fang (2013) and Feng and Zhao (2014) use the Medical Expenditure Panel Survey to measure the Markov process governing health shocks.\footnote{They focus on other issues. Jeske and Kitao (2009) examine U.S. healthcare subsidies and show that the tax is regressive. Fang and Gavazza (2011) construct a life cycle model of medical expenditure and find that EHI leads to dynamically inefficient investment in health. Aizawa and Fang (2013) develop a labour search model and use it to examine the ACA, with particular focus on the policy’s effect on the uninsured rate. Feng and Zhao (2014) study the impact of health policy on labour supply decisions.}

We take all policies as given, and explore the effect of insurance mandates on occupational misallocations of two types: Relative to a benchmark regime, some highly skilled individuals with adverse health risks choose to become workers rather than entrepreneurs, and some individuals with moderate managerial talent but good health choose to run a firm rather than become workers. This misallocation affects more than just the individuals involved because entrepreneurs create jobs. Antunes, Cavalcanti and Villamil (2008a) show that in the absence of health shocks, fewer high talent entrepreneurs running larger firms may lead to higher wages and output, making both entrepreneurs and workers better off. We show that health shocks can lead to changes in occupational choice that reduce wages net of taxes, change the distribution of firm size, and have large and heterogeneous effects on welfare. We use this model to evaluate aspects of the ACA reform. We discuss the ACA in section 2.2.

The literature on health policy, and firm and employee decisions, is large. For example, Garthwaite, Gross and Notowidigdo (2014) examine the effect of employer-sponsored health insurance in creating “employment locks” where agents pursue full-time jobs primarily to secure health insurance. Their focus is on the effect of health insurance on labour supply and they find micro econometric results consistent with a significant employment lock. In contrast, Fairlie, Kapur and Gates (2011) focus on “entrepreneur locks” and examine whether the U.S. EHI system impedes business creation. Using innovative econometric methods, they find a negative effect of having a spouse without insurance for business creation and that business ownership rates increase at age 65 when individuals qualify for
Medicare. We examine another aspect of an entrepreneurship/worker lock with different methods. Using a general equilibrium model calibrated to U.S. data, we analyse the effects occupational misallocation due to EHI on macroeconomic variables such as GDP, the distribution of firm size, wages and welfare.

The paper also contributes to a broad literature that studies macroeconomic aspects of health policies. This literature originates from Grossman (1972) and includes Brugemann and Manovskii (2010), Cole, Kim and Krueger (2014), Feng and Zhao (2014), French and Jones (2004), Hansen, Hsu and Lee (2014), Hall and Jones (2007), Jeske and Kitao (2009), Kopecky and Koreshkova (2012), Pashchenko and Porapkkam (2012), and others. Our paper is most related to Jeske and Kitao (2009), who show that EHI subsidies constitute a regressive tax, and Cole, Kim and Krueger (2014) who also study labour and health insurance market mandates but focus on static insurance gains versus dynamic incentive costs when individuals can use effort to improve health.

In order to focus on EHI characteristics that can distort private agent’s occupational choice, we incorporate health risk and health insurance into a Lucas (1978) “span of control” model. Hence, our paper is related to the literature on entrepreneurship. Antunes, Cavalcanti and Villamil (2008) and Herranz, Krasa and Villamil (2014) study the effect of credit market frictions on entrepreneurship. Cagetti and De Nardi (2006), Guner, Ventura and Xu (2008), and Kitao (2008), Panousi (2008), Li (2002) focus on the impact of government policies related to capital accumulation on entrepreneurship. Instead, this paper investigates the impact of a labour market friction on entrepreneurship. Finally, the paper is related to a large body of literature examining the cause and implications of factor misallocation. See Restuccia and Rogerson (2013) and the articles therein. Our work complements this literature by identifying a new friction associated with health insurance mandates, which leads to occupational misallocation.

In summary, in order to analyse occupational misallocation our model has the following key features. Individuals are endowed with heterogeneous managerial talent and heterogeneous health shocks. Firms face different costs of administering insurance that depend on their size. Contracts are incomplete: wages cannot be conditioned on health shocks by law. Section 2.2 summarizes these stylized facts in the data and Section 2.3 builds a model consistent with the facts. Section
2.4 describes optimal behaviour and the equilibrium. Section 2.5 contains the model calibration and quantitative analysis is performed in section 2.6. Section 2.7 concludes.

Figure 2.1: Health expenditure in OECD countries, 1970-2012

2.2 Facts and Policies

We now summarize some stylized facts about U.S. healthcare and the policies we consider.

2.2.1 U.S. Health Care

Fact 1: Figure 2.1 shows that U.S. health expenditure is high relative to OECD countries.\footnote{The figure is from Kaiser (2011): http://kff.org/health-costs/issue-brief/snapshots-health-care-spending-in-the-united-states-selected-oecd-countries/ The OECD reports total health expenditure as a fraction of GDP, which is the sum of public and private health spending. The measure includes health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health, but does not include provision of water and sanitation. http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS}

54
In 2012 the U.S. spent 17.9% of GDP on healthcare, about twice the OECD average.

Fact 2: In contrast to most countries, the U.S. healthcare system is employment based.

In the U.S. over 90% of private health insurance coverage is employment based. Buchmueller and Monheit (2009) discuss two government decisions that cemented the link between employment and health insurance: (i) During World War II the U.S. imposed wage and price controls, and in 1943 the War Tabor Board ruled that the controls did not apply to fringe benefits such as health insurance. Many employers used insurance benefits to attract and retain workers. (ii) In 1954 the Internal Revenue Service ruled that health insurance premiums paid by employers were exempt from income taxation, providing a subsidy to EHI through the U.S. tax code.

![Figure 2.2: EHI offer rate by establishment size, MES data](image)

Fact 3: EHI is strongly correlated with firm size and offer rates are fairly stable over time.
Figure 2.2 shows that about 97% of firms with over 100 employees offer health insurance, about 80% of firms with 25-99 employees offer insurance, and only 40% of firms with less than 25 employees offer coverage.

*Fact 4:* The share of premiums paid by employers is approximately constant over time, averaging about 85 percent for individual coverage and 75 percent for family coverage.\(^5\)

*Fact 5:* Employment based health insurance has a premium based on a community rating.

The Employee Retirement Income Security Act of 1974 (ERISA), amended by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), requires employers to offer health plans at common prices to all employees. The common price is known as *community rating*, where insurers evaluate risk factors of a market population rather than an individual. In contrast, private health insurance is generally based on individual characteristics and is more expensive than employment based (group) insurance. Community ratings are one way to address a fundamental market incompleteness that arises, for example, because individuals cannot choose their genes. Adjusted community ratings permit lifestyle factors such as smoking status to be considered.

*Fact 6:* The administrative cost for EHI is about half of the cost of individually purchased policies: 15-20% versus 30-40% and administrative costs decline with firm size.

Swartz (2006) shows that the cost savings from administrative economies of scale and better risk pooling increase with group size. Premiums are based on two components: average expected medical expenses for people in the group and a “loading fee”. Expected medical expenses are the same regardless of whether the person is in a large or small group, but the loading fee falls as size increases for

three reasons: efficiencies in administration and marketing, lower risk of adverse selection in a bigger pool, and lower risk that a fraction of individuals will have very high costs.

Interviews conducted by the Employee Benefit Research Institute with large employers indicate that EHI remains a valuable tool in recruiting and retaining workers. The percentage of firms offering health insurance as an employee benefit has remained remarkably stable over time.

2.2.2 Policies

We focus on four health care policies:

1. Employer provided health insurance: EHI is offered as part of employee compensation. The key feature is that an employer offers a worker a total compensation package consisting of a monetary wage plus a health insurance benefit.

2. Indemnity: This policy replaces EHI with an indemnity contract under which the insurance provider agrees to pay for health expenditures incurred by the individual. The policy resembles a contingent claim and is efficient by design.

3. Health exchanges with an insurance mandate and subsidies: The 2010 Patient Protection and Affordable Health Care Act (ACA) reformed the U.S. EHI health care system in key ways. The goal of ACA is to increase the quality and affordability of health care, lower the rate of uninsured, and reduce the cost of healthcare.\(^6\) We focus on key features of the law:

   • Individual mandate: Requires most U.S. citizens and legal residents to have insurance.

   • Policy mandate: Insurance market and rating rules

     – Guaranteed issue: Insurance companies are required to cover all applicants regardless of pre-existing conditions.

---

Table 2.1: Income levels and insurance premium subsidies under ACA

<table>
<thead>
<tr>
<th>Income in % of FPL</th>
<th>Premium subsidy rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-150%</td>
<td>94%</td>
</tr>
<tr>
<td>150-200%</td>
<td>77%</td>
</tr>
<tr>
<td>200-250%</td>
<td>62%</td>
</tr>
<tr>
<td>250-300%</td>
<td>42%</td>
</tr>
<tr>
<td>300-350%</td>
<td>25%</td>
</tr>
<tr>
<td>350-400%</td>
<td>13%</td>
</tr>
</tbody>
</table>

- **Community rating**: Insurance companies must charge a single premium to all individuals in a market with the premium based on the risk factors of a market population, not an individual.

- **Coverage standards**: Policies must meet minimum coverage standards.

- **HIX**: Create health insurance exchanges (HIX) through which individuals who do not have access to EHI can purchase coverage.

- **Subsidies**: Premium and cost-sharing credits are available to individuals/families with income between 133-400% of the federal poverty level (FPL). The poverty level was $19,530 for a family of three in 2013. Employers will pay penalties for employees who receive tax credits for health insurance through an Exchange, with exceptions for small employers.

4. **Universal health coverage**: This policy extends health care coverage to all members. Taxation is the main source of funding, but individuals or employers may pay supplements. Some countries have private insurance and universal health care (e.g., Germany), while other countries have single payer systems where the government contracts for health care with private providers (Canada) or operates a national health care system directly (the United Kingdom).

2.3 **The Model: Economic Environment**

Consider a Lucas (1978) span of control model, where individuals differ in the ability to manage capital and labour. Productivity $x_i$ for each agent $i$ is drawn
from a common continuous cumulative probability distribution with \( x \in [0, \infty] \). Productivity is not hereditary and is publicly observed. Households receive an idiosyncratic labour productivity shock \( z \) that indicates the efficiency units per unit of work hours. They also face an idiosyncratic health expenditure shock \( m_i \), which follows a finite-state Markov process. For notational convenience, we drop agent superscript \( i \) and time subscript \( t \) whenever possible, and \( \prime \) denotes the future value of a variable.

We will show that two types of individuals emerge, workers and managers. We begin by providing an overview. In section 3.3 we provide the intuition for a critical value, \( x^* \), where individuals above this value choose to be managers and those below it are workers.

### 2.3.1 Preferences, endowments and technology

**Preferences:** Consumption by an agent in period \( t \) is \( c_t \), with utility given by \( U(c_t) \).

**Endowments:** Each individual is endowed with managerial talent, \( x \), and productivity \( z \). Assume that the distribution and realizations are public information. Each agent receives a medical spending shock \( m \). Agents are also endowed with an initial capital asset, \( a_0 \), which can be used as an input in production.

**Production:** Firms use efficiency labour (\( n \)) and capital (\( k \)) to produce a single consumption good, \( y \). Efficiency labour is \( n = \int z\hat{n} \), the sum of hours worked, \( \hat{n} \), weighted by the productivity of each worker, \( z \). Capital depreciates at a constant rate of \( \delta \). Managers can operate only one project. The functional form of the production function is:

\[
y = X k^\alpha n^\gamma \quad \text{where} \quad \alpha, \gamma > 0,
\]

where talent is given by \( X = x^{1-(\alpha+\gamma)} \).

