Essays on Information, Asset Bubbles, and Economic Policy



Andrea de Mauro

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

Abstract

This work consists of three self-contained essays tackling independent research questions pertaining to information constraints, asset price bubbles, economic policy, or a combination thereof.

The first chapter tackles the problem of growing income inequality by weighing the effectiveness of an array of fiscal policies currently available to policy-makers. To do so, I evaluate the impact of alternative forms of public expenditure in an overlapping generations framework where households' ability to invest in the education of their offspring is constrained by various sources of heterogeneity. I find that the optimal use of public funds consists in a combination of public grants aimed at improving educational quality and an increase in the upper limit of student loans. The subsequent panel data analysis corroborates my results and suggests that parental income remains the largest determinant of earnings and inequality.

In Chapter 2, I turn my attention to the challenges posed by imperfect information about the source of asset price movements. I develop a New Keynesian model where stock prices contain a fundamental and a bubble component, both unobservable to agents and the central bank, who are forced to speculate about their value. I use this model to analyse the implications for investors' behaviour and monetary policy. I find that, under perfect information, a systematic increase in interest rates in response to growing bubbles exacerbates market volatility and welfare losses. For high levels of noise, the central bank should reverse its policy stance. However, due to the continuous underestimation of the size of bubbles, policy responses are perpetually inadequate.

Finally, Chapter 3 explores the extent to which cognitive biases can explain market anomalies such as the formation of large financial bubbles. To this end, I develop a stochastic bubbles framework with imperfect information about the probability of a future market crash, where investors are overconfident about the accuracy of their knowledge. I put forward a number of generalisable propositions about the relationship between overconfidence and bubbles in general equilibrium frameworks. In particular, I find that overconfidence can play a key role in the existence of bubbles and is able to rationalise excessive market movements. I quantify my results and test the validity of the framework in a numerical analysis.

To my parents

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Francesco Zanetti, for his patience, guidance, and support. Without him, this thesis would not have seen the light of day. I also want to thank my MPhil supervisor, Professor Christopher Adam, who has been of great help during my first years at Oxford.

Many others have assisted in this work, starting with professors and colleagues in the Macroeconomics group of the Department of Economics, who have provided invaluable advice and guidance throughout the years. I am also especially grateful to Professor Xiaowen Lei for her generosity and for always making herself available to discuss my ideas at length.

Thank you to my family for their constant love and support, and for always believing in me. They have made this and so much more possible. I must also thank my friends, many of whom I have made along the way, for being there for me in the best of times, as well as the hardest.

Finally, I want to express my appreciation for the students and staff at Mansfield and St Antony's Colleges for offering me a home away from home for all of these years.

Table of Contents

1 Fiscal Policy and Educational Inequality **1**

1.1 Introduction 2

1.2 Literature Review 5

1.2.1 Educational Quality and Income Inequality 11

1.2.2 Liquidity Constraints and Income Inequality 18

1.3 Theoretical Framework 22

1.3.1 A Model of Income Inequality and Educational Heterogeneity . . . 24

1.3.1.1 Households 24

1.3.1.2 Firms 28

1.3.1.3 Government 29

1.3.1.4 Income Inequality 31

1.3.2 Extending the Framework to Account for Liquidity Constraints . . 32

1.3.2.1 Introducing Constrained Agents and Government Student Loans Programs 32

1.3.2.2 GSL Upper Limit and Government Budget 33

1.3.3 Calibration 34

1.4 Theoretical Results 36

1.4.1 Starting Point: A Basic Setup 37

1.4.1.1 The Impact of Idiosyncratic Shocks 37

1.4.1.2 Policy Shocks Under Educational Homogeneity and Perfect Liquidity Access 39

1.4.2 The Role of Liquidity Constraints 41

1.4.2.1 Uniform Policy Shocks With Liquidity Constraints 41

1.4.2.2 Asymmetric Policy Shock With Liquidity Constraints . . 44
1.4.3 The Role of Educational Heterogeneity 46
1.4.3.1 Uniform Policy Shocks Under Educational Heterogeneity 46
1.4.3.2 Asymmetric Policy Shocks Under Educational Heterogeneity 49
1.4.4 Accounting for Multiple Sources of Income Inequality 51
1.4.4.1 Effectiveness of Uniform Policies in the Complete Setup 51
1.4.4.2 Effectiveness of Asymmetric Policies in the Complete Setup 53
1.4.5 Implications 55
1.5 Empirical Analysis 56
1.5.1 Data 56
1.5.2 Empirical Strategy 59
1.5.2.1 Step One: State-Level Panel Analysis 60
1.5.2.2 Step Two: Effects on Federal-Level Income Inequality . . 63
1.5.3 Empirical Results 64
1.6 Concluding Remarks 69
Appendix 1.A Tables 80

2 Dealing with Uncertainty: Speculative Bubbles and Monetary Policy 83
2.1 Introduction 84
2.2 The Model 90
2.2.1 Households 91
2.2.2 Firms 94
2.2.3 Monetary Policy 95
2.2.4 Equilibrium 96
2.2.4.1 Market Clearing 96
2.2.4.2 Steady State 98
2.2.4.3 Log-linearised Equilibrium 99
2.2.5 Bubbles, Fundamentals, and Signal Extraction 99
2.2.5.1 Underlying Dynamics and Signaling 100
2.2.5.2 Signal Extraction 102
2.2.5.3 Closed-Form Solutions 104

2.3 Impulse Response Analysis . . . 105
2.3.1 Calibration . . . 106
2.3.2 Bubble and Fundamental Shocks . . . 107
2.4 Monetary Policy Design . . . 112
2.4.1 The Central Bank's Dilemma . . . 113
2.4.2 Monetary Policy in an Uncertain World . . . 116
2.5 Concluding Comments . . . 120
Appendix 2.A Kalman Filter . . . 125
2.A.1 Full Information Equivalent . . . 126

3 Stock Market Overconfidence and Rational Bubbles 128
3.1 Introduction . . . 129
3.2 Evidence on Overconfidence and Stock Market Bubbles . . . 137
3.2.1 Evidence on Market Efficiency . . . 137
3.2.2 Evidence of Investors' Overconfidence . . . 139
3.2.3 Evidence on the Effects of Overconfidence on Stock Market Bubbles 142
3.3 The Model . . . 144
3.3.1 Households . . . 145
3.3.1.1 Information Structure . . . 146
3.3.1.2 Maximization Problem . . . 151
3.3.2 Firms . . . 152
3.4 Equilibrium . . . 153
3.4.1 Market Clearing . . . 154
3.4.2 Balanced Growth Path . . . 156
3.5 The Impact of Overconfidence . . . 159
3.5.1 Revisiting the Conditions for Bubble Existence . . . 159
3.5.2 Comparative Statics . . . 163
3.6 Information, Dynamic Efficiency, and Welfare . . . 169
3.6.1 Overconfidence and Pre-Crash Efficiency . . . 170
3.6.2 Welfare Analysis . . . 172
3.7 Numerical Analysis . . . 176
3.7.1 Calibration . . . 176

3.7.2 Overconfidence in a Simulated Economy 178
3.7.3 Empirical Strategy and Results 180
3.8 Concluding Remarks 186
Appendix 3.A Truncated Normal Distribution 197
Appendix 3.B Proofs 197

Chapter 1

Fiscal Policy and Educational Inequality

Chapter Abstract

This paper provides a theoretical and empirical framework to assess the effectiveness of alternative forms of public expenditure aimed at reducing income inequality. We evaluate the impact of fiscal policy shocks in a two-period overlapping-generations model in which altruistic parents invest in the education of their children. Agents may differ in a number of ways, which affects their capacity to complete such investment, and it is shown that simultaneously accounting for various sources of household heterogeneity is crucial to determining the net effect of fiscal measures. When differences in educational quality are introduced, subsidies aimed at reducing tuition fees are revealed to exacerbate the dispersion of the income distribution by excessively benefitting students having access to better schooling. Improving educational quality through public grants, even uniformly, has a substantial negative impact on inequality, and combining this with student loans is shown to be the optimal use of public funds. The subsequent panel data analysis of income disparities across U.S. states confirms our results and reveals that parental income remains the largest determinant of earnings and inequality.

1.1 Introduction

Income inequality has been at the forefront of public and academic debates for decades, both in developing and industrialised economies. In particular, after a significant fall in the post-World War II period, many developed countries have experienced a fast rise in income disparities in the past thirty years. The United States has been especially affected by this trend, with inequality today reaching levels that had not been seen since the 1930s. This has given rise to a renewed focus on the issue in the economic literature in recent years, which in turn has led a number of policy-makers in the U.S. to declare income inequality one of the greatest challenges facing the country (Reeves and Rodrigue, 2014).

Nevertheless, inequality remains largely misunderstood and economists are still seeking to identify the fundamental drivers behind its rise in recent decades. This has led policy-makers to form misinformed proposals which, when implemented, have largely failed to decelerate the expansion of earning gaps, let alone reverse it. One policy around which a consensus seems to have formed in the policy field is to increase educational attainment through a homogeneous decrease in tuition fees, the hope being that a fall in educational gaps would reduce income disparities at a later stage in life. Nevertheless, little to no evidence has been presented to support such policy, and while the intuition behind it seems correct, other factors may come into play and ultimately produce unexpected effects. More generally, the literature on the causes of inequality seems to be segregated into different strands that have largely ignored the possible inter-connections between various factors and how this may affect their policy conclusions. Furthermore, these policies are often intuitively inferred from these studies rather than formally tested. Hence, it is not only crucial to first identify the possible roots of the rise in income inequality to form sound policy recommendations, but it is just as important to consider them jointly in order to formally understand the ultimate

impact of the proposed measures.

It therefore remains to be determined what the most efficient policies to reduce rising income disparities are. This is the question this paper seeks to answer. It does so by jointly considering the two major drivers of inequality identified in earlier studies – namely, the increasing limitations of children from poorer backgrounds to access secondary and tertiary education and the widening heterogeneity in the quality of education offered by schools – and by formally evaluating the effectiveness of key policies in mitigating their consequences.

This paper is formed of two core parts. Firstly, we built a theoretical framework to consider how the impact of various fiscal policies may be affected by the successive introduction of educational heterogeneity and liquidity constraints. We do so by examining fiscal policy shocks in a two-period overlapping generations model where agents accumulate human capital when young and invest in the education of their children when adult. Liquidity constrained agents, who are unable to complete such investment, are then introduced, and we subsequently consider, for the first time in a theoretical framework, heterogeneity in the quality of schooling by allowing agents to receive different levels of educational quality. In each case, we evaluate the impact of the policies put forward by the literature on income inequality, namely, government grants directly allocated to schools to improve the quality of the education they provide – from the educational quality strand of the literature – as well as public subsidies aimed at reducing tuition fees and government student loans – taken from the liquidity constraints literature.

Secondly, we seek to empirically test the findings from our theoretical model by evaluating the effects of the aforementioned policies in a panel data analysis that takes advantage of time and spatial variations in U.S. state-level data, and controls for various other factors while addressing a number of methodological issues raised in earlier studies. We first

estimate these effects on state-level median income before using inequality decomposition techniques to identify causal mechanisms.

Our theoretical and empirical analyses point to similar results. Firstly, our theoretical framework allows us to show that ignoring qualitative educational heterogeneity leads to misleading conclusions. In fact, a setup where agents all receive the same quality of education will indicate that increasing public subsidies or student loans will reduce income inequality while improving educational quality will be entirely inefficient. This holds true both when we consider liquidity constraints and when we don't, since in the first case subsidies improve access to education for constrained agents, while in the second case subsidies reduce the reliance of parents' educational investment on idiosyncratic shocks. Introducing qualitative educational heterogeneity will nearly reverse our conclusions. In fact, public subsidies will now be shown to exacerbate income disparities, as agents from higher educational backgrounds will be able to save proportionally more and will therefore expand their educational investment by a larger margin than poorer parents. This result also holds when constrained agents are considered. On the other hand, improving the quality of education, even uniformly, will benefit poorer agents to a larger extent, therefore reducing income inequality. The beneficial effect of student loans is preserved. Secondly, our empirical analysis corroborates the above results. Public expenditure aimed at improving educational quality largely contributes to decreasing income inequality, while government subsidies have the opposite effect. Most of all, however, parents' earnings remain the largest driver of income inequality, despite the policies put in place in the past.

The rest of the paper is divided as follows. Section 1.2 reviews the leading causes of income inequality identified in the literature, most significantly liquidity constraints and educational quality. Section 1.3 outlines our theoretical framework, whose results are pre-

sented in Section 1.4. Section 1.5 describes the data and empirical strategy we use, as well as our empirical results. Finally, Section 1.6 concludes.

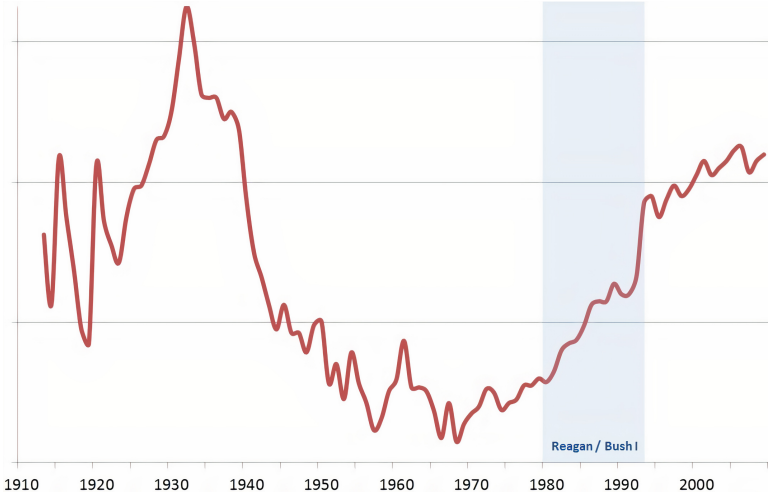
1.2 Literature Review

The rise in income inequality experienced by industrialised economies in the past three decades has sparked a vast literature on the causes and consequences of economic disparities and how to tackle them. Prior to that, in the 1960s and 1970s, the consensus among economists was largely influenced by Simon Kuznets' (1955) theory, which stated that rising economic inequalities were only a temporary feature of industrializing economies and that market forces would eventually allow those at the bottom of the income distribution to catch up through a rebalancing of sectors, without the need for government intervention. The popularity of this idea was anchored in the falling disparities experienced since the Great Depression, which seemed to confirm the inherent dynamic relationship between economic growth and income inequality (Figure 1.1).

Since the second half of the 1970s and, more significantly, since the 1980s, the fast and relentless pick up in the growth of disparities has led to a progressive but complete reconsideration of the mechanisms behind inequality. The idea of the existence of a Kuznets curve has now been rejected by most economists, as some of the most industrialised economies have seen the largest increases in disparities, while several developing economies have been able to combine industrialization with a reduction in income gaps. In particular, the U.S. has seen the share of income of the top decile go from an average of 33% in the 1950s-70s to nearly 50% in 2010 (Saez, 2013).

In addition, while prior to the 1980s high levels of inequality had often been seen as

Figure 1.1: Evolution of Income Inequality in the United States From 1913-2009



Source: Berruyer (2010), adapted from U.S. Census Bureau

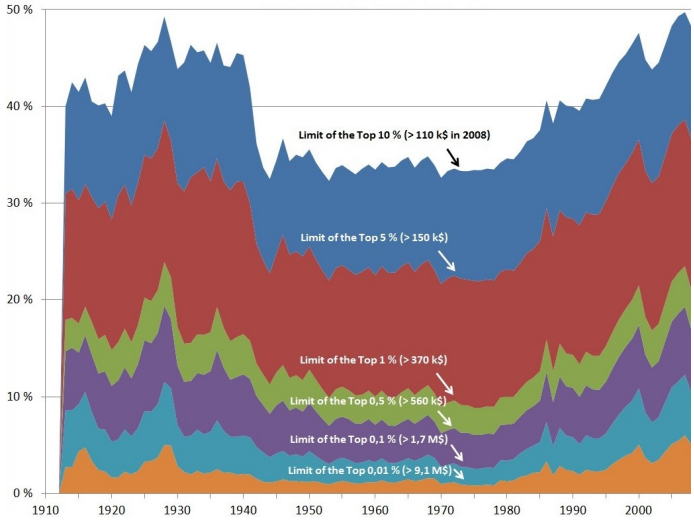
a necessary, temporary, and unavoidable cost allowing nations to industrialise and reach higher standards of living that could eventually be shared more equally amongst the population, the rejection of the Kuznets curve sparked a debate as to whether and how much inequality is desirable. In fact, some inequality is often seen as necessary in order to create the right incentives to promote risk-taking, investment, education, and innovation (Castells-Quintana and Royuela, 2017). However, a large body of literature has convincingly shown that high levels of inequality, such as those seen in the vast majority of advanced and developing economies, including the United States, not only hurt those at the bottom of the distribution, but makes the entire population worse-off. In fact, inequality has been shown to reduce human capital due to important adverse health and educational effects, resulting in slower economic growth (Karlsson et al., 2010; OECD, 2013; OECD, 2014). A second negative impact on economic growth comes from the fact that wealthier individuals have a lower marginal propensity to consume, resulting in higher savings and lower aggregate demand (Bivens, 2017). Inequality has also been shown to exacerbate economic and political instability through increased polarisation and social unrest, as well as reduced social cohe-

sion (Satz and White, 2021). Another unintended consequence that has been highlighted in recent years is the negative impact inequality has on climate change, which in turn is likely to excessively affect the most vulnerable parts of the population and worsen economic disparities even further (Hsiang et al., 2017). In fact, it has been suggested that the resulting vicious cycle may become one of the biggest challenges governments will be faced with in the years to come. Another strand of this literature has also sought to establish that lower inequality and higher growth need not be mutually exclusive outcomes. This paper contributes to this debate by showing that this may indeed not be the case following the implementation of certain inequality-reducing policies, while other measures may give rise to a trade-off between the two objectives.

Furthermore, it has been shown that the fall in income inequality that followed the Great Depression and World War II was not the result of a reduction in structural disparities, but rather of the fiscal and regulatory policies put in place at the time and reversed under the Reagan and Bush administrations (Piketty, 2014). Consequently, it has become widely accepted that policy intervention is needed in order to limit or reverse the polarization of the income distribution. Nevertheless, identifying the optimal policies to achieve this requires an understanding of the fundamental drivers behind recent trends. The vast literature that has developed in recent years has gone in many different directions, sending mixed signals to policy-makers.

One of the most popular early hypotheses regarding the roots of inequality, which overwhelmingly remains a consensus among policy-makers, is that gaps in educational attainment are one of the leading drivers of gaps in earnings, if not the main one. Most of the policies aiming to reduce inequality implemented in the United States, as well as through development programs around the world, have sought to improve school enrolment rates

Figure 1.2: Evolution of the Income Shares of the Top Decile in the United States From 1913-2008

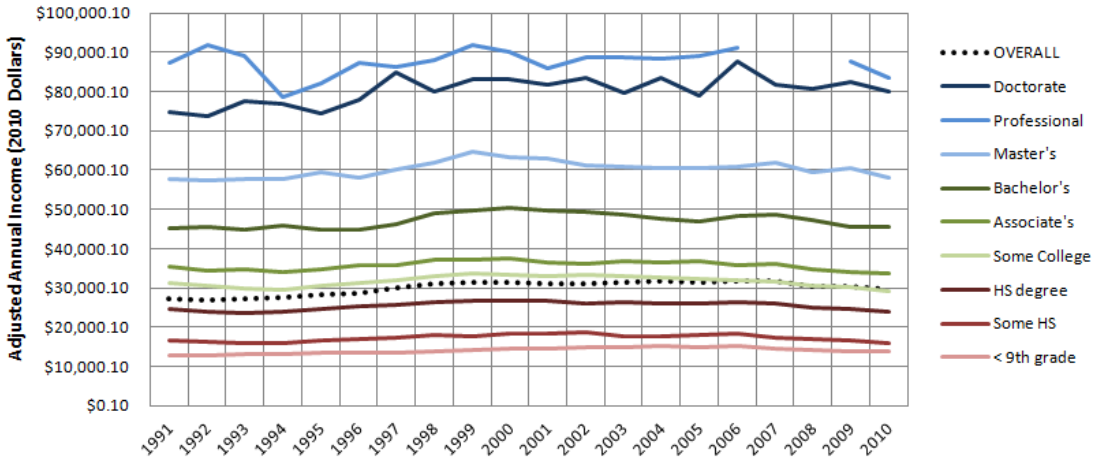


Source: Berruyer (2010), using data from Piketty and Saez

(Hanushek and Wößmann, 2007). This has been successfully achieved, as the share of 25 year-olds and over holding a high school degree in the U.S. went from about half of the population in the early 1970s to nearly 90% in 2014, and the share holding a bachelor's degree or more reached one third in 2014, up from less than 10% in the 1970s – although this is partly the result of a general trend (U.S. Census Bureau, 2014). Nevertheless, despite this, income inequality has kept rising over the same period, therefore raising doubts that education has any role to play in the reduction of income gaps.

In fact, some economists have used this observation to argue that what has driven the past trends in inequality is that a few influential people such as CEOs hold so much power that they have been able to attract most of the new wealth created over the years, and that education has no role in determining the evolution of earnings disparities (Krugman 2015; Piketty and Saez, 2014; Klein, 2015; den Haan, 2011; Syll, 2014). This can however be questioned by looking at the evolution of the decomposition of the share of income held by the top decile. Indeed, Figure 1.2 shows that, while it is true that the very top of the income distribution has seen a large increase in its share since the 1980s, the top 10% altogether

Figure 1.3: Evolution of Median Income by Educational Attainment in the United States From 1991-2010



Source: U.S. Census Bureau

and all of its sub-groups experienced similar trends over the same period. By contrast, the bottom 90% saw their share fall, with all sub-groups either stagnating or falling for the most part (Alvaredo et al., 2016). Therefore, there must be other factors driving income disparities than CEOs and a few influential people concentrating all of the gains from economic growth in their hands.

Others have argued that rising disparities have been driven by skill-biased technological change (SBTC). Namely, a boom in technological change has led to an increase in demand for better-educated workers, resulting in a relative increase in their wages compared to their lower-skilled counterparts (see for instance Autor et al., 1998; Goldin and Katz, 2008). A crucial issue with this hypothesis is that it is not been corroborated by key wage patterns over the last decades (Card and DiNardo, 2002; Schmitt et al., 2013; Krugman, 2015). In particular, the idea that a rising wage gap between different levels of education has driven rising inequality over the last 30 years clashes with the evidence that median income by educational attainment has stagnated for all education levels over the same period (Figure 1.3). As a response to this criticism, alternative formulations of the SBTC hypothesis have

been put forward. The most recent one has argued that computerisation has led to the polarisation of the job market, whereby higher- and lower-skilled workers have benefited at the expense of the middle of the labour force (see for instance Acemoglu and Autor, 2012). Nevertheless, this hypothesis has again suffered a number of criticisms, including weak evidence of U-shaped job polarisation and the fact that it struggles to explain wage patterns since the 1990s (see Schmitt et al., 2013, for a review of the criticisms). While we do not argue that skill-biased technological change did not play any role in inequality trends over the last decades, the lack of conclusive evidence suggests that other important factors are at play.

Two major strands of the literature on the causes of rising inequalities stand out and are increasingly becoming widely accepted among academics, while corroborating the above observations. The first one focuses on educational quality, arguing that the reason for the inefficient role of shrinking disparities in educational attainment to reduce income gaps is the fact that schooling quality also matters (Livingstone, 1997; Rao and Jani, 2008). In addition, the stagnating trends of median income uncovered in Figure 1.3 – which have been used to suggest education did not play any role in driving inequality – can also be explained by the fact that, within each educational level, disparities in educational quality have been rising (Blau and Kahn, 2005; Hanushek and Wößmann, 2007). The second major strand has focused instead on the limited access to education for children from disadvantaged backgrounds, which they argue is reflected in the high stickiness levels of social mobility at the bottom end of the distribution, where nearly half of the children born in families from the bottom quintile remain there when they become adults (Isaacs, 2008; Reeves and Joe, 2016). This is due to the stringent credit and liquidity constraints faced by poorer households, which affect their capacity to invest in the education of their children and alter the

distributions of skills, income and wealth across generations (Lochner and Monge-Naranjo, 2011*a*). As discussed below, the importance of credit constraints has been significantly rising since the 1980s, becoming a major driver of income disparities and of their transmission across generations (Cameron and Heckman, 1998, 1999; Belley and Lochner, 2007). What's more, access to high-quality education is even less accessible to poorer children, a phenomenon which may have been further exacerbated by the increase in the degree of educational quality (OECD, 2016).

These two strands of the literature seem to have a complementary relationship in explaining the trends observed in income inequality, with the rising heterogeneity in the quality of education seemingly explaining the rise in the income share of the top decile, while the increasing importance of credit and liquidity constraints explaining the fall in the share of the bottom of the distribution. However, to the best of our knowledge, their inter-connection has never been explored. We review below both the theoretical and empirical contributions of these two strands, as well as the extent to which their policy conclusions have been formally tested.

1.2.1 Educational Quality and Income Inequality

Evidence of the high and expanding degree of heterogeneity in the quality of education in the United States has been documented across the literature. These have not only been recorded through the rise of shorter, and generally much less valued alternatives to four-year bachelor's programs such as associate degrees, but also within each type of studies (Haskins and Sawhill, 2008). These growing educational gaps have been identified by many economists as playing a key role in the evolution of earnings disparities in recent decades. We start by reviewing the different measures of quality adopted by the literature and the

degree to which they are inter-changeable, before outlining the main findings on the links between schooling quality and earnings, and schooling quality and income inequality. Finally, we review what has been done to identify policies that could reduce income disparities through the educational quality channel.

The literature on educational quality can broadly be divided into two main groups, defined by their measure of quality, with one group focusing on the output side of education, the other on its inputs. The first looks at students' competences, quantified through standardised tests undertaken identically across different regions, which seek to measure core abilities in key areas such as mathematics, reading or science. These surveys include the OECD's Programme for International Student Assessment (PISA), which is run every three years on fifteen-year-olds across various countries, the International Adult Literacy Survey (IALS), also collected mostly in advanced economies, and the Trends in International Mathematics and Science Study (PIRLS). While these studies have the advantage of using generally more detailed micro-datasets, they often fail to track individuals over time, therefore making it difficult to uncover relationships between schooling and earnings or income inequality, as well as the impact of policies, as they often resort to examining cross-sectional correlations at given points in time (Checchi and van de Werfhorst, 2014; see also Solga, 2014). On the other hand, the second group has been focusing on the inputs made available to schooling facilities, such as pupil-to-teacher ratios and teachers' experience, salary, or education (Hanushek, 1986; Eide and Showalter, 2005; Liu, 2001). These measures allow for longer time dimensions, more precise measurements of educational quality, and the tracking of specific establishments or regions over time. Despite this, the use of this type of data has been relatively less explored than the previous one, and this paper seeks to illustrate the usefulness of looking at schooling inputs when assessing policies.

The validity of the use of schooling resources measures in analyzing the relation between educational quality and earnings or inequality relies on the positive association between students' competences and schooling facilities. In other words, it must first be verified that schooling inputs can effectively be translated into educational outputs, as otherwise any correlation between earnings and inputs may simply be the result of an endogenous process. This has already been accomplished by various earlier papers. In fact, since the influential Coleman report (Coleman, 1966), which argued that schooling facilities are ineffective in raising students' competences, a number of studies have contradicted this, arguing that the two are highly significantly and positively correlated (Welch, 1966; Watchel, 1976; Rizzuto and Watchel, 1980). Card and Krueger (1992) have argued that this contradiction may be explained by the fact that standardised tests were somewhat imprecise measures of educational quality that have improved over time. Nowadays, we should therefore be confident in using either standardised tests or schooling facilities as measures of quality.

Having said this, we are now able to review the literature having examined the relation between schooling quality – measured as both schooling facilities and students' competences – and earnings. Understanding this is the first step to examining the relation between quality and earnings disparities. The link between educational attainment and income has been extensively covered through the Mincerian literature, which investigates earnings as being determined by years of schooling and experience (Mincer, 1974). Studies that have sought to integrate educational quality into their analysis of income have done so by extending Mincerian equations, therefore remaining within the Mincerian approach and adding a measure of schooling quality (Glomm, 1997). Using a similar framework, a significant amount of papers in the early literature have concluded that educational quality plays a crucial role in determining earnings at an adult age, therefore contradicting the second con-

clusion from the Coleman report (1966) that the two are unrelated (Welch, 1966; Morgan and Sirageldin, 1968; Johnson and Stafford, 1973; Watchel, 1976; Rizzuto and Watchel, 1980). Nevertheless, Card and Krueger (1992) raised a fundamental endogeneity issue that all of these papers had ignored. In fact, the positive relation between schooling facilities and earnings may simply be driven by parents' earnings, which may both raise the amount of resources allocated to their children's schools and correlate with their offspring's income given the stickiness in social mobility. It is therefore necessary to control for parents' earnings to avoid this endogenous process, which we do in this analysis. Their findings clearly indicate that doing so does not alter the conclusion that the inter-state variation in returns to education are substantially caused by differences in educational quality, with individuals having attended establishments with higher teacher-pupil ratios and teachers' pay earning more in their adulthood. Nevertheless, neither Card and Krueger (1992) nor any of the earlier authors take their analysis further by deriving the effects of schooling quality on median income or income inequality, both of which we undertake in this analysis.

In fact, while there has been a significant focus on the relation between quality and earnings, much of the empirical literature has avoided going one step further to estimate the links between schooling quality and income disparities. This is mostly due to the methodological difficulties involved in doing so, as applying inequality decomposition techniques often requires a number of conditions to be met (Checchi and van de Werfhorst, 2014). Most of the studies having tried to estimate this relationship have done so by regressing inequality measures – most commonly Gini coefficients – on the dispersion of test scores as their measure of quality, therefore avoiding decomposition methods. Their results have nevertheless been very encouraging. Blau and Kahn (2005), for instance, find that an increase in test scores by one standard deviation in the United States results in wages being on average 10%

and 8% higher for men and women, respectively. The rest of the literature has mostly focused on cross-sectional results based on a similar approach, with Freeman and Schettkat (2001), Hanushek and Wößmann (2007), and Solga (2014) all finding that differences in the dispersion of educational quality are able to explain a significant amount of cross-country variations in income inequality.

Two important issues ignored in these analyses are raised by Checchi and van de Werfhorst (2014). First, there is a potential endogeneity issue when using data on students' competences, as test scores may be driven by unobservable effects such as children's competitiveness or talent, which prompts them to use an instrumental variables approach. Secondly, students' competences are typically measured in the same period as earnings, which presents obvious inference issues, and is caused by a lack of data and tracking across time. They tackle this by adopting a country/cohort analysis, matching data on income inequality and education based on cohorts' year of birth. Their findings indicate that the impact of educational quality on income inequality is twice as large as that of educational attainment.

Nevertheless, their empirical strategy still relies on regressing the dispersion of their education measure on income inequality, therefore avoiding the use of inequality decomposition techniques. While these methods have been used in studies unrelated to educational quality (see Morduch and Sicular, 2002), the present literature has had to resort to much weaker identification strategies due to its use of test scores as a measure of quality, which leads to the presence of unobservable effects that inevitably bias decomposition results. Our use of state-level data allows us to avoid issues related to unobservable effects in our policy analysis, therefore enabling us to undertake an inequality decomposition approach for the first time in this literature. Furthermore, given the availability of this type of data across long time periods, we are able to look at education and earnings measures in their relevant

periods of collection by appropriately lagging variables. We can also control for parents' earnings using the same approach, which Checchi and van de Werfhorst (2014) and other studies were unable to achieve.

The theoretical literature on educational quality, on the other hand, remains extremely scarce. In fact, overlapping-generations (OLG) and dynamic stochastic general equilibrium (DSGE) models taking into account the role of education usually do so using the Mincerian approach, just like in the empirical literature. Human capital is most commonly modelled as depending on inborn abilities and years of schooling, and agents are sometimes allowed to accumulate human capital across time or inherit that of their parents (Ben-Porath, 1967; Heckman, 1999; Dabla-Norris and Matovu, 2002). Households' income is then made dependent on human capital, and inequality measures can in turn be derived if individuals are heterogeneous, which can be achieved through the introduction of idiosyncratic shocks and/or agent disaggregation (Troch, 2014). However, very little has been done to build an extended Mincer equation incorporating educational quality – for notable exceptions, see for instance Glomm and Ravikumar (1998) – and, to the best of our knowledge, no study has evaluated the consequences of accounting for heterogeneity in the quality of education, let alone its impact on income disparities and on inequality-reducing policies. We therefore contribute to this area of research through our theoretical framework, whose focus on quality, inequality, and inequality-reducing policies is predominantly new.

Finally, we turn our attention to studies having considered policies seeking to tackle income inequality by addressing the schooling quality dispersion channel. Most of the literature has not sought to formally investigate the impact of such policies, but rather has resorted to inferring them from the above studies. What's more, no study has sought to explore how policy conclusions may be impacted when considering several potential drivers of income

inequality. While a number of studies have attempted to empirically identify the effects of various policy changes relating to educational quality, most of them have limited themselves to looking at how these might affect educational attainment or achievement. Furthermore, given the emphasis of most studies on students' competences rather than schooling facilities, there has been a lack of focus on the effects of public expenditure on education. In fact, most policies that have been considered are institutional reforms such as changes in the compulsory start and end ages of primary and secondary education, the age at which students are forced to start specializing, teachers' autonomy in setting learning programs, the use of standardised methods of testing, and so on (Braga et al., 2011). The rare studies having considered their impact on income inequality have shown that some of them are able to effectively reduce it, while others may worsen it (Checchi and van de Werfhorst, 2014). None of these studies has however been able to use inequality decomposition in their analysis.

While these reforms are obviously important factors affecting the efficiency of students' learning experiences, they do not capture the fundamental aspect determining the quality of education offered by schools, namely governments' allocation of public funds to educational establishments, which requires them to identify the most efficient policies to achieve their set goals given budgetary constraints. We therefore expand on the empirical literature by considering fiscal educational policies and their impact not only on educational attainment and quality, but also, ultimately, on median income and income inequality. Furthermore, we contribute to the theoretical literature by constructing a framework allowing us to examine the role of policies affecting schooling quality, and to compare their efficiency in reducing income disparities to that of alternative fiscal measures.

1.2.2 Liquidity Constraints and Income Inequality

Credit and liquidity constraints have been increasingly identified as playing a key role in driving the evolution of inequality by creating an opportunity gap between children born at the bottom of the income distribution and their peers, due to the inability of their parents to invest in their secondary and tertiary education (Carneiro and Heckman, 2002). One branch of the literature has therefore focused on the impact of credit constraints on educational attainment, although less attention has been given to their effect on income inequality specifically.

The empirical literature on credit and liquidity constraints has sought to estimate their importance in determining school attendance, using two main approaches (Lochner and Monge-Naranjo, 2011*a*). The first one, which has been far more popular, consists in evaluating the role of parental income in the determination of schooling attainment and/or earnings. The second has sought to infer credit constraints from the gap between instrumental variables (IV) estimates of returns to education and ordinary least squares estimates (Lang, 1993; Card, 1999). However, this approach has been heavily criticised given the endogeneity issue raised by self-selection and the likelihood of IV estimates being higher regardless of constraints (Carneiro and Heckman, 2002). We therefore focus on the first method. The main issue in using it is untangling the contribution of other factors, such as racial discrimination in the case of earnings or ability, from that of parental income (Braga et al., 2011).

A number of studies have sought to uncover this relationship, but their results have drastically changed over time, indicating a rise in the importance of social background in determining college attendance and earnings. In fact, studies undertaken for the United States using data from the 1980s or earlier have found no significant relationship between parental income and schooling attendance, after controlling for a number of potentially biasing fac-

tors (Card and Krueger, 1992; see Cameron and Heckman, 1998, 1999; Carneiro and Heckman, 2002). On the other hand, later studies that used data from the 1990s and early 2000s found a much stronger and significant effect of social background. For instance, Belley and Lochner (2007) find that, after controlling for factors such as ethnicity and adolescent cognitive results, individuals from richer backgrounds had a significant 16 percentage-points higher probability of attending college than youth from poorer backgrounds, which corresponds to a doubling from the 1980s estimate. Black and Devereux (2010) find similar results. While the link between credit constraints and income inequality could be inferred from this, none of these studies seeks to quantify nor formally verify this relationship, which is something this paper will seek to achieve.

Credit and liquidity constraints have also been vastly explored in the DSGE literature, as they provide a valuable and realistic means to introduce non-Ricardian behavior, which is necessary to analyse the role of fiscal policies (Roeger and in't Veld, 2009). Nevertheless, despite the opportunity they offer, constraints have never been exploited in a theoretical framework, as far as we know, to analyse their impact on income inequality nor to evaluate policies aiming to address them. We contribute to this literature by providing a framework to accomplish this, and we illustrate how OLG-based DSGE models provide a useful approach for analyzing fiscal policies and their impact in the presence of heterogeneous agents.

Policies aiming to reduce the role of parental income in determining school attendance, earnings, and inequality, revolve around the issue of student financing, often in the context of tertiary education. Policy instruments available to the government can be separated in two categories: public subsidies aiming to partly or fully finance tuition fees, therefore reducing the costs privately incurred by students or parents, and government student loans (GSL) programs, which provide direct financing to students.

Three main combinations of these two policies can be used by the government, as discussed in Chapman (2006). First, the government can decide to fully (or almost fully) subsidise tuition fees, therefore leaving little need for student loans to cover education costs. This is the option chosen by many European countries and advocated by a number of prominent policy-makers in the U.S., although its social equitability has been questioned, since subsidies usually affect all students homogeneously. The opposite solution would be to let students incur all costs privately, and for the government to simply ensure the repayment of private student loans in case of default, although it has been shown that such policy significantly reduces school attendance because of debt aversion. The third, more balanced option would be to provide partial subsidies to reduce tuition fees while providing government student loans. This is what the United States has chosen and, to a fuller extent, a number of other countries such as the UK, Australia, and Germany. In these countries, GSL programs have mostly been provided on an income-contingent basis, meaning that students are only required to start repaying their loans once they are employed.

In the United States, public lending programs, which form the majority of funding sources, are characterised by an upper limit on the amount that each student is able to access, and a much stronger ability to enforce repayments than in the private sector (Lochner and Monge-Naranjo, 2011*b*). Nevertheless, the upper limit on per-student funding has barely progressed in the last decades, resulting in a large and increasing fraction of students reaching their upper limit, hence explaining the failure of GSL programs to reduce the importance of credit constraints in school attendance since the 1990s (Berkner, 2000; Titus, 2002). We consider in our analysis the effects of raising the GSL upper limit and increasing public subsidies.

Another important issue that the government needs to consider is how to deal with

agents' heterogeneity in their abilities and social background. GSL programs are able to target their loans to specific groups, but they are only able to address one of these two issues at a time, with merit-based programs tackling the former and need-based loans focusing on the latter (Boldrin and Montes, 2004). The U.S. generally gives more importance to need-based programs, given that public subsidies are too low for poorer students to afford tertiary education by their own means (Belley et al., 2011). Subsidies, on the other hand, are mostly administered uniformly. This also applies to policies aiming to alter educational quality, as public grants are sometimes provided uniformly and sometimes targeted on schools that have fallen behind. In our theoretical analysis, we analyse and contrast the effects of public subsidies, grants and loans, both uniformly – for the first two policies only – and asymmetrically – for all three policies.

The need to consider these three policies both in the context of educational heterogeneity and credit constraints is further highlighted by the strengthening interdependence between parental income and educational quality. While in the 1980s this link was found to be nearly inexistent, using data from the 1990s-2000s revealed that individuals born in the top quartile were 11% more likely than those from the bottom quartile to attend a four-year college institution (Belley and Lochner, 2007; Haskins and Sawhill, 2008). Consequently, policy conclusions drawn from the schooling quality literature could fundamentally change when taking into account the inability of a large number of agents to access secondary and tertiary education levels altogether. This would in fact preclude them from benefitting from policies aiming to improve schooling facilities, hence reducing their effectiveness to tackle income inequality, if not reverse their effect. At the same time, policies aiming to facilitate access to education might unintentionally have the opposite net effect on inequality when taking into account heterogeneity in educational quality, as some agents whose access did not require

such policies may excessively benefit from them. This paper will formally discuss these issues in detail for the first time.

1.3 Theoretical Framework

This section outlines a theoretical framework that allows us to analyse the effects and transmission mechanisms of various policies when different sources of inequality are taken into account. More precisely, our interest is to evaluate how the effects of public subsidies, grants and government student loans change when taking into account the fact that agents do not only differ in their level of educational attainment but also in terms of the quality of the education they receive, and/or that some agents are unable to invest in the education of their children. We first present in detail the setup enabling us to account for educational heterogeneity and to evaluate the impact of grants and subsidies. Once we have done this, we add an additional friction to the model in the form of liquidity constrained agents in the second part of this section, which will allow us to analyse the impact of student loans. Section 1.4 will then analyse the impact of subsidies, grants and loans when (1) agents all receive the same level of educational quality and there are no constraints, (2) when we consider constrained agents but ignore educational heterogeneity, (3) when we add educational heterogeneity but ignore credit constraints, and (4) when both are taken into account. Adopting this approach of increasing complexity will allow us to easily and intuitively understand how policies are affected by these two different sources of inequality, and will give us a good grasp of the transmission mechanisms at play by the time we present results from the complete setup.

The model is essentially a two-period overlapping-generations framework where agents

choose between human capital accumulation and leisure when young, and are able to invest in the education of their offspring when adult. Some agents will have access to high-quality education while others will receive low-quality one, hence resulting in human capital disparities and, in turn, income inequality. Once we introduce liquidity constraints, some agents will be unable to invest in the education of their children, who will in turn have access to income-contingent loans. We analyse the impact of changes in the government‘s fiscal policies by looking at one-time policy shocks and their repercussions across different generations. This paper also illustrates the usefulness of using such framework to analyse the effectiveness of measures aimed at reducing income inequality and, more generally, to evaluate fiscal policies.

In the present version of the model, heterogeneity in educational quality is produced exogenously by assigning discrete values to different dynasties. This simplification was adopted to allow for the analysis of one-time policy shocks using impulse response functions, hence giving a much clearer approach to analysing the impact of fiscal policies over time. This also enables us to track the evolution of different families across several time periods by avoiding inter-generational overlaps between different levels of educational quality. Reproducing the same analysis in a Bewley-type framework (Bewley, 1977), i.e., this time allowing agents to choose the quality of education received by their children among a continuous range of possible values and having idiosyncratic shocks as the only source of heterogeneity, produced similar results. This is because non-ergodic models effectively produce no social mobility in the steady state, hence resulting in similar conclusions to those obtained by exogenously imposing an additional source of inequality (Galor and Zeira, 1993; Lochner and Monge-Naranjo, 2011*a*). This paper presents the model with exogenous educational heterogeneity given the more extensive and clearer analytical possibilities it offers

and the absence of loss of generality. Furthermore, in Bewley-type models, agents do not differ in any fundamental way – i.e., they are identical ex-ante and differ ex-post because of idiosyncratic shocks – which means that the government has to rely on perfect information to implement any policy. In reality however, governments are unable to tailor their policies to individuals and rely instead on policies applied uniformly across certain groups (Troch, 2014). In this sense, this framework hence reflects the policy problems faced by governments more appropriately than Bewley models.

1.3.1 A Model of Income Inequality and Educational Heterogeneity

1.3.1.1 Households

There is a continuum of agents $a \in (0, 1)$. Agents live two periods, have a single child each, and their life overlaps with that of their offspring in the second period. In the first period, when young, their parent chooses for them the amount to invest in their education and, in turn, children must decide how much of these resources to dedicate to leisure and how much to effectively allocate to building up their human capital.[1] In the second period, agents are in their adulthood and inelastically supply the human capital they accumulated when young to firms and receive a wage in return. They must decide how much of this income to invest in the education of their child for a given level of tuition fees and consume the rest. Furthermore, agents are born in either of two types of families: one having access to high-quality education, the other having access to low-quality one, labelled H and L, respectively. The portion of agents of type $j = \{H, L\}$ in the population is denoted by μ_j.

[1] Note that the leisure decision when young is analogous to the consumption choice in the first period of a standard two-period OLG model. The division of agents' life into a schooling period and a working period has been explored in earlier studies, such as Huggett et al. (2011).

Agents' budget constraint when young therefore takes the following form:

$$l^j_{i,t-1} + h^j_{i,t-1} = (q^j_{t-1})^{\gamma_1}(s^j_{i,t-1})^{\gamma_2} + a^j_{i,t-1} \tag{1.1}$$

where $j = H, L$ is agent i's type; $l^j_{i,t}$ and $h^j_{i,t}$ are the amounts the agent chooses to dedicate to leisure and the accumulation of human capital, respectively; q^j_t is the quality of education received by the agent of type j, and it is such that $q^H_t > q^L_t$, which is the only intrinsic difference between the two types of household; f^j_t is the rate of tuition fees paid by the parent; $s^j_{i,t}$ is the amount invested by the parent in the education of the agent, and it can be treated as the level of educational attainment of the agent; $\gamma_1 \in (0, 1)$ and $\gamma_2 \in (0, 1)$, which ensures concavity – the higher educational quality or attainment, the smaller their marginal contribution to human capital – and follows the past literature (Dabla-Norris and Matovu, 2002). $a^j_{i,t}$ is the inborn ability and social environment of the agent, which is subject to an idiosyncratic shock and evolves according to:

$$a^j_{i,t} = \rho^j_a a^j_{i,t-1} + \varepsilon^j_{i,t} \tag{1.2}$$

where ρ^j_a is the rate of persistence of inborn abilities across generations, reflecting the lack of perfect social mobility within each type of household; and $\varepsilon^j_{i,t}$ are uniformly distributed stochastic shocks affecting agents. The right-hand side of equation (1.1) is effectively the gross amount of human capital accumulated, i.e., not net of leisure, which we define as:

$$H^j_{i,t} = (q^j_t)^{\gamma_1}(s^j_{i,t})^{\gamma_2} + a^j_{i,t} \tag{1.3}$$

Henceforth, when mentioning human capital we refer to its gross value. This extended

Mincerian functional form for human capital is inspired from previous attempts at modelling educational attainment and quality, such as in Glomm (1997) and Glomm and Ravikumar (1992) and, more generally, from widely used human capital formulations such as in Dabla-Norris and Matovu (2002) Heckman (1999).[2] Households' budget constraint when adult is given by:

$$c^j_{i,t} + (1 - p^j_t) f^j_t s^j_{i,t} = (1 - \tau^j_t) w^j_t h^j_{i,t-1} + T^j_t \tag{1.4}$$

where $c^j_{i,t}$ is consumption; p^j_t is the share of tuition fees subsidised by the government; w^j_t is the wage rate; τ^j_t is an income tax rate collected by the government; and T^j_t is a lump-sum transfer. We can rearrange equations (1.1), (1.3), and (1.4) to get a lifetime budget constraint:

$$c^j_{i,t} + (1 - p^j_t) f^j_t s^j_{i,t} = (1 - \tau^j_t) w^j_t (H^j_{i,t-1} - l^j_{i,t-1}) + T^j_t \tag{1.5}$$

Furthermore, tuition fees rates are assumed to be a function of educational quality, reflecting the case of the U.S., taking the following linear form:

$$f^j_t = \delta q^j_t \tag{1.6}$$

Households' utility depends on how much they consume in their adulthood and on the amount of leisure they choose when young. In addition, agents care about their child's utility at a discounted rate. Their expected lifetime utility is therefore defined as:

$$E_t U^j_{i,t} = u(c^j_{i,t}, l^j_{i,t-1}) + \beta E_t U^j_{i,t+1} \tag{1.7}$$

where $U^j_{i,t}$ is their lifetime utility, with the time subscript t indicating the period in which

[2]None of these studies, however, have looked at heterogeneity in the quality of education.

they are adult; $u(\cdot)$ is the utility they get from their own consumption and leisure; and β is the inter-generational discount rate. We assume the following logarithmic functional form defining the relation between utility, consumption and leisure:

$$u(c^j_{i,t}, l^j_{i,t-1}) = log(c^j_{i,t}) + \chi log(l^j_{i,t-1}) \tag{1.8}$$

where χ is the weight put on leisure. Then, substituting lifetime utility forward in equation (1.7), we obtain:

$$E_t U^j_{i,t} = \sum_{s=0}^{\infty} E_t \beta^s \left[log(c^j_{i,t+s}) + \chi log(l^j_{i,t+s-1}) \right] \tag{1.9}$$

which effectively gives an infinitely-lived agents' problem.

Households therefore have to choose $c^j_{i,t}$, $l^j_{i,t-1}$, and $s^j_{i,t}$ to maximise (1.9) subject to (1.3) and (1.5). Solving this problem yields the following first-order conditions for each agent type:

$$\frac{E_t c^j_{i,t+1}}{c^j_{i,t}} = \beta E_t \frac{\gamma_2 (1 - \tau^j_{t+1}) w^j_{t+1} (q^j_t)^{\gamma_1} (s^j_{i,t})^{\gamma_2 - 1}}{1 - p^j_t} \tag{1.10}$$

$$\frac{c^j_{i,t}}{l^j_{i,t-1}} = \frac{(1 - \tau^j_t) w^j_t}{\chi} \tag{1.11}$$

Equation (1.10) is the Euler equation for consumption, which gives agents' consumption profile. Agents will optimally invest in their child's education an amount of resources that equates their marginal utility from consuming an additional unit to their marginal utility from investing an additional unit in their child's education. Equation (1.11) defines agents' inter-temporal choice between consumption and leisure. Again, their optimal decision will equate their marginal utility of consumption to their marginal utility of leisure. Next, we aggregate these equations across all agents within each j type. For any variables $x^H_{i,t}$ and

$x^L_{i,t}$, after appropriately sorting agents by type, we define their aggregate value as follows:

$$x^L_t \equiv \int_0^{\mu_L} x^L_{i,t} di \qquad x^H_t \equiv \int_{\mu_L}^1 x^H_{i,t} di \tag{1.12}$$

We therefore get the following aggregate equations for each agent type:

$$c^j_t + (1 - p^j_t) f^j_t s^j_t = (1 - \tau^j_t) w^j_t (H^j_{t-1} - l^j_{t-1}) + T^j_t \tag{1.13}$$

$$H^j_t = (q^j_t)^{\gamma_1} (s^j_t)^{\gamma_2} \tag{1.14}$$

$$\frac{E_t c^j_{t+1}}{c^j_t} = \beta E_t \frac{\gamma_2 (1 - \tau^j_{t+1}) w^j_{t+1} (q^j_t)^{\gamma_1} (s^j_t)^{\gamma_2 - 1}}{1 - p^j_t} \tag{1.15}$$

$$\frac{c^j_t}{l^j_{t-1}} = \frac{(1 - \tau^j_t) w^j_t}{\chi} \tag{1.16}$$

1.3.1.2 Firms

Output is produced by an infinitesimal number of identical firms, which behave in a perfectly competitive manner, and can hence be collectively treated as a representative firm. They employ both types of agents, renting their human capital in exchange for a wage rate. Following Galor and Zeira (1993), the production process is fragmented across the two types of labour, implying that they receive different wage rates. Production for each type is undertaken using a Cobb-Douglas technology of the form:

$$y^j_t = A_t (h^j_{t-1})^{\alpha_j} \tag{1.17}$$

where y^j_t is the output of agent-type j; α_j is the share of type j's human capital in the production process; A_t is aggregate productivity, which evolves according to an autoregressive

process of order 1 and is subject to aggregate stochastic shocks ε_t^A:

$$ln(A_t) = \rho_A ln(A_{t-1}) + \varepsilon_t^A \tag{1.18}$$

where ρ_A is the persistence of productivity across different periods. Total output is given by the weighted sum:

$$y_t = \sum_j \mu_j y_t^j \tag{1.19}$$

Firms choose to employ a certain amount of each type of human capital to maximise the profit they make from each production process. Given perfect competition, the wage rate simply equates the marginal product of labour:

$$w_t^j = \alpha_j A_t (h_{t-1}^j)^{\alpha_j - 1} \tag{1.20}$$

1.3.1.3 Government

The government will play a central role in this economy. In fact, it won't simply be limited to collecting taxes and spending revenues on transfers, as is the case in most of the literature, but it will have access to different types of fiscal policies, the effects of which is the main focus of our analysis.

In fact, we are interested in testing the most commonly accepted policy proposal put forward by the literature and policy-makers to reduce income inequality, namely, making education more affordable. We therefore allow the government to subsidise a portion p_t of tuition fees. In practice, this can easily be interpreted in the context of the public education system, where the local and/or federal government is able to directly affect the costs of education incurred by individuals.

To determine the effects of raising government subsidies, we make p_t subject to a policy shock υ_t^j and assume that the amount of subsidies chosen every period by the government partly depends on its choice in the previous period, as a means of comparison with other types of shocks and to reflect the stickiness of tuition fees:

$$ln(p_t^j) = \kappa ln(p_{t-1}^j) + \theta_j^p + \upsilon_t^j \tag{1.21}$$

where κ is the rate of persistence of public subsidies; θ_j^p is an adjustment parameter which defines the steady state value of p_t^j; and υ_t^j is a normally-distributed stochastic policy shock equal to 0 in the steady state.

We are also interested in examining the impact of the second major policy proposal stemming out from the literature, namely increasing the amount spent by the government to improve the quality of schooling. This can easily be measured in practice by looking at the amount spent by governments on teaching facilities and school employees' salaries and, in this model, we interpret it as grants directly allocated to schools translating into a temporarily higher quality of education. In other words, we treat grants g_t^j allocated to schools of type j as policy shocks affecting educational quality, which in turn is assumed to be sticky, implying that a grant is only progressively depleted:

$$ln(q_t^j) = \rho_j^q ln(q_{t-1}^j) + \theta_j^q + \phi_j \lambda g_t^j \tag{1.22}$$

where ρ_j^q is the persistence of educational quality of type j across time; θ_j^q is an adjustment parameter determining the steady state level of educational quality for each agent-type; ϕ_j is a scaling parameter which we will use in the case of uniform shocks to ensure that the educational quality of both types of agents jumps by an equal amount; λ determines the

efficiency of public spending; and g_t^j is a policy shock representing public grants, which is equal to 0 in the steady state.

Finally, we want to analyse the effects of a third and more direct type of policy put forward by the literature, lump-sum transfers, which we also treat as stochastic shocks, as a means of comparison.

We assume that the government balances its budget in each period. To achieve this, it collects an income tax from both types of agents to finance public subsidies and grants (Lucas, 1990; Barro, 1991; Glomm and Ravikumar, 1992, 1998; Dabla-Norris and Matovu, 2002). The government budget constraint is therefore given by:

$$\sum_j \tau_t^j w_t^j = \sum_j p_t^j f_t^j + \sum_j g_t^j \tag{1.23}$$

1.3.1.4 Income Inequality

There exist various methods to measure disparities in a given variable. In this study, we are interested in examining the drivers of income inequality. We therefore chose to use the most commonly adopted and accepted measure in the literature, the Gini coefficient, for comparison purposes. Ranging from 0 (perfect equality) to 1 (perfect inequality), the Gini ratio is calculated by ordering the income of agents – or groups of agents – in ascending order and applying the following formula:

$$G_t = \frac{1}{n}\left[n + 1 - 2\frac{\sum_{i=1}^{n}(n+1-i)Y_t^i}{\sum_{i=1}^{n} Y_t^i}\right] \tag{1.24}$$

where G_t is the Gini coefficient; n is the total number of individuals; and Y_t^i is total income of agent i, such that $Y_t^1 < Y_t^2 < \cdots < Y_t^n$.

The Gini ratio will play a key role in this analysis, as we will be interested in looking at the effect of various policies on this measure. Later, in the empirical analysis, we will be focusing on decomposing the contribution of different factors to income inequality, and we will therefore introduce a comparable measure, the Theil index.

1.3.2 Extending the Framework to Account for Liquidity Constraints

We now extend the above model to account for the second major driver of income inequality identified in the literature, namely the inability of some agents to invest in their children's education. This will allow us to simultaneously analyse educational heterogeneity and liquidity constraints and, more importantly, how each affects the effectiveness of policies aimed at reducing inequality. This extension also enables the introduction of government student loans, whose structure is described below.

1.3.2.1 Introducing Constrained Agents and Government Student Loans Programs

Liquidity constrained individuals differ from Ricardian agents in that they are unable to invest in their child's education in the second period of their life. Young agents are therefore unable to use their parents' contribution to build up their human capital above their inborn ability level and are therefore reliant on government student loans to attend school.

Based on the case of the United States, student loans are income-contingent (they are offered to students in the period of their studies and repayments are made when they enter the job market in the second period), need-based (they are only provided to agents from constrained families), interest-free, and repayments are fully enforced. Furthermore, given that in practice a large fraction of students hits the upper limit of the GSL amount they are entitled to and that nearly all students from socially disadvantaged backgrounds attending

college are dependent on GSL programs to finance their studies, the variable d_t is defined as the upper limit of the per-student loan allowance, and it is assumed that demand for loans is inelastic. Therefore, constrained agents' educational attainment will rely entirely on the GSL limit. For simplicity, we assume that these individuals do not have access to private lenders. If agents receive a positive loan amount, they then gain access to high-quality education with probability ψ or to low quality education with probability $(1-\psi)$.[3]

Non-Ricardian agents choose consumption and leisure to maximise their lifetime utility, which takes the same form as in equation (1.9), subject to:

$$l^C_{i,t-1} + h^C_{i,t-1} = \left[\psi(q^H_{t-1})^{\gamma_1} + (1-\psi)(q^L_{t-1})^{\gamma_1}\right] p^C_{t-1} d^{\gamma_2}_{t-1} + a^C_{i,t-1} \tag{1.25}$$

$$c^C_{i,t} + (1-p^C_{t-1})\left[\psi f^H_{t-1} + (1-\psi) f^L_{t-1}\right] d_{t-1} = (1-\tau^C_t) w^C_t h^C_{i,t-1} + b_t + T^C_t \tag{1.26}$$

where b_t refers to the amount of unemployment benefits offered to agents in case no loans are provided; and d_{t-1} is the amount borrowed in period $(t-1)$ and reimbursed at t. We then obtain a condition analogous to (1.11), which we aggregate across constrained agents, alongside their human capital equation and lifetime budget constraint.

1.3.2.2 GSL Upper Limit and Government Budget

For comparison purposes, we assume that the upper limit on per-student loan allowances follows a first-order Markov process across generations:

$$d_t = \rho_d d_{t-1} + \theta^d + \varepsilon^d_t \tag{1.27}$$

[3]For simplicity and tractability of dynasties, we assume that constraints are independent across generations, so that obtaining a student loan does not allow constrained agents to invest in their child's education.

where ρ_d is the persistence of shocks to the GSL upper limit; and θ^d defines the steady state value of d_t. The government funds student loans in every period through tax revenues and the repayment of loans issued in the previous period. The government's budget constraint therefore becomes:

$$\sum_k \tau_t^k w_t^k = \sum_j p_t^j f_t^j + p_t^C \left[\psi f_t^H + (1-\psi) f_t^L\right] + \sum_j g_t^j + b_t + \Delta d_t \tag{1.28}$$

where $\Delta d_t = d_t - d_{t-1}$, $j = \{H, L\}$, and $k = \{H, L, C\}$. The production process remains fragmented, with a third input introduced, which allows output to remain positive in the event that constrained agents are unable to accumulate any human capital.

1.3.3 Calibration

Having set up the model that will serve as the basis of our study, we now need to calibrate the various parameters that we have introduced before moving to simulating policy shocks. Given that our analysis focuses on the case of the United States, we have sought to select parameter values that will make our model emulate the U.S. economy as closely as possible. At the same time, we have sought to isolate the effect of heterogeneity in the quality of education by preserving the same coefficients across high and low types, such as discount rates, returns to education, and inter-temporal rates of persistence. Table 1.1 in Appendix 1.A presents the values chosen for our calibration.

Parameters relating to the household and firm problems were calibrated following previous studies relating to the U.S. economy, when possible, with γ_1 set equal to γ_2 so as to have a compound parameter of returns to education. In addition, χ was added to have agents weigh leisure when young by less than consumption when adult, which appears to be

a reasonable assumption. Given that the top percentile of the income distribution saw their share of U.S. income increase in the past decades, we assume that 10% of the population receives high-quality education. Looking at tertiary education, around 30% of Americans aged 25 years old or over (34% of 25-29 year-olds) held a college degree in 2014 (U.S. Census Bureau, 2014). In a setup without constrained agents, this translates into 30% of unconstrained agents receiving high-quality education, so that μ_H is set to 0.3. When including constrained agents, their share of the population is set to 70%, with high-quality and low-quality constrained agents forming respectively 10% and 20% of the total population.

Parameters relating to our policy shocks were calibrated as follows. The persistence of shocks was set the same for all policy variables, so as to make inter-temporal comparisons of impulse response functions easier. Adjustment parameters were set arbitrarily to have p_t^j equal to 10% in the steady state for both agent-types – as we want to isolate the effect of heterogeneity in educational quality, and subsidies enter the lifetime budget constraint linearly anyway – and q_t^H and q_t^L equal to 0.7 and 0.3, respectively. Standard deviations of stochastic terms were chosen so that all policy variables would jump by an equal amount to allow for magnitude comparisons. The results below present stochastic shocks leading to 0.3-point jumps in p_t^H, p_t^L, q_t^H, q_t^L, d_t, T_t^H, T_t^L and/or T_t^C. The spending effectiveness parameter is set to 1 for generalisability purposes, but it should be clear that setting its value lower than one will simply reduce the impact of grants on the quality of education, since part of the resources invested by the government will be wasted through inefficient investments. Parameters on the persistence of idiosyncratic shocks to inborn ability were set based on social mobility data. In fact, inborn abilities are not only understood as the 'genetic' predisposition of individuals but also as the social environment in which they are raised.

Finally, the probability of constrained agents accessing high-quality education when the GSL upper limit is above zero was set to 15%, based on social mobility data. Unemployment benefits can be fixed and set arbitrarily as they do not play a significant role in our dynamic analysis. Finally, note that θ^d is set to zero in the case where the government does not provide loans.

1.4 Theoretical Results

This section outlines the results obtained from simulating policy shocks in the theoretical framework outlined in Section 1.3. The purpose of this analysis is to uncover the impact of public subsidies, grants, and government student loans on income inequality. To do this, we evaluate these shocks under four setups of the above framework presented in order of increasing complexity, which will also allow us to understand how educational heterogeneity and liquidity constraints affect the effectiveness of each of these policies.[4] As explained in Section 1.2, we are interested in both uniform and asymmetric shocks, since in practice some policies tend to be administered homogeneously across all agents while others are targeted toward specific groups. Since we want to compare the magnitudes of these effects, we consider both types of shocks for all policies. Furthermore, in order to isolate the effect of these policies on income inequality, we apply a flat tax rate throughout this analysis. Our framework also allows us to evaluate the effectiveness of a more blunt form of redistribution, lump-sum transfers.

We start by analysing the effect of subsidies and grants in a basic setup where neither constrained agents nor differential levels of educational quality are present. This will allow

[4]Note that government student loans can only be analysed when non-Ricardian agents are present in our model, since we have assumed that these are allocated on the basis of need and are therefore only provided to constrained agents.

us to then isolate the impact of liquidity constraints and educational heterogeneity on policy outcomes when we introduce each of these frictions separately in a second and third setup, respectively. Finally, we consider the more realistic case of an economy that incorporates both features. Results from this final setup will be tested empirically in the next section.

1.4.1 Starting Point: A Basic Setup

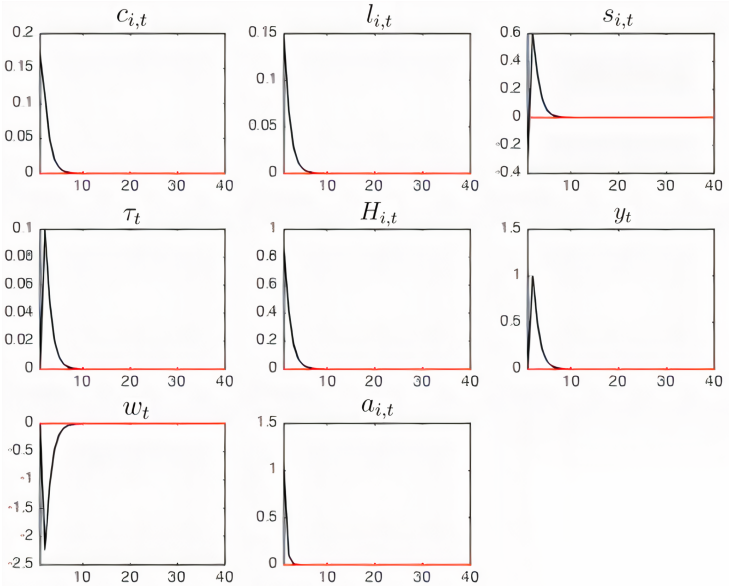
We first want to understand what happens in a setup without liquidity constraints where agents do not differ in the quality of the education they receive. Note that they do differ in terms of their educational attainment because of idiosyncratic shocks to their inborn abilities, which affect their capacity to invest in the education of their child.

1.4.1.1 The Impact of Idiosyncratic Shocks

To understand how agents react to being born with a different level of ability than their parents and other agents, we first look at the impact of idiosyncratic shocks in the case where all agents are otherwise identical. These shocks reflect the fact that agents may be born with different abilities, while their stickiness represents the lack of perfect social mobility across generations. Impulse response functions (IRFs) are reported in Figure 1.4.

A single individual being born with higher initial abilities than her parent will have a higher individual level of gross human capital $H_{i,t}$, which will allow her to use a smaller share of resources to build up her net human capital $h_{i,t}$ so that her level of leisure when young increases. Notice that gross human capital increases to a much larger extent than leisure, so that net human capital increases. The parent of the agent born in the period of the shock will invest less in her child's education and increase her consumption level, knowing that her offspring does not need to build up her human capital by as much as she had to to

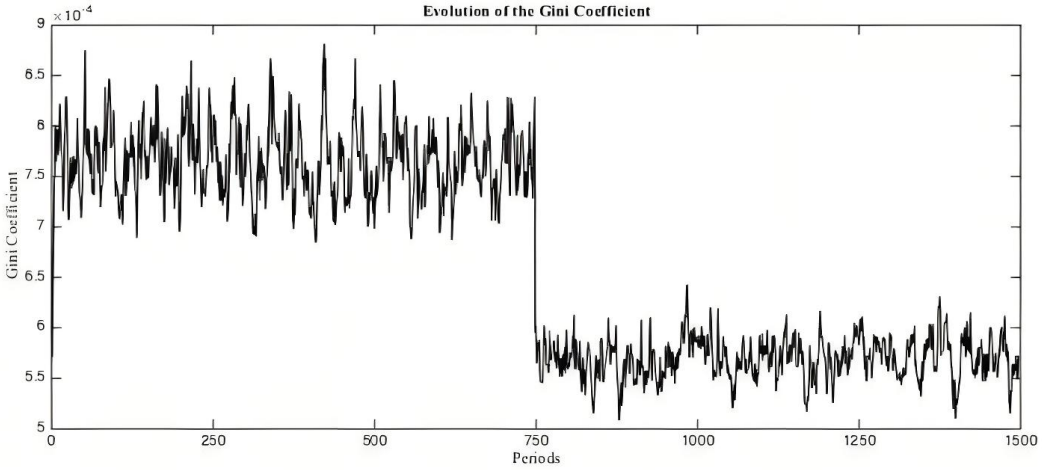
Figure 1.4: Dynamic Response of the Basic Setup to an Idiosyncratic Shock



reach a similar level of income. This decrease in educational investment will however not be enough to offset the increase in inborn abilities. Keeping other agents' abilities fixed, total output will remain unchanged in the period of the shock since adults in that period have predetermined levels of human capital.

In the following period, the agent's higher net human capital will both increase aggregate output and decrease the wage rate given decreasing returns to individual inputs. In turn, since the expenditure side of the government's budget is not altered, the tax rate has to increase to maintain a balanced budget. Despite this, the agent's total income increases. While the individual will consume more than had there not been a positive idiosyncratic shock, she will consume less than her parent. This is because her child's inborn abilities are lower than hers – as inborn abilities start returning to their steady state – so that, in her desire for inter-generational consumption smoothing, she will increase the amount invested in the education of her child. The economy returns to its steady state level at a rate given by social mobility.

Figure 1.5: Evolution of the Gini Coefficient Following a Shift in Government Subsidies

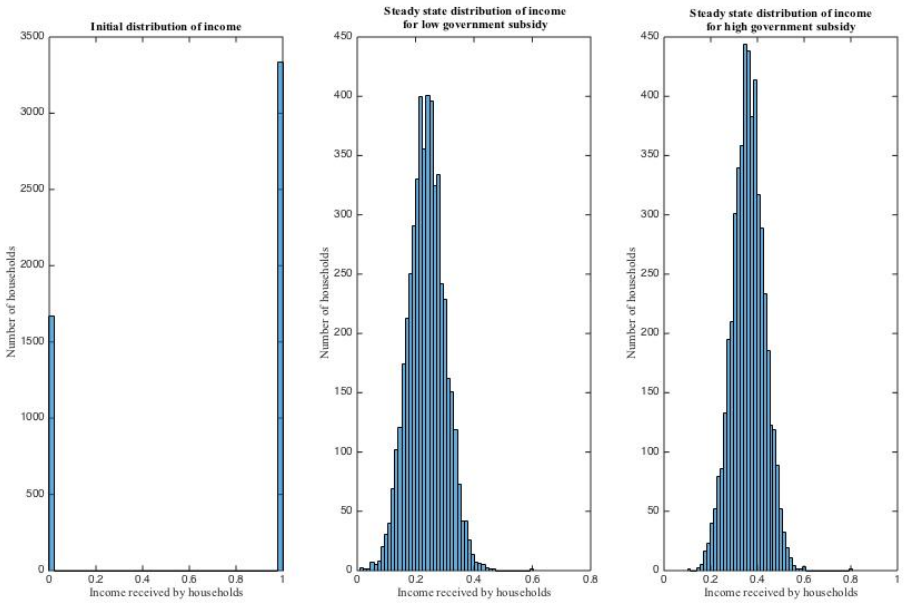


1.4.1.2 Policy Shocks Under Educational Homogeneity and Perfect Liquidity Access

We now turn to analyzing the effects of government subsidies and grants on income inequality, in the context of qualitative educational homogeneity. The only source of income disparities being idiosyncratic shocks, we simulate an economy subject to uniformly distributed idiosyncratic shocks and observe the effect of subsidies p_t and grants g_t. These findings will then be compared to the results we obtain when introducing liquidity constraints and educational heterogeneity, which will allow us to isolate their respective impact.

Our first important result in this setup, which reflects the view commonly accepted by policy-makers, is that an increase in the portion of tuition fees subsidised by the government reduces income inequality. Figure 1.5 illustrates this by simulating an economy with a low level of subsidies ($p_t = 0.1$) from periods 0 to 749. The government permanently raises subsidies to 0.4 in period 750, which allows us to observe the fall in the Gini coefficient between these two states. The policy change also reduces the volatility of the Gini coefficient across generations. This is because, as noted above, the only source of inequality, idiosyncratic shocks, creates dispersion in individuals' total income, which in turn produces inequality in

Figure 1.6: Evolution of the Distribution of Income Following a Shift in Government Subsidies



the capacity of parents to invest in the education of their child. Hence, increasing the share of tuition fees financed by the government reduces the weight of parental income – which in turn depends on idiosyncratic shocks – in determining educational attainment. Government subsidies therefore reduce the dispersion of educational attainment and stabilise social mobility by clustering agents towards a middle class.