**Factor remuneration:** Firms rent capital at the common market rate \( r(1+\Delta) \), where \( r \) is the risk-free rate and \( \Delta \geq 0 \). We assume that the intermediary charges
a proportional cost $\Delta$ per unit of funds loaned to the firm. As usual, this wedge above the risk-free rate accounts for intermediation costs and a risk premium.

The firm offers a worker a compensation package $\tilde{w}$ that includes a monetary wage $w$ and a term that accounts for the expected cost of insurance. Each firm offers employment-based health insurance (EHI) with given probability $p_E$. The firm’s expected cost of providing EHI directly is $p_E \left[ 1 + g(n)q_A^j \right] q_E$, where $g(n)$ is a decreasing function of $n$ because it is more costly for a small firm to offer health insurance than bigger firms due to economies of scale. The variable $q_E$ is the fair price of insurance and $q_A^j$ is the expected administrative cost of insuring workers in firm size bin $j$. Under actuarially fair insurance, the cost of insurance equals the expected health shock, which is the price that would be charged in a perfectly competitive market

$$q_E = \sum_{s=0}^{s} E \left[ \pi_s \phi_s m_s \right], \quad (2.2)$$

where $\pi_s$ is the probability of the shock given state of the world $s$, $\phi_s$ is the co-insurance rate, and $m_s$ is the health shock. We introduce an expected administration cost $q_A$ that affects the economies of scale that firms face $g(n)$,

$$q_E(n) = g(n)q_A^j + \sum_{s=0}^{s} E \left[ \pi_s \phi_s m_s \right], \quad (2.3)$$

where shock $q_A$ is normally distributed. Although the idiosyncratic administrative cost is uncertain for firms, the mean cost is decreasing as firm size increases. This is due to the presence of $g(n)$, a decreasing function of $n$ such as $\frac{n^a}{n}$, which captures the effect evident in figure 2.2: it is more costly, on average, for a small firm to offer health insurance than bigger firms due to economies of scale.

In a competitive market without commitment, thr expected utility of the wage package, $\tilde{w}$, must always be the same for workers, regardless of wheter the recieve insurance or not. Hence, if a firm does not offer health insurance it must raise wages by an amount $b$ where $b$ is the monetary compensation that would make the worker indifferent between having insurance or being given a higher wage.\footnote{Idiosyncratic uncertainty stems from the fact firms do not know the health status of individuals they employ, heterogeneity in U.S. state laws, and bargaining power.} As

\footnote{This is consistent with evidence from Olson (2002) as well as Dey and Flinn (2005).}
workers are risk averse it follows that the compensation payment $b$ will be higher than the fair price of insurance $q_E$. However, the cost of providing the insurance for employers will vary depending on the size of the firm, $n$, as well as an idiosyncratic administration cost, $q_A^j$.

Due to the presence of the idiosyncratic administration cost $q_A^j$, offering insurance may not always be cheaper for an individual firm than offering a higher wage. It follows that a firm will offer health insurance if the cost of doing so is less than the compensation payment,

$$w + \left[1 + g(n)q_A^j\right] q_E < w + b$$ (2.4)

The idiosyncratic administration cost is important in determining whether a particular firm offers insurance, nonetheless larger firms are much more likely to offer health insurance as they benefit from economies of scale captured through the decreasing concave function $g(n)$.

In the model we express the total wage package for workers as

$$\tilde{w} = w + i_E \left[1 + g(n)q_A^j\right] q_E + (1 - i_E)b$$ (2.5)

where $i_E = \{0, 1\}$. Firms choose $i_E = 1$ when the cost of providing EHI is lower than the compensation payment, and $i_E = 0$ otherwise. Probability $p_E$ that the firm offers insurance is the value such that the expected value of the two payments is $\tilde{w}$ when $b = q_E$.

If we assume that the average idiosyncratic shock $q_A^j$ has the same mean and distribution across all firm sizes, we can infer an estimate the value of $g(n)$ and hence the role that economies of scale have on the decision to offer health insurance. It follows that there will be a critical value of the idiosyncratic shock $\hat{q}_A$ that determines whether a firm offers insurance or not. We obtain this critical value by rearranging equation (37)

$$\hat{q}_A = \frac{1}{g(n)} \left[ \frac{b}{q_E} - 1\right]$$ (2.6)

The optimal value of $b$ differs across individuals due to heterogeneity in individuals’ previous health costs. The firm can calculate $b$ based on the average expected health costs. The model assumes complete information.
If the realized idiosyncratic shock is lower (higher) than the critical value $q_A^i < \hat{q}_A$ ($q_A^i > \hat{q}_A$), then a firm will offer (not offer) insurance. Notice that by substituting values of $g(n)$ into the above equation we can see that as firm size increases the critical level increases. This means that larger firms are more likely to offer health insurance as it will take a significantly higher idiosyncratic health cost, compared to smaller firms, to exceed the critical value.

### Link between $p_E(n)$ and $i_E$

We model whether or not the firm offers insurance, where $i_E = 1$ indicates that the firm offers insurance and $i_E = 0$ indicates that the firm does not offer insurance. The firm’s choice to offer insurance is the least costly compensation alternative, given its cost structure. Figure 2.2 and the data in Section 2.2.5 on firm size indicate that large firms offer insurance with higher probability than small firms. We display the data again for convenience:

<table>
<thead>
<tr>
<th>Firm size (j bins)</th>
<th>$n &lt; 10$</th>
<th>$10 - 24$</th>
<th>$25 - 99$</th>
<th>$100 - 999$</th>
<th>$n &gt; 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_E(n)$</td>
<td>0.336</td>
<td>0.625</td>
<td>0.816</td>
<td>0.943</td>
<td>0.992</td>
</tr>
<tr>
<td>Administrative cost, $g(n)$</td>
<td>0.3</td>
<td>0.21</td>
<td>0.132</td>
<td>0.0849</td>
<td>0.06</td>
</tr>
</tbody>
</table>

In the model, we assume that total labour compensation is given by

$$\tilde{w} = w + p_E \left[ 1 + g(n)q_A^i \right] q_E + (1 - p_E)b_E$$

In the data we do not observe idiosyncratic administrative cost shock $q_A^i$, and therefore we cannot establish whether an individual firm chooses to offer insurance, which corresponds to the decision $i_E = 1$. However, we know that the firm will choose the least costly of its two options, and this will provide the link between unobserved firm choice $i_E$ and observed probability $p_E(n)$.

### Health insurance market:

There are two types of insurance, EHI and private insurance:

**EHI:** Households have access to EHI with probability $\hat{p}_E$, which is determined by $i_E$. We differentiate between $p_E$ and $\hat{p}_E$ because workers are randomly matched
with firms of different sizes, but each worker has the same probability of receiving an EHI offer. Insurance covers a fraction $\phi(m)$ of total medical expenditures, where $\phi(\cdot) \in [0,1]$. The EHI premium is denoted by $\pi_E$ and is not dependent on the individual’s prior health history or any individual states. This accounts for the community rating practice in the U.S. where group health insurance cannot price-discriminate among the insured based on such individual characteristics. A fraction $\psi \in [0,1]$ of the premium is paid by the employer as a subsidy.

**Private:** If the worker is not offered EHI (or declines the EHI offer), she has the option to purchase health insurance in the private market at premium $\pi_P(m)$ with coinsurance rate $\phi(m)$. This can happen if a household becomes a manager and does not offer (or has no access to) EHI.

Once the firm makes an offer to the worker, which is denoted as $i_E = 1$, the worker chooses either to obtain coverage (through EHI or purchase health insurance in the private market) or remain uninsured ($i_{HI}' = \{0,1\}$). Health insurance companies are competitive. The premiums for EHI and private plans are determined by the expected expenditures for each contract plus a proportional markup denoted by $\eta$. EHI has two advantages compared with private insurance:

(i) EHI receives a tax subsidy from the government, which is more cost-efficient for firms.

(ii) EHI has a more inclusive risk pool, which helps to share risk among the insured.

**Government:** The government runs a balanced budget each period and provides (only) two types of policies, which are financed through lump sum taxation, $\tau_y$.\(^{11}\)

- Public safety-net program, $T_{SI}$: This program guarantees each household a minimum consumption level of $c$. This reflects the option available to U.S. households to rely on public transfer programs such as food stamps,

---

\(^{10}\)In line with Jeske and Kitao (2009), we assume a segmented labour market where employers do not adjust wages if EHI coverage is declined.

\(^{11}\)Jeske and Kitao (2009) show that a distortionary tax with an EHI subsidy constitute regressive taxation. We focus the distortion that EHI induces in occupational choice, hence we abstract from distortionary taxation.
Workers: Medicaid, disability and unemployment insurance if substantial income and health spending shocks occur.

- In the baseline model, the government subsidises EHI at rate $\tau$.

### 2.3.2 Firm’s problem

The firm’s problem is:

$$\max_{n,k} Xk^\alpha n^\gamma - \tilde{w}n - rk$$  \hspace{1cm} (2.7)

where $\tilde{w}$, the average cost of hiring labour, includes monetary wage component $w$ and the expected cost of EHI or a compensation payment by the firm when EHI is not offered. See the appendix for the derivation of $n^*$ and $k^*$, for constrained and unconstrained borrowing.
2.3.3 EHI and talent misallocation

Figure 2.3 illustrates that misallocation can occur when there is a link between insurance and employment. Exogenous managerial ability \( x \) is on the horizontal axis, which determines the profit if an individual decides to become an entrepreneur and manage a firm. Idiosyncratic health risk is on the vertical axis. First consider a frictionless world, where there is no insurance distortion (or credit constraint).\(^{12}\) In this case a cutoff value \( x^* \) exists that differentiates entrepreneurs from workers. This is illustrated by the vertical red dashed line. On the right side of the vertical line an agent’s \( x \) is sufficiently high to yield greater profit from running a firm than from choosing to work at the market wage (i.e., only the vertical dotted line exists and the red and blue areas are not relevant). Choosing to be a worker is optimal on the left side of the line.

Now consider occupational choice when health insurance is employment-based and worker compensation includes a wage and health insurance package. Current U.S. law requires employers to offer a health plan at a price common to all employees. However, the value of health insurance to agents varies with their idiosyncratic health risk. Hence, the link between employment and health insurance creates a gap between the marginal cost and marginal benefit of health insurance. Figure 2.3 shows that two types of misallocation can occur: (i) Some healthy but low ability agents select into entrepreneurship, and (ii) some agents with high ability but poor health shocks select out of entrepreneurship. Consider a healthy agent who would choose to be a worker in the absence of employment-based health insurance. This individual receives a wage plus health insurance as a worker, and does not value the firm’s health insurance greatly but cannot get additional compensation if he declines the insurance. This individual may find it more attractive to become an entrepreneur to get a higher return and either self-insure or get insurance in the private market. This is the blue area. Now consider an individual with high managerial ability but an unfavourable health shock. It may be advantageous for this individual to work for a firm to get group health insurance. This is the red

\(^{12}\)We focus on how health care policy affects occupational choice. ACA also imposes an employer mandate that requires firms with over 50 employees to provide EHI, which could distort a firm’s labour demand decision. Aizawa and Fang (2013) look at this issue and their results suggest that the effect is quite small, as does recent data in Garrett and Kaestner (2014).
Overall the graph shows that some individuals that are healthy but less skilled become entrepreneurs, while others that are less healthy but highly skilled leave entrepreneurship. These misallocations relative to a frictionless world are caused by the link between health insurance and employment. We call this “talent misallocation” as the individuals in the blue region that are healthy but less-skilled would be workers absent the EHI friction, while those with bad health shocks but high ability in the red region would run firms. We will quantify the effects of counterfactual policy experiments related to the ACA on this misallocation.

2.4 Optimal behaviour and equilibrium

The timing of the economy is given as follows.