Figure 1.6 illustrates the evolution of the steady state distribution of income. In fact, starting from a distribution where two-thirds of agents hold all of the available income (panel 1 of Figure 1.6), the economy naturally moves towards a hump-shaped distribution where most agents are clustered towards a middle class (panel 2). Nevertheless, the degree of equality and the level of median income depend on government subsidies. Indeed, moving to higher government subsidies will create a more equal distribution, even more clustered towards the median agent. In addition, median income itself will increase, moving from slightly above 0.2 to slightly below 0.4 (panel 3).

Finally, introducing aggregate technological shocks or improving the quality of education by increasing the amount of government grants does not affect income inequality. This is because, in the absence of educational heterogeneity, both of these shocks will affect all agents equally.

The policy conclusion stemming from an analysis that does not take into account disparities in the quality of education and access to liquidity is therefore that the most effective measure to reduce income inequality is to decrease the level of tuition fees paid by individuals. On the other hand, improving the quality of education may be beneficial in itself through its effects on aggregate human capital and output, but it will be entirely ineffective in reducing income disparities. In the remainder of this section, we successively add two frictions to this setup and analyse how they affect our conclusions.

1.4.2 The Role of Liquidity Constraints

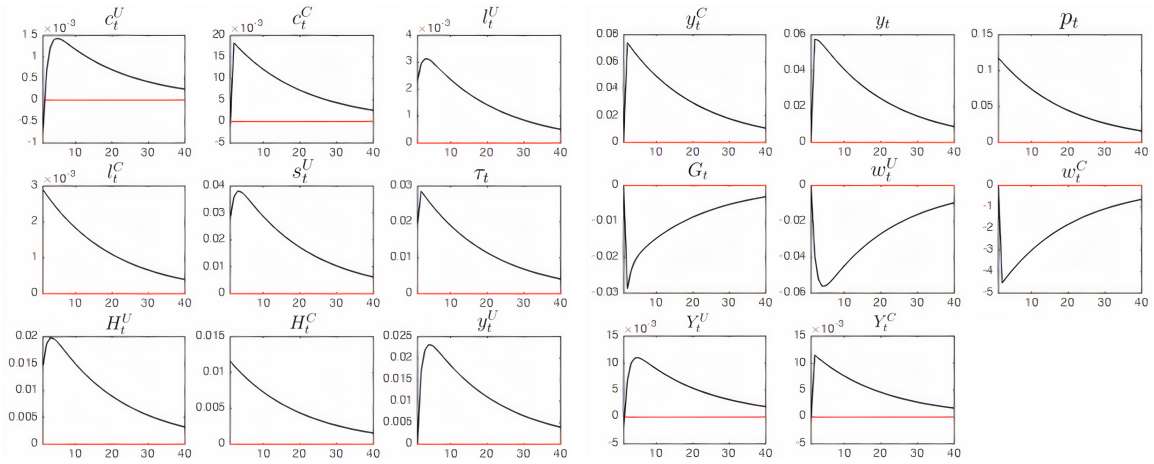
Next, we introduce non-Ricardian agents to evaluate the extent to which liquidity constraints hinder or improve the effectiveness of public subsidies and grants. In addition, we are able to analyse the impact of raising the GSL upper limit. Notice that since qualitative educational heterogeneity is absent from this setup, unconstrained agents are considered jointly throughout.[5]

1.4.2.1 Uniform Policy Shocks With Liquidity Constraints

As discussed in Section 1.2, governments can combine two policies to try and address the issue of liquidity constraints – subsidies and student loans. Before comparing the effects of

[5] As previously discussed in Section 1.3, this section uses impulse response functions to analyse the impact of various shocks on income inequality, given that idiosyncratic shocks are no longer the only source of inequality and IRFs allow to evaluate the impact of one-time shocks and to isolate various effects more easily. Using simulations gives the same results, so this does not involve any loss of generality.

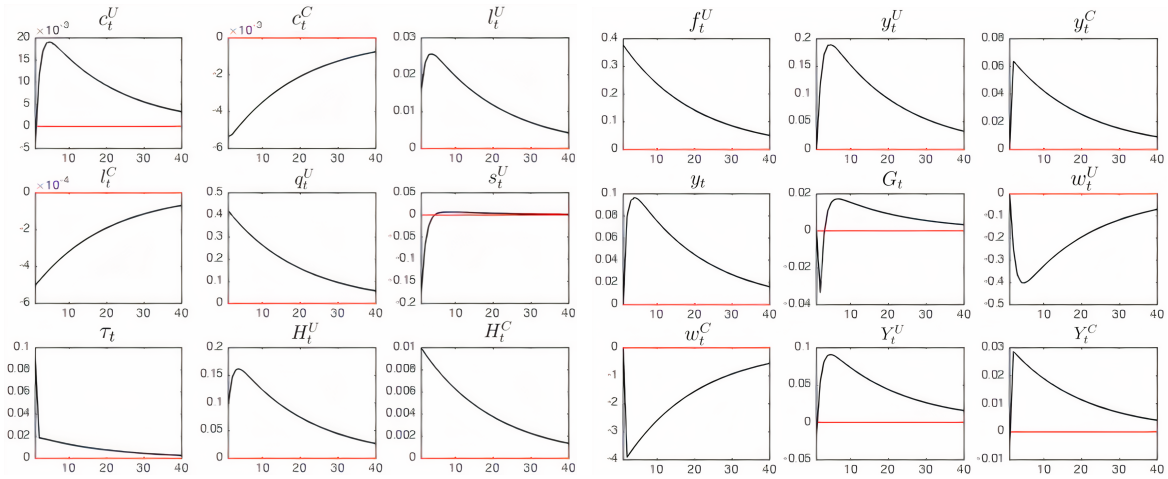
Figure 1.7: Dynamic Impact of a Uniform Shock to Government Subsidies (p_t) in the Presence of Liquidity Constraints



need-based student loans and asymmetrically allocated subsidies, we first want to evaluate how the introduction of liquidity constraints may affect the effectiveness of uniform shocks to subsidies and grants, compared to the basic setup simulations. We start with the case where the GSL upper limit is strictly positive, and then briefly discuss the case where it is set to zero.

Figure 1.7 shows the dynamic response of the model to a uniform shock in p_t, where subscripts C and U indicate the aggregate variables of constrained and unconstrained agents, respectively. While both agent-types will benefit from this policy change, constrained agents will do so disproportionally. An increase in the portion of tuition fees financed by the government will trigger unconstrained agents to invest more in the education of their children. In turn, the young at the time of the shock will be able to accumulate more capital and, although to a lesser extent, have more leisure than their parents. In their adulthood, their wages will be lower, given decreasing returns to single inputs, and they will face a higher tax rate. Despite this, they will see their total income (Y_t^U) increase thanks to their higher human capital. They will therefore consume more than their parents and invest more in their own child's education. As subsidies return to their original level, this trend will be reversed

Figure 1.8: Dynamic Impact of a Uniform Shock to Government Grants (g_t) in the Presence of Liquidity Constraints



and variables will converge back to their steady state. Constrained agents will experience a similar dynamic response but with larger magnitudes. This is because the education they acquire, given the non-zero GSL upper limit, depends to a much larger extent on public subsidies. The policy shock will therefore have a similar net effect on income inequality as in the basic setup, although through a different transmission mechanism.

A uniform shock in public grants (g_t) going towards the improvement of educational quality, on the other hand, will no longer have a null effect (Figure 1.8). Note however that all agents have access to the same level of educational quality, and that these effects will simply be triggered by differential smoothing behaviors. In fact, a homogeneous increase in grants will cause an upward-jump in both agent-types' human capital, given the positive GSL upper limit. The tax rate will incur a one-time increase as grants are fully repaid in the second period but will not immediately reach its steady state value as it will need to compensate for the fall in the wage rates. Despite this, both agents' total income will grow. Nevertheless, constrained and unconstrained agents will not be able to manage this increase in resources in the same manner. In fact, the latter will be able to smooth their

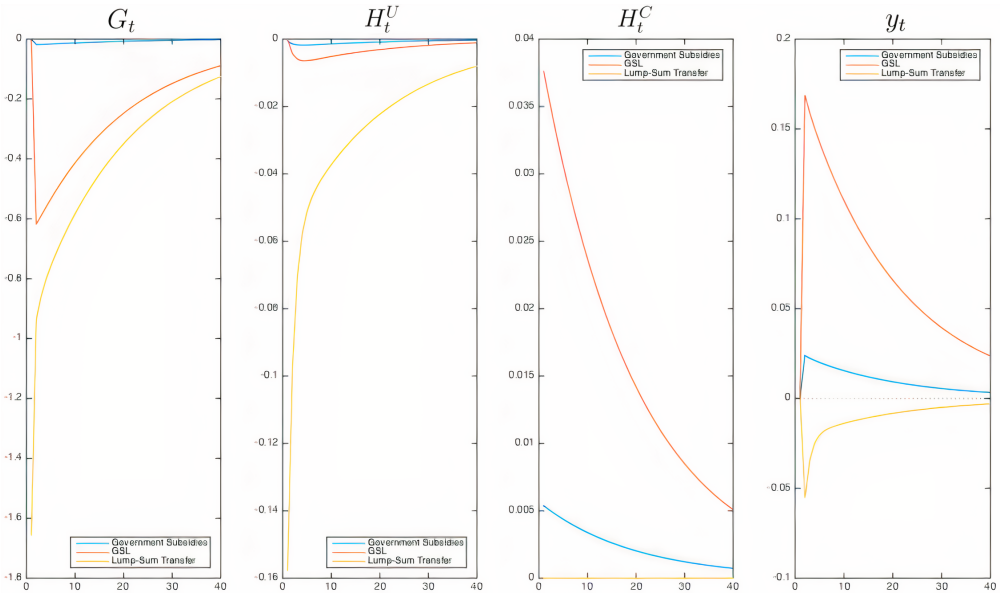
return to the steady state by temporarily decreasing the amount they invest in their child's education to compensate for the higher tuition fees, which will enable them to increase their consumption level while transferring some resources to the next generation. Constrained agents, on the other hand, cannot undertake such inter-generational transfer. What's more, they will not even be able to increase their consumption level due to the higher cost of the education they acquired by taking a loan, which they must reimburse in their adulthood. The income of constrained agents will therefore reach its peak in the period following the shock, before progressively returning to its steady state, following the depletion rate of the public grant. Unconstrained agents, on the other hand, will be able to smooth the increase in their income across generations, therefore reaching their peak after a few periods. Hence, in the period after the shock, non-Ricardian agents will be able to increase their income to a greater extent than Ricardian individuals, therefore leading to a fall in inequality. As the increase in their income starts falling while that of unconstrained agents keeps growing, the Gini coefficient will quickly return to its steady state level before rising. Once the income of the unconstrained reaches its peak, inequality starts returning to its original level.

This exercise therefore provides a first illustration of how liquidity constraints are able to alter the effectiveness of policies aimed at improving the quality of education. Note that in the case where the GSL upper limit is set to zero, both subsidies and grants will increase inequality as constrained agents won't have access to education.

1.4.2.2 Asymmetric Policy Shock With Liquidity Constraints

We now turn to considering the impact of three asymmetric shocks: an increase in government subsidies only directed toward non-Ricardian agents (p_t^C), a rise in the GSL upper borrowing limit (d_t), and lump-sum transfers from unconstrained to constrained individu-

Figure 1.9: Dynamic Impact of Asymmetric Policy Shocks in the Presence of Liquidity Constraints



als (T_t^C). It should be obvious that all three shocks will decrease income inequality as by definition unconstrained agents will not benefit from these policies while they will see their taxes increase, but they will each do so with different magnitudes and with different effects on the rest of the economy.

As shown in Figure 1.9, lump-sum transfers will have the most drastic effect of all three policies, reaching a negative impact on income inequality more than two and a half times larger than what an increase in the GSL limit will achieve at its peak. This is no surprise, as we are directly transferring income from the top 30% of the population to the bottom 70%. Public subsidies, on the other hand, will only have a minor effect on the Gini coefficient, relative to the two other policies.

Beyond the fact that lump-sum transfers are generally difficult to implement in practice, there are a number of reasons why policy-makers concerned with income inequality may not want to choose this policy over the other two. While transfers have the greatest impact on inequality, they hold little other benefit to the economy since they produce small inter-

generational effects. In fact, the negative effect of lump-sum transfers on inequality is mostly the result of a deterioration in the conditions of unconstrained agents. While non-Ricardian agents will be unable to invest any of their incoming transfers to improve their children's human capital, unconstrained agents will see their human capital fall very sharply. This will result in a significant fall in output. By contrast, an increase in the GSL upper limit will allow constrained agents to increase their human capital while Ricardian agents will only suffer a minor fall in theirs caused by the increase in the tax rate. As a result, not only will this policy trigger a fall in income inequality, but it will do so without having to experience a trade-off in terms of output growth.

1.4.3 The Role of Educational Heterogeneity

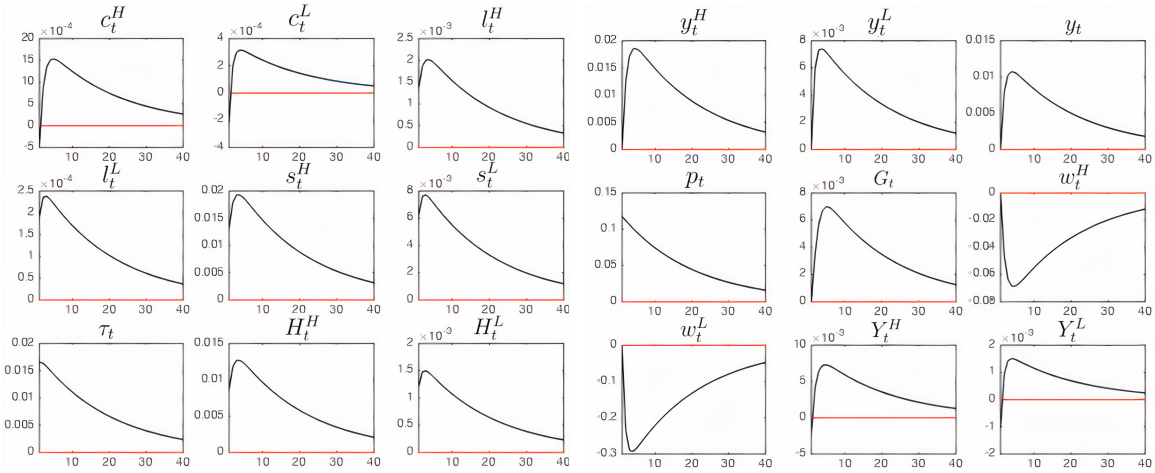
We now add differential levels in the quality of education to the basic setup, to evaluate how educational heterogeneity impacts our conclusions regarding the effect of various policy shocks on overall income inequality. Note that liquidity constraints are absent from this setup, and will be re-introduced in the next sub-section.

1.4.3.1 Uniform Policy Shocks Under Educational Heterogeneity

First, we turn to reconsidering the effects of uniformly increasing the share p_t of tuition fees financed by the government and the amount of grants g_t allocated to schools, when agents differ in the quality of the education they receive and have equal access to liquidity.

Figure 1.10 presents the first key result suggested by this paper: a higher level of government subsidies will actually increase income inequality. While both types of agents will gain from such policy – therefore raising median income – we find that agents having access to higher-quality education will disproportionately benefit from it. For any agent type,

Figure 1.10: Dynamic Impact of a Uniform Shock to Government Subsidies (p_t) Under Educational Heterogeneity

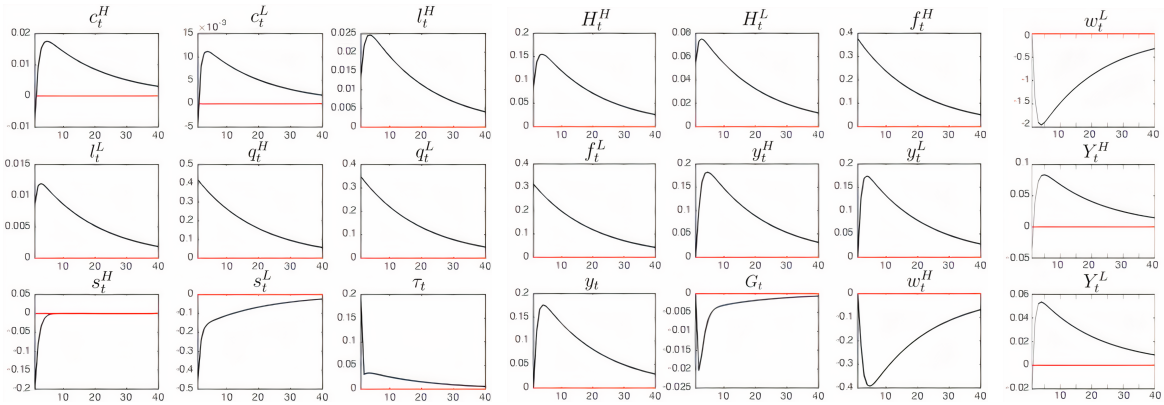


higher subsidies make education effectively cheaper, so that parents will decide to invest a greater amount in their child's educational attainment. Agents studying at the time of the shock will therefore see an increase in their human capital – and a smaller increase in their leisure amount – resulting in higher total income in their adulthood. Parents at the time of the shock will observe the highest increase in the tax rate, which will actually reduce their total income and their consumption level. On the other hand, adults in the next period will experience an increase in human capital that will out-weight the higher tax rate, resulting in higher total income, which will allow them to both consume more and invest a greater amount than their parents did. The economy will then slowly return to its steady state.

What leads income inequality to increase is the following. High-type agents are subject to relatively higher tuition fees, given the higher quality of their education. Therefore, the increase in government subsidies will allow them to save a greater amount of resources than low-type agents, which will allow their children to experience a higher increase in their level of human capital and total income level, and in turn to both consume and invest more. This will result in higher income inequality until the steady state is reached again.

This is a very important result given the persistence of most policy-makers concerned

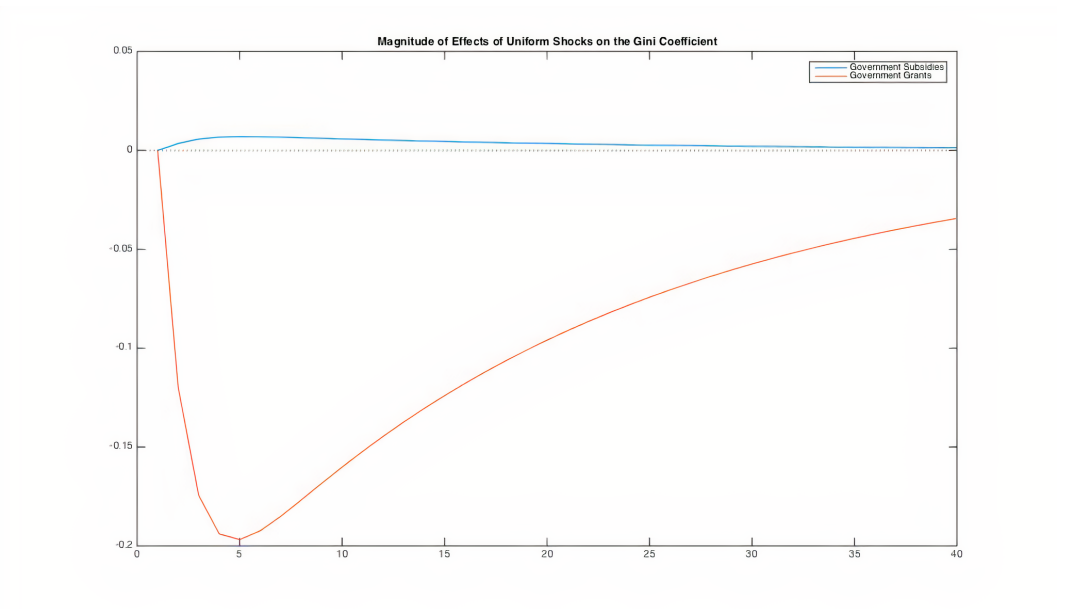
Figure 1.11: Dynamic Impact of a Uniform Shock to Government Grants (g_t) Under Educational Heterogeneity



with rising income inequality to simply reduce tuition fees indiscriminately. Our analysis shows that, because of disparities in the quality of education offered by different establishments, such a policy is instead likely to effectively subsidise the richest students.

Figure 1.11 shows that our conclusions for the case of a uniform increase in government grants differ from the basic setup. In fact, as highlighted in Figure 1.12, increasing the amount spent by the government on improving the quality of education for all agents will decrease inequality by a much greater magnitude than the increase in disparities caused by the increase in subsidies. This is because agents receiving lower quality education will benefit to a greater extent due to the non-convexities exhibited by the human capital equation. In fact, the increase in grants will boost the quality of education of both agents by the same amount. This will cause, for both agent-types, a boost in their human capital, and this despite the decrease in educational investment caused by the jump in tuition fees. Total income will therefore increase for both agents, together with consumption and leisure. Despite the equal jump in educational quality for both agent-types, human capital will increase to a greater extent for low-type individuals. This comes from the concavity of returns to education and the fact that low-type agents start from a lower level of quality. In turn, their

Figure 1.12: Magnitude of the Dynamic Responses of the Gini Coefficient to Uniform Policy Shocks



total income will increase by more than that of high-type agents, so that income inequality is reduced.

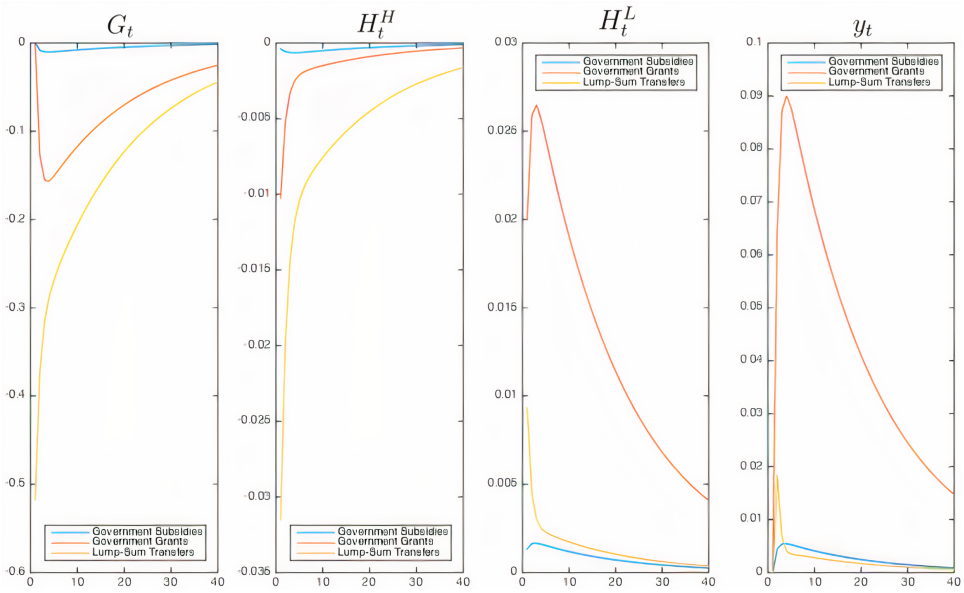
This is the second key result suggested by this analysis. Policy-makers committed to reducing income inequality should focus their efforts on improving the quality of education rather than simply reducing its costs.

1.4.3.2 Asymmetric Policy Shocks Under Educational Heterogeneity

We now consider the impact of asymmetric increases in government grants and subsidies directed toward low-type agents, and lump-sum transfers from high- to low-types. Given the absence of constrained agents in this setup, GSL cannot be analysed.

First, while an increase in government subsidies p_t^L does have a negative impact on the Gini coefficient, its effect is relatively small. By contrast, the decrease in disparities resulting from an asymmetric shock in government grants g_t^L is fifteen times larger. Yet, a lump-transfer T_t^L of similar magnitude will have the greatest impact, with a magnitude more

Figure 1.13: Dynamic Impact of Asymmetric Policy Shocks Under Educational Heterogeneity



than three times larger than that of grants and fifty times larger than subsidies, as shown in Figure 1.13.

While lump-sum transfers will no longer cause a fall in output, it will still be preferable to employ a policy that is more targeted to tackling the source of inequality. In fact, while low-quality agents are able to increase the human capital of their offspring following a lump-sum transfer thanks to their higher income, they are unable do so to a significant extent nor to maintain such increase for many generations, while high-type agents will see a disproportionate fall in their human capital. This will result in very small and temporary gains in total output. Conversely, an increase in government grants will have much more durable and larger positive benefits while limiting adverse effects on high-type agents. In fact, low-type agents will experience a very large increase in their human capital, more than two and a half times larger at its maximum than following a lump-sum transfer, and much more durable. This is because investing in the quality of education offers the double benefit of 'naturally' higher human capital and higher income. In addition, the fall in the human

capital of high-type agents is limited in size and duration, due to the fact that they are simply affected by a temporarily higher tax rate. This limited adverse effect on high-types is essentially the reason why the Gini coefficient does not fall by as much as with lump-sum transfers, but it does not mean that poorer individuals are relatively better off under the latter policy. Furthermore, the vast increase in low-type agents' human capital results in a large increase in total output. Note that the Gini coefficient still decreases by more than the increase in output.

It therefore seems more beneficial for policy-makers to choose government grants over both lump-sum transfers and public subsidies as a strategy to decrease income inequality in the presence of educational heterogeneity. It is also worth noting that under all three policy scenarios, inequality-reducing policies have a positive impact on output, which corroborates recent findings in the literature that the relation between inequality-reducing measures and growth can be positive and complementary.

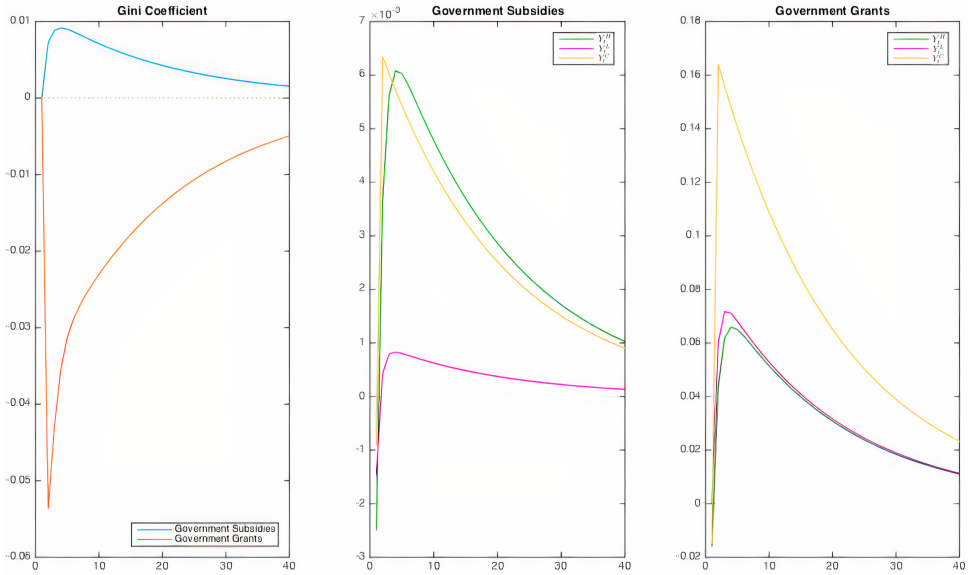
1.4.4 Accounting for Multiple Sources of Income Inequality

Finally, we put together the two frictions analysed above to obtain a more realistic setup that incorporates both liquidity constrained agents and educational heterogeneity. The results we obtain will be tested empirically in Section 1.5.

1.4.4.1 Effectiveness of Uniform Policies in the Complete Setup

Figure 1.14 illustrates what happens when uniform shocks to government subsidies and grants, respectively, are introduced in our complete setup. The general responses of variables analysed in the last two sub-sections will still hold, unless indicated otherwise, so we present impulse response functions for the most relevant variables. Panel 1 shows the reaction of the

Figure 1.14: Dynamic Impact of Uniform Policy Shocks in the Complete Setup



Gini coefficient to both shocks, while panels 2 and 3 show the impact of each policy on the total income of each group. It is important to note that, altogether, the 70% of the population that is liquidity constrained will have a higher total income than the 20% of unconstrained agents having access to low-quality education, which corroborates empirical stylized facts.

We have previously seen that a homogeneous subsidy (p_t) shock will benefit constrained agents to a greater extent than unconstrained agents when educational quality is the same for all, while it will benefit high-quality agents disproportionally when accounting for educational heterogeneity. These two findings seem to drive our results in a setup with all three types of agents. High-quality agents will benefit more than their low-type counterparts, as they will manage to reinvest a greater amount of resources in the education of their children. Subsidies will benefit constrained agents by facilitating their access to schooling. While their total income will rise by more than the average of types of unconstrained agents, high-quality individuals will still benefit to a greater extent – except for the first period after the shock as constrained agents are unable to transfer part of their gains to their child. Therefore, government subsidies will have an exacerbating effect on income inequality by making

low-quality agents fall further behind while high-quality agents will benefit the most.

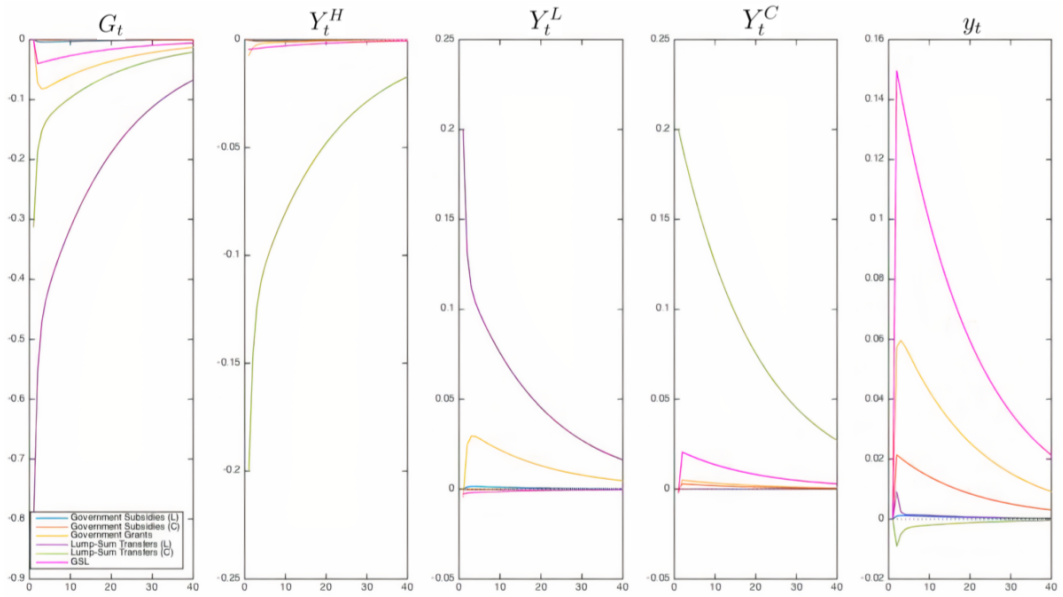
We have also seen that a uniform shock in grants (g_t) will have a greater effect on low-quality agents than high-quality ones, therefore dragging down intra-unconstrained agents' inequality. Implementing such a shock in the presence of constrained agents but not educational heterogeneity produced ambiguous results. In this case however, constrained agents gaining access to education through borrowing will mostly go into low-quality education, while a few others will manage to obtain high-quality education. Since non-Ricardian agents are obligated to repay the GSL upper limit in their adulthood – contrary to unconstrained agents who will decrease their educational investment – they will reduce their leisure when young to accumulate more human capital and earn higher income in order to be able to repay their student loans while facing increased tuition fees reflecting the improved quality of education. This will also allow them to smooth the reduction in their consumption/leisure levels across their lifetime. This result matches empirical findings. In addition, constrained agents will see their income jump to the greatest extent, followed by low-quality and high-quality agents.

As a result, a uniform increase in subsidies will decrease intra-unconstrained agents' inequality, disparities between the constrained and unconstrained, as well as overall inequality. Government grants therefore appear to be a much more efficient homogeneous policy than government subsidies.

1.4.4.2 Effectiveness of Asymmetric Policies in the Complete Setup

Finally, we consider the six asymmetric shocks that this framework allows us to analyse: increases in government subsidies tailored to low-quality agents and constrained agents (p_t^L and p_t^C); an increase in grants going to low-quality schools (g_t^L); an increase in the upper

Figure 1.15: Dynamic Impact of Asymmetric Policy Shocks in the Complete Setup



Note: The legend applies to all panels. The IRFs for the two lump-sum transfers are superimposed in panel 2.

limit of government student loans (d_t); and lump-sum transfers from high-quality to low-quality agents (T_t^L), and from high-quality to constrained households (T_t^C).

Looking at their effect on income inequality, it is clear that lump-sum transfers will produce the most drastic impact. Lump-sum transfers to low-quality agents will be most effective as, on aggregate, these agents have the smallest share of aggregate income. Nevertheless, transfers will only have a minor effect on output, if not negative, given that they produce little inter-generational boost to human capital for its beneficiaries, and a large fall in human capital for high-type agents, as shown in Figure 1.15.

Government subsidies, whether allocated to low-type agents or to constrained individuals, will have little effect on both inequality and output. Grants and GSL, on the other hand, will not only produce a significant reduction in inequalities, but they will also allow a large boost in human capital, as the former will directly improve the quality of education while the latter will allow constrained agents to improve their level of educational attainment. At the same time, both policies will have little adverse impact on high-types, as they will only

incur a temporary increase in their tax rate. Which policy performs best depends on policy-makers' objectives. If interested in decreasing inequality, asymmetric grants will allow to reduce income gaps the most, while GSL will allow a larger boost in output through the increase in human capital of constrained agents.

1.4.5 Implications

This analysis highlights a number of important results. First of all, it illustrates how crucial it is to consider the various potential causes of income inequality jointly, as opposed to the nearly segregated analyses undertaken in the literature. In fact, when considering the policy proposals formulated to address a specific issue while ignoring other potential sources of heterogeneity, results will almost inevitably indicate that those policies are the most efficient, compared to those proposed by other strands of the literature. However, these conclusions may be highly biased and lead to harmful policy recommendations.

One key result shown across this analysis is that reducing the cost of education by homogeneously reducing tuition fees – one of the most popular proposals among policy-makers – will actually have a detrimental effect on inequality when taking into account that agents differ in the quality of the education they receive. What happens effectively is that the education of the richest students is subsidised to a larger extent. Furthermore, even when applying this policy only on children from more disadvantaged backgrounds, subsidies will only have a minor contribution to reducing income inequality, compared to using the same amount of public funds on different measures. Hence, while subsidies do not make any group worse-off over time, it may not be efficient for the government to use them as their primary strategy to reduce inequality.

On the other hand, even when applied uniformly, improving the quality of schools

through public grants will have a shrinking effect on inequality, as students from lower-quality schools will benefit the most. When applied asymmetrically, increases in grants and the GSL upper limit will be the most efficient policies, both to decrease inequality and boost output growth. The optimal use of public funds would therefore be a mix of asymmetric policy shocks in grants and GSL upper limits, which would effectively improve the conditions of, respectively, agents attending low-quality schools and students from constrained families.

We now move to testing some of these results empirically by taking advantage of spatial and temporal variations in different types of state-level spending in the United States.

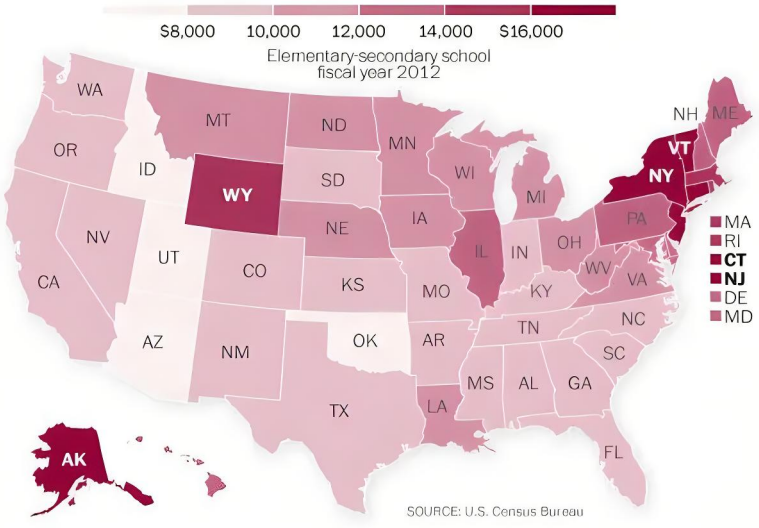
1.5 Empirical Analysis

Our theoretical analysis has produced a number of important results regarding the effects of fiscal educational policies on income inequality. In order to verify the validity of our framework and findings, we test our results empirically in this section. The main objective of this analysis is to evaluate the role of public subsidies and grants in the United States, as well as the importance of credit and liquidity constraints in determining income disparities.

1.5.1 Data

The empirical analysis presented in this section relies on spatial and temporal variations in state-level expenditure and educational costs, controlling for a number of factors. As discussed in Section 1.2, a large portion of the literature looking at the role of educational quality uses students' competence data such as test scores. Nevertheless, this leads to a number of issues that have forced earlier studies to resort to inferior identification strategies.

Figure 1.16: State-Level Public School Expenditure per Pupil, 2012



Source: Berman (2014), based on data from the U.S. Census Bureau

Furthermore, test scores have been deemed less precise in evaluating the quality of teaching (Card and Krueger, 1992; Braga et al., 2011), as opposed to more objective fiscal indicators.

As a measure of government grants, this study uses U.S. data on state-level public education expenditure per student enrolled in elementary and secondary public schools. The advantage of this data, which is collected annually by the U.S. Census Bureau, is that it clearly identifies spending going toward improving schooling quality, as it includes spending on teachers' salaries and wages, pupil support, school administration, school maintenance, food services, and instructional equipment. In particular, it is not aimed at reducing the tuition fees privately incurred by parents. Expenditure per pupil varies extensively across states – ranging from $6,206 for Utah to $19,552 for New York in 2012, as shown in Figure 1.16 – and time (Berman, 2014; U.S. Census Bureau, 2015).

The amount of public subsidies aimed at reducing the cost of education incurred by individuals is calculated based on data on tuition fees, which also varies greatly across states and years. Our index proxies state subsidies by looking at the difference between the total cost of education and the amount effectively charged by public establishments, which corre-

sponds to the portion subsidised by the government. While we only have data on 'effective' tuition fees, we can proxy the total cost of education by identifying the state where fees are the highest in any given year and taking this as the total average cost incurred by schools. The index for government subsidies is then given by:

$$p_{i,t} \equiv F_t - f_{i,t} \tag{1.29}$$

where $p_{i,t}$ is public subsidies for state i in period t; $f_{i,t}$ is average state-level tuition fees paid by individuals; and F_t are the highest state-level tuition fees at time t.[6] Since data on the price of education is not available for elementary and secondary education, we use data on college tuition fees. Nevertheless, this appears to be more appropriate when looking at policies aimed at increasing educational attainment given that 90% of the population has access to secondary education but only one third attends college. In addition, using state-level data allows us to exploit their long time dimension, and we appropriately lag college subsidies and expenditure on primary and secondary education to capture the effect of these two policies on median income. This is a step forward from most of the earlier literature, which looked at cross-sectional samples measuring earnings and educational quality in the same period (see Solga, 2014).

To the best of our knowledge, data on government student loans programs is not publicly available. Nevertheless, as part of our identification strategy, we will need to control for parental income. This will have the benefit of allowing us to evaluate the role of credit constraints in determining disparities in median income and inequality.

Moreover, as outlined in the next sub-section, we control for a number of factors in our

[6]Results remain identical with opposite sign if we simply include tuition fees as an explanatory variable instead of our index for state subsidies, since F_t is the same for all states. Nevertheless, the latter provides more clarity when interpreting our results, as we are interested in government policies.

empirical model. We collected our data on median income, primary and secondary expenditure, tuition fees, and our control variables from various sources. Public expenditure for pupils enrolled in elementary and secondary public schools is publicly accessible from 1993 to 2013 for all 50 states through the Public Education Finances report, published annually by the U.S. Census Bureau. The Census also provides median households income data by state from 1984 to 2014 through the American Community Survey, which also collects data on domestic migration. Decompositions of states' population by ethnicity, on the other hand, are provided by the Henry J. Kaiser Family Foundation. Information on the average tuition fees charged by public colleges for four-year degrees is available for the academic years 2004-05 to 2015-16 from The College Board. State-level unemployment rates are collected by the U.S. Bureau of Labour Statistics. Finally, data on patent counts by state is made available by the U.S. Patent and Trademark Office.

Our empirical models therefore use a balanced panel data set including all 50 U.S. states for varying time spans, ranging from 10 to 21 years.

1.5.2 Empirical Strategy

The use of state-level data departs from most of the earlier literature on the links between education and inequality, which exploits student test scores instead. As a consequence, these studies are unable to make use of inequality decomposition methods to derive the contribution of each factor of interest based on earnings regressions, as their educational variables are correlated with unobservable effects such as students' competitiveness (Checchi and van de Werfhorst, 2014). Instead, their regressions evaluate the effect of the variation in their regressors directly on income disparities, therefore losing explanatory power by resorting to this more modest strategy. Focusing on the supply side of schooling allows us to avoid such

issues and therefore permits a more rigorous two-step approach that is able to exploit the space dimension.

This analysis will therefore consist of two main steps: first, we estimate the impact of government grants and subsidies on median income across states and time using the panel data set outlined above, controlling for a number of factors; next, we derive the effect of these policies on federal-level income disparities using decomposition techniques that will determine the proportional contribution of each variable.

1.5.2.1 Step One: State-Level Panel Analysis

The first step consists in regressing our two measures of interest, state grants and subsidies, on state-level median household income, using an empirical model of the following form:

$$y_{i,t} = \gamma g_{i,t-15} + \beta p_{i,t-5} + \boldsymbol{X_{i,t-r}}\boldsymbol{\alpha} + v_t + \lambda_i + \varepsilon_{i,t} \tag{1.30}$$

where $y_{i,t}$ is median household income for state i at time t; $g_{i,t-15}$ is public spending on elementary and secondary education, lagged by fifteen years; $p_{i,t-5}$ is college subsidies, lagged by five years; $\boldsymbol{X_{i,t-r}}$ is a vector of control variables, whose lag r may vary from one variable to another; v_t and λ_i are time and state fixed effects, respectively; and $\varepsilon_{i,t}$ is the error term. Following part of the earlier literature, we do not use log transformations – although results remain largely unchanged when doing so. Lags on the two policy variables were chosen to reflect the length of the U.S. educational system, so that all agents benefitting from policies in the lagged period have passed the working-age at time t.

There are a number of factors discussed in the literature that we may want to control for. First, some studies have suggested that returns to education in the U.S. may be smaller for households identified as belonging to an ethnic minority, due to racial discrimination

on the job market resulting in lower income for a given education level. This might bias our coefficients on educational variables downwards, and we therefore successively test the inclusion of the state-level share of all ethnic minorities and of African-American minorities only – since some minorities do not appear to suffer from lower returns to education (Joo et al., 2016). Secondly, there is a potential endogeneity issue which the early literature has heavily emphasised (Card and Krueger, 1992): in fact, it may be that higher state-level expenditure per pupil on educational quality is driven by the higher median income of parents, as public schools seek to compete with private ones by meeting the higher expectations of wealthier parents, and states have larger budgets to meet those demands. The coefficient on state expenditure would then be biased upwards as it captures this effect. At the same time, this may bias the coefficient on state subsidies downwards, as states with richer households may not be willing to spend as much to make public college less expensive. Therefore, we include a lag of the dependent variable in our controls, and test the following specification:[7]

$$y_{i,t} = \delta y_{i,t-15} + \gamma g_{i,t-15} + \beta p_{i,t-5} + \boldsymbol{X_{i,t-r}\alpha} + v_t + \lambda_i + \varepsilon_{i,t} \tag{1.31}$$

This allows to capture to capture the effects of parental income and, at the same time, offers the opportunity to evaluate the impact of parents' credit constraints on earnings and inequality, using the same approach as most of the literature on the matter but in a different context and controlling for educational policy effects. State fixed effects will in addition control for the fact that some states are permanently wealthier than others.[8]

Two further issues arise in this analysis. First, it will be the case that some individu-

[7]We estimate equation (1.31) using both OLS and GMM estimators, to correct for potential Nickell-bias issues.

[8]Furthermore, note that both our measures of public spending are computed in per-student form, which makes it unnecessary to control for population growth.

als will work in a different state than the one they were educated in. While the share of Americans changing residence has kept decreasing over the past decades, with as many as 9 in 10 Americans not having changed location in 2008 (Cohn and Morin, 2008), we control for net migration as a robustness check. Net migration should be expected to bias policy coefficients upwards as domestic migrants are often wealthier than those who are unable to move. Secondly, asymmetric macroeconomic conditions may bias the estimated effect of educational policies on income by affecting median state income directly. For instance, if the unemployment rate increases due to an adverse macroeconomic shock, the benefit of educational policies on income will appear to be lower, even though the change in unemployment was not due to a change in policy. State-level unemployment rates are therefore included as controls. As a further robustness check, we include a lagged value of the unemployment rate to ensure that changes in educational policies are not simply the result of macroeconomic shocks affecting states' budgets.

In addition, we are interested in testing whether technological progress is important in driving median income and earnings inequality and, if so, whether it is skill biased. Nevertheless, given the high degree and speed of knowledge transmission within the United States, we expect to find a small effect on disparities from technological change. While we would expect richer states to contribute more to technological advances, therefore raising issues of causality, state fixed effects should address this concern.

Finally, we are able to analyse the long-term impact of our policies and control variables on median income and inequality by retrieving the coefficients from a steady-state transformation of equation (1.31):

$$y_i^* = \frac{\gamma}{1-\delta} g_i^* + \frac{\beta}{1-\delta} p_i^* + \boldsymbol{X_i^* \alpha^*} + \frac{v^*}{1-\delta} + \frac{\lambda_i}{1-\delta} \tag{1.32}$$

where x_i^* indicates the steady-state value of any variable $x_{i,t}$; and $\boldsymbol{\alpha}^* = (1-\delta)^{-1}\boldsymbol{\alpha}$. We can then compare these coefficients to those obtained by estimating in equation (1.30).

1.5.2.2 Step Two: Effects on Federal-Level Income Inequality

The second step of the analysis consists in decomposing the contribution of each variable to overall income inequality. Modern methods of inequality decomposition were pioneered by Shorrocks (1980, 1982, 1984), and have become a central element of inequality studies (for a review of inequality decomposition techniques, see Heshmati, 2004).

While it may be tempting to use the Gini coefficient as our measure of income disparities, it has been shown that decomposing inequality through a regression-based approach using the Gini coefficient can lead to misleading results. As best explained by Morduch and Sicular (2002), this is because "the Gini decomposition is best interpreted as quantifying the factors that explain observed inequality conditional on the value of the overall Gini coefficient", so that it carries little information as to what led income inequality to reaching that level in the first place. A more appropriate measure of the contribution of each factor to income inequality is the Theil-T index, which measures the relative entropy of income, and whose natural decomposition rule is given by (Morduch and Sicular, 2002; Troch, 2014):

$$T_t = \frac{1}{n}\sum_{i=1}^{n}\frac{y_{i,t}}{\mu_t} ln\left(\frac{y_{i,t}}{\mu_t}\right) \tag{1.33}$$

where T_t is the Theil-T index in period t; n is the number of individuals or, in this case, states; $y_{i,t}$ is the median income of state i in period t; and μ_t is the average median income across states in period t. This measure not only allows us to decompose the contribution of different factors to overall inequality, but it also explains how inequality reached such value

in the first place.

The empirical proportional contribution of variable j to aggregate income inequality is then given by:

$$C_t^j = \theta^j \frac{\frac{1}{n}\sum_{i=1}^{n} x_{i,t}^j ln\left(\frac{y_{i,t}}{\mu_t}\right)}{\frac{1}{n}\sum_{i=1}^{n} y_{i,t} ln\left(\frac{y_{i,t}}{\mu_t}\right)} \tag{1.34}$$

where $x_{i,t}^j$ is the appropriately-lagged jth element of the vector of explanatory variables $x_{i,t}$, and θ^j is the coefficient of this variable estimated from the panel data analysis. Note that this measure varies across time, so that we will be able to analyse the evolution of the contribution of each variable. Standard errors are given by:

$$\sigma(C_t^j) = \sigma(\theta^j) \frac{\frac{1}{n}\sum_{i=1}^{n} x_{i,t}^j ln\left(\frac{y_{i,t}}{\mu_t}\right)}{\frac{1}{n}\sum_{i=1}^{n} y_{i,t} ln\left(\frac{y_{i,t}}{\mu_t}\right)} \tag{1.35}$$

where $\sigma(\cdot)$ indicates the standard error. Note therefore that the significance level of the proportional distribution to overall inequality of each variable will be the same as the significance level of their coefficient in the panel regression, $\theta^j/\sigma(\theta^j)$.

1.5.3 Empirical Results

In this sub-section, we present our findings from the panel analysis and the inequality decomposition regarding the determinants of household income disparities across U.S. states. We also undertake a number of robustness checks to reinforce the validity of our findings. Results from the panel analysis and inequality decomposition are provided, respectively, in the 'estimates' and 'Theil-T' columns of Table 1.2 in the Appendix. All of the regressions were estimated using ordinary least squares on all 50 U.S. states using fixed effects, for comparison with earlier studies.[9] Models (1) to (4) provide estimation results for equation

[9]GMM estimates were also computed to correct for potential Nickell-bias problems caused by the inclusion of lagged income. Results were not fundamentally altered so we avoid reproducing them here.

(1.30), while models (5) to (6) estimate equation (1.31).

We start by looking at models (1) and (2), which include our two policy variables without any controls. Model (2) seems to confirm from the onset one of the fundamental findings of this paper: public expenditure per pupil provides a very large, highly significant, negative contribution to overall income inequality in the United States, while public subsidies put a smaller, yet significant, upward pressure on inequality. Not only that, but making college more affordable actually reduces median income, which suggests that financing such policy comes at the expense of those who are still unable to access college education. Results from model (1) are reported to show the consistency of the significance of public expenditure in a larger sample. Expenditure per pupil alone seems to explain a quarter of inter-state variations in median income. These results also suggest that policies affecting primary and secondary education, which are attended by nearly 90% of the population, will have greater effects on inequality than measures affecting college education, which is only attended by 30% of Americans, despite subsidies supposedly increasing tertiary enrolment.

In terms of public subsidies, our theoretical analysis is able to provide an explanation for the first effect: their positive impact on overall inequality seems to be caused by children from wealthier backgrounds excessively benefitting from lower tuition fees. Children from more modest families will also benefit, but to a lower extent, while the income of those having access to college will move away from that of individuals who will still not be able to afford a tertiary education, therefore driving income disparities up. The second observed effect from public subsidies – the decrease in median income – seems to be caused by more qualified workers – and a higher number of them – crowding out less qualified individuals on the labour market. The two effects from expenditure per pupil, on the other hand, perfectly match our theoretical analysis, with higher educational quality benefitting children

from disadvantaged background to a greater extent and increasing median income by raising human capital.

Next, we add the unemployment rate and the state-level share of minorities in models (3) and (4), to control for macroeconomic conditions and racial discrimination on the labour market, respectively. In particular, model (3) uses the share of all minorities while model (4) incorporates the share of African-American minorities only. We also include a measure of patent counts by state in both models, as we are interested in the contribution of technological progress to income inequality.

The coefficient on the unemployment rate is highly significant, with an increase in unemployment making median income fall by a large extent. The effect on overall inequality is substantial but much smaller than that of educational policies, perhaps because a significant portion of macroeconomic shocks affecting the U.S. are sufficiently symmetric across states to be captured by time fixed effects. The inclusion of the unemployment rate raises both the absolute value and the significance of the coefficient on public subsidies.

Patents also have a highly significant and positive correlation with median income and they have a positive effect on income disparities. This suggests that technological progress is skill-biased, although, as expected, the size of their proportional contribution to overall inequality is only 2.5% and 3.4% in models (3) and (4), respectively, given the speed of knowledge transmission within the United States and the relatively small degree of heterogeneity across states.[10]

A higher share of minorities in a state is correlated with higher median income, as shown by model (3). This result does not seem to reflect the effect of some minorities having lower

[10] Although this was not reported in Section 1.4 to save space, positive technological shocks in all the setups of our theoretical model that incorporated either educational heterogeneity or liquidity constraints led to higher income disparities, as wealthier individuals were more positively affected.

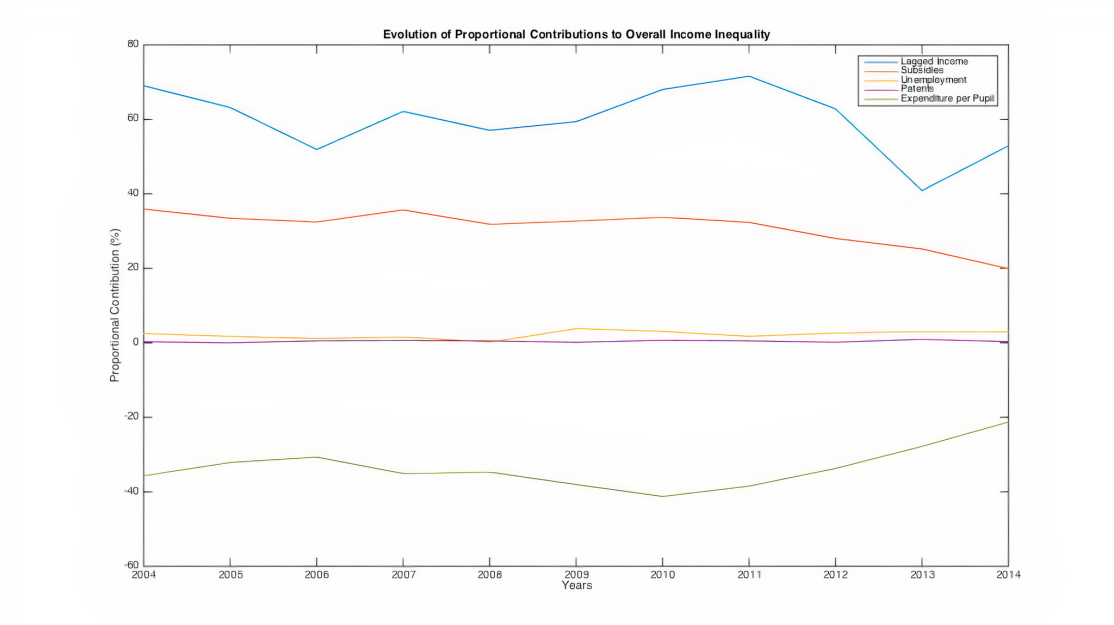
returns to education due to discrimination on the labour market, and we therefore include in model (4) the share of African-Americans only. Despite the coefficient on the latter not being significant, its negative sign and positive contribution to overall inequality seems to capture the effect that we want to control for. As argued by Card and Krueger (1992), findings of insignificant coefficients on minority dummies in a panel earnings equation highlights the effectiveness of educational policies aiming to improve schooling quality, as the U.S. has actively sought to reduce racial gaps in returns to education in the past decades.

In order to identify the causal direction of the effect of educational policies on median income, models (5) and (6) also include a lagged value of the dependent variable to control for parents' income, which will also indicate the contribution of credit constraints to income inequality.

Our prior regarding the bias in the policy variables seems to be confirmed by models (5) and (6), as the coefficient on expenditure per pupil is greatly decreased and so is that on public subsidies, in absolute terms. Their proportional contribution to overall inequality is also reduced, as that of lagged median income is as high as 62%, suggesting that credit constraints are the biggest determinant of income disparities across states. Nevertheless, both policies remain highly significant and have higher proportional contributions than control variables. Note in addition that once we control for parental income, public subsidies have a larger effect on inequality than expenditure. Furthermore, both patent counts and unemployment have much smaller coefficients once we control for parental income. Their effect on median income remains highly significant, yet their proportional contributions to inequality become very small.

We report in model (7), Table 1.3, the steady-state coefficients retrieved from transforming model (6) using equation (1.32). The long-term impact of subsidies, grants and our

Figure 1.17: Evolution of Proportional Contributions to Overall Income Inequality



Note: The proportional contribution of Subsidies is presented here without lags to extend the time-series. Expenditure per Pupil and Lagged Income were constructed assuming that the coefficients found in model (6) could be extended to previous periods.

control variables on median income are all larger than in the short run, and their contribution to overall inequality is substantially increased. This result holds even compared to the coefficients of model (4).

Finally, we are able to look at the evolution of proportional contributions across time. Figure 1.17 plots an extended view of the proportional contributions of each variable from 2004 to 2014, based on model (6). Keeping in mind that the sum of contributions from all variables, fixed effects, and the residual term equals 100% in each period, this figure allows us to analyse the relative evolution of each factor in driving income inequality.

A number of notable results must be highlighted. First, it is clear that credit constraints and public subsidies have been the largest contributors to income inequality over the past decade, while expenditure per pupil has been the only factor driving it downwards. Unemployment, technological progress, and the share of African-Americans, by contrast, have had very small contributions over the years. Secondly, the contributions of all three major

variables have declined over the past few years – although that of constraints has peaked up again in 2013-14. It therefore appears that some structural changes have affected the determinants of income inequality over the years, which may however be partly due to the 2007-8 financial crisis, as illustrated by the small rise in the contribution of the unemployment rate, which would instead suggest that these are cyclical changes.

We also undertake a number of robustness checks to verify the validity of our results, which we report in Table 1.3. First, we want to see whether individuals changing states between their studies and their adulthood may alter our results. We therefore include a measure of net migration for each state, with results presented in model (8). Our previous findings remain largely unchanged, and the coefficient on net migration is insignificant, with only a minor contribution to overall inequality and the expected sign. Secondly, to verify that educational policies are not simply driven by changes in states' budgets caused by macroeconomic shocks, we include a lagged value of the unemployment rate. Again, this does not alter our results and the coefficient is small and insignificant.

1.6 Concluding Remarks

This paper has sought to identify the most effective policies to reduce rising income disparities in the United States. It has done so by analysing the impact of a number of fiscal measures on a model with liquidity constraints and educational heterogeneity. It has also demonstrated the importance of jointly considering the various potential drivers of income disparities when seeking to provide policy recommendations aimed at reducing them. In fact, taking into account one issue at a time – as has been done by most of the literature thus far – has led to a number of misled conclusions.

With this in mind, we have reconsidered the main policy suggestions put forward by the two leading branches of the literature – one that has focused on schooling quality, and the other on credit and liquidity constraints – by developing a new framework to think about fiscal policy when agents differ in a number of fundamental ways. Our analysis suggests that once educational heterogeneity is taken into account, one of the major measures proposed by the credit constraints literature, namely reducing the costs of education, has the opposite effect to the one expected, as students from wealthier backgrounds will excessively benefit from it. Government grants aiming to improve the quality of education, on the other hand, will benefit poorer students the most, even when applied uniformly, and will therefore reduce income inequality whenever we take into account educational heterogeneity. When considering asymmetric policies aiming to benefit non-Ricardian agents or students receiving low-quality education, lump-sum transfers appear to be most efficient at reducing income inequality, but they come at an important cost for richer families and carry little inter-generational benefits to their beneficiaries. It is also the only policy that has the potential of giving rise to a trade-off between lower inequality and higher growth. Government grants and government student loans programs, on the other hand, appear to be the most effective policies to reduce income disparities when applied heterogeneously, the former benefitting students having access to low-quality education the most, and the latter lifting constrained agents by improving their access to education. These two policies also allow to significantly and durably improve output growth. A combination of the two therefore constitutes the optimal fiscal measure when educational heterogeneity and credit constraints are taken into account.

In order to test the validity of our framework and results, we have undertaken a two-step panel analysis of the policy drivers of U.S. income inequality. To the best of our knowledge,

this is the first paper that is able to evaluate the impact of policies aiming to address different sources of income disparities, and we are able to do so using modern inequality decomposition techniques, thanks to the supply-side nature of our measure of educational quality. Furthermore, the long temporal dimension of this type of data allows us to appropriately capture the effect on earnings of policies affecting individuals in their youth. Our analysis confirms that public subsidies have had a positive contribution to income inequality, while expenditure per pupil has reduced it. Furthermore, we have shown that credit constraints have been the largest determinant of income gaps.

Our empirical results therefore confirm our theoretical conclusions that the most effective policy mix to tackle income disparities would be to allocate larger funds to improving schooling facilities and teaching quality, while facilitating the access to secondary and tertiary education by raising government student loans' upper limits on per-student allowances rather than public subsidies. These findings come at a time when the goal of inequality reduction has gained vast momentum in the United States and in other industrialised economies across the political spectrum. Yet, policy-makers have received mixed signals from academics and researchers as to the best strategy to achieve this and seem to be coming to a consensus that a uniform reduction in college tuition fees is the most efficient option available.

Shifting the focus of inequality-reduction strategies toward educational quality and student loans will require a number of further considerations that would allow to optimise the implementation of these policies. In particular, the United States has seen a constant exacerbation in the amount of student debt, which is becoming a major national issue. While raising the GSL upper limit will enable a reversal of the growing importance of private student loan providers who tend to charge much higher interests, students from poorer fam-

ilies are the ones carrying the largest amounts of student debt, which often takes them a lifetime to repay, therefore hindering their capacity to invest in the education of their own children. Student debt therefore contributes to raising the inter-generational effects of credit constraints which, as shown in this paper, are the largest contributor to income disparities. In light of this, despite their lack of efficiency in reducing inequalities even when applied asymmetrically, public subsidies could be used to reduce the debt burden of children from poorer backgrounds and therefore improve social mobility.

The theoretical and empirical frameworks developed in this paper would allow to further investigate the student debt problem, as well as many other policy questions relating to income disparities. Whatever the issue at hand, however, analyzing it while taking into account the various drivers of household heterogeneity will be crucial to formulating sound policy recommendations able to provide meaningful and durable solutions to the growing problem of income inequality.

References

Acemoglu, D. and Autor, D. (2012), 'What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology', *Journal of Economic Literature* **50**(2).

Alvaredo, F., Atkinson, T., Piketty, T. and Zucman, G. (2016), 'The world wealth and income database'.
URL: *http://www.wid.world*

Autor, D. H., Katz, L. K. and Krueger, A. B. (1998), 'Computing Inequality: Have Computers Changed the Labor Market?', *The Quarterly Journal of Economics* **113**(4).

Barro, R. (1991), 'Economic Growth in A Cross Section of Countries', *Quarterly Journal of Economics* **106**(May), 407–42.

Belley, P., Frenette, M. and Lochner, L. (2011), 'Post-Secondary Attendance by Parental Income in the U.S. and Canada: What Role for Financial Aid Policy?', *National Bureau of Economic Research, Working Paper 17218* .

Belley, P. and Lochner, L. (2007), 'The Changing Role of Family Income and Ability in Determining Educational Achievement', *Journal of Human Capital* **1**(1), 37–89.

Ben-Porath, Y. (1967), 'The Production of Human Capital and the Life Cycle of Earnings', *Journal of Political Economy* **84**, 449–472.

Berkner, L. K. (2000), 'Trends in Undergraduate Borrowing: Federal Student Loans in 1989-90, 1992-93, and 1995-96', *National Center for Education Statistics Report* **2000**(151).

Berman, L. K. (2014), 'Here's how much each state spends on public school students', *The Washington Post* .

Berruyer, M. (2010), 'Income inequality in the u.s.', *The Crises* .
URL: *http://www.the-crises.com/income-inequality-in-the-us-1/*

Bewley, T. (1977), 'The Permanent Income Hypothesis: A Theoretical Formulation', *Journal of Economic Theory* **16**(2), 252–292.

Bivens, J. (2017), 'Inequality Is Slowing U.S. Economic Growth', *Economic Policy Institute* .

Black, S. E. and Devereux, P. J. (2010), 'Recent Developments in Intergenerational Mobility', *National Bureau of Economic Research, Working Paper No. 15889* .

Blau, F. and Kahn, L. (2005), 'Do Cognitive Test Scores Explain Higher U.S. Wage Inequality?', *Review of Economics and Statistics* **87**(1), 184–193.

Boldrin, M. and Montes, A. (2004), 'The Intergenerational State: Education and Pensions', *Federal Reserve Bank of Minneapolis, Staff Report 336* .

Braga, M., Checchi, D. and Meschi, E. (2011), 'Institutional Reforms and Educational Attainment in Europe: A Long Run Perspective', *Institute for the Study of Labor, Discussion Paper No. 6190* .

Cameron, S. and Heckman, J. J. (1998), 'Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males', *Journal of Political Economy* **106**(2), 262–333.

Cameron, S. and Heckman, J. J. (1999), Can tuition policy combat rising wage inequality?, *in* M. H. Kosters, ed., 'Financing College Tuition: Government Policies and Educational Priorities', American Enterprise Institute Press, pp. 67–124.

Card, D. (1999), 'The Causal Effect of Education on Earnings', *Handbook of Labor Economics* **3**(A).

Card, D. and DiNardo, J. E. (2002), 'Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles', *Journal of Labor Economics* **20**(4).

Card, D. and Krueger, B. (1992), 'Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States', *The Journal of Political Economy* **100**(1), 1–40.

Carneiro, P. and Heckman, J. J. (2002), 'The Evidence on Credit Constraints in Post-Secondary Schooling', *Economic Journal* **112**(482), 705–734.

Castells-Quintana, D. and Royuela, V. (2017), 'Tracking Positive and Negative Effects of Inequality on Long-Run Growth', *Empirical Economics* **53**, 1349–1378.

Chapman, B. (2006), Income contingent loans for higher education: International reforms, *in* E. Hanushek and F. Welch, eds, 'Handbook of the Economics of Education', Vol. 2, pp. 1435–1503.

Checchi, D. and van de Werfhorst, H. G. (2014), 'Educational Policies and Income Inequality', *Institute for the Study of Labor, Discussion Paper No. 8222* .

Cohn, D. and Morin, R. (2008), 'Who Moves? Who Stays Put? Where's Home?', *Pew Research Center* .

Coleman, J. S. (1966), 'Equality of Educational Opportunity', *Government Printing Office* .

Dabla-Norris, E. and Matovu, J. M. (2002), 'Composition of Government Expenditures and Demand for Education in Developing Countries', *IMF Working Paper* **02**(78).

den Haan, W. (2011), 'Why do we need a financial sector?'.
URL: *https://voxeu.org/article/why-do-we-need-financial-sector*

Eide, E. R. and Showalter, M. H. (2005), 'Does Improving School Quality Reduce the Probability of Unemployment?', *Contemporary Economic Policy* **23**(4), 578–584.

Freeman, R. and Schettkat, R. (2001), 'Skill Compression, Wage Differentials and Employment: Germany vs U.S.', *Oxford Economic Papers* **53**(3), 582–603.

Galor, O. and Zeira, J. (1993), 'Income Distribution and Macroeconomics', *Review of Economic Studies* **60**(1), 35–52.

Glomm, G. (1997), 'Parental Choice of Human Capital Investment', *Journal of Development Economics* **53**(1), 99–114.

Glomm, G. and Ravikumar, B. (1992), 'Public Versus Private Investment in Human Capital: Endogenous Growth and Income Inequality', *Journal of Political Economy* **100**, 818–834.

Glomm, G. and Ravikumar, B. (1998), 'Flat-Rate Taxes, Government Spending on Education, and Growth', *Review of Economic Dynamics* **1**(January), 306–325.

Goldin, C. D. and Katz, L. F. (2008), *The Race between Education and Technology*, Harvard University Press.

Hanushek, E. A. (1986), 'The Economics of Schooling: Production and Efficiency in Public Schools', *Journal of Economic Literature* **24**(3), 1141–1177.

Hanushek, E. A. and Wößmann, L. (2007), 'The Role of Education Quality in Economic Growth', *World Bank Policy Research Working Paper* **4122**.

Haskins, R. and Sawhill, V. (2008), Education and Economic Mobility, *in* R. H. Isaacs, J. B. and I. V. Sawhill, eds, 'Getting Ahead or Losing Ground: Economic Mobility in America. Economic Mobility Project', The Pew Charitable Trusts and Brookings Institution.

Heckman, J. J. (1999), 'Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programs', *National Bureau of Economic Research, Working Paper 6230* .

Heshmati, A. (2004), 'A Review of Decomposition of Income Inequality'. Institute for the Study of Labor, Discussion Paper No. 1221.

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D. J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K. and Houser, T. (2017), 'Estimating Economic Damage from Climate Change in the United States', *Science* **356**(6345), 1362–1369.

Huggett, M., Ventura, G. and Yaron, A. (2011), 'Sources of Lifetime Inequality', *American Economic Review* **101**(7), 2923–54.

Isaacs, J. B. (2008), Economic Mobility of Families Across Generations, *in* R. H. Isaacs, J. B. and I. V. Sawhill, eds, 'Getting Ahead or Losing Ground: Economic Mobility in America. Economic Mobility Project', The Pew Charitable Trusts and Brookings Institution.

Johnson, G. E. and Stafford, F. P. (1973), 'Social Returns to Quantity and Quality of Schooling', *Journal of Human Resources* **8**(Spring), 139–55.

Joo, N., Reeves, R. V. and Rodrigue, E. (2016), 'Asian American Success and the Pitfalls of Generalization', *The Brookings Institution, Social Mobility Memos* .

Karlsson, M., Nilsson, T., Lyttkens, C. H. and Leeson, G. (2010), 'Income Inequality and Health: Importance of a Cross-Country Perspective', *Social Science & Medicine* **70**(6), 875–885.

Klein, E. (2015), 'Krugman: Soaring Inequality Isn't About Education; It's About Power', *Vox* .

Krugman, P. (2015), 'Knowledge Isn't Power', *The New York Times* .

Kuznets, S. (1955), 'Economic Growth and Income Inequality', *American Economic Review* **45**(March), 1–28.

Lang, K. (1993), 'Ability Bias, Discount Rate Bias, and the Return to Education', *Boston University, Department of Economics Discussion Paper* .

Liu, A. Y. C. (2001), 'Flying Ducks? Girls' Schooling in Rural Vietnam', *Asian Economic Journal* **15**(4), 385–404.

Livingstone, D. W. (1997), 'The Limits of Human Capital Theory: Expanding Knowledge, Informal Learning and Underemployment', *Policy Options, July/August* p. 9–13.

Lochner, L. and Monge-Naranjo, A. (2011*a*), 'Credit Constraints in Education', *National Bureau of Economic Research, Working Paper 17435* .

Lochner, L. and Monge-Naranjo, A. (2011*b*), 'The Nature of Credit Constraints and Human Capital', *American Economic Review* **101**(6).

Lucas, R. E. (1990), 'Why Doesn't Capital Flow from Rich to Poor Countries?', *The American Economic Review* **80**(2), 92–96.

Mincer, J. (1974), 'Schooling, Experience and Earnings', *National Bureau of Economic Research* .

Morduch, J. and Sicular, T. (2002), 'Rethinking Inequality Decomposition, with Evidence from Rural China', *The Economic Journal* **112**(1), 93–106.

Morgan, J. N. and Sirageldin, I. (1968), 'A Note on the Quality Dimension in Education', *Journal of Political Economy* **76**(September/October), 1069–77.

OECD (2013), 'Reducing Income Inequality While Boosting Economic Growth: Can It Be Done?', *Economic Policy Reforms* **2013**.

OECD (2014), 'Trends in Income Inequality and its Impact on Economic Growth', *OECD Social, Employment and Migration Working Paper No.* **163**.

OECD (2016), 'Economic Policy Reforms 2016', *Going for Growth: Interim Report* .

Piketty, T. (2014), *Capital in the Twenty-first Century*, Harvard University Press.

Piketty, T. and Saez, E. (2014), 'Inequality in the Long Run', *Science* **344**(6186).

Rao, R. and Jani, R. (2008), 'School Quality, Educational Inequality and Economic Growth', *International Education Studies* **1**(2).