1. Households enter each new period with assets $a$ and health insurance status $i_{HI}$.
2. Idiosyncratic shocks $x, z$ and $m$ are drawn by nature.
3. Households make an occupation decision: entrepreneur ($I_e = 1$) or worker ($I_e = 0$).
4. Workers randomly match with firms. Firms decide whether to offer insurance and $i_E$ determines the EHI offering status (EHI availability for workers).
5. Capital and labour markets clear and production takes place.
6. Households choose: health insurance ($i_{HI}' = \{0, 1\}$), consumption ($c$), borrowing/saving ($a'$), and supply labour inelastically. Managers and workers decide on health insurance purchases.

2.4.1 Firm manager

Firms are distinguished by their productivity realization $x$. Agents with sufficient ability to become managers choose the level of capital and the number of employees to maximise profit subject to a technological constraint and exogenously given
health care policy. The benefits component of EHI exists for historical reasons and clearly it would be more efficient to use an insurance pool. In order to simplify the exposition, first consider the problem of a manager with talent \( x_i \) for a given level of capital \( k \) (i.e., the labour input choice only)

\[
\max_n X k^\alpha n^\gamma - \tilde{w} n
\]

(2.8)

where \( \tilde{w} = [w + p_E (1 + g(n)) q_E + (1 - p_E) q_E] \) is the firm’s per capita labour cost and \( g(n) \) is the administrative cost of organizing EHI at the firm level.

Note that \( n^* \) will depend on the size of the firm, which depends on the functional form of the markup on health insurance \( g(n) \). For example we could use a functional form such as \( \frac{g}{n} \), which is a decreasing function of firm size \( n \). However, we assume that the function \( g(n) \) can be approximated by decreasing steps in firms size where \( j \) is the number of intervals and \( \lambda \) is the value that \( g(n) \) takes at the specific interval \( j \) where \( \lambda \) is.

The variable \( \lambda \) reflects what we infer from the data and is presented in figure 2.2. Recall that the administration costs represents the notion that the cost of group health insurance is decreasing in firm size because the fixed cost component is spread over a larger base.

Consider the simple case where \( j \) equals two. Economies-of-scale occur for sufficiently large firms and not for small firms. Hence, for small firms, \( \lambda \) is equal to 1.

The optimal \( n^* \) for small firms therefore can be expressed as

\[
\max n x^i k^\alpha n^\gamma - [i_E (w + (1 + \lambda(n) q_A - \psi) q_E) + (1 - i_E) (w + b_E)] n
\]

(2.9)

The FOC is given as

\[
n'(k, x, w) = \gamma x^i k^\alpha n^{\gamma-1} - [i_E (w + (1 + \lambda'(n) q_A - \psi) q_E) + (1 - i_E) (w + b_E)] = 0
\]

(2.10)

Different \( n^* \) will exist depending on the size of the firm. Crucially, this will depend on how \( \lambda \) is distributed. We will assume that \( \lambda \) decreases over a number of intervals \( j \). Consider the simple case where \( j \) equals two; there are economies-of-scale for sufficiently large firms and not for small firms. Hence, for small firms, \( \lambda \) is equal
to 1. The optimal \( n^* \) for small firms therefore would be expressed as

\[
n^*_{\text{SMALL}}(k, x, w) = \left[ \frac{\gamma x^i k^\alpha}{i_E (w + (1 + q_A - \psi) qE) + (1 - i_E) (w + bE)} \right]^{\frac{1}{1 - \gamma}}.
\]

For a large firm which can benefit from economies of scale \( n^* \) is expressed as

\[
n^*_{\text{LARGE}}(k, x, w) = \left[ \frac{\gamma x^i k^\alpha}{i_E (w + (1 + \lambda q_A - \psi) qE) + (1 - i_E) (w + bE)} \right]^{\frac{1}{1 - \gamma}},
\]

where \( \lambda \in (0, 1) \). Naturally, in this simple case there is an incentive for firms sufficiently close to the point where it becomes a large firm to employ more workers in order to obtain the savings from economies-of-scale. 13 The superscript \( j \) indicates that there are multiple steady-state variables depending on the distribution of the savings due to economies of scale \( \lambda \).

Substituting \( n^* \) into (2.9) yields the manager’s profit function for a given level of capital:

\[
y^*_{j}(k, x, w) = x k^\alpha \left[ \frac{\gamma x^i k^\alpha}{i_E (w + (1 + \lambda q_A - \psi) qE) + (1 - i_E) (w + bE)} \right]^{\frac{1}{1 - \gamma}} \tag{2.11}
\]

However, in the following analysis we drop the subscript \( j \) for convenience.

Now consider the choice of capital. Let

- \( a \) denote the amount of self-financed capital; and
- \( l \) denote the amount of funds borrowed from a bank.

Both sources of funds are used to raise capital, with \( k(\cdot) = a(\cdot) + l(\cdot) \). There is no commitment problem regarding bank loan repayment, so the two sources of funds have the same cost.

### 2.4.1.1 Remark on random matching

Workers supply labour inelastically at the given the wage package \( \tilde{w} \). They enter the market and are randomly matched to firms. Workers receive EHI with

---

13 It can be shown that the marginal savings for a firm employing one more worker would not be greater than the marginal cost of employing one more worker.

14 This will adjust with EHI offering status, since EHI benefits from a tax subsidy.
probability \( \hat{p}_E \), which is determined by shock \( i_E \). We differentiate between \( p_E \) and \( \hat{p}_E \) because each worker has the same probability of receiving an EHI offer. Consider two firms, one big and one small. The bigger firm offers insurance with 90% probability and the smaller with 50% probability. From the worker’s point of view, probability \( \hat{p}_E \) is a weighted average of the two firms. In general, \( \hat{p}_E = \frac{\int I_e n^* p_E(n^*)d\Psi(s)}{\int I_e n^* d\Psi(s)} \). Equivalently, \( \hat{p}_E = \int [\int I_e n^* d\Psi(s)]p_E(n^*)d\Psi(s) \), where the weight is given by the term in brackets.\(^{15}\)

### 2.4.1.2 Capital

Now consider the choice of capital. Let

- \( a \) denote the amount of self-finance; and
- \( l \) denote the amount rented from the capital market.

Both sources of funds are used to raise capital, with \( k(\cdot) = (a(\cdot) - oop) + l(\cdot) \), where \( oop \) denotes out of pocket medical expenses. The entrepreneur can either use personal funds net of out-of-pocket medical spending \( (a - oop) \) or rent capital from the market \( (l) \). The two sources of funds have the following costs. The entrepreneur owns capital and therefore the opportunity cost of \( a \) is only the foregone interest the entrepreneur could have received from the capital market. This amount is given by \( ra \). In addition, the entrepreneur may rent capital in the market, at cost \( (1 + \Delta)rl \), \( l \leq \bar{l} \). Here \( \bar{l} \) is an upper limit on borrowing. We will first consider the case where this borrowing constraint does not bind.

**Self-financed firm:** When initial assets are sufficient to run a business without renting new capital from the market (i.e., \( l = 0 \)), the manager of the firm solves the problem:

\[
\nu(a, x, i_E; w, r) = \max_{k \geq 0} y(k, x, \tilde{w}) - rk - \tilde{w}n(k, x, \tilde{w}) \quad (2.12)
\]

\(^{15}\)We model the way firms offer EHI as a preference shock \( i_E \), an approach also used by Aizawa and Fang (2013). In the appendix we consider a cost shock, which is an alternative approach that endogenises the insurance offer.
This gives the optimal physical capital level:

$$\nu(a, x, w, r) = \max_{k \geq 0} Xk^\alpha \left[ \frac{\gamma Xk^\alpha}{\bar{w}} \right]^\frac{1}{1-\gamma} - rk - \bar{w}n$$  \hfill (2.13)

$$k^*(x, w, r) = \left[ X \left( \frac{\gamma}{w} \right)^\gamma \left( \frac{\alpha}{r} \right)^{1-\gamma} \right]^\frac{1}{1-\gamma}$$  \hfill (2.14)

The manager’s profit at the optimal level of capital is:

$$\nu(k^*, x, w) = Xk^\alpha \left[ \frac{\gamma Xk^\alpha}{\bar{w}} \right]^\frac{1}{1-\gamma} - \bar{w}(k^*, x, \bar{w}) - rk^*$$  \hfill (2.15)

The manager’s consumption is determined as follows.

$$c + a' + (1 - i_{HI}\phi(m)) m + \bar{\pi} \leq (1 + r - \delta)a + \nu - \tau_y + T_{SI} + \tau_s i_{E} i'_{HI} \pi_E$$  \hfill (2.16)

where

$$\bar{\pi} = \begin{cases} \pi_E & i'_{HI} = 1, i_E = 1 \\ \pi_p(m) & i'_{HI} = 1, i_E = 0 \\ 0 & i'_{HI} = 0 \end{cases}$$  \hfill (2.17)

$$T_{SI} = \max \{ 0, \zeta + \tau_s - \tau_s i_{E} i'_{HI} \pi_E + (1 - i_{HI}\phi(m)) m - (1 + r - \delta)a - \nu(k^*, x, \bar{w}) \}$$  \hfill (2.18)

$$a' \geq -\bar{a}.$$  \hfill (2.19)

The budget constraint is standard: consumption, saving/borrowing, uncovered (out of pocket) medical expenses, and insurance premia cannot exceed asset market returns, firm profit, lump sum taxes, government transfers, and the insurance subsidy. Lump-sum tax, $\tau_y$, is collected to finance a consumption floor $\zeta$ and EHI subsidy $\tau_s$. The premium that the manager pays for insurance, $\bar{\pi}$, has two components: $i'_{HI}$ is the entrepreneur’s choice to buy health insurance for himself for next period and $i_E$ is the variable that indicates that the employer must provide health insurance to the employee. We focus on three cases: the entrepreneur
purchases insurance for himself and the employees, the entrepreneur purchases insurance only for himself, and the entrepreneur purchases no insurance. The government defrays the cost of EHI by providing subsidy $\tau_s i_{E_{HI}} \pi_E$. $T_{SI}$ denotes a transfer from the government as specified in Hubbard et al. (1995), where $\nu$ are firm profits, and the firm’s borrowing is determined by the optimal $k^*$ as explained in the appendix.

**Firm with assets borrowed from the market:** When managers do not have enough personal assets to operate the firm, they can rent $l$ from the capital market at rate $(1 + \Delta)r$. The firm’s problem is given as follows.

$$\hat{\nu}^*(\tilde{k}, x, w) = \max_k X \tilde{k}^\alpha \tilde{n}^\gamma - \tilde{w} \tilde{n} - \hat{r} \left( \tilde{k} - (a - oop) \right)$$  \hspace{1cm} (2.20)

where

$$\hat{r} = \begin{cases} r & \text{if } \tilde{k} \leq a - oop \\ (1 + \Delta)r & \text{if } \tilde{k} > a - oop \end{cases}$$  \hspace{1cm} (2.21)

$$\hat{n}^*(\tilde{k}, x, w) = \left[ \frac{\gamma X \tilde{k}^\alpha}{\tilde{w}} \right]^{\frac{1}{1 - \gamma}}.$$  \hspace{1cm} (2.22)

### 2.4.2 Workers

Workers maximise expected discounted utility of consumption

$$\max_{\{c_t, a_{t+1}, \delta_{HI, t+1}\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t U(c_t)$$

subject to the following budget constraint:

$$c + a' + (1 - i_{HI}(m)) m + \pi \leq (1 + r - \delta)a + \tilde{w}z - \tau_y + T_{SI} + \tau_s i_{E_{HI}} \pi_E$$  \hspace{1cm} (2.23)
where

\[ \tilde{\pi} = \begin{cases} 
\pi_E (1 - \psi) & i_{HI}' = 1, i_E = 1 \\
\pi_P(m) & i_{HI}' = 1, i_E = 0 \\
0 & i_{HI}' = 0 
\end{cases} \]

(2.24)

\[ \tilde{w} = \begin{cases} 
w + c_E & i_E = 0 \\
w & i_E = 1 
\end{cases} \]

(2.25)

\[ T_{SI} = \max \{ 0, \gamma + \tau_y - \tau_s i_{E} i_{HI}' \pi_E + (1 - i_{HI}\phi(m)) m - [(1 + r - \delta)a + \tilde{w}] \} \]

(2.26)

\[ a' \geq -\bar{a} \]

(2.27)

The worker’s budget constraint indicates that consumption, saving/borrowing, out of pocket medical expenses, and insurance premia cannot exceed asset market returns, total labour compensation, lump sum taxes, government transfers, and the insurance subsidy. The insurance premium, \( \tilde{\pi} \), again has two components: \( i_{HI}' \) is the agent’s choice to buy health insurance for himself for next period where \( i_E \) is the variable that indicates that EHI is offered. There are three cases: the worker gets EHI but must pay the remaining \( 1 - \psi \) of the premium not paid for by the firm, the worker purchases insurance directly in the private market, or the worker purchases no insurance. The government defrays the cost of EHI by providing subsidy \( \tau_s i_{E} i_{HI}' \pi_E \). Again \( T_{SI} \) is a transfer from the government that is analogous to the firm specification except that firm profits, \( \nu \), are replaced by employee total compensation \( \tilde{w}z \).