Reeves, R. V. and Joe, N. (2016), 'How Much Social Mobility Do People Really Want?', *The Brookings Institution* .

Reeves, R. V. and Rodrigue, E. (2014), 'One Year's Progress on President Obama's 'Defining Challenge of Our Time'', *The Brookings Institution, Social Mobility Memos* .

Rizzuto, R. and Watchel, P. (1980), 'Further Evidence on the Returns to School Quality', *Journal of Human Resources* **15**(Spring), 240–254.

Roeger, W. and in't Veld, J. (2009), 'Fiscal Policies with Credit Constrained Households', *European Commission, Directorate-General for Economic and Financial Affairs, Economic Paper* **357**.

Saez, E. (2013), 'Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2009 and 2010 estimates). Updated version of Striking It Richer: The Evolution of Top Incomes in the United States', *Pathways Magazine, Stanford Center for the Study of Poverty and Inequality* p. 6–7.

Satz, D. and White, S. (2021), 'What Is Wrong With Inequality?', *IFS Deaton Review of Inequalities* .

Schmitt, J., Schierholz, H. and Mishel, L. (2013), 'Don't Blame the Robots Assessing the Job Polarization Explanation of Growing Wage Inequality', *Economic Policy Institute* .

Shorrocks, A. F. (1980), 'The Class of Additively Decomposable Inequality', *Econometrica* **48**(3), 613–625.

Shorrocks, A. F. (1982), 'Inequality Decomposition by Factor Components', *Econometrica* **50**(1), 193–211.

Shorrocks, A. F. (1984), 'Inequality Decomposition by Population Subgroups', *Econometrica* **52**(6), 1369–1385.

Solga, H. (2014), 'Education, Economic Inequality and the Promises of the Social Investment State', *Socio-Economic Review* **12**(2), 269–297.

Syll, L. S. (2014), 'Piketty and the Limits of Marginal Productivity Theory', *Real-World Economics Review* **69**.

Titus, M. (2002), 'Supplemental Table Update for Trends in Undergraduate Borrowing: Federal Student Loans in 1989-90, 1992-93, and 1995-96'.

Troch, T. (2014), 'Wealth Inequality in Dynamic Stochastic General Equilibrium Models', *Charles University of Prague* .

U.S. Census Bureau (2014), 'Educational Attainment in the United States in 2014 – Detailed Tables'.
URL: *census.gov/hhes/socdemo/education/data/cps/2014/tables.html*

U.S. Census Bureau (2015), 'Per Pupil Spending Varies Heavily Across the United States', *News Release* .

Watchel, P. (1976), 'The Effects on Earnings of School and College Investment Expenditures', *Revue of Economics and Statistics* **58**(August), 326–31.

Welch, F. (1966), 'Measurement of the Quality of Schooling', *American Economic Review* **56**(May), 379–92.

Appendix 1.A Tables

Table 1.1: Calibrated Parameters

β	0.95	Inter-generational discount	ρ_T	0.95	Tax persistence
χ	0.9	Weight on leisure	σ_T	0.54772	Tax standard deviation
γ_1	0.7	Returns to quality	ρ_H^q	0.95	High-quality persistence
γ_2	0.7	Returns to attainment	ρ_L^q	0.95	Low-quality persistence
δ	0.9	Quality-fees correlation	σ_H^q	0.5972	High-quality standard dev.
μ_H	0.3	Portion of high types	σ_L^q	0.83255	Low-quality standard dev.
μ_L	0.7	Portion of low types	σ^q	0.5972	Uniform standard deviation
α_H	0.6	High-quality labour share	θ_H^q	-0.0178	High-quality adjustment
α_L	0.4	Low-quality labour share	θ_L^q	-0.0602	Low-quality adjustment
ρ_A	0.95	Technological persistence	ϕ_H	1	High-quality scaling
σ_A	1	Technology standard dev.	ϕ_L	2.3	Low-quality scaling
κ	0.95	Subsidy persistence	λ	1	Spending effectiveness
θ_H^p	-0.1151	High-type subsidy adj.	ρ_a^H	0.1	High-type mobility
θ_L^p	-0.1151	Low-type subsidy adj.	ρ_a^L	0.2	Low-type mobility
σ_p	1.1774	Subsidy standard deviation	σ_a	1	Idiosyncratic standard dev.
Additional parameters from the extended model					
ψ	0.15	Access to high-quality schools	b	0.0001	Unemployment benefits
θ^d	0.005	GSL upper limit adjustment	μ_C	0.7	Portion of constrained
σ_d	0.3	GSL standard deviation	ρ_d	0.95	GSL persistence

Table 1.2: Results From the Panel Analysis

	(1)		(2)		(3)		(4)		(5)		(6)	
	Estimates	Theil-T	Estimates	Theil-T	Estimates	Theil-T	Estimates	Theil-T	Estimates	Theil-T	Estimates	Theil-T
Expenditure per Student	2.63*** (0.62)	-97.5	1.69*** (0.68)	-88.0	1.53** (0.66)	-85.3	1.56*** (0.64)	-85.8	0.15** (0.07)	-24.8	0.19*** (0.07)	-31.5
Subsidies			-0.79* (0.61)	52.3	-1.05** (0.56)	82.9	-0.92** (0.56)	80.1	-0.13** (0.07)	28.4	-0.15*** (0.06)	32.7
Patent Counts					0.47*** (0.20)	2.5	0.63*** (0.16)	3.40			0.08*** (0.02)	0.4
Unemployment Rate					-1972.72*** (437.9)	11.7	-1647.15*** (460.6)	-9.8			-374.49*** (65.7)	2.2
All Minorities					114.3** (58.81)	-0.03						
African-Americans							-1.60 (1.58)	0.0005			-0.11 (0.21)	0.0001
Lagged Income									0.91*** (0.02)	62.0	0.88*** (0.02)	59.9
r-squared	0.251		0.229		0.381		0.361		0.867		0.871	
Observations	800		300		300		300		300		300	

Note: Asterisks (, **, ***) are significance levels (10%, 5%, 1%), given by one-tailed t-tests. Values in brackets are standard errors. The dependent variable is state median households income. The Theil-T columns give the average proportional contribution of each variable to aggregate income inequality over the sample period, expressed in percentage-points. Expenditure per Student is state-level total elementary-secondary expenditure per pupil enrolled in a public school, lagged by 15 periods. Subsidies is an index measuring the amount spent by the state to subsidise college tuition fees, lagged by 5 periods. Lagged Income is the dependent variable lagged by 15 periods. All Minorities is the share of individuals identified as non-white in a state, while African-Americans is the share of individuals identified as African-American. Patent Counts is the number of patents granted by state. Unemployment Rate is the percentage of the labour force that is unemployed.*

Table 1.3: Long-Term Model and Robustness Checks

	(7)		(8)		(9)	
	Estimates	Theil-T	Estimates	Theil-T	Estimates	Theil-T
Expenditure per Student	1.58	-88.1	0.20*** (0.07)	-33.1	0.19*** (0.07)	-31.5
Subsidies	-1.25	84.3	-0.16*** (0.06)	34.9	-0.15** (0.06)	32.7
Patent Counts	0.67	3.35	0.09*** (0.02)	0.5	0.09*** (0.03)	0.5
Unemployment Rate	-3120.75	18.1	-392.44*** (71.4)	2.3	-353.65*** (68.1)	2.1
African-Americans	-0.92	0.0008	-0.11 (0.22)	0.0001	-0.12 (0.22)	0.0001
Lagged Income			0.88*** (0.02)	59.9	0.88*** (0.02)	59.9
Net Migration			0.89 (0.71)	0.31		
Lagged Unemployment					-90.45 (158.6)	0.23
r-squared	0.871		0.871		0.872	
Long Term	Yes		No		No	
States, Observations	50, 300		50, 300		50, 300	

Note: Asterisks (, **, ***) are significance levels (10%, 5%, 1%), given by one-tailed t-tests. Values in brackets are standard errors. The dependent variable is state median household income. Lagged Unemployment is the unemployment rate lagged by 10 periods. Net Migration is the difference between the number of people having immigrated to the state from a different U.S. state and the number of people having emigrated from it. The remaining variables are defined as in Table 5.1. Model (7) is the long-term transformation of model (6).*

Chapter 2

Dealing with Uncertainty: Speculative Bubbles and Monetary Policy

Chapter Abstract

We develop a New Keynesian model featuring imperfectly observable bubble-like market episodes. Stock prices contain a fundamental and a bubble component, which agents are unable to observe. Instead, they are forced to speculate about their present and future value based on a noisy signal in order to make consumption and investment decisions. This framework is used to analyse the implications of imperfect information for the design of monetary policy. We find that, under perfect information, a systematic increase in interest rates in response to a growing bubble will exacerbate market instability and welfare losses. For a high enough level of noisiness in stock markets, the central bank will reverse its policy stance. Nevertheless, investors and the central bank perpetually under-estimate the size of bubbles, implying that the latter's policy response is always insufficient.

2.1 Introduction

Imperfect information about the source of asset price movements forms the basis of most of the risk present in financial markets. It is arguably the main reason asset prices tend to diverge from their fundamental value and is precisely what gives bubbles their harmful potential – it is not only impossible to predict when a bubble will burst, it is also very difficult to know when large market movements are even a bubble. This is because, at any point in time, the present and future value of the two components of asset prices – fundamentals and bubbles – cannot be observed, hence creating opportunities for speculation and unexpected market shifts.

Perhaps the most salient example is the U.S. housing bubble. While the housing market experienced a colossal boom in the first half of the 2000s, nearly no one was able to grasp the nature of the massive increase in housing prices or the extent of the deviations from fundamentals experienced across the market. In fact, the rapid expansion of complex financial instruments such as collateralised debt obligations (CDOs) likely amplified the noisiness of the information available to market participants and led them to ignore the many signs pointing to the real causes of the housing boom. This resulted in important losses for investors and many foreclosures of financial institutions, suggesting they were either unaware they were investing in a bubble, or could not grasp its size or forecast when it would burst. Even the Federal Reserve failed to understand the size and ramifications of the housing bubble, and board members continued to expect a quick market recovery well after the onset of the crash (FOMC, 2008).

Economists responded to the 2008 financial crisis by paying significantly more attention to bubbles. This led to the rise of the rational bubble literature which, unlike standard macroeconomic models, allows for asset prices to diverge from their fundamental value

and for the existence of bubbles in a macroeconomic context.[1] Nevertheless, bubbles in these models are perfectly observable and deliberately purchased by agents in order to relax borrowing constraints or diversify their holdings, with little to no risk involved.[2]

Crucially, this literature has failed to address perhaps the central issue pertaining to rational bubbles: how do investors and central banks deal with their inability to observe them when making investment and policy decisions. This question is precisely what this paper will seek to answer. It will show how imperfect information about the source of asset price movements impacts investors' behaviour, and analyse the implications for the conduct of monetary policy.

To this end, we develop a New Keynesian model with overlapping generations wherein households purchase assets whose price contains a fundamental and a bubble component. Due to imperfect information, agents are unable to observe these two components and are forced to speculate about their present and future value by solving a signal extraction problem.[3] The incentive for agents to successfully extract information from the noisy signal they receive is twofold. First, the beliefs they form about the source of asset price movements will determine their consumption and investment decisions. Second, the underlying dynamics of fundamentals and bubbles are very different: the former will permanently raise asset prices, the latter will only do so temporarily. Since the return to purchasing assets is given by next period's price, it is crucial for them to understand the source of stock market movements. Furthermore, following Galí (2014), the central bank systematically reacts to changes in the

[1] Earlier contributions include Samuelson (1958) and Tirole (1985). For a comprehensive review of the rational bubble literature, see Martin and Ventura (2018).

[2] See for instance Martin and Ventura (2012), Miao et al. (2015), Galí (2014,2017), Ideka (2017), and Miao and Wang (2018).

[3] In contrast with Adam et al. (2007), where agents are unable to observe the model's parameters and progressively learn about them, here agents only have imperfect information about the shocks, similar to Blanchard et al. (2013), Peng (2016), and Forni et al. (2017).

size of the bubble, which we assume it cannot directly observe.[4] This paper analyses the impact of imperfect information on agents' behaviour and its effects on the optimal design of monetary policy. The model provides four main results.

First, as would be expected, the presence of imperfect information causes agents to under-react to fundamental and bubble shocks. In fact, uncertainty about the source of market movements forces households to be more careful when making investment decisions, which lowers their lifetime consumption capacities. In turn, this creates important welfare losses, which worsen as the level of noise in the stock market rises.

Second, investors and the central bank tend to significantly underestimate the size of bubbles. This implies that the central bank's response to a growing bubble will always be insufficient, and it will be unable to fully compensate consumers for their losses relative to the perfect information case. Furthermore, this raises concerns about the real-world capacity of central banks to appropriately respond to bubbly episodes prior to market crashes.

Third, under perfect information, 'leaning against the wind' policies (i.e., raising interest rates when a bubble grows in size) will further destabilise stock markets and worsen welfare losses. This is because bubbles have a dual effect on consumers. On one hand, it will make the old generation better off by increasing the returns it gets from selling its assets, and the young generation worse off as it faces higher prices when purchasing stocks (the 'asset price channel'). On the other hand, it will exert upward pressure on wages and downward pressure on dividends, therefore resulting in a positive income shock on the young and a negative one on the old (the 'dividends channel'). Under perfect information, the asset price channel dominates, leading to higher asset price and consumption volatility. In order to

[4]The literature on imperfect information and monetary policy provides various rationales for assuming that the central bank does not have full information, and shows that the fit of New Keynesian models to the data is improved when the central bank's information set matches that of private agents. See for instance Levine et al. (2012).

compensate the young generation and promote consumption and market stability, the central bank will be required to lower interest rates following a bubble shock. This result provides support for some recent findings in the rational bubble literature, although it contrasts with the consensus among policy-makers, which is discussed in more detail below.

Finally, imperfect information will increase the weight of the dividends channel by impacting the old generation to a greater extent, hence reducing the fall in interest rates that accompany a bubble shock. For a high enough degree of noisiness, the central bank will reverse its policy stance and start 'leaning against the wind', despite further increasing market volatility. This finding will be used to offer a potential reconciliation between seemingly incompatible views expressed in the literature and among policy-makers.

Related literature. The literature on rational bubbles did not fully take off until the 2008 financial crisis. Two notable exceptions are the seminal papers by Samuelson (1958) and Tirole (1985). The former used an overlapping generations endowment economy to show how bubbles could help improve the problem of dynamic inefficiency. The latter extended this framework and introduced bubbles in a production economy for the first time. After the crisis, the number of contributions soared. Martin and Ventura (2012) integrated financial frictions in Tirole's (1985) framework and showed that bubbles can act to dampen the effects of these frictions. Similarly, Miao et al. (2015), Ideka (2017), and Miao and Wang (2018) generated bubbles that allow agents to relax borrowing constraints, and Galí (2014, 2017) modeled bubbles as a store of value. In spite of the intrinsic unobservability of the components of asset prices discussed above, this literature has always modeled bubbles as being perfectly observable and allowed agents to deliberately purchase them simply to relax some constraint or help them diversify their source of wealth. No paper in this literature has sought to analyse the effects of imperfect information regarding the components of asset

prices. The present paper contributes to this literature by introducing unobservable bubbles and by analyzing their implications for the design of monetary policy. Furthermore, agents' inability to observe the source of asset price movements is able to rationalise some of the seemingly irrational behaviours observed in stock markets that were not explained in past papers.

The imperfect information literature also fails to address this gap in our understanding of bubbles. While some papers have analysed the effects of imperfect information on asset pricing in general, most have failed to consider bubbles in their models, and none have integrated their framework into a general equilibrium setup. Forni et al. (2017), for instance, develop an asset pricing model with news shocks that are imperfectly observable and affect fundamentals with a lag. Bubbles are however ruled out in this framework as asset prices and fundamentals always coincide. Peng (2016) develops a useful framework where agents are unable to observe the fundamental value of stocks, which may be fully revealed in any period with some unknown probability. Over time, beliefs progressively converge to the fundamental value, unless it is suddenly revealed, in which case the market crashes. This imperfect information mechanism is however analysed on its own and is not integrated into a general equilibrium framework allowing for macroeconomic effects and for policy analysis. In fact, the literature on monetary policy and imperfect information is quite limited. One exception is Levine et al. (2012), who developed a simple New Keynesian model where private agents and the central bank are unable to observe current output and inflation (see also Nimark, 2006). Bubbles and asset pricing are however ignored in these models. To the best of our knowledge, the present paper is the first to analyse imperfectly observable bubbles in a general equilibrium framework.

My results on optimal monetary policy are related to the heated debate over how central

banks should respond to growing bubbles. We briefly summarise the key points of this debate to put our results into context. Prior to the financial crisis, the consensus among policy-makers was that central banks should solely focus on inflation stabilization and the output gap and should not respond to market episodes (see for instance Bernanke and Gertler, 1999 and 2001, Bernanke, 2002, and Kohn, 2006).[5] While the rational bubble literature developed rapidly after the crisis, the discussion over the role of monetary policy in relation to bubbles remained absent from it for a long time. Meanwhile, this question became a burning concern among policy-makers, who seemed to reach a new consensus: speculative bubbles can be extremely destabilizing for inflation and for the output gap, so central banks should react by raising interest rates ('leaning against the wind') in order to deflate these market events.

In recent years, the rational bubble literature has started contributing to this discussion. The first paper that extensively analysed the role of monetary policy in dealing with bubbles was Galí (2014). In this paper, he built a New Keynesian model with overlapping generations and studied the optimal policy reaction to changes in the size of a bubble. Since then, a number of papers have also addressed the issue. Asriyan et al. (2016) looked at the effects of monetary policy on bubbles at the Zero Lower Bound (ZLB) in a flexible price framework. Barlevy et al. (2017) analysed how monetary policy impacts bubbles through interest rates outside the ZLB. Galí (2017), Ideka (2017), and Dong et al. (2017) studied the interaction between monetary policy and bubbles in New Keynesian models with long-lived agents. Some of these papers provide support for the 'leaning against the wind' view, while others have questioned the economic reasoning behind it and have argued that such policy stance may end up exacerbating the volatility of bubbles. In fact, Galí (2014) argued that, contrary to fundamentals, which are given by a stream of payoffs discounted by the interest

[5]Rare exceptions include Cecchetti et al. (2000) and Borio and Lowe (2002), who argue that central banks should respond to market events in some circumstances.

rate, bubbles do not contain payoffs that need to be discounted. Instead, the only equilibrium condition on the latter is that its expected growth rate equals the interest rate. Hence, while leaning against the wind policies may result in an apparent fall in asset prices, this may simply be due to the fact that the increase in the size of the bubble implied by such policy is more than offset by a fall in fundamentals. This paper will provide a potential reconciliation between these two opposing views.

Structure. The remainder of the paper is structured as follows. Section 2 introduces a New Keynesian rational bubble framework with imperfect information. Agents' learning process and impulse responses are analysed in Section 3. Section 4 discusses the optimal design of monetary policy under perfect and imperfect information. Section 5 concludes.

2.2 The Model

We embed uncertainty about the source of asset price movements into a rational bubble framework with nominal rigidities à la Galí (2014). Households purchase assets whose price contains a fundamental component and a bubbly one. These components are unobservable to all agents in the economy, who must resort to form beliefs about their present and future value by solving a signal extraction problem. Agents have strong incentives to do so as fundamental shocks have permanent effects on asset prices while bubble shocks dissipate over time. An overlapping generations structure with asset creation-destruction is adopted to ensure that assets are traded in equilibrium.[6] Given my interest in understanding the effects of imperfect information on the conduct of monetary policy, nominal rigidities are included in the model.

It is useful to introduce some notation at this stage. For any variable x_t, agents' belief

[6] See Galí (2017) and Martin and Ventura (2018).

about its value at time j, based on information set Ψ_t^Ω at time k, is given by:

$$E_k\left[x_j|\Psi_k^\Omega\right] \equiv E_k^\Omega(x_j)$$

for $\Omega = \{P, I\}$, where:

$$\Psi_t^P = \{\boldsymbol{Y_\tau}, \boldsymbol{m_\tau}, \boldsymbol{X_\tau}, \boldsymbol{\Sigma_\tau} : 0 \leqslant \tau \leqslant t\}$$

$$\Psi_t^I = \{\boldsymbol{Y_\tau}, \boldsymbol{m_\tau} : 0 \leqslant \tau \leqslant t\}$$

are the information sets under perfect and imperfect information, respectively; $\boldsymbol{X_t}$ is a vector of variables that are unobservable under imperfect information and observable under perfect information; $\boldsymbol{\Sigma_t}$ is the corresponding variance-covariance matrix; $\boldsymbol{Y_t}$ is the vector of variables that are observable under both perfect and imperfect information; and $\boldsymbol{m}_t$ is the vector of parameters of the model, including the standard deviations of the shocks affecting the economy. Note that under perfect information, $E_t^P(x_t) = x_t$. In what follows, we use $E_t^\Omega(x_t)$ when referring to variables whose observability depends on the information set.

In this section, we first present the optimization problem of the various agents and the associated first-order and market clearing conditions. We then introduce fundamental and bubble shocks and discuss the signal extraction process.

2.2.1 Households

We assume a simple economy with no capital where consumers live for two periods and supply labour inelastically. An agent born at t receives an endowment of $\delta \in [0, 1)$ units of assets, whose price is $Q_{t,t}$. Each period, a fraction δ of assets is destroyed. The total quantity of assets therefore remains constant and equal to one. All assets (new and

pre-existing) can be traded on a stock market and contain a fundamental component and a bubbly one, which are stochastic and unobservable to the agent. In addition, agents are endowed with the knowledge to produce a differentiated good. They create a firm when they are young with this purpose, which only becomes productive after one period. They receive dividends in the second period of their life and the firm is dissolved when the agent dies.

Young consumers have access to two types of securities: bonds, B_t, and a variety of assets. Their budget constraint is given by:

$$\int_0^1 \frac{P_t(i)C_{1,t}(i)}{P_t}di + \frac{1}{P_t}B_t + \sum_{k=0}^{\infty} Q_{t,t-k}Z_{t,t-k} = W_t + \delta Q_{t,t} \tag{2.1}$$

where $P_t \equiv \left(\int_0^1 P_t(i)^{1-\epsilon}di\right)^{\frac{1}{1-\epsilon}}$ is the aggregate price index, $C_{1,t}(i)$ is the quantity of good i consumed by the young, with price $P_t(i)$, $Z_{t,t-k}$ is the quantity purchased of the assets introduced by the cohort born at $t-k$, whose price at t is $Q_{t,t-k}$.

When old, agents consume all their wealth. They receive payoffs from their maturing bonds, returns from selling their surviving assets, and dividends from their firm:

$$\int_0^1 \frac{P_{t+1}(i)C_{2,t+1}(i)}{P_{t+1}}di = D_{t+1} + \frac{1}{P_{t+1}}B_t(1+i_t) + (1-\delta)\sum_{k=0}^{\infty} Q_{t+1,t-k}Z_{t,t-k} \tag{2.2}$$

where $C_{2,t}(i)$ is the quantity of good i consumed by the old generation, D_t is dividends, and i_t is the nominal interest rate.[7]

Households must solve three problems: optimally allocate their consumption across goods; maximise their lifetime utility subject to (2.1) and (2.2); and extract information about the stochastic components of asset prices. The latter is discussed in Section 2.2.5 as it

[7]The budget constraints presented here closely resemble the ones introduced by Galí (2014). However, they differ in two crucial ways: in Galí, consumers can choose to purchase *bubbly* assets whose price is perfectly *observable*. In this paper, by contrast, agents purchase *generic* assets whose price contains *unobservable* bubbly and fundamental components.

also affects the central bank's problem. The former is standard and results in the following demand functions:

$$C_{1,t}(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} C_{1,t}$$

$$C_{2,t+1}(i) = \left(\frac{P_{t+1}(i)}{P_{t+1}}\right)^{-\epsilon} C_{2,t+1}$$

which imply:

$$C_{1,t} = \int_0^1 \frac{P_t(i)C_{1,t}(i)}{P_t} di$$

$$C_{2,t+1} = \int_0^1 \frac{P_{t+1}(i)C_{2,t+1}(i)}{P_{t+1}} di$$

where $C_{1,t}$ and $C_{2,t}$ are aggregate consumption of the young and of the old, respectively, and can be substituted into the budget constraints.

Consumers maximise their lifetime utility:

$$log(C_{1,t}) + \beta E_t log(C_{2,t+1})$$

subject to (2.1) and (2.2). The associated first-order conditions are:

$$1 = \beta(1 + i_t)E_t \left(\frac{P_t}{P_{t+1}} \frac{C_{1,t}}{C_{2,t+1}}\right) \tag{2.3}$$

$$Q_{t,t-k} = (1 - \delta)\beta E_t \left(\frac{C_{1,t}}{C_{2,t+1}} Q_{t+1,t-k}\right) \tag{2.4}$$

The first expression is the Euler equation, which determines the optimal consumption choice between the first and second periods of a consumer's life. The second expression implies that the value of an asset corresponds to the agents' expectation of the discounted price at which they will be able to sell it in the future.

Finally, let us define the real interest rate R_t as:

$$R_t = (1 + i_t)E_t \left(\frac{P_t}{P_{t+1}} \right)$$

2.2.2 Firms

Households are endowed with the knowledge to produce differentiated goods and they create a firm in the first period of their life with that purpose. The firm becomes productive in the next period and for one period only. It is then dissolved when its owner dies and a new firm inherits its 'i' subscript.

Firms employ labour in order to produce their differentiated good using a constant returns to scale technology:

$$Y_t(i) = N_t(i)$$

They are monopolistic competitors operating under nominal rigidities. They must therefore set the price of goods produced at t, P_t^*, in the previous period in order to maximise their value:

$$E_{t-1} \left[\Lambda_{t-1,t} Y_t(i) \left(\frac{P_t^*}{P_t} - W_t \right) \right]$$

subject to the demand constraint:

$$Y_t(i) = \left[\frac{P_t(i)}{P_t} \right]^{-\epsilon} C_t$$

where $\Lambda_{t-1,t} = \beta \frac{C_{1,t-1}}{C_{2,t}}$ is the relevant discount factor and $C_t = C_{1,t} + C_{2,t}$.

Optimal price setting requires:

$$E_{t-1}\left[\Lambda_{t-1,t}Y_t(i)\left(\frac{P_t^*}{P_t}-MW_t\right)\right]=0 \tag{2.5}$$

where $M=\frac{\epsilon}{\epsilon-1}$ is the mark-up.

2.2.3 Monetary Policy

Following Galí (2014), the central bank systematically reacts to changes in the inflation rate and in the size of the bubble. In contrast to it, however, we follow much of the information literature and assume informational symmetry across all agents in the economy, including the central bank.[8] This implies that the central bank cannot directly observe the bubble and instead forms beliefs about its size and reacts to these expectations instead.

The central bank therefore sets the nominal interest rate according to:[9]

$$1+i_t=RE_t(\Pi_{t+1})\left(\frac{\Pi_t}{\Pi}\right)^{\phi_\pi}E_t^\Omega\left(Q_t^B\right)^{\phi_b} \tag{2.6}$$

where $\phi_\pi>0$, and $E_t^\Omega\left(Q_t^B\right)$ is the central bank's belief about the size of the bubble, Q_t^B, at time t. Again, note that under perfect information $E_t^P\left(Q_t^B\right)=Q_t^B$. Section 2.4 will analyse the consequences of choosing alternative ϕ_b values, which determines the degree and direction of the central bank's reaction to bubble fluctuations.

[8] See Collard and Dellas (2004), Nimark (2006), Collard et al. (2009), Levine et al. (2010), and Levine et al. (2012).

[9] We have also tested the alternative rule: $1+i_t=R(\Pi_t/\Pi)^{\phi_\pi}(E_t^\Omega\left(Q_t^B\right)/Q^B)^{\phi_b}$. Given that results were qualitatively unchanged and rule (2.6) produced less algebraically heavy results, we did not retain the alternative rule in the present version of the paper.

2.2.4 Equilibrium

Having discussed the constraints and first-order conditions faced by households and firms, as well as the policy rule followed by the central bank, we now derive the market clearing conditions of the economy and express the equilibrium in terms of deviations from the steady state.

2.2.4.1 Market Clearing

Goods market clearing for each good requires: $Y_t(i) = C_{1,t}(i) + C_{2,t}(i)$. Let us define aggregate output as $Y_t = \left(\int_0^1 Y_t(i)^{1-\frac{1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}$. The demand conditions we derived in the consumers' problem imply the following aggregate goods market clearing condition:

$$Y_t = C_{1,t} + C_{2,t} \tag{2.7}$$

Combining this expression with the consumer's budget constraints yields the following income-side condition:

$$Y_t = D_t + W_t \tag{2.8}$$

Labor market clearing requires:

$$\begin{aligned} 1 &= \int_0^1 Y_t(i) di \\ &= Y_t \end{aligned} \tag{2.9}$$

where the second equality comes from the fact that all firms choose the same price and

produce the same quantities in equilibrium, which also implies:

$$E_{t-1}\left\{\frac{1}{C_{2,t}}(1-MW_t)\right\}=0 \tag{2.10}$$

On securities markets, the quantity of bonds and assets traded in equilibrium are:

$$B_t=0$$

$$Z_{t,t-k}=\delta(1-\delta)^k$$

for $k=0,1,2,\ldots$ Aggregate asset prices, Q_t, can now be defined as:

$$\begin{aligned} Q_t &\equiv \delta\sum_{k=0}^{\infty}(1-\delta)^k Q_{t,t-k} \\ &= \underbrace{\delta Q_{t,t}}_{\equiv U_t} + \underbrace{\delta\sum_{k=1}^{\infty}(1-\delta)^k Q_{t,t-k}}_{\equiv G_t} \end{aligned}$$

For convenience, the second line decomposes asset prices into the price of newly-created assets, U_t, and of pre-existing assets, G_t. Applying these definitions to the first-order condition (2.4), we obtain:

$$Q_t=G_t+U_t=\beta E_t\left\{\left(\frac{C_{1,t}}{C_{2,t+1}}\right)G_{t+1}\right\} \tag{2.11}$$

Note that similar to Q_t, both G_t and U_t contain stochastic fundamental and bubbly components, which we present in Section 2.2.5. We now define the steady state of the model and express the above conditions in terms of deviations from it.

2.2.4.2 Steady State

In the absence of uncertainty, optimal price setting (2.10) implies that firms set wages at a constant rate:

$$W = \frac{1}{M}$$

Hence, dividends are also constant:

$$D = 1 - \frac{1}{M}$$

Applying these two results to the consumer's budget constraints, steady state consumption can be expressed as:

$$C_1 = \frac{1}{M} - G$$

$$C_2 = 1 - \frac{1}{M} + G$$

Combining these two expressions with (2.3) and the definition of the real interest rate yields:

$$R = \frac{1}{\beta}\left(\frac{1 - \frac{1}{M} + G}{\frac{1}{M} - G}\right)$$

The steady state price of pre-existing assets is therefore given by:

$$G = \frac{\left(1 - \frac{1}{M}\right)(G + U)}{\frac{\beta}{M} - (1 + \beta)G - U} \equiv H(G, U)$$

Note that if $\partial H(G,U)/\partial G < 1$, the steady state is locally stable.

2.2.4.3 Log-linearised Equilibrium

Let us now re-express the equilibrium conditions of the model in terms of deviations from the steady state. In what follows, we use lower cases to refer to log-linearised variables. The equilibrium is characterised by the following equations:

$$c_{1,t} + \beta R c_{2,t} = 0 \tag{2.12}$$

$$c_{1,t} = E_t(c_{2,t+1}) - r_t \tag{2.13}$$

$$c_{2,t} = (1 - \Gamma)d_t + \Gamma g_t \tag{2.14}$$

$$\begin{aligned} q_t &= Rg_t + (1 - R)u_t \\ &= E_t g_{t+1} - r_t \end{aligned} \tag{2.15}$$

$$r_t = \phi_\pi \pi_t + \phi_b E_t^\Omega(q_t^b) \tag{2.16}$$

$$E_{t-1} w_t = E_{t-1} d_t = 0 \tag{2.17}$$

where $\Gamma = \frac{\epsilon G}{\epsilon G + 1}$. Equation (2.12) is the goods market clearing condition; equation (2.13) is the Euler equation; (2.14) is derived from the budget constraint of the old; (2.15) comes from expression (2.11); (2.16) is the policy rule; and equation (2.17) is derived from (2.8) and (2.10).

2.2.5 Bubbles, Fundamentals, and Signal Extraction

We now introduce shocks into the model by rewriting the asset price equation (2.15) as:

$$g_t = Rg_{t-1} + (1 - R)u_{t-1} + r_{t-1} + \xi_t \tag{2.18}$$

where ξ_t is a shock to (or forecast error of) the price of pre-existing assets, which contains fundamental and bubbly stochastic components that are unobservable to agents. Furthermore, we conjecture that the price of newly-created assets also includes similar elements. We now discuss these components in detail.

2.2.5.1 Underlying Dynamics and Signaling

Let us define:

$$\xi_t = E_t^{\Omega}(\xi_t^f) + E_t^{\Omega}(\xi_t^b) \tag{2.19}$$

$$u_t = E_t^{\Omega}(u_t^f) + E_t^{\Omega}(u_t^b) \tag{2.20}$$

where ξ_t^b and u_t^b are bubble shocks to the price of pre-existing and newly-created assets, respectively, and ξ_t^f and u_t^f are their fundamental counterparts. Using (2.15), we also define the aggregate bubble and aggregate fundamentals, respectively, as:

$$E_t^{\Omega}(q_t^b) = RE_t^{\Omega}(\xi_t^b) + (1-R)E_t^{\Omega}(u_t^b) \tag{2.21}$$

$$E_t^{\Omega}(q_t^f) = RE_t^{\Omega}(\xi_t^f) + (1-R)E_t^{\Omega}(u_t^f) \tag{2.22}$$

Signaling. Crucially, agents are able to observe ξ_t and u_t, but not their components. Instead, they receive a noisy signal about the composition of each shock:

$$s_t^{\xi} = \xi_t^f + \nu_t \tag{2.23}$$

$$s_t^{u} = u_t^f + \omega_t \tag{2.24}$$

where $\nu_t \sim N(0, \sigma_\nu^2)$ and $\omega_t \sim N(0, \sigma_\omega^2)$ are noise shocks which prevent agents from per-

fectly observing the composition of ξ_t and u_t. These signals are relevant if noise shocks are small, but become misleading if they are large. Note that when $\sigma_\nu = \sigma_\omega = 0$, the signals are perfectly informative and the model reduces to the full information benchmark.

Underlying shock dynamics. Agents are particularly incentivised to extract information from these signals given that bubble shocks and fundamental shocks have opposite implications for future asset prices. In fact, shocks to fundamentals are conjectured to cause a progressive but permanent increase in the pre-existing component of asset prices, while bubble shocks will slowly deflate following the initial jump in prices:

$$\Delta\xi_t^f = \rho^f \Delta\xi_{t-1}^f + \varepsilon_t \tag{2.25}$$

$$\xi_t^b = \rho^b \xi_{t-1}^b + \eta_t \tag{2.26}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $\eta_t \sim N(0, \sigma_\eta^2)$ are i.i.d. shocks. Newly-created assets, on the other hand, are affected by simple one-off shocks:

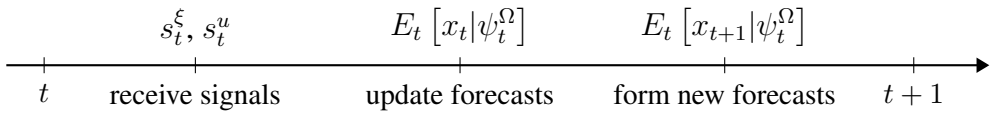
$$u_t^f = \varsigma_t \tag{2.27}$$

$$u_t^b = \zeta_t \tag{2.28}$$

where $\varsigma_t \sim N(0, \sigma_\varsigma^2)$ and $\zeta_t \sim N(0, \sigma_\zeta^2)$ are also i.i.d. shocks. Given the simplicity of shocks to newly-created assets, our analysis below focuses more heavily on shocks to pre-existing assets.

Evidently, our model does not seek to closely replicate the real-world dynamics of bubbles and fundamentals. Instead, we deliberately simplify these dynamics in order to create a laboratory for the analysis of unobservable bubbles and their implications for agents'

Figure 2.1: Timeline



Note: For any variable x_t, $E_k\left[x_j|\psi_k^{\Omega}\right]$ is the belief about x_t in period j based on the information set available at k, ψ_k^{Ω}, which contains the signals, s_t^{ξ} and s_t^{u}.

behaviour and the design of monetary policy. The conjecture that bubbles slowly deflate following a boom reflects the common understanding that a bubble is a temporary deviation of asset prices from their fundamental value. Crucially, by making such assumptions, we create a strong incentive for agents to disentangle the source of observed asset price movements. In fact, agents purchase assets when young but are uncertain about the price they will be able to sell them for in the following period. If the initial change in asset prices is due to a bubble shock, consumers will be selling their assets at a lower price than what they purchased them for.

2.2.5.2 Signal Extraction

In order to extract information from the signals they receive, consumers and the central bank use Kalman filtering. The intuition behind it is quite straightforward and is illustrated in Figure 2.1.

Every period, agents receive a public signal about the composition of asset prices, which they use to form beliefs about the present value of bubbles and fundamentals, and to forecast their future value. In the following period, they receive a new signal and use it to update these forecasts and to make new predictions. The extent to which agents rely on the signal when updating beliefs is determined by the Kalman gain, which in turn depends on noise volatility.

To see this, let us stack the variables that are unobservable under imperfect information

into a vector $\boldsymbol{X_t}$, and re-express equations (2.25)-(2.28) as:

$$\boldsymbol{X_t} = \boldsymbol{F}\boldsymbol{X_{t-1}} + \boldsymbol{B}\boldsymbol{\tau_t}$$

where $\boldsymbol{X_t} = \begin{bmatrix} \xi_t^f & \xi_{t-1}^f & \xi_t^b & u_t^f & u_{t-1}^f & u_t^b \end{bmatrix}'$, $\boldsymbol{\tau_t} = \begin{bmatrix} \varepsilon_t & \eta_t & \nu_t & \varsigma_t & \zeta_t & \omega_t \end{bmatrix}'$, and $\boldsymbol{F}$ and $\boldsymbol{B}$ are transition matrices. Relevant variables that are observable under imperfect information are captured by the vector $\boldsymbol{Z_t}$. Equations (2.19)-(2.20) and (2.23)-(2.24) are re-expressed as:

$$\boldsymbol{Z_t} = \boldsymbol{H}\boldsymbol{X_t} + \boldsymbol{C}\boldsymbol{\tau_t}$$

where $\boldsymbol{Z_t} = \begin{bmatrix} \xi_t & s_t^{\xi} & u_t & s_t^{u} \end{bmatrix}'$, and $\boldsymbol{H}$ and $\boldsymbol{C}$ are transition matrices.

The full set of equations describing the Kalman Filter is presented in Appendix 2.A. From Pasricha (2006), these can be simplified to:

$$\boldsymbol{F}\boldsymbol{X_{t|t}} = \boldsymbol{K_t}\boldsymbol{Z_t} + (\boldsymbol{F} - \boldsymbol{K_t}\boldsymbol{H})\boldsymbol{X_{t|t-1}} \tag{2.29}$$

$$\boldsymbol{X_{t+1|t}} = \boldsymbol{F}\boldsymbol{X_{t|t}} \tag{2.30}$$

where $\boldsymbol{X_{j|k}} \equiv E[\boldsymbol{X_j}|\boldsymbol{Z_k}]$ is the belief about period j based on the signal received at k, and:

$$\boldsymbol{K_t} = \boldsymbol{F}\boldsymbol{\Sigma_{t|t-1}}\boldsymbol{H'}[\boldsymbol{H}\boldsymbol{\Sigma_{t|t-1}}\boldsymbol{H'} + \boldsymbol{L_t}]^{-1}$$

is the Kalman gain, where $\boldsymbol{\Sigma_t}$ is the variance-covariance matrix associated with $\boldsymbol{X_t}$, and $\boldsymbol{L_t} = E[\boldsymbol{C}\boldsymbol{\tau_t}\boldsymbol{\tau_t'}\boldsymbol{C'}]$ captures the volatility of noise shocks. (2.29) is the updation equation and (2.30) is the forecast equation.

These equations imply that higher noise volatility will cause agents to put less weight

on the signals they receive when updating their forecasts. This in turn implies that the size of innovations to the agents' beliefs (the difference between updates and forecasts, $\boldsymbol{X_{t|t} - X_{t|t-1}}$) will be smaller, but these innovations will be more volatile across time. This result will become particularly relevant for the monetary policy analysis of Section 2.4.

2.2.5.3 Closed-Form Solutions

First, note that the underlying dynamics of fundamental and bubble shocks described in (2.25)-(2.28) hold regardless of whether agents are able to observe them or not. On the other hand, following a shock, agents make consumption and investment decisions based on their beliefs about these dynamics, which can be expressed as:

$$\Delta E_t^{\Omega}(\xi_t^f) = \rho \Delta E_{t-1}^{\Omega}(\xi_{t-1}^f) + \underbrace{[E_t^{\Omega}(\xi_t^f) - E_{t-1}^{\Omega}(\xi_t^f)]}_{\equiv \varepsilon_t^{\Omega}} \tag{2.31}$$

$$E_t^{\Omega}(\xi_t^b) = \rho E_{t-1}^{\Omega}(\xi_{t-1}^b) + \underbrace{[E_t^{\Omega}(\xi_t^b) - E_{t-1}^{\Omega}(\xi_t^b)]}_{\equiv \eta_t^{\Omega}} \tag{2.32}$$

$$E_t^{\Omega}(u_t^f) = \underbrace{[E_t^{\Omega}(u_t^f) - E_{t-1}^{\Omega}(u_t^f)]}_{\equiv \zeta_t^{\Omega}} \tag{2.33}$$

$$E_t^{\Omega}(u_t^b) = \underbrace{[E_t^{\Omega}(u_t^b) - E_{t-1}^{\Omega}(u_t^b)]}_{\equiv \varsigma_t^{\Omega}} \tag{2.34}$$

where ε_t^{Ω}, η_t^{Ω}, ζ_t^{Ω}, and ς_t^{Ω} are the innovations to agents' beliefs about the value of asset price components. As discussed above, they reflect the extent to which agents use the signals to update their beliefs in each period. Note that under perfect information, these beliefs coincide with the shocks introduced in equations (2.25)-(2.28).

We can now write the closed-form evolution of pre-existing assets as:

$$\begin{aligned} g_t = \chi g_{t-1} + \xi_t + [\phi_b - \epsilon G(1+\beta)] R\eta^{\Omega}_{t-1} + (\phi_b + 1) \\ (1-R)\zeta^{\Omega}_{t-1} - \epsilon G(1+\beta) R\varepsilon^{\Omega}_{t-1} + (1-R)\varsigma^{\Omega}_{t-1} \end{aligned} \tag{2.35}$$

where $\chi = R[1 + \epsilon G(1+\beta)]$. Note that noise volatility enters this expression through the innovation terms. In addition, due to nominal rigidities, the central bank is able to influence the pre-existing component of asset prices through the policy parameter, ϕ_b.

Equilibrium inflation is given by:

$$\begin{aligned} \pi_t = -\frac{1}{\phi_\pi} \Bigg\{ & \left[\phi_b - \chi\epsilon G(1+\beta)\right] R E^{\Omega}_{t-1}(\xi^b_{t-1}) \\ & - \epsilon G(1+\beta) R \Big\{ \chi g_{t-1} + (1+\rho) E^{\Omega}_{t-1}(\xi^f_{t-1}) - \rho E^{\Omega}_{t-2}(\xi^f_{t-2}) \\ & + \Big[\phi_b - \Big(1 + 2G(1+\beta)\Big)\Big] R\eta^{\Omega}_{t-1} - R[1 + 2\epsilon G(1+\beta)]\varepsilon^{\Omega}_{t-1} \\ & + (\phi_b + 1)(1-R)\zeta^{\Omega}_{t-1} + (1-R)\varsigma^{\Omega}_{t-1} \Big\} \Bigg\} \end{aligned} \tag{2.36}$$

In the next two sections, we present the results of the model. We first analyse how bubble and fundamental shocks impact the economy under perfect and imperfect information. We then discuss the trade-offs faced by the central bank when setting policy, and the way in which noise affects its optimal policy decision.

2.3 Impulse Response Analysis

In this section, we discuss the way consumers and the central bank react to bubble and fundamental shocks, and analyse the impact of imperfect information on their decisions.

Table 2.1: Parameter values

Parameter	Value	Description	Calibration target
β	1	Discount factor	Galí (2014)
ϵ	6	Elasticity of substitution	Price (wage) markup of 20%
ϕ_π	1.5	Policy parameter for inflation	Standard
ρ^b, ρ^f	0.891	Persistence, bubble and fundamental shocks	See Section 2.3.1
σ_η, σ_ζ	0.25	Standard deviation, bubble shocks	See Section 2.3.1
σ_ε, σ_ς	0.16	Standard deviation, fundamental shocks	See Section 2.3.1

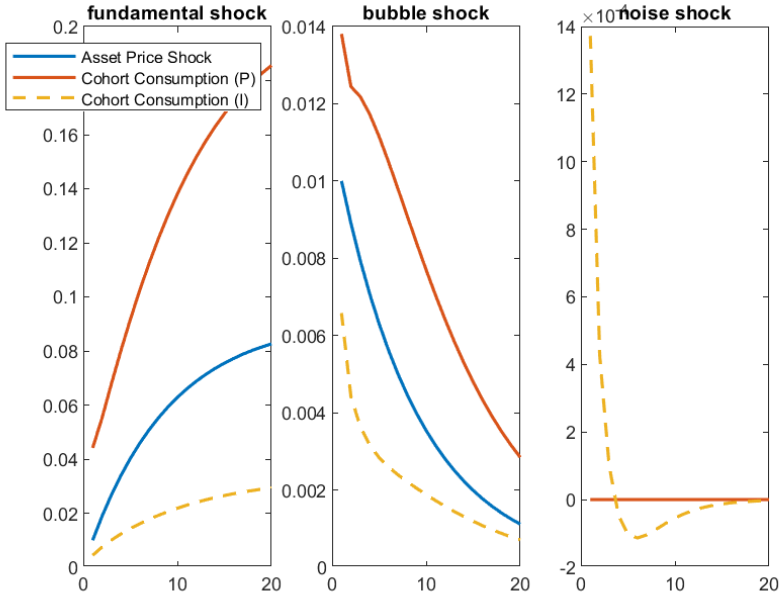
2.3.1 Calibration

We start by calibrating the model. There are two sets of parameters. The first contains standard parameters, namely $\{\beta, \epsilon, \phi_\pi, M, \chi, \Gamma\}$. The second set includes parameters pertaining to bubble and fundamental shocks, $\{\phi_b, \rho^b, \rho^f, \sigma_\eta, \sigma_\zeta, \sigma_\varepsilon, \sigma_\varsigma, \sigma_\nu\}$.

Most of the standard parameters of this model mirror Galí's (2014) calibration choices. Namely, the discount factor is $\beta = 1$ and the elasticity of substitution, ϵ, is set so that the price markup equals 20%. From these parameters and the steady state value of interest rates and asset prices, we can derive: $M = 1.2$, $\chi = 0.8$, and $\Gamma = 0.375$. The policy parameter on inflation, ϕ_π, is set to 1.5, which is standard in the New Keynesian literature.

We take a different approach with the parameters pertaining to bubble and fundamental shocks. Given the limited breadth of evidence about these coefficients, we are forced to calibrate them by combining different results from the literature. We undertake a sensitivity analysis in the next section to test the extent to which our results depend on the calibration of these parameters. My starting point is to set the standard deviation of pre-existing bubble shocks, σ_η, equal to 0.25, following Ideka (2017). We use the empirical analysis from Liu et al. (2017) to calibrate the relative volatility of bubbles and fundamentals, $\frac{\sigma_\eta}{\sigma_\varepsilon}$ and $\frac{\sigma_\zeta}{\sigma_\varsigma}$, to 2.47. Finally, from Galí (2014), we assume that shocks to pre-existing and newly created assets are equally volatile, such that $\sigma_\eta = \sigma_\zeta$ and $\sigma_\varepsilon = \sigma_\varsigma$. For simplicity, we assume that $\rho^b = \rho^f = \rho$, which we set equal to 0.891. In Section 2.4, we will study the implications of

Figure 2.2: Learning About Shocks



Note: The orange and yellow curves plot cohort consumption under perfect (P) and imperfect (I) information, respectively. The blue curves plot the response of the innovation term in the pre-existing asset pricing equation, ξ_t, which does not depend on the information set.

choosing alternative values for the policy parameter on the bubble, ϕ_b, and noise volatility, σ_ν. For now, we set $\phi_b = 0.35$ irrespective of the information set – to isolate the effect of imperfect information – and $\sigma_\nu = 0.79$, which implies a reasonably high degree of noisiness.

2.3.2 Bubble and Fundamental Shocks

We now analyse the reaction of the economy to fundamental, bubble, and noise shocks to the value of pre-existing assets.

Learning about shocks. First, let us provide some intuition about the way in which agents progressively learn about the nature of asset price movements. Figure 2.2 plots the responses of ξ_t and cohort consumption to fundamental, bubble, and noise shocks under perfect and imperfect information.

Following a shock to the fundamental value of assets, agents are at first unsure whether

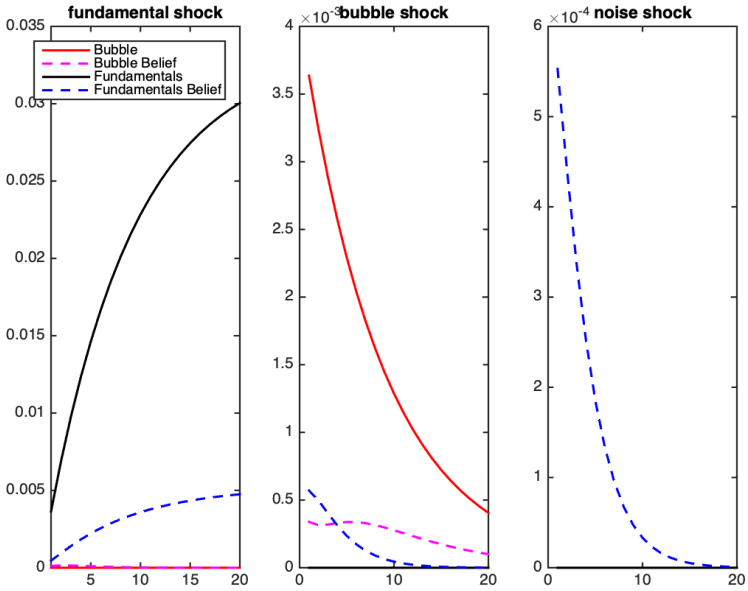
they are witnessing a fundamental, bubble, or noise shock. Their consumption therefore jumps significantly less than under perfect information. As asset prices keep increasing, they progressively learn that it is a fundamental shock and they increase their consumption level further. By then it is too late for them to reach their full information consumption level while preserving consumption smoothing.

A similar cost is incurred following a bubble shock. At first, agents are concerned this might just be a noise shock, so their consumption does not jump as much as under perfect information. As they learn the true nature of the shock – since the asset pricing innovation term starts returning to its original level – they progressively decrease their consumption level.

Finally, in the event of a noise shock, agents do not react at all under perfect information, as fundamentals and bubbles remain unchanged. However, under imperfect information, agents are unsure whether the change in the signal they receive is due noise or to a fundamental or bubble shock. Their consumption jumps and then quickly falls as they learn the true nature of the shock. To compensate for the unjustified increase in consumption, they temporarily bring it below its original level for a few periods.

Quantitative misinformation. Even once agents have learnt the true nature of a shock, their beliefs about its size are still incorrect. Figure 2.3 compares agents' beliefs about the size of the bubble and fundamentals under imperfect information, $E_t^I(q_t^b)$ and $E_t^I(q_t^f)$, with their actual values, q_t^b and q_t^f. Following a bubble shock for instance, agents initially believe it is more likely that they are witnessing a fundamental shock. Once they find out which shock has hit the economy, they keep underestimating the size of the bubble until the shock dissipates. This – together with consumption smoothing – explains why consumption never reaches its full information level even once agents understand the source of asset price

Figure 2.3: Beliefs Versus Underlying Dynamics



Note: The black and red curves plot the underlying dynamics of fundamentals and bubbles, respectively. The blue and pink dashed curves plot agents' beliefs about the underlying dynamics.

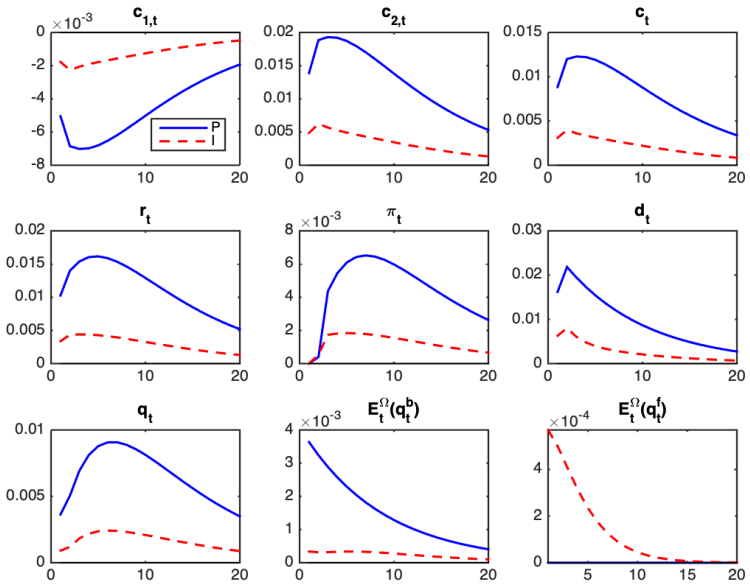
movements.

This result has crucial implications for policy-making. Since the central bank significantly underestimates changes to the size of the bubble, its reaction (whether it seeks to raise or lower interest rates in response) will always be insufficient. Given that consumers will also react inappropriately by under-reacting to shocks, the central bank will not be able to fully compensate for their losses relative to the full information case. This result concurs with the idea that the Fed and investors misjudged the size of the housing bubble in the run up to the 2008 financial crisis, and it raises concerns about the ability of central banks and investors to appropriately respond to growing bubbles.

Response to bubble shocks. We now discuss the full results of the model. Figure 2.4 plots the impulse responses of the economy to a bubble shock.[10] Under perfect information,

[10]In Figures 2.4 and 2.5, the red curves in the $E_t^{\Omega}(q_t^b)$ and $E_t^{\Omega}(q_t^f)$ panels capture agents' beliefs about bubbles and fundamentals under imperfect information. The blue curves, on the other hand, show the underlying dynamics of q_t^b and q_t^f, which hold regardless of the information set.

Figure 2.4: Impulse Responses of the Economy to a Bubble Shock

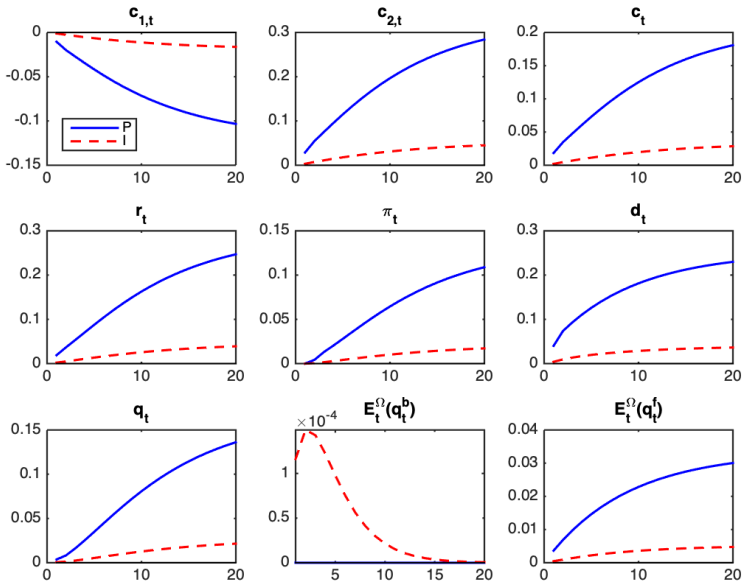


Note: Responses under perfect information are plotted in blue. The imperfect information case is captured by the red curves.

asset prices will temporarily rise. This in turn will increase the price at which the old generation can sell its assets, which allows them to raise their consumption. On the other hand, it will increase the price faced by the young when purchasing assets. They are hence forced to reduce their consumption in order to keep purchasing stocks and ensure that they receive returns from their sale in the next period. Following their initial jump, asset prices keep increasing for several periods due to their autoregressive nature. As time goes by, this effect is slowly outweighed by the deflation of the bubble and asset prices start to fall. Meanwhile, given the current assumption that $\phi_b > 0$, together with the increase in the inflation rate, the central bank reacts by significantly raising the interest rate. While the increase in total consumption pushes wages up and dividends down, the large increase in the interest rate ultimately leads dividends to increase, hence further boosting the consumption of the old.

Response to fundamental shocks. In the event of a fundamental shock, variables are permanently affected and respond by slowing converging to a new level. The intuition be-

Figure 2.5: Impulse Responses of the Economy to a Fundamental Shock

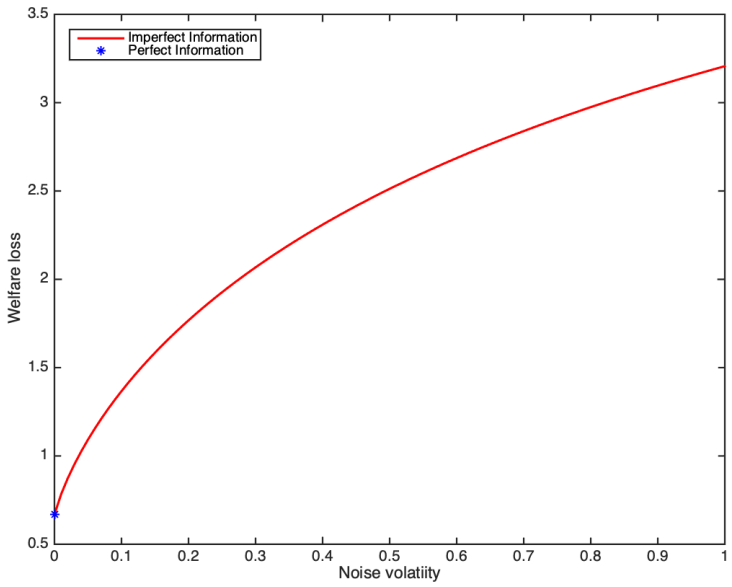


Note: Responses under perfect information are plotted in blue. The imperfect information case is captured by the red curves.

hind the results in Figure 2.5 is similar to the case of bubble shocks. Note that the increase in asset prices will permanently raise the old generation's consumption level, while the young will need to permanently reduce theirs in order to keep purchasing assets at the higher price.

The impact of imperfect information. As already discussed above, imperfect information acts by dampening the response of consumers and the central bank to bubble and fundamental shocks. Agents witness the increase in asset prices and must decide the quantity of assets they wish to purchase. Since they are unsure about the source of these price changes and consequently about the future value of stocks, they will purchase fewer assets and reduce their consumption to a lesser extent when young. This leads to a smaller increase in the consumption of the old, who still benefit from the price change but by less given that they have fewer assets to sell. Since agents underestimate the size of the bubble and fundamentals, they will also underestimate the future discounted value of assets, which implies a smaller increase in asset prices. As discussed above, the central bank will raise the interest

Figure 2.6: Imperfect Information and Welfare Loss



rate insufficiently in response to the increase in the inflation rate and the perceived bubble size.

Imperfect information creates large welfare losses by forcing consumers and the central bank to react sub-optimally compared to the full information case. Figure 2.6 plots the loss function against noise volatility (see Section 2.4 for details). The higher the degree of imperfect information faced by consumers and the central bank, the higher the welfare costs incurred by agents.

2.4 Monetary Policy Design

In this section, we first discuss the dilemma faced by the central bank when choosing the direction and degree of its response to changes in the size of the bubble in the perfect information benchmark. We then introduce imperfect information and analyse how the central bank's optimal response is affected. As mentioned above, the focus of this analysis is the

policy coefficient on the size of the bubble, ϕ_b.

2.4.1 The Central Bank's Dilemma

In order to analyse the central bank's optimal choice of policy, we first derive the loss function it will seek to minimise. To do so, we follow Nimark (2006) and Levine et al. (2012) and derive a more generalised version of the welfare criterion proposed by Galí (2014), hence allowing for shock volatilities to affect policy decisions. As is standard in the macroeconomic literature, we use the unconditional mean of an individual's lifetime utility, which to a second order is approximated by:

$$EU_t \approx -\frac{1}{2}\left(varc_{1,t} + \beta varc_{2,t+1}\right) + t.i.p.$$

where $t.i.p.$ are terms independent of policy.

Let us re-express $c_{1,t}$ and $c_{2,t}$ as:

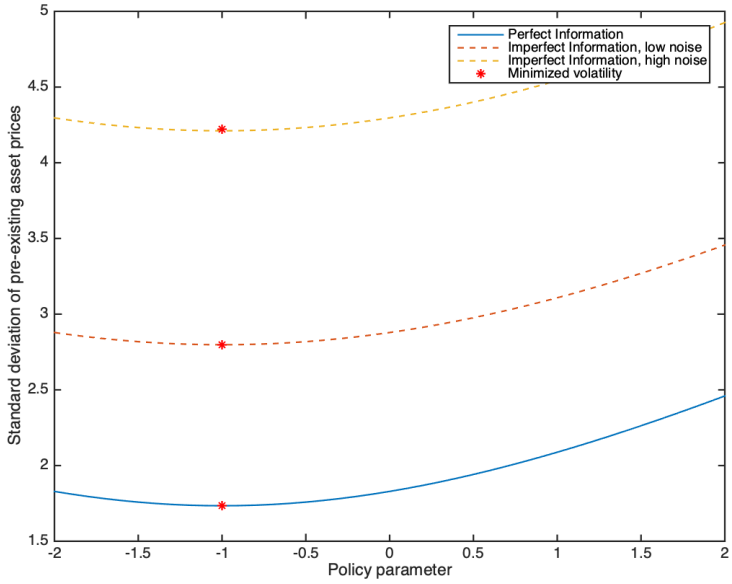
$$c_{1,t} = -\frac{1}{\beta R(\epsilon - 1 - \epsilon G)}(\beta R d_t + \beta R \epsilon G g_t)$$

$$c_{2,t} = (1 - \Gamma)(\beta R d_t + \beta R \epsilon G g_t)$$

The central bank therefore needs to choose ϕ_b in order to minimise the volatility of consumption of the young and of the old, which in turn requires minimizing the volatility of dividends and of the pre-existing component of asset prices. This creates a dilemma for the central bank.

The asset price channel. To see this, we express asset prices as a function of innovations

Figure 2.7: Policy and Asset Price Volatility



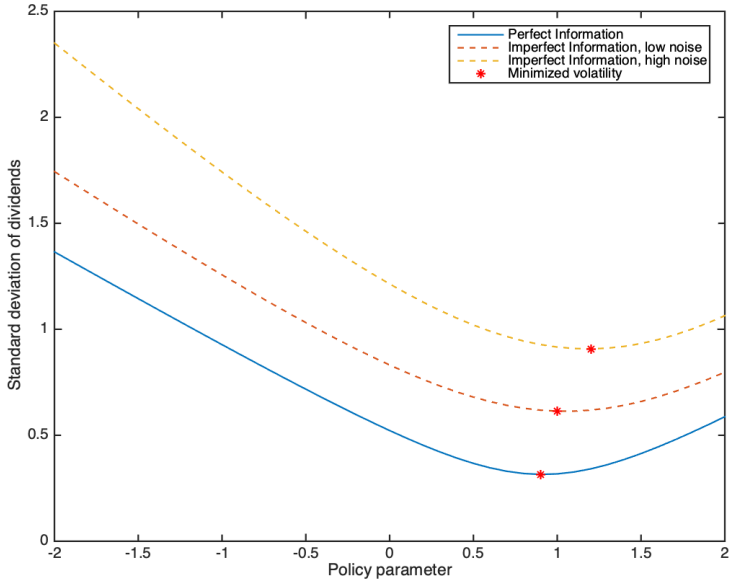
to its components:

$$\begin{aligned} g_t =& \chi E_{t-2} g_{t-1} + E_{t-1}\xi_t + \eta_t^{\Omega} + \varepsilon_t^{\Omega} + (\phi_b + 1) \\ & \left[R\eta_{t-1}^{\Omega} + (1-R)\zeta_{t-1}^{\Omega}\right] + R\varepsilon_{t-1}^{\Omega} + (1-R)\varsigma_{t-1}^{\Omega} \end{aligned} \tag{2.37}$$

Figure 2.7 plots the standard deviation of (2.37) against the policy parameter. In order to analyse the relation between these two terms, we first focus on the perfect information case. The role of noise is discussed in the next subsection.

As is evident from Figure 2.7, the central bank needs to set $\phi_b < 0$ in order to stabilise stock prices. In fact, asset prices will rise following a bubble shock. As a consequence, the young generation will face higher prices and will need to reduce its consumption in order to keep purchasing assets, hence increasing the volatility of $c_{1,t}$. By contrast, the old generation will be able to sell its assets at a higher price and therefore consume more, which raises the volatility of $c_{2,t}$. By lowering interest rates following a bubble shock, the central bank will improve demand conditions, therefore causing wages to grow and dividends to fall. This

Figure 2.8: Policy and Volatility of Dividends



will compensate the young for their losses and reduce the old's gains, hence stabilizing consumption and improving welfare.

The dividends channel. Nevertheless, such a policy stance will clash with the second channel through which asset prices affect welfare. To illustrate this, let us also express d_t as a function of the innovation terms:

$$d_t = \frac{1}{\beta R}\Big\{\big[\phi_b - \epsilon G(1+\beta)\big]R\eta_t^{\Omega} + \big[\phi_b - \epsilon G\big](1-R)\zeta_t^{\Omega} - \epsilon G(1+\beta)R\varepsilon_t^{\Omega} - \epsilon G(1-R)\varsigma_t^{\Omega}\Big\} \tag{2.38}$$

Minimizing the volatility of dividends requires setting $\phi_b > 0$, as shown in Figure 2.8. In fact, a bubble shock will improve demand conditions by raising total consumption, hence driving wages up and dividends down. The young will be able to consume more, while the old generation will be negatively affected. The central bank will be able to raise dividends and compensate the old for their losses by raising interest rates following a bubble shock. This will improve consumption smoothing and reduce welfare losses.

Optimal policy under perfect information. This creates a dilemma for the central bank, who must balance these two objectives when setting ϕ_b in order to maximise welfare. The optimal policy parameter under perfect information is given by:

$$\phi_b^P = \kappa + \theta \frac{R^2\sigma_\eta^2}{R^2\sigma_\eta^2 + (1-R)^2\sigma_\zeta^2} \tag{2.39}$$

where $\kappa \equiv -\left[\frac{(\beta R\epsilon G)^2}{1-\chi^2} - \epsilon G\right] / \left[1 + \frac{(\beta R\epsilon G)^2}{1-\chi^2}\right] < 0; \theta \equiv [2\beta\epsilon G] / \left[\left(1 + \frac{(\beta R\epsilon G)^2}{1-\chi^2}\right)(1+\beta\gamma)\right] > 0$; and $\gamma \equiv (1-\Gamma)(\epsilon - 1 - \epsilon G)$. The calibration discussed in Section 2.3 suggests setting $\phi_b^P = -0.37$ under perfect information, implying that the asset price channel will outweigh the dividends channel since bubble shocks will adversely impact the young generation to a greater extent than the old. In turn, this calls for the central bank to lower interest rates in order to compensate them and stabilise consumption.

2.4.2 Monetary Policy in an Uncertain World

This section has thus far assumed that agents can perfectly observe the source of asset price movements. We now analyse the implications of introducing noise in the economy – i.e., setting $\sigma_v > 0$ – for the central bank's dilemma and the design of monetary policy.

As pointed out in Section 2.2.5, noise enters the model by raising the volatility of the innovations to agents' beliefs, defined in equations (2.31)-(2.34). As shown by (2.37) and (2.38), these volatilities enter the standard deviations of asset prices and dividends. Consequently, it is expected that noise volatility will raise the variance of g_t and d_t, hence destabilizing consumption and lowering welfare.

To see this, let us go back to Figure 2.7, which plots the standard deviation of asset prices for high and low noise volatility and for the perfect information benchmark.[11] For any

[11]The low volatility curves in Figures 2.7 and 2.8 use $\sigma_\nu^2 = 0.1$, which implies a relative volatility of

given value of ϕ_b, imperfect information raises the volatility of stock prices, which is still minimised when $\phi_b < 0$. Noise volatility will have similar effects on dividends, as shown in Figure 2.8. The higher the degree of imperfect information, the higher the volatility of dividends for any given policy coefficient chosen by the central bank.

Optimal policy under imperfect information. What are the implications for the optimal design of monetary policy? The optimal policy parameter under imperfect information can be expressed as:

$$\phi_b^I = \kappa + \theta \frac{R^2 \hat{\sigma}_\eta^2}{R^2 \hat{\sigma}_\eta^2 + (1-R)^2 \hat{\sigma}_\zeta^2} \tag{2.40}$$

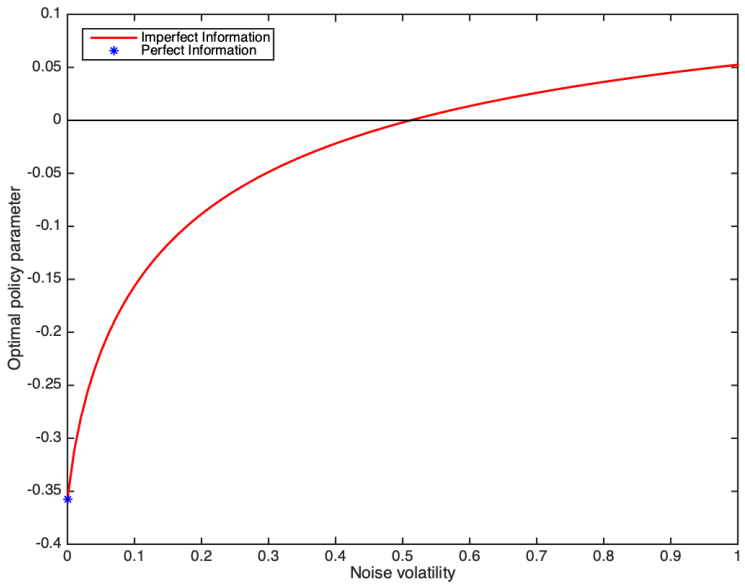
where $\hat{\sigma}_\eta^2 \equiv var\left(\eta_t^I\right)$ and $\hat{\sigma}_\zeta^2 \equiv var\left(\zeta_t^I\right)$. For pre-existing assets, an increase in noise volatility relative to the volatility of the signal, $\frac{\sigma_\nu}{\sigma_s}$, will raise the volatility of agents' beliefs about the value of η_t, which in turn will raise the optimal policy parameter. An increase in the noisiness of the signals pertaining to newly-created assets will raise $\hat{\sigma}_\zeta^2$ and hence lower ϕ_b^I.

The intuition behind this can be summarised as follows. An increase in the degree of noisiness in asset markets means that agents are less certain about the source of asset price changes they experience following a shock. In other words, the prediction they make about the value of the stocks they will sell in the next period becomes less accurate. This forces the young generation to purchase fewer assets and consume more of their income, which in turn lowers the consumption capacities of the old, hence destabilizing consumption and lowering welfare. In response, the central bank raises its policy parameter in order to lower its compensation to the young generation.