2.4.3 Government

The government runs a balanced budget with a lump-sum tax \( \tau_y \):

\[ \tau_y = \int (T_{SI} + \tau_s i_{E} i_{HI}' \pi_E) d\Psi(s) \]

where \( \Psi(s) \) represents the distribution of agents in equilibrium as explained below.
2.4.4 The household’s problem

Let $I_e$ indicate occupational choice, where if $I_e = 1$ the household is an entrepreneur and if $I_e = 0$ the household is a worker. We can write the household’s problem recursively as follows.

$$V(a, x, z, m, i_{HI}) = \max_{\{a', c', i'_{HI, I_e}\}} \left[ I_e V_e + (1 - I_e) V_w + \beta E V(a', x', z', m', i'_{HI}) \right]$$

subject to

$$c + a' + \text{oop} + \tilde{\pi} \leq (1 - \bar{r} - \delta)a + inc - Tax \quad (2.28)$$

where

$$\tilde{\pi} = \begin{cases} 
\pi_E (1 - \psi) & i'_{HI} = 1, i_E = 1 \\
\pi_P (m) & i'_{HI} = 1, i_E = 0 \\
0 & i'_{HI} = 0
\end{cases} \quad (2.29)$$

$$Tax = \tau_y - T_{SI} - \tau s i_E i'_{HI} \pi_E \quad (2.30)$$

$$T_{SI} = \max \{ 0, \zeta + \tau_y - \tau s i_E i'_{HI} \pi_E + \text{oop} - [(1 - \delta)a + inc] \} \quad (2.31)$$

$$inc = \begin{cases} 
ra + \bar{\omega} z + (1 - i_E) q_E & \text{if } I_e = 0 \\
ra + \nu(k, x; \bar{r}, \bar{w}) & \text{if } I_e = 1
\end{cases} \quad (2.32)$$

$$\text{oop} = (1 - i_{HI} \phi(m)) m \quad (2.33)$$

$Tax$ is the lump sum tax net of social insurance benefit (if applicable) and the health care subsidy, $inc$ is the earnings of the worker or entrepreneur, and $oop$ is out of pocket medical expense.

The value functions $V_e$ and $V_w$ are defined as follows:

$$V_e = p_E (n^*) U(c | i_E = 1) + (1 - p_E (n^*)) U(c | i_E = 0)$$

$$V_w = \tilde{p}_E U(c | i_E = 1) + (1 - \tilde{p}_E) U(c | i_E = 0).$$

$\tilde{p}_E$ and $p_E$ reflect the random matching between workers and firms, as explained
in section 2.4.1.2.

## 2.4.5 Health insurance

There are two kinds of insurance, private and employer based group insurance. The latter benefits from pooling and tax advantages, while private insurance has higher administrative costs. The cost of providing insurance for the firm is given as:

\[ q_E = \psi \pi_E \]  

(2.34)

The EHI premium equals the average cost of providing insurance:

\[ \pi_E = (1 + \eta) \int i_E i'_{HI} \phi(m) m d\Psi(s) \]  

(2.35)

The premium for private insurance equals:

\[ \pi_P(m) = (1 + \eta) \frac{E [\phi(m') m' | m]}{1 + r - \delta}. \]  

(2.36)

Markup \( \eta \) applies to both EHI and private insurance, consistent with MEPS data.

## 2.4.6 Steady state equilibrium

We characterize the steady state equilibrium. Denote the equilibrium aggregate variables by \( \Phi = \{r, w, \pi_E, \hat{\pi}_E, \tau_y\} \). Individual state variables \( s = \{a, x, z, m, i_{HI}\} \) denote asset holding \( a \in A \), managerial ability \( x \in X \), labour productivity \( z \in Z \), health spending shock \( m \in M \) and insurance status \( i_{HI} \in I \). Let \( S = A \times X \times Z \times M \times I \) denote the entire state space.

### 2.1 The steady state equilibrium for the economy is given by aggregate variables \( \Phi \), allocations \((c, a', i'_{HI}, I_e)\) for households characterized by \( s = (a, x, z, m, i_{HI})\) and the distribution of agents over the state space \( S \) given by \( \Psi(s) \), \( s \in S \), such that:

1. Given \( \Phi \), allocations \((c, a', i'_{HI}, I_e)\) solve the household’s optimization problem.
2. The health insurance market is competitive.

3. The asset market clears: \( \int k \, d\Psi(s) = \int a \, d\Psi(s) \).

4. The labour market clears: \( \int \mathcal{L} \, n \, d\Psi(s) = \int (1 - \mathcal{I}) \, \hat{n} \, z \, d\Psi(s) \).

5. The goods market clears.

6. The government balances its budget: \( \tau_y = \int (T_{SI} + \tau_s i_{E} i_{HI} \pi_{e}) \, d\Psi(s) \).

7. Distribution \( \Psi(s) \) is time-invariant. The law of motion for the distribution of agents over the state space \( S \) satisfies \( \Psi = F_{\Psi}(\Psi) \), where \( F_{\Psi} \) is a one-period transition operator on the distribution.

### 2.4.7 Analysis of competitive equilibrium

The following proposition states that there exists a cutoff value that differentiates entrepreneurs from workers based on managerial ability, as illustrated in figure 2.3.

**2.1** Let denote \( x^* \) the cutoff value such that an agent with \( x \geq x^* \) becomes an entrepreneur; otherwise the agent is a worker. The cutoff value is a function of \( (a, z, m, i_{HI}) \).

The proof follows from Antunes, Cavalcanti and Villamil (2008b), where the credit friction causes \( x^* \) to decrease with an agent’s assets. In their case loans are given by \( l = k - a \), at rate \( r \). The ability to borrow allows some low asset but high ability agents to become entrepreneurs. In our case \( l = k - \tilde{a} \), where \( \tilde{a} = a - oop \) and \( \tilde{r} = (1 + \Delta)r \), and EHI allows some individuals with poor health shocks and high ability to become entrepreneurs.

**2.2** The cutoff value is decreasing in \( a \), if \( \Delta > 0 \).

**Proof.** See the Appendix.

We show that when EHI is a mandated benefit, this distorts the cutoff value. The following proposition states that less healthy agents need a higher \( x^* \) to become entrepreneurs.
2.3 In the presence of EHI, cutoff value $x^*(a, z, m, i_{HI})$ increases with the size of $m$.

Proof. See the Appendix.

The cutoff value that we compute in the equilibrium is illustrated in figure 2.4.

2.5 Calibration

Preferences: Household preferences are given by $\sum_{t=0}^{\infty} \beta^t U(c_t)$, where $U(c) = \frac{c^{1-\rho} - 1}{1-\rho}$.

The coefficient of relative risk aversion $\rho$ is set to 1.5 in the baseline economy, which follows estimates in the literature. We also consider $\rho = 3$ as a robustness check.

The subjective time discount factor $\beta$ is set to 0.94 so that the aggregate capital-output ratio is 2.42 in the stationary equilibrium, consistent with U.S. data.
Labour Productivity: We assume that stochastic labour productivity $z$ follows a first-order autoregressive process: $\ln z_t = \rho_z \ln z_{t-1} + \varepsilon_{z,t}$, where $\varepsilon_{z,t} \sim N(0, \sigma^2_z)$. In line with the literature, we choose the value for coefficient $\rho_z$ and the residual variance $\sigma^2_z$ to be 0.94 and 0.02 respectively.\textsuperscript{16} To facilitate the computation, we approximate this process by a five state Markov process using the method of Tauchen and Hussey (1991). The calibrated Markov process is represented by finite states:

$$z \in \{0.646, 0.798, 0.966, 1.169, 1.444\}$$

and a transition matrix

$$\Pi_z = \begin{bmatrix}
0.731 & 0.253 & 0.016 & 0.000 & 0.000 \\
0.192 & 0.555 & 0.236 & 0.017 & 0.000 \\
0.011 & 0.222 & 0.533 & 0.222 & 0.011 \\
0.000 & 0.017 & 0.236 & 0.555 & 0.192 \\
0.000 & 0.000 & 0.016 & 0.253 & 0.731
\end{bmatrix}.$$  

Entrepreneurial ability and technology: The entrepreneur is endowed with managerial ability $x$ and operates a firm with a neo-classical production function $Xk^\alpha n^\gamma$, where $X = x^{1-(\alpha+\gamma)}$. We assume that managerial ability is distributed log-normal with mean $\mu_x$ and variance $\sigma^2_x$, so that $\log(x) \sim N(\mu_x, \sigma^2_x)$. We choose $\alpha$ to match the capital share of 0.34 for the U.S economy for the period 1960-2000. We choose $\gamma$ to match the fraction of entrepreneurs in the economy. We find $\mu_x$ and $\sigma^2_x$ to match the fraction of firms at different levels of employees and the mean size of establishments, which are listed in Table 2.3. In line with Guner, Ventura and Xu (2008), we truncate the distribution of $x$ and approximate it with 40 grid points.

Health spending shocks and health insurance: We use Medical Expenditure Panel Survey (MEPS) data to estimate health expenditure shocks and health insurance. We focus on the working population and use seven states for health expenditures. In line with Jeske and Kitao (2009), we divide data into bins of size (20%, 20%, 20%, 20%, 15%, 4%, 1%). The first bin contains all agents whose health

\textsuperscript{16}See Storesletten et al. (2004) and Hubbard et al. (1994).
expenditures fall in the bottom twenty percentiles, while the last bin has agents inside the first percentile of the distribution. We represent each bin using the mean expenditure in that bin and normalize them in terms of the average earnings in 2003 (based on MEPS 2003, the average wage income of all heads of households is $32,800). To this end, health spending follows a finite state Markov chain, with \( m \in \{0.000, 0.006, 0.022, 0.061, 0.171, 0.500, 1.594\} \). The transition matrix for \( m \) is estimated by counting the fraction of agents who move into each bin in the following year.

\[
\Pi_m = \begin{bmatrix}
0.542 & 0.243 & 0.113 & 0.061 & 0.032 & 0.007 & 0.002 \\
0.243 & 0.330 & 0.242 & 0.117 & 0.056 & 0.011 & 0.001 \\
0.119 & 0.224 & 0.296 & 0.232 & 0.098 & 0.025 & 0.006 \\
0.058 & 0.130 & 0.225 & 0.347 & 0.201 & 0.035 & 0.005 \\
0.043 & 0.079 & 0.140 & 0.263 & 0.371 & 0.090 & 0.014 \\
0.030 & 0.063 & 0.080 & 0.203 & 0.359 & 0.200 & 0.065 \\
0.008 & 0.024 & 0.073 & 0.106 & 0.269 & 0.286 & 0.233
\end{bmatrix}
\]

We calibrate the coinsurance rate for each of the seven shocks from the MEPS data, which is given as follows.