This result can be illustrated by plotting the policy parameter chosen by the central bank

$\frac{\sigma_\nu}{\sigma_s} = 0.89$, where $\sigma_s^2 = \sigma_\varepsilon^2 + \sigma_\nu^2$ is the volatility of the signal. The high volatility curves, on the other hand, use $\sigma_\nu^2 = 0.9$ and a relative volatility of 0.99.

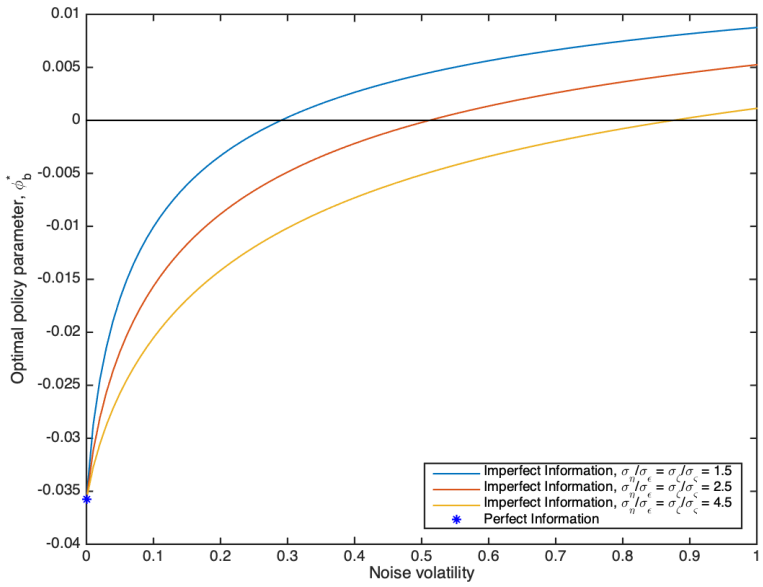
Figure 2.9: Optimal Policy and Noise



to minimise the loss function, ϕ_b^{Ω}, against the relative volatility of noise pertaining to pre-existing assets (Figure 2.9). When $\sigma_\nu = 0$, agents have full information and the model reverts back to the perfect information benchmark, where $\phi_b^P = -0.37$. As noise starts increasing, the central bank raises its policy parameter. Eventually, the old will need to be compensated to such an extent that the central bank will reverse its policy stance and start 'leaning against the wind'.

Sensitivity analysis. Given the limited breadth of evidence regarding the calibration of fundamental and bubble shock parameters, we perform a sensitivity analysis to verify the extent to which the above result depends on the chosen parameter values. We therefore plot the optimal policy parameter against relative noise volatility for alternative bubble-to-fundamental volatility ratios (Figure 2.10). The result that the central bank starts reversing its policy stance for a high enough degree of noisiness seems robust. In fact, it holds for bubbles being up to 5.5 times more volatile than fundamentals. While not many empirical estimations of this ratio exist, it is more than twice the value found by Liu et al. (2017)

Figure 2.10: Sensitivity Analysis



for the period 1991-2011, whose findings suggest that bubble volatility is about 250% that of fundamentals. Further research is needed to confirm such magnitudes and their possible range over time.

Discussion. This bring us back to the ongoing debate regarding the role central banks should play when dealing with growing bubbles described in Section 3.1. While central banks have typically conducted monetary policy assuming that interest rates should 'lean against the bubble', recent studies by Galí (2014 and 2017) have argued that such a policy stance is likely to further destabilise markets. My findings may be able to reconcile these two views. In fact, studies that have looked at the effects of monetary policy on bubbles did so within a perfect information framework. Under full information, our model also finds that the central bank should lower interest rates in response to a growing bubble. Policy-makers, on the other hand, operate within a noisy world. When noise is high in this model, the central bank also believes it should raise interest rates when bubble shocks hit. Hence, both views hold but for different levels of noise. However, since the central bank's response will

inevitably be sub-optimal under imperfect information given that it cannot directly observe the size of the bubble, the normative question of how the monetary authority should adjust its response once it internalises the fact that it does not have full information remains open.

My model does provide some clues. The policy parameter which minimises the welfare loss under perfect information is no longer optimal in the presence of noise given that consumers will be reacting differently to bubble shocks. Crucially, the sub-optimality of the central bank's response is due to policy-makers underestimating the size of bubble shocks, as shown in Section 2.3. Hence, for any given policy parameter ϕ_b^I, the central bank will adjust the interest rate insufficiently as it underestimates the size of the shock it is reacting to. Needless to say, it would be misguided to conclude that central banks should overreact to any observed market movement, given that they are unable to observe their source. Such a policy stance would lead to negative externalities that would likely outweigh its benefits. Further research is required in order to shed light on this important question.

2.5 Concluding Comments

This paper sought to contribute to the literature on rational bubbles by modeling imperfectly observable bubbles in a general equilibrium framework for the first time, and analyzing their implications for the conduct of monetary policy. We found that imperfect information forces agents to dampen their responses to fundamental and bubble shocks, hence producing large welfare losses that go hand-in-hand with the level of noise in stock markets. The central bank seeks to counteract these losses but faces a dilemma, between compensating the young and the old generation, when determining its optimal policy response. Under perfect information, 'leaning against the wind' policies will end up exacerbating market

instability and welfare losses. Imperfect information, on the other hand, calls for the central bank to compensate the old generation to a greater extent. For a high enough level of noisiness, the central bank will start 'leaning against the wind'. Nevertheless, investors and the central bank will always underestimate the size of a bubble, implying that any policy response will be sub-optimal.

This model offers a useful laboratory for understanding the implications of unobservable bubbles for agents' behaviour and monetary policy. To this end, it deliberately makes a number of simplified assumptions that facilitate this analysis. However, this also creates some caveats. First, this framework is ill-suited for quantitative analysis and the study of business cycles as output and employment remain constant in equilibrium, which rules out bubble-driven fluctuations. Second, the underlying dynamics adopted here do not allow for bubbles to burst, as they assume instead that bubbles can only progressively deflate over time. Future research should address these issues. These caveats notwithstanding, the analysis above may be useful by providing a framework to think about imperfect information in financial markets and its implications for policy-makers.

References

Adam, K., Marcet, A. and Nicolini, J. P. (2007), 'Stock Market Volatility and Learning', *CEPR Discussion Paper 6518* .

Asriyan, V., Fornaro, L., Martin, A. and Ventura, J. (2016), 'Monetary Policy for a Bubbly World', *National Bureau of Economic Research No. w22639* .

Barlevy, G., Gale, D. and Allen, F. (2017), 'On Interest Rate Policy and Asset Bubbles', *Society for Economic Dynamics* .

Bernanke, B. S. (2002), 'Asset Price "Bubbles" and Monetary Policy', *Remarks, New York Chapter of the National Association for Business Economics, New York* **103**(7), 3045–70.

Bernanke, B. S. and Gertler, M. (1999), 'Monetary Policy and Asset Price Volatility', *Presented at the Federal Reserve Bank of Kansas City Annual Conference* .

Bernanke, B. S. and Gertler, M. (2001), 'Should Central Banks Respond to Movements in Asset Prices?', *American Economic Review* **91**(2), 253–257.

Blanchard, O. J., L'Huillier, J.-P. and Lorenzoni, G. (2013), 'News, Noise, and Fluctuations: An Empirical Exploration', *American Economic Review* **103**(7), 3045–70.

Borio, C. and Lowe, P. (2002), 'Asset Prices, Financial and Monetary Stability: Exploring the Nexus', *BIS Working Paper 114* .

Cecchetti, S. G., Genberg, H., Lipsky, J. and Wadhwani, S. (2000), 'Asset Prices and Central Bank Policy', *The Geneva Report on the World Economy* **2**.

Collard, F. and Dellas, H. (2004), 'The New Keynesian Model with Imperfect Information and Learning ', *mimeo, CNRS-GREMAQ* .

Collard, F., Dellas, H. and Smets, F. (2009), 'Information and the Business Cycle', *Journal of Monetary Economics* **56**(S1), S38–S56.

Dong, F., Miao, J. and Wang, P. (2017), 'Asset Bubbles and Monetary Policy', *mimeo, Boston University* .

FOMC (2008), 'Minutes of the federal open market committee'.
URL: *https://www.federalreserve.gov/monetarypolicy/fomcminutes20080130ep.htm*

Forni, M., Gambetti, L., Lippi, M. and Sala, L. (2017), 'Noise Bubbles', *American Economic Journal* **9**(4), 122–152.

Galí, J. (2014), 'Monetary Policy and Rational Asset Pricing Bubbles', *American Economic Review* **104**(3), 721–752.

Galí, J. (2017), 'Monetary Policy and Bubbles in a New Keynesian Model with Overlapping Generations', *CREI Working Paper* .

Ideka, D. (2017), 'Monetary Policy, Inflation and Rational Asset Price Bubbles', *mimeo, Bank of England* .

Kohn, D. L. (2006), 'Monetary Policy and Asset Prices', *At 'Monetary Policy: A Journey from Theory to Practice', Frankfurt Germany* .

Levine, P., Pearlman, J., Perendia, G. and Yang, B. (2010), 'Endogenous Persistence in an Estimated DSGE Model under Imperfect Information', *Department of Economics and University of Surrey Discussion Papers* **0310**.

Levine, P., Pearlman, J. and Yang, B. (2012), 'Imperfect Information, Optimal Monetary Policy and Informational Consistency', *Department of Economics and University of Surrey Discussion Papers* **1012**.

Liu, R., Chui-Man Hui, E., Lv, J. and Chen, Y. (2017), 'What Drives Housing Markets: Fundamentals or Bubbles?', *The Journal of Real Estate Finance and Economics* **55**, 395–415.

Martin, J. and Ventura, A. (2012), 'Economic Growth with Bubbles', *American Economic Review* **102**(6), 3033–3058.

Martin, J. and Ventura, A. (2018), 'The Macroeconomics of Rational Bubbles: A User's Guide', *Annual Review of Economics* **10**, 505–539.

Miao, J. and Wang, P. (2018), 'Asset Bubbles and Credit Constraints ', *American Economic Review* **108**(9), 2590–2628.

Miao, J., Wang, P. and Xu, Z. (2015), 'A Bayesian Dynamic Stochastic General Equilibrium Model of Stock Market Bubbles and Business Cycles ', *Quantitative Economics* **6**, 599–635.

Nimark, K. (2006), 'Optimal monetary policy with real-time signal extraction from the bond market', *Reserve Bank of Australia Discussion Papers* **2006**(05).

Pasricha, G. K. (2006), 'Kalman Filter and its Economic Applications ', *MPRA Working Paper* .

Peng, Q. (2016), 'Noisy Rational Bubbles', *mimeo* .

Samuelson, P. A. (1958), 'An Exact Consumption-loan Model of Interest with or without the Social Contrivance of Money', *The Journal of Political Economy* **66**(6), 467–482.

Tirole, J. (1985), 'Asset Bubbles and Overlapping Generations', *Econometrica* **53**(6), 1071–1100.

Appendix 2.A Kalman Filter

The unobservable components of pre-existing assets are captured by the vector $\boldsymbol{X_t}$:

$$\boldsymbol{X_t} = \boldsymbol{F}\boldsymbol{X_{t-1}} + \boldsymbol{B}\boldsymbol{\tau_t}$$

where:

$$\boldsymbol{X_t} = \begin{bmatrix} \xi_t^f & \xi_{t-1}^f & \xi_t^b & u_t^f & u_{t-1}^f & u_t^b \end{bmatrix}', \boldsymbol{\tau_t} = \begin{bmatrix} \varepsilon_t & \eta_t & \nu_t & \varsigma_t & \zeta_t & \omega_t \end{bmatrix}',$$

$$\boldsymbol{F} = \begin{bmatrix} 1+\rho & -\rho & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \rho \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Agents observe the vector $\boldsymbol{Z_t}$:

$$\boldsymbol{Z_t} = \boldsymbol{H}\boldsymbol{X_t} + \boldsymbol{C}\boldsymbol{\tau_t}$$

where:

$$\boldsymbol{Z_t} = \begin{bmatrix} \xi_t & s_t^{\xi} & u_t & s_t^{u} \end{bmatrix}', \boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \boldsymbol{C} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Every period, agents predict the values of $\boldsymbol{X_t}$ and the associated variance-covariance matrix $\boldsymbol{\Sigma_t}$ using:

$$\boldsymbol{X_{t+1|t}} \equiv E[\boldsymbol{X_{t+1}}|\boldsymbol{Z_t}] = [\boldsymbol{F} - \boldsymbol{K_t}\boldsymbol{H}]\boldsymbol{X_{t|t-1}} + \boldsymbol{K_t}\boldsymbol{Z_t} \tag{2.41}$$

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{F}\left[\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{L}_t)^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\right]\boldsymbol{F}'$$

where $\boldsymbol{X}_{j|k}$ is the belief about period j based on the signal received at k and:

$$\boldsymbol{K}_t = \boldsymbol{F}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{L}_t]^{-1}$$

is the Kalman gain, where $\boldsymbol{L}_t = E[\boldsymbol{C}\boldsymbol{\tau}_t\boldsymbol{\tau}_t'\boldsymbol{C}']$ captures the volatility of the noise shock. The forecast made in the previous period is updated using the current signal following:

$$\boldsymbol{X}_{t|t} \equiv E[\boldsymbol{X}_t|\boldsymbol{Z}_t] = \boldsymbol{X}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{L}_t]^{-1}(\boldsymbol{Z}_t - \boldsymbol{H}\boldsymbol{X}_{t|t-1}) \quad (2.42)$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{L}_t)^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}$$

Following Pasricha (2006), equations (2.41) and (2.42) can be simplified to:

$$\boldsymbol{F}\boldsymbol{X}_{t|t} = \boldsymbol{K}_t\boldsymbol{Z}_t + (\boldsymbol{F} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{X}_{t|t-1}$$

$$\boldsymbol{X}_{t+1|t} = \boldsymbol{F}\boldsymbol{X}_{t|t}$$

2.A.1 Full Information Equivalent

Following Blanchard et al. (2013), the above equations can be expressed as a full information problem. Let us write the joint dynamics of $\boldsymbol{X}_{t|t}$ and $\boldsymbol{Z}_t$ as:

$$\boldsymbol{X}_{t|t} = \boldsymbol{F}\boldsymbol{X}_{t-1|t-1} + \boldsymbol{K}_t(\boldsymbol{Z}_t - \boldsymbol{C}\boldsymbol{F}\boldsymbol{X}_{t-1|t-1})$$

$$\boldsymbol{Z}_t = \boldsymbol{C}\boldsymbol{F}\boldsymbol{X}_{t-1|t-1} + \boldsymbol{Z}_t - \boldsymbol{C}\boldsymbol{F}\boldsymbol{X}_{t-1|t-1}$$

Let $\boldsymbol{\Sigma_s}$ be the variance-covariance matrix of $var_{t-1}[\boldsymbol{Z_t}]$ such that it can be factorised as $\boldsymbol{\Sigma_s} = \boldsymbol{GG'}$ for some matrix $\boldsymbol{G}$. Then, the following model is observationally equivalent to the original signal extraction problem:

$$\boldsymbol{\hat{X}_t} = \boldsymbol{F}\boldsymbol{\hat{X}_{t-1}} + \boldsymbol{K_t}\boldsymbol{G}\hat{\boldsymbol{\gamma}}_t$$

$$\boldsymbol{Z_t} = \boldsymbol{CF}\boldsymbol{\hat{X}_{t-1}} + \boldsymbol{G}\hat{\boldsymbol{\gamma}}_t$$

where $\boldsymbol{\hat{X}_t}$ and $\hat{\boldsymbol{\gamma}}_t$ are assumed to be perfectly observable and $\hat{\boldsymbol{\gamma}}_t$ is a vector of i.i.d. shocks.

Chapter 3

Stock Market Overconfidence and Rational Bubbles

Chapter Abstract

Motivated by the empirical evidence on behavioural biases on stock markets, we develop a rational bubble framework with imperfect information about the probability of a future market crash, where investors are overconfident about the accuracy of their knowledge. We use this model to draw a number of generalisable propositions about the ways in which overconfidence impacts stock market bubbles and the channels through which it occurs. We show that overconfidence is key to bubble existence whenever the range of reasonable beliefs that agents can hold includes the possibility that assets are too risky to be purchased. We also find that overconfidence in good times distorts asset prices upwards through a chain reaction of risk miscalculation, over-investing, and mispricing. The reverse is true during bad times. In addition, excessive optimism improves the efficiency of the pre-crash economy. Despite the potentially catastrophic costs of bubble bursts, our welfare analysis indicates that investors are better off maximising their overconfidence when they receive optimistic news, as the potential benefits of higher efficiency exceed these potential costs. Finally, we quantify our key results and test the empirical validity of our framework through an empirically-based numerical analysis.

3.1 Introduction

Historical patterns of investors' behaviour on financial markets pose a serious challenge to asset pricing theories that are based on fully rational expectations. In particular, the notion of efficient markets is especially hard to reconcile with significant bubbly episodes and the prolonged recessionary periods that follow their collapse.

In fact, according to the Efficient Market Hypothesis (EMH) first proposed by Fama (1970), all information, public or private, is entirely and promptly captured by the market. As a consequence, stock prices are always traded at their fair market value and the existence of bubbles – defined as a systematic deviation of asset prices from fundamentals – is always ruled out (Herdegen and Schweizer, 2015). The EMH also implies that no amount of technical analysis can allow investors to hold better information than their peers, given that all information is already reflected in stock prices (Grossman, 1976; Milgrom and Stockey, 1982; Tirole, 1982; Malkiel, 2003). In other words, they should not agree to disagree, nor should they expect to make abnormally high profits based on their knowledge, as it is not possible for them to beat the market.

Numerous studies, however, have documented historical phenomena conflicting with the EMH. In particular, periods of significant optimism are typically accompanied by excessively high prices, and are often followed by prolonged periods of pessimism and excessively low prices (Galí, 2014; Daniel and Hirshleifer, 2016; CFI, 2022). One such episode occurred during the dot-com craze that saw optimistic investors flood into the fast-expanding Internet industry. During this time, what started as a gradual increase in Internet stocks resulted in a 400% expansion of the Nasdaq composite index between 1995 and 2000, which was followed by an 80% drop in less than two years. Contrary to the rationale proffered by the EMH, these large swings were not mirrored by changes in fundamentals. In reality, it

appears that stock prices are not always traded at their fair market value and bubbles indeed exist in markets. In fact, the size of the boom has been attributed in large part to the excessive purchase of stocks of a large pool of unprofitable Internet companies, while the recession that followed led to the undervaluation of a certain number of profitable firms (Lamont and Thaler, 2001; Cochrane, 2002; Ofek and Richardson, 2003; Stiglitz, 2003; Rad, 2018). In addition, investors expect to make abnormally high profits during bubbly episodes, resulting in excessive trading volumes. The ensuing recessionary periods have then coincided with abnormally low expected profits and trading volumes, which standard expectations models do not adequately explain (Daniel et al., 1998; Scheinkman and Xiong, 2003).

The mounting evidence against the validity of the EMH has led to a number of questions. First of all, why do investors hold such extreme beliefs about their future returns and is it possible to rationalise their seemingly irrational behaviour? In addition, how can we explain the fact that investors take large risks during bubbly periods by purchasing assets that are significantly overvalued and whose value may suddenly collapse, while on the other hand exhibiting excessive pessimism during bad times? In this paper, we posit that imperfect information about the future value of assets opens the door to behavioural biases, as expectations become inevitably entangled with emotion and personal beliefs. In particular, we explore overconfidence as a plausible explanation for the phenomena described above. Namely, we want to understand whether overconfidence can explain the size of bubbles during booms and the excessively low prices observed during recessionary periods, as well as the mechanisms giving rise to these patterns.

The presence of overconfidence in financial markets has been widely documented by literature in psychology and behavioural finance. Despite market participants receiving frequent and measurable feedback on their investments, investors appear to constantly overes-

timate the accuracy of the information they hold and their capacity to analyze it. In turn, this leads them to overestimate their ability to predict the future (Odean, 2002; Scheinkman and Xiong, 2003; Ben-David et al., 2007; Daniel and Hirshleifer, 2016).[1] While we are not suggesting that overconfidence can be used as an explanation for all stock market anomalies, we do propose that it is very likely to play a central role, especially when it comes to asset pricing.

We develop a tractable two-period overlapping generations model with stochastic bubbles where agents face imperfect information. The central feature of the framework is that investors do not observe the probability of bubble crashes and overestimate the accuracy of the signal they receive. The novelty of the model lies in its ability to study the impact of overconfidence on bubbles in a general equilibrium setting. The use of a stylised framework allows us to draw generalisable propositions on the role of this behavioural bias in the existence and dynamics of rational bubbles, and to isolate the channels through which this happens.

Agents purchase bubbly assets when young and sell them when old. These risky assets may suddenly lose their value in any period, resulting in the old generation receiving no returns to their investments. The probability of this happening is unobservable, and agents must instead estimate it based on the noisy signal they receive and whose accuracy they overestimate. Hence, if agents receive good news suggesting a lower-than-expected crash probability, they become overoptimistic and underestimate the likelihood that assets will lose their value next period. On the other hand, if they receive bad news, they overestimate this likelihood and become overly pessimistic. As discussed below, this setup is equivalent to a beauty contest with a continuum of agents facing asymmetric information and higher-

[1]The evidence on overconfidence in financial markets is reviewed in Section 3.2.

order beliefs. The model is then used to understand the various ways in which overconfident beliefs and bubbles interact, leading to a number of key results.

First, we revisit the conditions for bubble existence in the presence of imperfect information and biased beliefs, and find that overconfidence can be key in determining whether a bubbly equilibrium can arise. In fact, under certain conditions, there exists a threshold level of confidence such that agents must be overconfident enough about positive news or not overly confident about negative news. In other words, bubbles can only exist if they are *perceived* as not being too risky.

We also find that overconfidence leads to excessive prices. When agents receive a signal, they overestimate its accuracy, leading them to overreact to it. When the signal is favourable (unfavourable), this overreaction makes them over-(under-) invest in assets, resulting in higher (lower) prices. We also rationalise agents' behaviour as we are able to explain the anomalies described above using a model with individuals making optimal choices given the biased information they hold. The chain reaction explaining these phenomena is that overconfidence leads agents to underestimate (overestimate) the risks they are taking in good (bad) times, which in turn makes them miscalculate their risk premium, resulting in their overreaction and the mispricing of the asset.

Furthermore, overconfidence is able to explain the market's overreaction to a shift in news sentiment. For instance, if the signal goes from being favourable to unfavourable, the shift in asset prices will be drastically larger in the presence of overconfidence than with standard imperfect information. This is because the bubble will be excessively large under optimistic news and excessively small under pessimistic news, resulting in a much heftier drop in the market.

The use of stochastic bubbles gives us the opportunity to analyse the net impact of over-

confidence on welfare in a more realistic manner than with deterministic bubbles. In fact, stochastic bubbles not only enable us to capture the benefits of giving agents an additional store of value, but we are also able to account for the costs of potential bubble bursts. We show that overconfidence about good news improves the efficiency of the pre-crash economy by reducing the problem of overaccumulation. The reverse is true in scenarios where agents are overconfident about bad news. Despite the very large costs incurred in the event of a crash, we find that agents are better off underestimating the risks they take, as the potential benefits of higher efficiency outweigh the potential costs of bubble bursts.

We then take our model to the data to quantify our key results and to test the empirical validity of our calibrated framework. We find that, on average, investors believe their information to be nearly four and a half times more precise than it truly is. As a result, the bubbly component of asset prices is 10.48% larger during optimistic times, and 2.58 times more volatile across the entire sample. Furthermore, the drop in the value of bubbles following a reversal in news sentiment is three times larger than under standard imperfect information. This suggests that overconfidence plays an important role in determining the first two moments of the bubbly component of stock prices. Finally, we find that the model is able to produce bubble fluctuations of similar magnitude to what can be estimated using empirical data, hence indicating that our results may hold empirical relevance. This finding is further supported by a robustness check.

Related Literature: The concept of overconfidence has been studied for decades by the various literature in psychology, which has found ample evidence of its widespread existence across time and space.[2] In the context of financial markets, the exploration of this psychological phenomenon as a potentially important factor in the determination of asset

[2]See for instance Stevenson (1981), Cooper et al. (1988); Hogarth and Karelaia (2012) Alpert and Raiffa (2013), Lichtenstein et al. (2013), and Moore and Healy (2008).

prices started with the emergence of behavioural finance. This branch of the literature offered a deviation from traditional finance, whose foundations relied on the Efficient Market Hypothesis and Modern Portfolio Theory, both of which were based on the idea that markets are efficient and investors are fully rational (Fama, 1970). Instead, behavioural finance argued that markets do not meet the efficiency criteria and are heavily affected by psychological and behavioural factors that are able to explain market phenomena and investors' behaviour. This literature hence explored whether investors exhibit overconfidence in their trading decisions, consistently finding this to be the case.[3]

Despite the extensive coverage of overconfidence in the behavioural finance literature, the study of its impact on stock market bubbles and the extent to which it could explain boom-bust cycles has been limited. A number of papers have analysed its effect on asset pricing dynamics. This includes Daniel et al. (1998) and Daniel and Hirshleifer (2016), who construct a series of one-equation asset pricing models where agents receive a private signal about future payoffs but are overconfident about its accuracy. They show that overconfidence leads to an initial increase (decrease) in expected prices when agents receive a favourable (unfavourable) private signal before converging back to their initial level as agents learn from a public signal they subsequently receive. They also show that self-attribution biases make agents respond more slowly to the public signal, keeping prices high for longer.[4] A related paper is Scheinkman and Xiong (2003), which uses overconfidence as a means to create disagreement between two groups of investors who are overconfident about the precision of two different signals in a single-asset model. They find that investors are willing to purchase assets at a price higher than their own valuation of future dividends, as they expect

[3]See Section 3.2 for a more extensive coverage of this evidence.

[4]Self-attribution bias is a concept referring to individuals' tendency to attribute negative outcomes to bad luck and positive ones to their own skills and knowledge.

other buyers to be willing to pay an even higher price in the future.[5] [6] Nevertheless, none of these papers are able to capture the general equilibrium effects and transmission mechanisms of overconfidence, as they all operate within simple one-equation or partial equilibrium models.[7] In addition, nearly all of these papers do not have well-defined bubbles or simply look at asset prices as a whole. Hence, the key contribution of this paper to the behavioural finance literature is to analyse the impact of overconfidence on well-defined bubbles in an overlapping generations framework allowing for general equilibrium effects.

The study of bubbles in a general equilibrium context, on the other hand, has been vastly covered by the rational bubbles literature. It originated with the seminal papers by Samuelson (1958) and Tirole (1985), who developed overlapping generations models allowing for the existence of bubbly equilibria. It rapidly expanded following the 2007-8 crisis to analyse, amongst other things, the impact of monetary policy on bubbles (Galí, 2014; Asriyan et al., 2016; Barlevy et al., 2017; Galí, 2017; Ideka, 2017), the role of fiscal policy (Teulings, 2016), as well the interaction of bubbles with the labour market (Cahuc and Challe, 2009; Hashimoto and Im, 2019). However, a recurrent criticism that has affected the credibility of this literature has been the fact that rational (deterministic) bubbles can only produce beneficial effects on the economy and agents by relaxing borrowing constraints and/or offering an additional store of value. In addition, a well-known finding across the literature is that bubbles actually make the economy more efficient by lessening the problem of overaccumulation of capital. This contrasts with the potentially catastrophic consequences of bubbly

[5]For more papers on the relation between overconfidence and stock prices in a partial equilibrium model, see de Long et al. (1990), Kyle and Wang (1997), Benos (1998), Odean (2002), Bernardo and Welch (2004), and Caballé and Sákovics (2003). For the empirical evidence, see Section 3.2.

[6]Note that some of the papers in the literature use overconfidence and overoptimism interchangeably, which might lead to some confusion. In this paper, we define overoptimism as overconfidence about good news and overpessimism as overconfidence about bad news.

[7]A notable exception of a paper integrating overconfidence in a general equilibrium framework is Kantorovitch (2021). In his paper, agents receive a noisy signal about firm productivity and are overconfident about its precision. However, the model focuses on the efficiency of capital allocation across firms, with asset pricing taking a secondary role and not allowing for the existence of well-defined bubbles.

episodes observed in the real world.

A strand of the literature that has sought to address this limitation has developed so-called stochastic bubbles. These bubbles may burst in any period with a positive probability and are not expected to reappear once they have collapsed. Originally developed by Blanchard (1979), Blanchard and Watson (1982), and Weil (1987), this strand recently expanded with a number of contributions by Binswanger (2005), Hirano and Yanagawa (2017), Guerron-Quintana et al. (2019), and Guerron-Quintana et al. (2021). Such bubbles are still able to reduce market inefficiencies prior to bursting, but they can cause serious damage to the generations affected by the bubble crash and, crucially, they introduce risk into agents' decision-making process.

Nevertheless, attempts at integrating the lessons found within the behavioural finance literature into a rational bubbles framework have been limited to say the least. As far as we know, no paper has attempted to incorporate imperfect information about the returns to assets – a concept often explored by the behavioural finance literature in partial equilibrium models. By extension, the notion of overconfidence – which, as a prerequisite, requires the presence of imperfect information – and its role in the formation of rational bubbles has never been studied. This is despite ample evidence of the presence and importance of this cognitive bias in financial markets. Our contribution to the rational bubbles literature lies here. In particular, through this model, we revisit the conditions for bubble existence under imperfect information and overconfidence; we analyse and quantify the impact of overconfidence on rational bubble equilibria and the mechanisms through which this occurs; and we determine the net welfare impact of overconfidence under uncertainty.

Structure: The remainder of the paper is structured as follows. In the next section, we review the evidence and empirical literature on the Efficient Market Hypothesis, over-

confidence, and bubbles. Section 3.3 describes the model setup, including the information structure. Section 3.4 discusses the model's equilibrium conditions and presents a phase diagram able to fully capture the dynamics of the economy. The impact of overconfidence on bubble existence and dynamics is discussed in Section 3.5. In Section 3.6, we focus on the implications for dynamic efficiency and we construct a welfare analysis. Section 3.7 presents the numerical analysis. Section 3.8 concludes.

3.2 Evidence on Overconfidence and Stock Market Bubbles

This section aims to provide a short review of the literature surrounding the three major themes discussed above – market (in)efficiency, overconfidence, and stock market bubbles. We first delve into the historical patterns suggesting that the implications of the EMH do not hold in practice, including evidence of the existence of large bubbles. We then review the psychological and financial literature documenting the presence of overconfidence in financial markets. Finally, we discuss findings corroborating the idea that overconfidence affects stock market bubbles.

3.2.1 Evidence on Market Efficiency

As discussed in Section 3.1, the EMH stipulates that stocks are always traded at their fair market value, implying that stock market bubbles can never exist. This is because stock prices fully capture all publicly and privately available information, such that agents should not agree to disagree, nor should they expect to make large profits from their investments, as consistently beating the market is not possible.

The first piece of evidence against the EMH therefore comes from the very existence of

stock market bubbles. Identifying whether a market swing is due to an underlying change in fundamentals or is simply driven by investors' sentiment is not an easy feat. In fact, according to the Discounted Cash Flow (DCF) model, fundamentals are defined as the expected stream of future discounted cash flows. This, however, is not directly observable. A number of studies have gotten around this problem by approximating this stream of payments using various measures of earnings – such as current, average, or forecast earnings – or by developing more complex methods, usually involving sequences of unit root tests specifically designed for bubble detection (Phillips et al., 2011; Contessi and Kerdnunvong, 2015). They find many instances where market swings appear to significantly diverge from fundamental changes, signalling the presence of a bubble (Phillips et al., 2015; Rad, 2018). One of the most notable examples in recent history is the dot-com bubble, which at its peak in the year 2000 saw the value of all Nasdaq Internet stocks shoot higher than the GDP of most nations at the time, despite a large number of Internet firms being unprofitable, or at the very least severely overvalued (Shiller, 2003). All in all, stock market bubbles have been documented across a wide range of countries and decades, contradicting the claim that stock prices must always equal their fair market value.

Furthermore, the boom and bust cycle correlates with a number of behavioural anomalies. For instance, it has been shown that investors expect to make abnormally high profits during bubbly episodes, resulting in excessive trading volumes, while the ensuing recessionary periods have coincided with abnormally low expected profits and market volumes (Daniel et al., 1998; Scheinkman and Xiong, 2003). For instance, Cochrane (2002) finds a strong correlation between the value of the Nasdaq index and Nasdaq trading volumes during the dot-com boom and bust, and a similar relation in the NYSE in 1929. Standard rational expectations models do not adequately explain these anomalies, as extensively dis-

cussed by Daniel and Hirshleifer (2016). In addition, these patterns clash with the EMH's idea that investors should not agree to disagree. In fact, ample evidence points to their conviction that they are able to outperform the market during booms, and that they become excessively pessimistic during recessions.

3.2.2 Evidence of Investors' Overconfidence

Stock market participants receive frequent and measurable feedback on their investments, first and foremost through price changes and dividend payments. This should allow them to quickly learn and adjust their expectations accordingly. Nonetheless, many studies have underscored the pervasiveness of overconfidence in financial markets over the years. In particular, they point to the fact that financial experts and investors seem to constantly overestimate their capacity to analyse the information they receive and to make predictions about the future. In turn, this results in frequent risk miscalculation and biased expectations (Odean, 2002; Scheinkman and Xiong, 2003; Ben-David et al., 2007).

The first indication of the existence of overconfidence comes from a strand of the behavioural finance literature investigating the so-called 'active investing puzzle'. Namely, it seeks to understand investors' obstinacy to take part in active investing, as opposed to following more passive strategies, such as holding index funds. This is particularly surprising due to the fact that very few investors are actually able to outperform the market on average. For instance, between 1999 and 2019, the average equity fund earned a 4.25% return, while the S&P 500 averaged a 6.06% return (Dalbar, 2020). The average individual investor, on the other hand, made a loss (Daniel and Hirshleifer, 2016). At first glance, it seems puzzling or even irrational that so many individuals, most of whom professionals, would persist in making choices that are likely to offer a significantly lower average payoff, despite receiving

constant feedback on their decisions.

The behavioural finance literature has argued that overconfidence is the key driver of this phenomenon (Odean, 2002; Barber and Odean, 2001; Statman et al., 2006). For starters, the knowledge that the average investor is bound to underperform the market is unlikely to deter an investor who overestimates her ability to produce high returns (French, 2008). In an anonymised survey of 300 fund managers, Montier (2006) reported that 74% of them believed to possess above-average investing skills. The other 26% thought they were average. In other words, virtually none of them had placed themselves in the bottom half of the population. Similarly, Ben-David et al. (2007) and Glaeser et al. (2012) document the presence of overconfidence amongst financial experts, such as corporate financial officers, investment bankers, and professional traders. The evidence even extends to individual investors, whose financial literacy and technical means are commonly outmatched. In fact, Malkiel (2003) found that they are overly confident in their capacity to spot better-performing active fund managers. In addition, the rapid expansion of online trading in recent decades, which has allowed individuals with little to no financial training to take part in active investing, has been highlighted by Choi et al. (2002) as a sign of widespread overconfidence on markets. Additional evidence comes from experimental markets, where investors have been found to overestimate the accuracy of their information, with more overconfident individuals even underperforming their less confident peers (Biais et al., 2005).

This leads to a crucial question: what form is overconfidence most likely to take in financial markets? The psychology literature, which has widely covered the topic, suggests that the two most common ways for overconfidence to arise are over-placement – overestimating one's ranking in a population – and overprecision – overestimating the precision of one's knowledge (Stevenson, 1981; Cooper et al., 1988; Hogarth and Karelaia, 2012).