<table>
<thead>
<tr>
<th>Health spending</th>
<th>0.000</th>
<th>0.006</th>
<th>0.022</th>
<th>0.061</th>
<th>0.171</th>
<th>0.500</th>
<th>1.594</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi(m) )</td>
<td>0.341</td>
<td>0.532</td>
<td>0.594</td>
<td>0.645</td>
<td>0.702</td>
<td>0.765</td>
<td>0.845</td>
</tr>
</tbody>
</table>

Consistent with the data in section 2.2, the probability of providing EHI is increasing with firm size. In addition, administrative costs decrease with firm size.

<table>
<thead>
<tr>
<th>Firm size</th>
<th>0 &lt; 10</th>
<th>10 – 24</th>
<th>25 – 99</th>
<th>100 – 999</th>
<th>n &gt; 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_E(n) )</td>
<td>0.336</td>
<td>0.625</td>
<td>0.816</td>
<td>0.943</td>
<td>0.992</td>
</tr>
<tr>
<td>Administrative cost, ( g(n) )</td>
<td>0.3</td>
<td>0.21</td>
<td>0.132</td>
<td>0.0849</td>
<td>0.06</td>
</tr>
</tbody>
</table>

**Government:** The minimum consumption floor \( c \) is calibrated so that the model has 20% of households with net worth of less than $5,000 in the benchmark economy. The payroll tax is 12%. Lump-sum tax \( \tau_y \) is chosen in equilibrium to balance the overall government budget.
Table 2.2: Parameter values, baseline economy

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
<th>Description</th>
<th>Comments/observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.94</td>
<td>Discount factor</td>
<td>target K/Y ratio 2.42</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.3207</td>
<td>Capital share</td>
<td>target K share of 0.34</td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.5, 3</td>
<td>Risk aversion</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.4693</td>
<td>Frac. of entrepreneurs</td>
<td>Antunes et al. (2008)</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>-0.3667</td>
<td>Mean of distribution of $x$</td>
<td>mean size of firms</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>2.302</td>
<td>Std. dev of distribution of $x$</td>
<td>size distribution of firms</td>
</tr>
<tr>
<td>$m$</td>
<td></td>
<td>Health spending shock</td>
<td>MEPS</td>
</tr>
<tr>
<td>$\phi(m)$</td>
<td></td>
<td>Coinsurance rate</td>
<td>MEPS</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.1</td>
<td>Markup of health insurance</td>
<td>MEPS</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.8</td>
<td>Employer contribution of EHI</td>
<td>MEPS</td>
</tr>
<tr>
<td>$g(n)$</td>
<td></td>
<td>Cost of providing EHI</td>
<td>MEPS</td>
</tr>
<tr>
<td>$p_E(n)$</td>
<td></td>
<td>Probability of providing EHI</td>
<td>MEPS</td>
</tr>
<tr>
<td>$\hat{p}_E$</td>
<td>0.558</td>
<td>% covered by EHI</td>
<td>MEPS</td>
</tr>
<tr>
<td>$\zeta$</td>
<td></td>
<td>Consumption floor</td>
<td>20% hhs with wealth &lt; $5000</td>
</tr>
<tr>
<td>$\tau_s$</td>
<td>12%</td>
<td>Payroll tax</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>6%</td>
<td>Capital depreciation</td>
<td></td>
</tr>
</tbody>
</table>

2.6 Quantitative Analysis

In this section, we first present the performance of our benchmark model. We then explain the design of policy experiments. This is followed by a detailed analysis of our counter-factual policy experiments. Finally, we provide some remarks on our numerical exercises.

2.6.1 Baseline Economy

Our model succeeds in matching several aspects of the macroeconomy, including the distribution of firm size measured by the number of employees and observed patterns of health insurance coverage. Table 2.3 summarizes the performance of our model. In the benchmark, entrepreneurs account for 5.33% of the population, which is slightly below the target of 7%. This underestimate of entrepreneurship is attributed to the fact that our model of occupational choice does not account for other reasons that individuals choose to become entrepreneurs such as the utility.
Table 2.3: Benchmark

<table>
<thead>
<tr>
<th>Statistics</th>
<th>U.S. Data</th>
<th>Baseline Economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual real interest rate</td>
<td>4.0</td>
<td>4.33</td>
</tr>
<tr>
<td>Aggregate capital share</td>
<td>0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Capital output ratio</td>
<td>2.5</td>
<td>2.42</td>
</tr>
<tr>
<td>% of entrepreneurs</td>
<td>7.0</td>
<td>5.33</td>
</tr>
<tr>
<td>Mean size of the firm</td>
<td>17.09</td>
<td>17.76</td>
</tr>
<tr>
<td>% firm at 0-9</td>
<td>70.7</td>
<td>74.98 ($\bar{x}_1 = 1.55$)</td>
</tr>
<tr>
<td>% firm at 10-19</td>
<td>14.0</td>
<td>10.24 ($\bar{x}_2 = 2.05$)</td>
</tr>
<tr>
<td>% firm at 20-49</td>
<td>9.4</td>
<td>9.38 ($\bar{x}_3 = 2.38$)</td>
</tr>
<tr>
<td>% firm at 50-99</td>
<td>3.2</td>
<td>2.53 ($\bar{x}_4 = 2.82$)</td>
</tr>
<tr>
<td>% firm at 100+</td>
<td>2.6</td>
<td>2.87 ($\bar{x}_5 = 3.63$)</td>
</tr>
<tr>
<td>Health insurance take-up ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>75.7</td>
<td>73.75</td>
</tr>
<tr>
<td>EHI offered</td>
<td>99.0</td>
<td>97.9</td>
</tr>
<tr>
<td>EHI not offered</td>
<td>35.5</td>
<td>32.8</td>
</tr>
</tbody>
</table>

Note: The number in parenthesis is average ability level $x_i$ in each firm size group $i = 1, 2, 3, 4, 5$.

value from “being your own boss.”\textsuperscript{17} Hence our analysis provides a lower bound. On average, firms hire 17.76 employees in our benchmark, very close to 17.09 in the data. The model is also successful in reproducing the fraction of firms with the selected levels of employment. Average ability in each firm group increases with size, and firms in the largest size group are more than twice as productive ($\bar{x}_5 = 3.63$) as those in the smallest group ($\bar{x}_1 = 1.55$). In terms of health insurance coverage, our model has a take-up ratio of 73.75%, compared with 75.7% in the MEPS data.\textsuperscript{18} The take-up ratio is the share of agents with health insurance coverage.

\textsuperscript{17}De Nardi, Doctor and Krane (2007), table 2.1, find that entrepreneur’s earnings are 3 to 4 times the earning of others in the Survey of Consumer Finances. In our baseline economy the ratio is about 6 (see table 2.4 below). Our model underpredicts the fraction of entrepreneurs, which leads to a higher earnings ratio relative to SCF data.

\textsuperscript{18}Employment-based insurance involves three factors: a worker must be employed by a firm that offers coverage, the worker must be eligible for coverage, and the worker must choose to take-up coverage.
2.6.2 Policy designs

In this section we report the results of four policy experiments: (i) expand EHI from the current level of 62% to full coverage, (ii) replace EHI with private indemnity insurance, (iii) supplement the current system with health insurance exchanges, and (iv) replace the current EHI system with universal health insurance. Tables 2.4 and 2.5 report key statistics across the policy experiments.

2.6.2.1 Expansion of EHI

This experiment requires all firms to offer EHI, expanding the program from the 62% level in the data to 100% coverage of workers. The first two columns of Table 2.4 show that there is a tradeoff: When EHI is expanded to 100% more people are insured, and this makes agents more willing to bear the risk of entrepreneurship. This raises the cost of workers for firms. Recall (2.34) gives the average cost of providing insurance. This effect would tend to depress average firm size, which drops from 17.76 workers to 16.95 in table 2.5. On the other hand, all individuals now have insurance at low cost (taxes drop to 1.89% in table 2.4 from 2.2%), hence individuals have more funds to put in the firm. We should expect to see more entrepreneurs, and table 2.5 shows that the percentage of entrepreneurs increases from 5.33% in the baseline to 5.56%. Overall, we see more entrepreneurs running smaller firms that are less productive. The average ability for each size group, $\bar{x}$, is reported in parenthesis in table 2.5 and falls from $x_5 = 3.63$ for the largest firm group to $x_1 = 1.51$ for the smallest group. Table 2.5 shows that EHI expansion leads to a fall in the percentage of firms in the three highest groups (i.e., more small firms) and a decline in productivity of the smallest firm groups ($\bar{x}_1$ and $\bar{x}_2$ fall to 1.51 and 1.98 from the baseline values 1.55 and 2.05). Productivity falls because some individuals with lower managerial talent become entrepreneurs; they no longer need to either self-insure to cover medical shocks or buy more expensive health insurance, and they can use the funds to open firms.

2.6.2.2 No EHI: Private indemnity contract

This experiment considers the polar opposite case where there is no EHI and all insurance is purchased on the private market (if any). The take up rate falls from
the baseline level of 73.75% to 23.0% in table 2.4. This is not surprising since private insurance is disadvantaged relative to mandated EHI. The last column of table 2.5 shows that the percentage of entrepreneurs falls from the baseline level of 5.33% to 4.94% because risk has increased and the potential assets available to invest in the firm have decreased (most agents self insure). Average firm size increases to 19.26 from 17.76 workers and output per firm increases to 36.54 from 33.78. Overall, we see fewer entrepreneurs running larger firms that are more productive.

2.6.2.3 Health insurance exchange

Under the health insurance exchanges, such as those established in the ACA reform, individuals who are not covered by EHI can purchase insurance at subsidised premium rates that are independent of an individual’s health risk. The health insurance exchange changes the premium equations (2.17), (2.24) as follows:

\[
\tilde{\pi} = \begin{cases} 
\pi_{E} & i'_{HI} = 1, \overline{i}_{E} = 1 \\
\pi_{HIX}(1 - \tau_{HIX}) & i'_{HI} = 1, \overline{i}_{E} = 0 \\
0 & i'_{HI} = 0 
\end{cases}
\] (2.37)

The health insurance exchange premium is \(\pi_{HIX}\) and \(\tau_{HIX}\) is the subsidy rate from table 2.1. HIX does not involve firm administrative costs \(g(n)\), but instead specifies a loading cost: the ACA requires the medical loss ratio to be at least 80%, which translates into a upper bound on the loading cost of 0.25. The lump-sum tax increases relative to the baseline (3.06% versus 2.2% in table 2.4) and the percentage of entrepreneurs increases from 5.33% to 5.68% in table 2.5. There are significantly more very small firms, but average productivity within size groups does not change relative to the baseline.

2.6.2.4 Universal health insurance

In this experiment the government provides insurance as, for example, in Canada. The government now pays the individual’s premium and there are no subsidies. Recall that in the baseline the government defrays the worker’s cost of EHI by
Table 2.4: Aggregate variables

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Base</th>
<th>EHI exp</th>
<th>Universal</th>
<th>no EHI</th>
<th>HIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance take-up</td>
<td>73.75</td>
<td>97.56</td>
<td>97.55</td>
<td>23.0</td>
<td>97.36</td>
</tr>
<tr>
<td>real r (%)</td>
<td>4.33</td>
<td>4.30</td>
<td>4.34</td>
<td>4.30</td>
<td>4.38</td>
</tr>
<tr>
<td>wage</td>
<td>0.80</td>
<td>0.80</td>
<td>0.89</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td>Worker earnings</td>
<td>1.19</td>
<td>1.17</td>
<td>1.26</td>
<td>1.28</td>
<td>1.17</td>
</tr>
<tr>
<td>Entrepreneur earnings</td>
<td>7.76</td>
<td>7.42</td>
<td>8.07</td>
<td>8.36</td>
<td>7.29</td>
</tr>
<tr>
<td>Aggregate output</td>
<td>100</td>
<td>100.1</td>
<td>99.9</td>
<td>100.2</td>
<td>99.7</td>
</tr>
<tr>
<td>% at (\bar{c})</td>
<td>3.53</td>
<td>1.73</td>
<td>1.93</td>
<td>8.44</td>
<td>1.03</td>
</tr>
<tr>
<td>Welfare (%CEV)</td>
<td>-0.03</td>
<td>0.37</td>
<td>0.23</td>
<td>-1.60</td>
<td>-1.60</td>
</tr>
<tr>
<td>% with CEV(&gt;0)</td>
<td>41.69</td>
<td>68.71</td>
<td>61.98</td>
<td>2.62</td>
<td>2.62</td>
</tr>
<tr>
<td>tax/earn %</td>
<td>2.2</td>
<td>1.89</td>
<td>6.64</td>
<td>3.40</td>
<td>3.06</td>
</tr>
</tbody>
</table>

providing a subsidy given by the last term in (20), the worker’s budget constraint. See (21) and recall that \(\psi\) is the employer’s contribution to the EHI premium. Under universal health insurance the government provides the insurance directly and pays for it through the tax system, eliminating the firm’s part of insurance payments and firm administrative cost \(g(n)\). Entrepreneur earnings increase because the health care burden is taken “off the backs” of employers. Worker earnings drop net of taxes because taxes as a percentage of earnings increase from 2.2% in the baseline to 6.64% under universal insurance (see table 2.4). The take up rate is nearly 100%, as expected. Table 2.5 shows that entrepreneurs fall from the baseline of 5.33% to 5.14%, and firm size increases from 17.76 to 18.47, and there are less firms in the smallest size group.

### 2.6.3 Size distribution

Table 2.5 shows how the alternative policies affect the size distribution of firms. EHI expansion, universal insurance and the private insurance indemnity (no EHI) reduce the percentage of smallest firms (0-9 employees) but expand the next group (10-19 employees). This group’s productivity falls from \(x_2 = 2.05\) to 1.98. HIX increases the size of smallest group, lowers the next three groups relative to the baseline, and slightly decreases the largest group from 2.87% in the baseline to 2.69%, with little change in \(\bar{x}\). One of the points of our analysis is that in a model with heterogeneity, averages and coarse firm bin sizes can mask important individ-
ual changes. Presumably the goal is not to increase the number of entrepreneurs, but rather to maximise consumption. This goal is accomplished by allocating individuals and capital to their most productive use. We now consider welfare analyses at the individual level to evaluate the consumption gains and losses from policy changes.

Table 2.5: Policy experiments

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Base</th>
<th>EHI exp</th>
<th>Universal</th>
<th>no EHI</th>
<th>HIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>K/Y</td>
<td>2.42</td>
<td>2.43</td>
<td>2.42</td>
<td>2.43</td>
<td>2.41</td>
</tr>
<tr>
<td>Aggregate output</td>
<td>100</td>
<td>100.1</td>
<td>99.9</td>
<td>100.2</td>
<td>99.7</td>
</tr>
<tr>
<td>Entrepreneur %</td>
<td>5.33</td>
<td>5.56</td>
<td>5.14</td>
<td>4.94</td>
<td>5.68</td>
</tr>
<tr>
<td>Ave x (all firms)</td>
<td>1.77</td>
<td>1.76</td>
<td>1.79</td>
<td>1.81</td>
<td>1.75</td>
</tr>
<tr>
<td>Output per firm</td>
<td>33.78</td>
<td>32.29</td>
<td>35.05</td>
<td>36.54</td>
<td>31.61</td>
</tr>
<tr>
<td>Output per worker</td>
<td>1.90</td>
<td>1.91</td>
<td>1.90</td>
<td>1.90</td>
<td>1.90</td>
</tr>
<tr>
<td>Ave firm size</td>
<td>17.76</td>
<td>16.95</td>
<td>18.47</td>
<td>19.26</td>
<td>16.60</td>
</tr>
<tr>
<td>% firm at 0-9</td>
<td>74.98 ((\bar{x}_1 = 1.55))</td>
<td>68.86 ((\bar{x}_1 = 1.51))</td>
<td>66.19 ((\bar{x}_1 = 1.52))</td>
<td>64.83 ((\bar{x}_1 = 1.54))</td>
<td>76.50 ((\bar{x}_1 = 1.55))</td>
</tr>
<tr>
<td>% firm at 10-19</td>
<td>10.24 ((\bar{x}_2 = 2.05))</td>
<td>17.02 ((\bar{x}_2 = 1.98))</td>
<td>18.47 ((\bar{x}_2 = 1.98))</td>
<td>19.21 ((\bar{x}_2 = 1.98))</td>
<td>9.63 ((\bar{x}_2 = 2.05))</td>
</tr>
<tr>
<td>% firm at 20-49</td>
<td>9.38 ((\bar{x}_3 = 2.38))</td>
<td>8.96 ((\bar{x}_3 = 2.38))</td>
<td>9.73 ((\bar{x}_3 = 2.38))</td>
<td>10.13 ((\bar{x}_3 = 2.38))</td>
<td>8.78 ((\bar{x}_3 = 2.38))</td>
</tr>
<tr>
<td>% firm at 50-99</td>
<td>2.53 ((\bar{x}_4 = 2.82))</td>
<td>2.42 ((\bar{x}_4 = 2.82))</td>
<td>2.63 ((\bar{x}_4 = 2.82))</td>
<td>2.74 ((\bar{x}_4 = 2.82))</td>
<td>2.38 ((\bar{x}_4 = 2.82))</td>
</tr>
<tr>
<td>% firm at 100+</td>
<td>2.87 ((\bar{x}_5 = 3.63))</td>
<td>2.74 ((\bar{x}_5 = 3.63))</td>
<td>2.98 ((\bar{x}_5 = 3.63))</td>
<td>3.10 ((\bar{x}_5 = 3.63))</td>
<td>2.69 ((\bar{x}_5 = 3.63))</td>
</tr>
</tbody>
</table>

Note: The number in parenthesis is average ability level \(x_i\) in each firm size group \(i = 1, 2, 3, 4, 5\).

2.6.4 Welfare analysis: conditional change

We measure the welfare cost or gain of a specific health policy by how much lifetime consumption, in percentage terms, an agent in state \((a, x, z, m, i_{HI})\) would gain or lose if the agent lived through the transition to the new policy, compared to the initial steady-state. Put differently, we ask how much an agent with wealth-productivity tuple \((a, x, z, m, i_{HI})\) in the initial steady-state would be willing to give up as a percentage of lifetime consumption to avoid the reform. This is a conditional welfare change because it is computed for an individual in a particular state. The welfare change is the consumption-equivalent variation (CEV), where the amount that an agent would pay to avoid the reform is the \(\varpi(a, x, z, m, i_{HI})\) that solves the equation:

\[
E_0\sum_{t=0}^{\infty}\beta^t u ([1 + \varpi(a, x, z, m, i_{HI})] c_t) = E_0\sum_{t=0}^{\infty}\beta^t u (\hat{c}_t)
\]
\( c_t^* \) denotes consumption in the initial state, while \( \hat{c}_t \) is consumption under the new policy. For the case of CRRA preferences, \( u(c) = \frac{c^{1-\sigma} - 1}{1-\rho} \), we can exploit the homogeneity of the utility function and the solution to the above equation is given by

\[
\varpi(a, x, m, i_{HI}) = \left[ \hat{V}(a, x, z, m, i_{HI}) + \frac{1}{(1-\rho)(1-\beta)} \right]^{\frac{1}{1-\sigma}} - 1.
\]

The conditional welfare change is computed for an individual that is in a particular state, thus we consider welfare plots for various states \( \varpi(a, x, z, m, i_{HI}) \).

Figure 2.5: Conditional welfare change, EHI expansion

Figure 2.5 shows the conditional welfare change for EHI expansion to 100% coverage. We consider two health shocks, high and low, when we expand EHI relative to the three insurance states for the individual: uninsured, baseline EHI insurance, and private insurance. The figure shows that expanding EHI increases the conditional welfare of high ability individuals (especially with high assets), and leads to welfare losses for low ability and poor agents. When the medical shock is high and individuals have baseline EHI or private insurance, we see that there are some welfare gains for the very poor with high ability, but overall EHI expansion largely favours high ability, high asset individuals because lump sum
taxes are inconsequential for these agents. Table 2.4 shows that the lump sum taxes required to fund the EHI expansion program are lower (1.89%) than in the baseline case (2.2%), thus expanded EHI reduces the tax on earnings. The policy benefits individuals with high ability and low assets because they now have insurance and more resources to invest in their firm. Table 2.4 shows that the earnings of entrepreneurs are much higher than the earnings of workers, and expanded EHI reduces the risk of health shocks. As a consequence, members of this high ability, low asset group may now switch their occupation from worker to entrepreneur. Finally, table 2.4 shows that when individual gains and losses are summed over all agents there is essentially no net welfare change, with 41.69% of individuals having a positive welfare gain (CEV > 0). The figure shows the distribution of gains and losses vary greatly, which a utilitarian sum treats equally.

Figure 2.6: Conditional welfare change, HIX

Figure 2.6 shows the conditional welfare change for the health insurance exchange policy (HIX). We again consider two health shocks, high and low, and introduce HIX relative to the three insurance states for the individual: uninsured, baseline EHI insurance, and private insurance. The figure shows that the distribution is fairly flat, but losses are tilted toward low ability and low asset individuals. This is the case because low ability agents will generally be workers, whether they
have high or low assets. Table 2.4 shows that workers’ wages and earnings fall under HIX and taxes rise as a percentage of earnings, which explains their welfare losses. Entrepreneur’s earnings also fall. Aggregate welfare losses are -1.60 under HIX in table 2.4, with only 2.62% of individuals with a positive welfare gain (who have extremely low or high assets).

Figure 2.7: Conditional welfare change, Universal Health Insurance

Figure 2.7 shows the conditional welfare change for universal health insurance. Again there are two health shocks, high and low, and we introduce universal health insurance relative to the three insurance states for the individual: uninsured, baseline EHI insurance, and private insurance. Overall, the figure shows that there are more positive welfare values relative to the previous two policies and losses are tilted toward low ability and low asset individuals. In general, high ability and high asset individuals gain from the policy. Low ability and low asset individuals suffer losses relative to the EHI baseline because now there is no subsidy and the lump sum taxes necessary to fund universal health insurance are high (6.64%) relative to the baseline (2.2%). Table 2.4 shows that there is an aggregate welfare gain of 0.37 and 68.71% of individuals have a positive welfare gain.
Figure 2.8 shows the conditional welfare change for the experiment where the current EHI baseline is dropped and there is only private insurance. Again there are two health care shocks, high and low, and we introduce private insurance (only) relative to the individual’s three insurance states: uninsured, baseline EHI insurance, and private insurance. This policy gives an aggregate welfare gain of 0.23 in table 2.4, and 61.98% of people have positive welfare gains. Consistent with the observed pattern, this policy produces welfare gains for high ability and high asset individuals, and the gains can be substantial for some. The subsidies help the very poor, but overall the policy reduces welfare for lower ability and asset individuals.

2.6.5 Welfare and risk aversion: stationary distribution

Figure 2.9 shows the welfare of all individuals for all four policy changes when $\rho = 1.5$ and figure 2.10 shows the results when $\rho = 3$. As we would expect, when $\rho$ increases from 1.5 to 3 welfare falls.
When $\rho = 1.5$, welfare is positive under universal health insurance except for the very poor with low ability. The program is expensive (the tax is 6.64% in...
and the very poor now pay the lump sum tax for insurance when they previously had subsidies in the baseline. This causes a decline in welfare for this group. Under the HIX policy virtually all of the very poor with low ability receive positive welfare gains ex ante due to the subsidised insurance. Higher ability and asset (except for the very top) individuals suffer declines in welfare because the program increases their taxes. EHI expansion has essentially no net welfare change. Now all individuals have access to EHI at lower cost, but the very poor already had a consumption floor through social insurance. Finally, most agents have welfare small losses under the private insurance indemnity, but a few have large gains. This occurs because few agents choose to buy private insurance, but the few poor agents with insurance benefit greatly. Overall, the welfare losses are largest for poor and low ability agents.

When agents are relatively tolerant to risk ($\rho = 1.5$), some are willing to accept the protection provided by the social insurance that gives consumption floor $c$. The take up rate is very low under private insurance, and many individuals choose to remain uninsured because private insurance is relatively expensive. When agents are more risk averse ($\rho = 3$), figure 2.10 shows that welfare switches to significantly negative values for the poor under universal insurance and HIX, and table 2.6 shows that taxes are high. Some individuals now value (slightly) the expanded EHI insurance regime, which provides insurance at the lowest cost. Under private insurance, the take-up rate is slightly higher when $\rho = 3$ and the poor value insurance because it is difficult for them to self-insure.

### 2.6.6 Policy Summary

The four distinct healthcare policies that we analysed provide insight into the recent U.S. Affordable Care Act (ACA), a reform with three stated goals: to increase the quality and affordability of health care, lower the rate of uninsured, and reduce the cost of healthcare. The policies we consider: (i) expand the current EHI system to 100% coverage, (ii) replace EHI with private insurance (only), (iii) add a health care exchange (HIX) to EHI, and (iv) replace EHI with universal insurance (only). Tables 2.4, 2.5 and 2.6 and the figures show that significant differences exist among the four policies relative to the baseline economy.
Table 2.6: Aggregate variables, $\rho = 3$

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Base</th>
<th>EHI exp</th>
<th>Universal</th>
<th>no EHI</th>
<th>HIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>K/Y</td>
<td>3.02</td>
<td>3.01</td>
<td>3.04</td>
<td>3.05</td>
<td>3.02</td>
</tr>
<tr>
<td>real r (%)</td>
<td>2.68</td>
<td>2.70</td>
<td>2.65</td>
<td>2.62</td>
<td>2.69</td>
</tr>
<tr>
<td>Aggregate output</td>
<td>100</td>
<td>100.1</td>
<td>100.3</td>
<td>100.4</td>
<td>100.1</td>
</tr>
<tr>
<td>Insurance take-up</td>
<td>76.52</td>
<td>99.98</td>
<td>99.98</td>
<td>39.17</td>
<td>97.13</td>
</tr>
<tr>
<td>Entrepreneur %</td>
<td>5.60</td>
<td>5.89</td>
<td>5.38</td>
<td>5.22</td>
<td>5.71</td>
</tr>
<tr>
<td>Ave $x$</td>
<td>1.73</td>
<td>1.70</td>
<td>1.74</td>
<td>1.75</td>
<td>1.74</td>
</tr>
<tr>
<td>Ave firm size</td>
<td>16.86</td>
<td>16.05</td>
<td>17.61</td>
<td>18.15</td>
<td>16.52</td>
</tr>
<tr>
<td>% at $\bar{c}$</td>
<td>2.75</td>
<td>1.18</td>
<td>7.11</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>Welfare</td>
<td>-0.08</td>
<td>-2.14</td>
<td>2.42</td>
<td>-2.06</td>
<td></td>
</tr>
<tr>
<td>% with CEV$\geq0$</td>
<td>38.42</td>
<td>35.51</td>
<td>93.5</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>tax/earn %</td>
<td>1.66</td>
<td>1.18</td>
<td>7.86</td>
<td>2.57</td>
<td>3.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Base</th>
<th>EHI exp</th>
<th>Universal</th>
<th>no EHI</th>
<th>HIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>% firm at 0-9</td>
<td>76.0 ($\bar{x}_1 = 1.53$)</td>
<td>70.34 ($\bar{x}_1 = 1.45$)</td>
<td>67.47 ($\bar{x}_1 = 1.48$)</td>
<td>66.67 ($\bar{x}_1 = 1.46$)</td>
<td>76.64 ($\bar{x}_1 = 1.54$)</td>
</tr>
<tr>
<td>% firm at 10-19</td>
<td>9.8 ($\bar{x}_2 = 2.05$)</td>
<td>16.11 ($\bar{x}_2 = 1.98$)</td>
<td>17.70 ($\bar{x}_2 = 1.98$)</td>
<td>18.11 ($\bar{x}_2 = 1.98$)</td>
<td>9.56 ($\bar{x}_2 = 2.05$)</td>
</tr>
<tr>
<td>% firm at 20-49</td>
<td>9.0 ($\bar{x}_3 = 2.38$)</td>
<td>8.51 ($\bar{x}_3 = 2.38$)</td>
<td>9.35 ($\bar{x}_3 = 2.38$)</td>
<td>9.61 ($\bar{x}_3 = 2.38$)</td>
<td>8.77 ($\bar{x}_3 = 2.38$)</td>
</tr>
<tr>
<td>% firm at 50-99</td>
<td>2.4 ($\bar{x}_4 = 2.82$)</td>
<td>2.28 ($\bar{x}_4 = 2.82$)</td>
<td>2.50 ($\bar{x}_4 = 2.82$)</td>
<td>2.55 ($\bar{x}_4 = 2.82$)</td>
<td>2.35 ($\bar{x}_4 = 2.82$)</td>
</tr>
<tr>
<td>% firm at 100+</td>
<td>2.7 ($\bar{x}_5 = 3.63$)</td>
<td>2.60 ($\bar{x}_5 = 3.63$)</td>
<td>2.85 ($\bar{x}_5 = 3.63$)</td>
<td>2.93 ($\bar{x}_5 = 3.61$)</td>
<td>2.68 ($\bar{x}_5 = 3.61$)</td>
</tr>
</tbody>
</table>

Note: The number in parenthesis is average ability level $x_i$ in each firm size group $i = 1, 2, 3, 4, 5$.

Our model incorporates distortions in the U.S. economy that we take as given. First, the ACA mandates benefits, requiring insurance premiums based on a community rating rather than individual risk characteristics, and coverage with minimum standards. Our expanded EHI experiment achieves 100% coverage by design, and under HIX and universal insurance nearly all individuals choose to have insurance. Second, there is a credit market friction, which we model as a standard interest rate wedge that is common across all policies. Third, EHI is advantaged relative to the other policies because in the U.S. EHI enjoys favourable tax treatment, has a more inclusive risk pool, and has administrative economies of scale.

Given the environment, our results indicate that universal insurance and private indemnity insurance (no EHI) give the highest welfare gains (0.37% and 0.23% in table 2.4) when risk aversion is 1.5, and only private insurance gives positive net welfare when $\rho = 3$ (2.42% in table 2.6). Of course, due to heterogeneity, the policies have very different effects at the individual level.

91
Consider first the insurance indemnity. This policy replaces EHI with an indemnity contract under which the insurance provider pays for the individual’s health expenditures. This contract is efficient by design, thus it is not surprising that it delivers positive net welfare gains. The take up ratio is 23% in table 2.4 and 39.17% in table 2.6 for this ex ante insurance contract, and significant ex post insurance occurs in the form of a higher $\bar{c}$ of 8.44% when $\rho = 1.5$ (table 2.4) and 7.11% when $\rho = 3$ (table 2.6) that is paid for through the tax system when agents are hit with bad medical shocks. Notably, capital increases under the indemnity when agents are more risk averse ($\rho = 3$), which allows individuals to both better self insure and expand firm size. The increase in firm size is evident in table 2.6 under the indemnity, where the smallest firm size declines from 74.98% to 66.67%, and all other firm sizes increase.

Tables 2.4 and 2.6 show that expanded EHI is the least costly policy in terms of taxes because administrative economies of scale lead to the lowest tax burden and the percentage of people requiring social insurance falls from 3.53% to 1.73% when $\rho = 1.5$ and to 2.75% when $\rho = 3$. In addition to being tax advantaged, EHI has a better risk pool and lower administrative costs than private insurance.

Health care exchanges are a key feature of the ACA. Thus our HIX policy considers the case where individuals who do not have EHI can purchase HIX at subsidies given by the ACA (see table 2.1). A wedge exists between EHI and HIX because the exchanges are effectively subsidised private insurance.\(^{19}\)

Under universal health insurance households pay the highest lump sum tax for insurance and more people are driven to social insurance ex post than under expanded EHI ($\bar{c}$ is 1.93% under UI but only 1.73% under expanded EHI). The reason for the increase in $\bar{c}$ is that higher taxes cause saving to decrease, and lower savings then cause more people with adverse shocks to hit the consumption floor because they are less able to absorb shocks.

\(^{19}\)In our model if EHI is not offered, then the individual gets additional monetary compensation to buy HIX or private insurance, which are not perfect substitutes for EHI.
2.7 Conclusion

This paper identifies a new friction and shows how alternative health care policies affect the macroeconomy and welfare. When insurance is mandated, talent misallocation can occur: Some individuals with high managerial talent but poor health become workers, while other individuals with moderate managerial talent but good health become entrepreneurs. Because entrepreneurs create jobs, the misallocation of a few key individuals affects the broader macroeconomy - firm size, output and wages. Understanding the nature of this misallocation is important because poorly designed health care policies can exacerbate distortions instead of correcting them.

Two extensions would provide additional insight into the analysis. First, we consider lump sum taxes because they do not distort choices. However, lump sum taxes have a large effect on welfare because the tax is more burdensome to poor agents than to rich. Clearly alternative tax policies would change the results. Progressive taxes could attenuate some of the extreme welfare gains of high asset individuals under some of the policies and raise the welfare of lower asset agents by lowering their taxes. However, progressive taxes would distort occupational choice by lowering entrepreneur’s earnings.

For example, under expanded EHI we found that higher ability agents enjoy the largest individual welfare gains. This “better treatment of high ability agents” is a standard result in optimal taxation for efficiency reasons – expanding the tax base by encouraging more productive individuals to work more permits marginal rates on less productive individuals to be lowered. In our model the analogue is that it is more efficient for higher ability individuals to run larger firms, ceteris paribus, and they must be compensated to do this. See Scheuer (2014) for an analysis of optimal taxation and entrepreneurship.

Another extension involves the labour market. As in Jeske and Kitao (2009), we assume a segmented labour market where employers do not adjust wages if EHI coverage is declined. Instead we could consider perfect compensation substitutability where workers sort to employers based on the demand for health insurance. This labour market structure might reduce occupational misallocation since a healthy agent can sort to a firm that offers monetary compensation but no EHI. Misallocation will continue to exist, however, as long as private health
insurance is not a perfect substitute for EHI.

Nevertheless it would be interesting to see how a different market structure affects occupational misallocation. A model that considers search and matching in the labour market could address this issue. For example, Fonseca, Lopez-Garcia and Pissarides (2001) provide a good benchmark, but they do not consider the impact of health policy on firm and employment decisions. Incorporating a search friction into our model would also allow us to analyse the interaction between entrepreneurship and unemployment. These issues go beyond the current paper and we leave them for future research.
References


Appendix

Derivation of $k^*$

**Unconstrained firm**  When initial assets are sufficient to run a business without resorting to credit finance (i.e., $l = 0$), the manager of the firm solves the problem:

$$\nu^j_i(a, x, i_E; w, r) = \max_{k \geq 0} y_i(k_i, x, w) - rk - \varphi \quad (2.38)$$

where $\varphi = [i_E (w + (1 + \lambda^j q_A - \psi) c_E) + (1 - i_E) (w + b_E)] \left[\frac{(1-\alpha)x^k}{i_E(w+(1+\lambda^j q_A-\psi)c_E)+(1-i_E)(w+b_E)}\right]^{\frac{1}{1-\gamma}}$

denotes the labour cost.

Substituting $n^j_i$ into profits $\nu^j_i$ gives

$$\nu^j_i(a, x, i_E; w, r) = (1-\gamma)(xk)^{\frac{\gamma}{\gamma-1}} \left[\frac{\gamma}{i_E (w + (1 + \lambda^j q_A - \psi) c_E) + (1 - i_E) (w + b_E)}\right]^{\frac{\gamma}{1-\gamma}} \quad (2.39)$$

$$k^*_i(x, w) = \left[ x \left(\frac{\gamma}{i_E (w + (1 + \lambda^j q_A - \psi) c_E) + (1 - i_E) (w + b_E)}\right) \right)^{\gamma} \left(\frac{\alpha}{r}\right)^{1-\gamma} \quad (2.40)$$

From equation (2.39), the manager’s profit at the optimal level of capital is:

$$\nu^j_i(k^*_i, x, w) = xk^\alpha \left[\frac{\gamma x^k}{i_E(w+(1+\lambda^j q_A-\psi)c_E)+(1-i_E)(w+b_E)}\right]^{\frac{\gamma}{\gamma-1}} -$$

$$[i_E (w + (1 + \lambda^j q_A - \psi) c_E) + (1 - i_E) (w + b_E)] n(k^*, x, w) - rk^*$$

The manager’s consumption is determined as follows.

$$c + a' + (1 - i_H)\tilde{m} + \tilde{\pi} \leq (1 + r)a + \nu_i(k^*, x, w) - Tax + TSI + \tau s i_E \pi E \quad (2.41)$$
where
\[
\tilde{\pi} = \begin{cases} 
\pi_E & i_{HI} = 1, i_E = 1 \\
\pi_P(m) & i_{HI} = 1, i_E = 0 \\
0 & i_{HI} = 0
\end{cases}
\] (2.42)

\[T_{SI} = \max \{0, \xi + Tax + \tilde{\pi} - \tau_s i_E \pi_E + (1 - i_{HI} \phi) m + (1 + r)(k - a) - \nu_i(k^*, x, w)\} \] (2.43)

\[a' \geq -\bar{a}. \] (2.44)

\[l \leq (1 - \Delta) \frac{\nu_i(a, x, w) + rk^*}{1 + r} - oop \] (2.45)

Note \(\tilde{\pi}\) is the amount that the manager pays for insurance, \(i_{HI}\) is the entrepreneur’s choice to buy health insurance for himself for next period, and \(i_E\) is the shock (whether the employer must provide insurance to employee). The government subsidises EHI purchases with \(\tau_s i_E \pi_E\). Equation (2.45) is a credit constraint for the firm, where \(oop\) is the out-of-pocket health shock of the entrepreneur and is defined as
\[oop = (1 - i_{HI} \phi(m)) m. \] (2.46)

Notice \(\nu_i(a, x, w) + (1 + r)k^*\) works as collateral, which yields the present value of the firm’s earning net of labour cost. We assume that there is a proportional cost of borrowing, which is represented by \((1 - \Delta)\). This constraint introduces interesting dynamics as the entrepreneur’s health insurance decision will affect its future available credit.

**Constrained firm** When managers do not have enough funds to operate the firm, they can borrow from the capital market at the risk free rate \(r\). However, they can borrow up to a limit of \(\bar{l}\). If the optimal level of capital \(k^*\) can be financed by borrowing, then the firm’s problem will be similar to the unconstrained one.

When managers are credit constrained, namely \(a + \bar{l} < k^*\), the firm will operate at the capital level of \(a + \bar{l}\). Notice the borrowing limit \(\bar{l}\) is endogenous, see equation (2.45). Accordingly, the credit constrained firms have borrowing that is determined
by the equation as follows.

\[ \tilde{\nu}^j(\tilde{k}, x, w) = x\tilde{k}^\alpha \tilde{n}^\gamma - i_E \left( w + \left( 1 + \lambda^j q_A - \psi \right) q_E \right) + (1 - i_E) (w + b_E) \hat{n} - ra - (1 + r)\bar{l} \]  

(2.47)

where

\[ \tilde{k}^j = a + \bar{l} = a + \frac{\tilde{\nu}^j(\tilde{k}, x, w) + ra + (1 + r)\bar{l}}{(1 + r)} (1 - \Delta) - oop \]  

(2.48)

\[ \tilde{n}^\gamma(\tilde{k}, x, w) = \left[ \frac{(1 - \alpha)x^\gamma \tilde{k}^\alpha}{i_E (w + \left( 1 + \lambda^j q_A - \psi \right) q_E) + (1 - i_E) (w + b_E)} \right]^{\frac{1}{1 - \gamma}}. \]  

(2.49)

Hence the credit constrained firms differ in their own capital holdings.

\[ \tilde{k}^j = \begin{cases} k^j & \text{if } a \geq k^* - \frac{\tilde{\nu}^j(\tilde{k}, x, w) + ra + (1 + r)\bar{l}}{(1 + r)}(1 - \Delta) + oop \\ a + \frac{\tilde{\nu}^j(\tilde{k}, x, w) + ra + (1 + r)\bar{l}}{(1 + r)}(1 - \Delta) - oop & \text{if } a < k^* - \frac{\tilde{\nu}^j(\tilde{k}, x, w) + ra + (1 + r)\bar{l}}{(1 + r)}(1 - \Delta) + oop \end{cases} \]

where \( \tilde{k}^j \) is the solution to equation (2.47).

**Proofs of propositions**

The proof follows Antunes, Cavalcanti and Villamil (2008b).

**Computation**

Given are values for parameters, the distribution \( \Gamma(x) \) for \( x \), \( \Omega_z \) for \( z \), \( \Omega_a \) for \( a \), and \( \Omega_m \) for \( m \). The numerical algorithm works as follows.

1. Set a tolerance \( \epsilon > 0 \).
2. Guess \( \Phi^0 = (r^0, w^0, \pi_E^0, \bar{p}_E^0, \tau_y^0) \). Solve for optimal household behaviour:

\[ f : (\theta; \Phi) \rightarrow (c, a', i'_{HI}, \mathcal{I}_e, n, k), \]

where \( \theta = \{a, x, z, m, i_E, i_{HI}\} \). We will use the method of value function iteration as follows.
(a) Guess value function $V^0(\theta; \Phi^0)$ and policy functions $f^0(\theta; \Phi^0)$.

(b) Update value and policy functions:

$$V^1(\theta; \Phi^0) = \max_{a', i'_{HI}, i'_c} \{I_c V_{c} + (1 - I_c) V_{w} + \beta \mathbb{E} [V^0(\theta'; \Phi^0)] \}$$

$$f^1(\theta; \Phi^0) = \arg \max V^1(\theta; \Phi^0)$$

(c) Stop if $\max \{|V^1 - V^0|, |f^1 - f^0|\} \leq \epsilon$. Otherwise, set $V^0 = V^1$, $f^0 = f^1$ and repeat step (b).

(d) Set $V^* = V^1$, and $f^* = f^1$.

3. Generate a large number of individuals, $N = 100000$. For each agent $j$ assign a vector of initial condition $(a^0_j, x^0_j, z^0_j, m^0_j, i^0_E, i^0_{HI}, 0, i^0_{HI})$, where $x^0_j \sim \Gamma(x)$, $z^0_j \in \Omega_z$, $m^0_j \in \Omega_m$, $i^0_{HI} = 0$.

4. Simulate the economy for $T$ periods, where $T$ is sufficiently large.

5. Calculate the following statistics from the simulated path $\{a^0_t, x^0_t, z^0_t, m^0_t, i^0_{IE,t}, i^0_{HI,t}, I_{we,t}, I_{ce,t}, n^0_t, k^0_j \}_{t=0}^T$:

$$LS^0 = \frac{\sum_{j=1}^N (I_{we} - I_{ce} n^0_t)}{N}$$

$$KS^0 = \frac{\sum_{j=1}^N (a^0_t - I_{ce} k^0_t)}{\sum_{j=1}^N a^0_t}$$

$$\pi^1_E = \frac{\sum_{j=1}^N (i^0_{IE} i^0_{HI} m^0_t)}{\sum_{j=1}^N (i^0_{IE} i^0_{HI})}$$

$$\hat{p}^1_E = \frac{\sum_{j=1}^N (I_{ce} p^0_t(n^0_t) n^0_t)}{\sum_{j=1}^N (I_{ce} n^0_t)}$$

and $\tau^1_y$ that balances the government’s budget.

6. Stop and set $(r^*, w^*, \pi^*_{E}, \hat{p}^*_{E}) = (r^0, w^0, \pi^0_{E}, \hat{p}^0_{E})$, if $\max \{LS^0, KS^0, |\pi^1_E - \pi^0_E|, |\hat{p}^1_E - \hat{p}^0_E|\} \leq \epsilon$.
\( \epsilon \). Otherwise, update aggregate variables (restart from step 2):

\[
\begin{align*}
\tau^0 &= \chi \tau^0 + (1 - \chi) \rho K S^0 \\
w^0 &= \chi w^0 + (1 - \chi) \rho L S^0 \\
\pi_E^0 &= \chi \pi_E^0 + (1 - \chi) \pi^1_E \\
\hat{p}_E^0 &= \chi \hat{p}_E^0 + (1 - \chi) \hat{p}_E^1 \\
\hat{r}_y^0 &= \chi \hat{r}_y^0 + (1 - \chi) \hat{r}_y^1
\end{align*}
\]

where \( \chi \in (0, 1) \) is the step for updating aggregate variables.
Summary and Conclusions

This thesis has been a collection of essays that focus on the role of risk on the decision to invest in wealth-enhancing investment. We have shown how the effect of uncertainty can be sufficiently strong that the preference to avoid an uncertain outcome alone can result in intergenerational persistent inequality. We have also shown how various government policies to insure against risk can have varying welfare effects on individuals with different managerial talents and different initial asset levels.

In chapter 1 we showed that although models which utilise capital market imperfections form the cornerstone of the modern theoretical inequality and growth literature, they do so without considering the impact of uncertainty on the agent’s decision making process. As the majority of investment projects that individuals can undertake to move higher up in the income distribution are inherently risky, capital market imperfections may not be enough to explain the formation and persistence of inequality within a market economy. We have argued that when agents are faced with such an investment, they are loss averse.

As a response to these observations, we presented a model where aspirationally-induced loss-averse agents face a decision to invest in human capital by borrowing from financial intermediaries. Although the pay-off for investing in human capital is beneficial on average, the outcome of the investment is unknown. Due to agents’ aversion to this uncertainty, agents with sufficiently low initial human capital are deterred from investing due to their relatively high susceptibility to fall below the individuals reference point. The probability of this is higher the lower the agents initial value of human capital as individuals do not have sufficient human capital to buffer themselves against a bad outcome. It transpired that the model containing loss aversion is isomorphic to a model with an imperfect credit market.

In chapter 2 we argued that although it is unquestionable that providing high quality healthcare is important to an economy, the structure of the health insurance system may affect the allocation of talent. For example, if Employer-sponsored Health Insurance (EHI) raises the level of wealth required to become an entrepreneur, then some of the most talented individuals may be deterred from starting a firm. If this is the case, the question becomes, do the benefits
of employer-based health insurance outweigh the potential negative cost of talent misallocation? This essentially comes down to the relative efficiency of employer-provided health insurance compared to other health insurance systems: which system can provide the highest quality healthcare at the lowest cost to society.

Subsequently we studied the quantitative effects of four policy experiments: the expansion of employer-based health insurance; private insurance; health insurance exchanges; and universal health coverage. Rather than simply looking at the rate of entrepreneurship take-up that results from the policy experiments we show that different policies have different outcomes for the overall macroeconomy due to talent misallocation. Further to this, we were able to specify the welfare effects of the different policy experiments for individuals across various levels of ability and health shocks.