However, a number of studies have stressed the latter as being the most widespread and persistent form. They have also argued that it is the source of over-placement and of other behavioural biases (see for instance Alpert and Raiffa, 2013; Lichtenstein et al., 2013; and, more importantly, Moore and Healy, 2008). These findings are consistent with the evidence from financial markets, which also points to investors overestimating the accuracy of their knowledge as being the most prevalent type of behavioural bias (Daniel et al., 1998; Odean, 2002; Scheinkman and Xiong, 2003; Daniel and Hirshleifer, 2016). As a consequence, market participants have a tendency to overestimate their ability to predict the future, resulting in incorrect beliefs about state probabilities and excessively narrow confidence intervals regarding future events (Ben-David et al., 2007). While the degree of overconfidence is likely to vary across a wide range of factors, such as country, gender, and time, the literature strongly suggests that a majority of individuals demonstrate excessive confidence about the information they hold (van Gasteren, 2016).

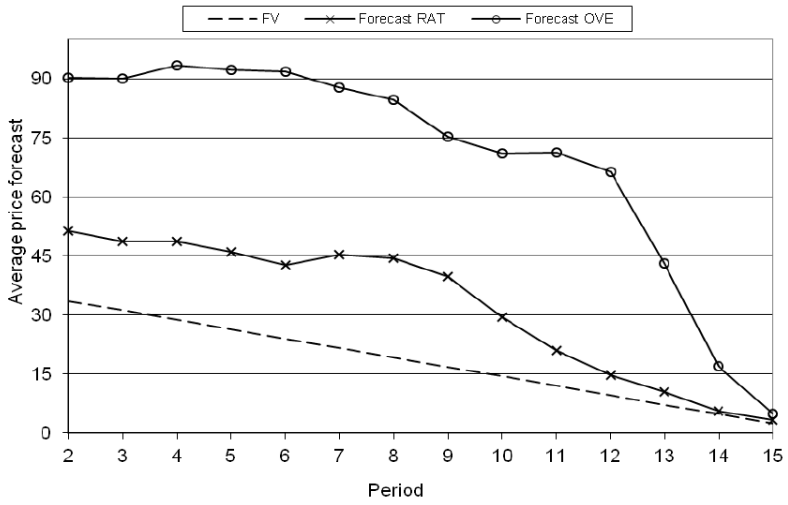
Another cognitive bias supporting the idea of persistent overconfidence is the self-attribution bias. It refers to investors' strong tendency to attribute successes to their own skills, while blaming negative outcomes on bad luck. This prevents individuals from adjusting their confidence level, even when their predictions turn out to be incorrect (Scheinkman and Xiong, 2003). This phenomenon has been widely documented in the literature, for instance by Wang et al. (2017), who find strong support that individuals exhibit this condition even in unambiguous contexts. Evidence of its presence has also been found more specifically in financial markets by Gervais and Odean (2001), Gervais et al. (2011), and Kim (2013).

3.2.3 Evidence on the Effects of Overconfidence on Stock Market Bubbles

The relation between overconfidence and bubbles has been analysed in both empirical and experimental settings. The bulk of the evidence comes from the latter, although some empirical studies have looked at the question as well. The difficulty with this approach is the lack of direct observability of either one of the variables of interest. Two papers are worth mentioning though. The first one is by van Gasteren (2016), who uses Malmendier and Tate's (2005) measure of overoptimism and Froot and Obsfeld's (1991) proxy for bubbles and finds a positive relationship between the two. The second one is Jlassi et al. (2014). While their study focuses on the impact of overconfidence on market volatility, some of its conclusions apply to this discussion. In particular, they find evidence of overconfidence in both up and down markets, which reinforces the idea that this cognitive bias not only impact the booming phase of the business cycle, but also its recessionary part. They also find that overconfidence was the main driver behind the large boom in the U.S. housing market of the 2000s, and that it significantly worsened and extended the subsequent global financial crisis.

The experimental literature on the subject, pioneered by Smith et al. (1988), has extensively analysed bubbles in experimental markets. In a key experiment, conducted by Michailova and Schmidt (2016), participants were divided into groups based on their level of overconfidence. This was measured in pre-experimental sessions by asking participants to answer a general knowledge survey and rate their level of confidence about each of their answers. Each group was then endowed with experimental assets and cash, and was given the opportunity to purchase and sell assets during 15 different periods. At the end of each trading session, dividends were distributed to them and they were asked to forecast next period's payoffs. The results, reproduced in Figure 3.1, were quite clear. Despite funda-

Figure 3.1: Overconfidence and Bubbles in Experimental Markets



Source: Michailova and Schmidt (2016)

Note: FV indicates fundamentals; Forecast OVE and Forecast RAT refer to the forecast of next period's dividends made by the more overconfident and less overconfident groups, respectively.

mentals linearly declining over time, all groups consistently overestimated them. However, the more confident groups expected much higher returns to their investments, resulting in a much larger bubble. They also tracked fundamental changes less accurately.

One question that remains to be addressed is that of the channels through which overconfidence affects bubbles. This is best answered in a general equilibrium framework, which we attempt to do in this paper. Nevertheless, some hypotheses can be formulated based on previous studies on the various effects of this cognitive bias. First of all, there is ample evidence that overconfidence can considerably amplify trading volumes during periods of market optimism, and reduce them during pessimistic times (Odean, 2002; Grinblatt and Keloharju, 2009; Kelley and Tetlock, 2013). The finance literature has also established a positive relationship between changes in trading volumes and shifts in stock prices (Karpoff, 1987; Pathirawasam, 2011). Putting these two findings together suggests that overconfidence could impact bubbles by distorting market volumes. It has also been shown that overconfident investors tend to overreact to news, whether it is positive or negative (de Long

et al., 1990; Shiller, 2003). Secondly, overconfidence has been shown to distort investors' perception of risk, leading them to take on too much of it during optimistic times (Chuang and Lee, 2006; Gervais et al., 2011). Coincidentally, excessive risk has been highlighted as a central factor in the creation of stock market bubbles, hence suggesting a second potential channel (Brunnermeier and Oehmke, 2012). Thirdly, Odean (2002) finds that overconfidence leads to the persistent mispricing of securities, as overly optimistic investors overstate the probability that they will be able to sell them back at an even higher price in the future.[8] This, in turn, is likely to push prices even higher. Hence, there appear to be a number of potential mechanisms that could explain the transmission of overconfidence to bubbles. Investors appear to overreact to news, over- or under-invest, miscalculate the risks they face, and misprice securities as a result of their biased beliefs. This paper will shed some light on which of these channels, if any, plays a role.

3.3 The Model

In this section, we develop a two-period overlapping generations model with rational bubbles à la Tirole (1985). Agents purchase intrinsically useless (bubbly) assets when young, and sell them to the next cohort when old. Following Blanchard (1979), Blanchard and Watson (1982), and Weil (1987), assets may suddenly lose their value at the start of any period with probability $\mu_t \in [0, 1]$. Once bubbles collapse, they are not expected to rise again. The first novelty of the model is that agents do not observe next period's probability of crash, and instead receive a noisy signal Z_t and form beliefs about it. The second is that households are overconfident about the precision of the signal they receive, resulting in biased beliefs about future state probabilities.

[8] See Biais et al. (2005) as well.

Table 3.1: Transition Matrix

	$Q_{t+1} > 0$	$Q_{t+1} = 0$
$Q_t > 0$	$1 - \mu_{t+1}$	μ_{t+1}
$Q_t = 0$	0	1

3.3.1 Households

The economy is populated by overlapping generations living for two periods. Population grows deterministically at rate n, such that the size of the workforce at any period t is given by $N_t = (1+n)N_{t-1}$, with $N_0 = 1$.[9] When young, households inelastically supply labour in return for a wage rate w_t, which they then use for consumption, savings, or to purchase x_t units of a risky asset at price Q_t. When old, they receive the risk-free rate r_{t+1} in return for their savings, and they sell their shares at price Q_{t+1} to the new cohort. Their budget constraint in each period of their life is therefore given by:

$$c_{1,t} + s_t + Q_t x_t = w_t \tag{3.1}$$

$$c_{2,t+1} = (1 + r_{t+1})s_t + Q_{t+1}x_t \tag{3.2}$$

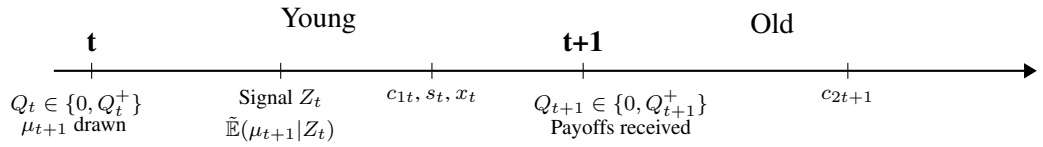
where $c_{1,t}$ and $c_{2,t}$ are consumption of the young and of the old, respectively, and s_t is savings. To remain consistent with the literature, the model exhibits short-sell constraints, which translate into agents not being able to buy negative amounts of the asset.[10]

Asset prices evolve according to the Markov process illustrated in Table 3.1. If assets are valued in period t, i.e., $Q_t > 0$, their value remains positive at $t+1$ with probability $(1 - \mu_{t+1}) \in [0, 1]$, or falls to 0 with probability μ_{t+1}. On the other hand, once assets are no

[9] The assumption of deterministic growth is a prerequisite for the existence of bubbles in an economy with positive interest rates. For more on this, see Tirole (1985).

[10] This assumption is rooted in the empirical observation that taking short positions in practice is not easily accessible to many investors, and presents additional risks that may deter risk-averse investors (Scheinkman and Xiong, 2003).

Figure 3.2: Timeline



longer valued, they remain worthless for the remainder of time. Note that, asymptotically, bubbles burst with certainty since the probability that they remain valued for T periods tends to 0 as T tends to infinity. While the true state of nature is determined by these probabilities, agents are unable to observe them and instead receive a noisy signal Z_t, which they use to form beliefs about μ_{t+1}, denoted by $\tilde{\mathbb{E}}(\mu_{t+1}|Z_t)$. This creates two layers of uncertainty for households: the first stemming from the positive probability that assets may suddenly lose their value in the following period; the second caused by the unobservability of the likelihood of this happening.

We discuss below agents' choices when young in chronological order, first laying out how beliefs about μ_{t+1} are formed, before solving households' maximisation problem.

3.3.1.1 Information Structure

The timeline of events across an individual's lifetime is summarised in Figure 3.2. At the beginning of each period, asset prices either stay positive, or fall to 0 and remain there for all remaining periods. If $Q_t > 0$, the probability μ_{t+1} that the bubble will burst in the following period is drawn from a truncated normal distribution such that $\mu_{t+1} \in [0, 1]$. In other words, $\mu_{t+1} \sim N(\mu, \sigma_\mu^2, 0, 1)$, where μ and σ_μ^2 are the first and second moments of a normal distribution. Households do not observe μ_{t+1}, and instead receive a noisy signal Z_t when young, such that:

$$Z_t = \mu_{t+1} + \upsilon_t \tag{3.3}$$

where υ_t is an iid noise shock that prevents agents from perfectly observing μ_{t+1}. υ_t is drawn from a truncated normal distribution such that $\upsilon_t \sim N(0, \sigma_\upsilon^2, -\mu_{t+1}, 1-\mu_{t+1})$, which ensures that $Z_t \in [0,1]$.[11] Young households then form beliefs about μ_{t+1} based on their signal, denoted by $\tilde{\mu}_{t+1|t} \equiv \tilde{\mathbb{E}}(\mu_{t+1}|Z_t)$, before making their consumption, savings, and investment decisions. When old, agents either sell their shares to the new cohort if asset prices remain positive, or are left with worthless assets if the bubble bursts. They use the remainder of their resources on consumption.

Let us denote the mean of the truncated normal distribution which μ_{t+1} is drawn from as $\bar{\mu} \equiv \mathbb{E}(\mu_{t+1}|0 < \mu_{t+1} < 1)$, and the variance of the truncated distribution which υ_t is drawn from as $\bar{\sigma}_\upsilon^2 \equiv Var(\upsilon_t| - \mu_{t+1} < \upsilon_t < 1 - \mu_{t+1})$.[12] Henceforth, we refer to $\bar{\sigma}_\upsilon^2$ as the variance of the signal and $p_\upsilon \equiv 1/\bar{\sigma}_\upsilon^2$ as its precision. Neither of them are observable to agents, who instead believe the signal precision to be $\tilde{p}_\upsilon$. Agents are said to be overconfident whenever they overestimate the accuracy of the new information they receive, i.e., whenever $\tilde{p}_\upsilon > p_\upsilon$.[13]

The model reduces to a perfect information (PI) framework if the precision of the signal p_υ goes to infinity or, equivalently, if $\bar{\sigma}_\upsilon^2 = 0$. In this case, the signal Z_t perfectly reveals μ_{t+1}.[14] On the other hand, whenever $p_\upsilon < +\infty$, agents cannot observe μ_{t+1} and the model exhibits imperfect information (II). Let us express agents' perceived signal precision as $\tilde{p}_\upsilon \equiv \kappa p_\upsilon$, where $\kappa \in [0, +\infty)$ captures individuals' degree of confidence. If $\kappa = 1$, the economy

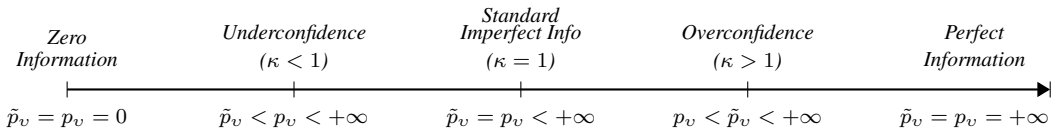
[11] An alternative probability distribution that could have been used is the uniform distribution, which allows values to be drawn from an arbitrary range. Nevertheless, all points within that interval are equally probable. This would render agents' priors uninformative and it would make it difficult for them to extract useful information from the signal they receive. Using a truncated normal distribution, by contrast, allows for points around the mean to be selected with higher probability, all the while ensuring that μ_{t+1} and Z_t lie between 0 and 1.

[12] See Appendix 3.A for more details on the two first moments of the truncated normal distribution.

[13] This definition of overconfidence in terms of signal precision is consistent with the one used by a majority of papers in the behavioural finance literature discussed in Sections 3.1 and 3.2.

[14] Note that the layer of uncertainty stemming from $\mu_{t+1} > 0$ still remains in this setting. The only difference is that μ_{t+1} is observable.

Figure 3.3: Information Sets



is in the standard imperfect information case, with the precision of the signal perceived by agents matching the true precision ($\tilde{p}_\upsilon = p_\upsilon$). If $\kappa > 1$, agents are overconfident (OC) about the precision of the signal, in the sense that they perceive it as being more accurate than it truly is, i.e., $p_\upsilon < \tilde{p}_\upsilon < +\infty$. Conversely, if $\kappa < 1$, they are underconfident (UC), such that $\tilde{p}_\upsilon < p_\upsilon < +\infty$. Figure 3.3 summarises these information sets and their relation to each other.[15] Based on the empirical evidence laid out in Section 3.2, we know that investors have a strong tendency to overestimate the accuracy of their information. Therefore, this analysis focuses on the case where $\kappa > 1$, and we compare it to the standard imperfect information and perfect information settings to isolate the impact of overconfidence. Depending on the signal Z_t they receive, agents can either be overly optimistic or overly pessimistic about the probability of a bubble crash, as we now explain.

In order to form beliefs about μ_{t+1}, agents use their *perceived* signal precision $\tilde{p}_\upsilon$ to compute a weighted average of their signal Z_t and their prior $\bar{\mu}$:

$$\tilde{\mu}_{t+1|t} = \omega_z Z_t + \omega_\mu \bar{\mu} \tag{3.4}$$

where $\omega_z = \frac{\tilde{p}_\upsilon}{\tilde{p}_\upsilon + p_\mu}$ and $\omega_\mu = \frac{p_\mu}{\tilde{p}_\upsilon + p_\mu}$ are the Bayesian weights assigned to the signal and the prior, respectively, with $p_\mu \equiv \frac{1}{\bar{\sigma}_\mu^2}$ and $\bar{\sigma}_\mu^2 \equiv Var(\mu_{t+1}|0 < \mu_{t+1} < 1)$.

The higher the level of confidence κ, the higher the weight ω_z agents put on the signal. However, whether this leads to an increase or a decrease in the perceived likelihood of

[15] Henceforth, we use the superscripts *PI*, *II*, *OC*, and *UC* to refer to the value of variables under these various information sets.

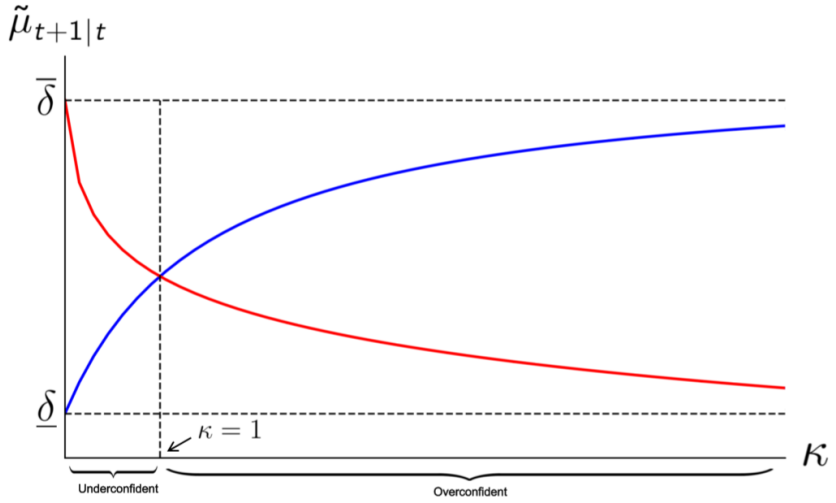
a bubble crash, $\tilde{\mu}_{t+1|t}$, depends on whether they receive a favourable or an unfavourable signal. Namely, when $Z_t < \bar{\mu}$, agents receive a signal suggesting a lower likelihood of a burst than their prior indicated, and we say that they receive good or optimistic news. On the other hand, when $Z_t > \bar{\mu}$, suggesting a higher-than-expected crash probability, we say that they receive bad or pessimistic news. The relation between $\tilde{\mu}_{t+1|t}$ and κ is determined by the following first-order derivative:

$$\frac{d\tilde{\mu}_{t+1|t}}{d\kappa} = \frac{p_\upsilon p_\mu}{(\tilde{p}_\upsilon + p_\mu)^2}(Z_t - \bar{\mu}) \tag{3.5}$$

Hence, $\frac{d\tilde{\mu}_{t+1|t}}{d\kappa} > 0$ iif $Z_t > \bar{\mu}$. In other words, the more confident agents are about an unfavourable signal, the more pessimistic they become about the probability of a crash. Conversely, $\frac{d\tilde{\mu}_{t+1|t}}{d\kappa} < 0$ iif $Z_t < \bar{\mu}$. Figure 3.4 illustrates the relationship between confidence and beliefs. Intuitively, if investors become more overconfident, they perceive the signal as being more precise and, consequently, they choose to put a higher weight ω_z on it. If they received bad news, this means that they are giving more importance to a signal that suggests a high μ_{t+1} draw, resulting in higher beliefs $\tilde{\mu}_{t+1|t}$ that the bubble will burst. Therefore, if agents are overconfident about good (bad) news, they become overly optimistic (pessimistic) in the sense that they underestimate (overestimate) the probability of a crash.

At any point in time, $\tilde{\mu}_{t+1|t}$ is bounded by Z_t and $\bar{\mu}$, which define the range of possible beliefs agents can hold. Specifically, the upper and lower bounds are given by $\overline{\delta} \equiv \max\{Z_t, \bar{\mu}\}$ and $\underline{\delta} \equiv min\{Z_t, \bar{\mu}\}$, respectively. Within that range, what agents actually end up believing is determined by their confidence level κ. In other words, and consistent with the literature reviewed in Section 3.2, investors cannot hold more extreme views than what their information and priors suggest. Note that $\overline{\delta}$ is reached when $\kappa = 0$ if $Z_t < \bar{\mu}$ and when κ goes to

Figure 3.4: Relationship Between Beliefs and Confidence



Note: $\overline{\delta} \equiv \max\{Z_t, \bar{\mu}\}$ *and* $\underline{\delta} \equiv min\{Z_t, \bar{\mu}\}$. *The blue curve captures the case where* $Z_t > \bar{\mu}$, *and the red one where* $Z_t < \bar{\mu}$.

infinity if $Z_t > \bar{\mu}$ – and conversely for $\underline{\delta}$ – as shown in Figure 3.4.

In the context of this model, overconfidence is captured by a representative investor overestimating her ability to process publicly available information whilst trying to estimate the likelihood of future states of nature. Interestingly, this is equivalent to having an asymmetric information model with a continuum of agents taking part in a beauty contest. In such models, originally developed by Keynes (1936), agents do not make decisions based on who they believe the best candidate to be, nor even on who they believe the majority of agents will genuinely pick as their favourite candidate. Instead, they have higher-order beliefs and take into account the fact that other agents will also be trying to guess who the majority is likely to pick. These models have been shown to constitute the best formalisation of sentiment-driven market fluctuations (Angeletos and La'O, 2009; Benhabib et al., 2015). Yet, Huo and Pedroni (2020) have demonstrated that this is exactly equivalent to having a single-agent model where excessive weight is put on a public signal, which is precisely what the representative agent in this model does. In fact, the optimal strategy in a beauty contest is to give additional importance to public rather than private information or priors in an at-

tempt to guess other agents' decisions. Hence, despite its simplicity, this setup is equivalent to having a continuum of investors with asymmetric information and higher-order beliefs.

3.3.1.2 Maximization Problem

Once households have determined their expectations regarding the likelihood of a bubble burst, they choose $c_{1,t}$, s_t, and x_t to maximise their expected lifetime utility:

$$u(c_{1,t}) + \beta\tilde{\mathbb{E}}_t u(c_{2,t+1})$$

subject to budget constraints (3.1) and (3.2). Solving this problem leads to two first-order conditions. The first is the Euler equation, which determines consumption choices across the two periods:

$$u'(c_{1,t}) = \beta(1 + r_{t+1})\tilde{\mathbb{E}}_t u'(c_{2,t+1}) \tag{3.6}$$

The second is an arbitrage condition, which defines asset prices as the present discounted value of future cash flows:

$$Q_t = \frac{\tilde{\mathbb{E}}_t Q_{t+1}}{1 + r_{t+1}} \tag{3.7}$$

An individual who purchases assets when young is able to sell them back at a positive price when old with probability $(1 - \mu_{t+1})$. If the bubble bursts, on the other hand, she is left with worthless assets and her only source of revenue in the second period becomes the risk-free return to her savings. Hence, the consumption of the old in each state is:

$$\begin{aligned} c^{+}_{2,t+1} &= (1 + r_{t+1})s_t + Q^{+}_{t+1}x_t \\ c^{0}_{2,t+1} &= (1 + r_{t+1})s_t \end{aligned} \tag{3.8}$$

with $c^{+}_{2,t+1} > c^{0}_{2,t+1}$.

Agents make their decisions in the first period of their life based on their perceived likelihood of a burst in the next period. We can therefore rewrite the first-order conditions (3.6) and (3.7) when $Q_t > 0$ as:

$$u'(c_{1,t}) = \beta(1 + r_{t+1}) \left[(1 - \tilde{\mu}_{t+1|t}) \tilde{\mathbb{E}}_t u'(c^{+}_{2,t+1}) + \tilde{\mu}_{t+1|t} \tilde{\mathbb{E}}_t u'(c^{0}_{2,t+1}) \right] \tag{3.9}$$

$$\frac{\tilde{\mathbb{E}}_t Q^{+}_{t+1}}{Q_t} = \frac{1}{1 - \tilde{\mu}_{t+1|t}} (1 + r_{t+1}) \tag{3.10}$$

Note that the fractional term on the right-hand side of equation (3.10) is the risk premium. This condition reflects the need for risk-averse agents to be compensated for the (perceived) level of risk they take when purchasing the asset. All else being equal, a higher probability of crash $\tilde{\mu}_{t+1|t}$ will require higher expected returns in order for agents to keep investing in the risky asset. In other words, the bubble will need to grow faster in expectations, so that $\frac{\tilde{\mathbb{E}}_t Q^{+}_{t+1}}{Q_t}$ must increase.

3.3.2 Firms

Output is produced by identical firms behaving in a perfectly competitive manner. Each firm chooses K_t units of capital and N_t units of labor to rent in order to maximize profits:

$$F(K_t, N_t) - r_t K_t - w_t N_t$$

where $F(\cdot)$ is the production function, r_t is the rental rate, and w_t is the wage rate. We assume, without loss of generality, that output is produced using a Cobb-Douglas technology:

$$F(K_t, N_t) = K_t^{\alpha} N_t^{1-\alpha}$$

where $\alpha \in (0, 1)$ is the share of capital in the production process.

The first-order conditions of the firm's problem simply equate the rental and wage rates to the marginal product of capital and labor, respectively:

$$r_t = \alpha k_t^{\alpha-1} \equiv r(k_t) \tag{3.11}$$

$$w_t = (1-\alpha) k_t^{\alpha} \equiv w(k_t) \tag{3.12}$$

where $k_t \equiv \frac{K_t}{N_t}$, $\frac{\partial r(k_t)}{\partial k_t} < 0$ and $\frac{\partial w(k_t)}{\partial k_t} > 0$.

3.4 Equilibrium

In this section, we characterize the equilibrium of the model. We start by describing the market clearing conditions, before showing that the equilibrium dynamics of the model can be captured by a two-equation system. We also demonstrate that the equilibrium conditions, saddle path, and steady states of the economy can be represented in a phase diagram that will become useful for interpreting our results in Section 3.5.

3.4.1 Market Clearing

Asset market clearing simply implies that the total quantity of assets in circulation must be equal to unity:

$$N_t x_t = 1$$

Capital market clearing, on the other hand, requires that all savings be used for the accumulation of physical capital. In other words:

$$K_{t+1} = N_t s_t \tag{3.13}$$

In order to derive analytically tractable results, it is useful at this stage to assume that utility is logarithmic, i.e., $u(\cdot) = log(\cdot)$.[16] This allows to derive an expression for savings from equation (3.9):

$$s_t = \frac{\beta}{1+\beta} w_t - Q_t x_t \tag{3.14}$$

where $\frac{\beta}{1+\beta}$ is the marginal propensity to save. Intuitively, savings are equal to non-consumed income minus the amount of resources allocated to asset purchases. Furthermore, we define the bubble per capita as $b_t \equiv Q_t x_t$, with $b_t \geq 0$.[17]

Expressing equation (3.13) in per capita form and combining it with (3.12) and (3.14) leads to the following expression for capital accumulation per capita:

$$k_{t+1} = (1+n)^{-1} \left[\frac{\beta}{1+\beta} w(k_t) - b_t \right] \tag{3.15}$$

Hence, bubbles divert resources away from capital accumulation. In addition, we can re-

[16] All results are generalisable to any utility function that is such that $\frac{\partial s_t}{\partial r_{t+1}} \geq 0$ when $\tilde{\mu}_{t+1|t} = \mu_{t+1} = 0$.

[17] With intrinsically useless assets, $b_t \geq 0$ simply captures the short-selling constraint and is equivalent to $x_t \geq 0$. For more on this, see footnote 8 in Tirole (1985).

express the arbitrage condition in terms of bubbles per capita using (3.10) and (3.11):

$$\frac{\tilde{\mathbb{E}}_t b_{t+1}}{b_t} = \frac{1}{1-\tilde{\mu}_{t+1|t}} \frac{1+r(k_{t+1})}{1+n} \tag{3.16}$$

where $\tilde{\mathbb{E}}_t b_{t+1} \equiv \tilde{\mathbb{E}}_t Q^+_{t+1} x_t$.

To better understand equation (3.16), it is useful to consider the case of deterministic bubbles, i.e., bubbles that will not burst with certainty and are not subject to imperfect information ($\tilde{\mu}_{t+1|t} = \mu_{t+1} = 0$). The risk premium goes to unity and the expression reduces to:

$$\frac{\tilde{\mathbb{E}}_t b_{t+1}}{b_t} = \frac{1+r(k_{t+1})}{1+n}$$

In order for deterministic bubbles per capita to remain constant, the aggregate bubble must grow at the same rate as the economy. The former grows at rate r_{t+1} as per equation (3.7), and the latter at rate n. If the aggregate bubble grows faster (slower) than n, the bubble per capita will grow (fall) over time. Once uncertainty about the future worthiness of bubbles is introduced, the risk premium – which becomes $\frac{1}{1-\tilde{\mu}_{t+1|t}} > 1$ – adds an additional factor in the determination of the rate of growth of bubbles per capita: agents must be compensated for the added risk they face with the expectation of higher returns in the form of higher expected bubble for any given r_{t+1}.

Together, equations (3.15) and (3.16) fully describe the dynamics of the economy. Next, we define the balanced growth path of the economy and show that the model can be visually represented through a phase diagram.

3.4.2 Balanced Growth Path

The economy is said to be at a steady state whenever both capital per capita and bubbles per capita are constant, i.e.,

$$\Delta k_{t+1} = 0 \tag{3.17}$$

$$\Delta b_{t+1} = 0 \tag{3.18}$$

Let us discuss when each of these conditions is satisfied.

First of all, imposing (3.17) on the expression for capital accumulation, (3.15), and solving for b_t allows to obtain a first expression for b_t in terms of k_t:

$$b_t = \frac{\beta}{1+\beta} w(k_t) - (1+n)k_t \equiv L^k(k_t) \tag{3.19}$$

which captures the combination of (b_t, k_t) points such that $\Delta k_{t+1} = 0$. Equation (3.19), which I henceforth refer to as the k locus, suggests a bell-shaped relationship between b_t and k_t, which I illustrate through the black curve in Figure 3.5. At any point below the k locus, too few resources are allocated to asset purchases for any given k_t, and capital increases as a result. The reverse is true for any point above the locus. Whenever the economy is on it, capital remains constant.

Secondly, equation (3.16) implies that $\Delta b_{t+1} = 0$ is satisfied when either $b_{t+1} = b_t = 0$ or $b_{t+1} = b_t > 0$. The former case simply captures the set of points such that the bubble is constant and equal to zero for any given k_t, i.e., $(b_t, k_t) = (0, k_t)\ \forall k_t$. This is represented by a horizontal line going along the x-axis in the (b_t, k_t) space (Figure 3.5). On the other hand, to see when the latter case holds, we impose (3.18) on the arbitrage condition per capita,

(3.16), which yields:

$$1 + r(k_{t+1}) = (1 - \tilde{\mu}_{t+1|t})(1 + n) \tag{3.20}$$

In other words, the bubble is constant and positive whenever the marginal expected return to savings is equal to the marginal expected return to assets. Substituting for $r(k_{t+1})$ in equation (3.20) using (3.11) and (3.15) and solving for b_t gives the set of (b_t, k_t) points such that $b_{t+1} = b_t > 0$:

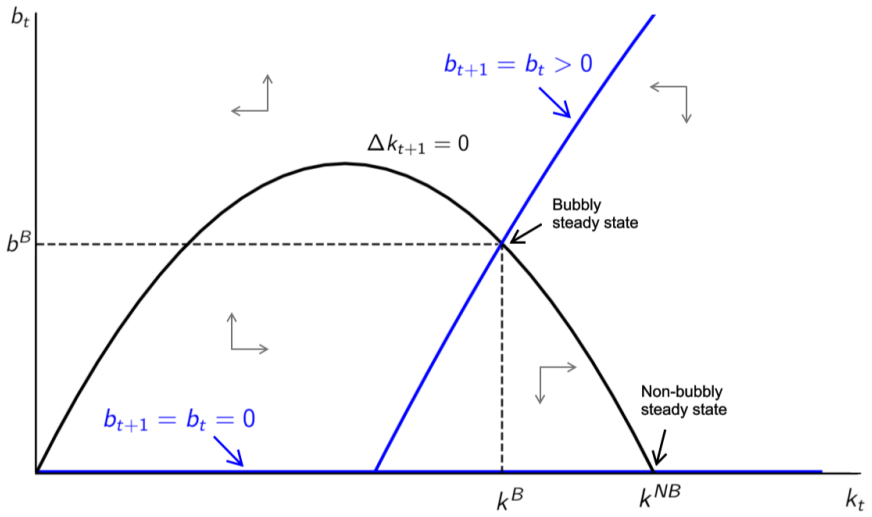
$$b_t = \frac{\beta}{1+\beta} w(k_t) - (1+n) \left[\frac{\alpha}{(1 - \tilde{\mu}_{t+1|t})(1+n) - 1} \right]^{\frac{1}{1-\alpha}} \equiv L^b(k_t) \tag{3.21}$$

This suggests an upward-sloping relationship between bubbles and capital, which is due to the former being a positive function of income and is represented in Figure 3.5. At any point to the left of this curve, the return to assets is higher than the return to savings, and the aggregate bubble grows faster in expectation than the economy. This results in a growing bubble per capita that eventually violated wealth constraints. At any point to the right of it, the aggregate bubble grows slower than the economy, eventually becoming insignificant. Collectively, the horizontal (non-bubbly) line and the upward-sloping (bubbly) curve can be referred to as the b locus.

The economy is at a steady state whenever the k locus ($\Delta k_{t+1} = 0$) crosses the b locus ($\Delta b_{t+1} = 0$), which happens in two instances in our framework.[18] First of all, when the k locus crosses the upward-sloping b locus, the economy is at the so-called bubbly steady state, which is such that $(b_t, k_t) = (b^B, k^B)\ \forall t$, with $b^B > 0$. On the other hand, when it crosses the horizontal b locus, the economy is at the non-bubbly steady state, where $(b_t, k_t) = (0, k^{NB})$ $\forall t$. Henceforth, we use the superscripts *B* and *NB* to refer to variables at the bubbly and non-

[18] There is a third, trivial case in which the economy is at a steady state: at the origin, where $(b_t, k_t) = (0, 0)$.

Figure 3.5: Phase Diagram



bubbly steady states, respectively. We can represent these two equilibria in a phase diagram, as shown in Figure 3.5. Note that it contains four quadrants, each with its own b_t and k_t dynamics, captured by arrows.

Furthermore, there exists a strictly upward-sloping and unique saddle path going through the bubbly steady state. Let b_0 be the size of the bubble in period 0. Then, for any k_0, the economy will converge to the bubbly steady state (b^B, k^B) if and only if $b_0 = b_0^B$, such that (b_0^B, k_0) is on the saddle path. If $b_0 \in [0, b_0^B)$, then the economy slowly converges to the non-bubbly steady state (b^{NB}, k^{NB}). If $b_0 > b_0^B$, the bubble keeps growing until wealth constraints are violated and all resources are depleted. If at any point the bubble bursts, b_t suddenly falls to 0 and the economy converges to the non-bubbly steady state.

It is useful to point out two properties of the bubbly steady state at this stage. Firstly, the interest rate r^B depends on agents' level of confidence κ through $\tilde{\mu}_{t+1|t}$, as is evident from equation (3.20) at the steady state: $1 + r^B = (1 - \tilde{\mu}_{t+1|t})(1 + n)$. Secondly, a well-known property of deterministic bubbles (such that $\tilde{\mu}_{t+1|t} = \mu_{t+1} = 0$) is that they make the economy efficient. Let us define the Golden Rule as the steady state at which the risk-free rate equates the rate of growth of the economy, i.e., $r^{GR} = n$. Then, starting from a non-

bubbly economy with capital overaccumulation ($r^{NB} < n$), the introduction of deterministic bubbles brings the economy to a bubbly steady state matching the Golden Rule. Introducing uncertainty about the future worthiness of bubbles ($\tilde{\mu}_{t+1|t} > 0$), on the other hand, brings the economy into inefficient territory, such that $r^B < n$.

3.5 The Impact of Overconfidence

In this section, the key implications of investors' overconfidence for rational bubbles are derived using the framework described above. Specifically, we want to understand whether such behavioural biases play a role in determining the very existence of bubbles. We also analyse how a change in confidence impacts their size and dynamics based on the type of signal investors receive.

3.5.1 Revisiting the Conditions for Bubble Existence

In order for rational bubbles to exist in equilibrium, it must be worth for agents to divert some resources away from savings to purchase assets. Put differently, the marginal expected return to investing in the bubble must strictly exceed the marginal return agents would get from savings if they saved *all* of their resources. This can be expressed as:

$$r^B > r^{NB} \tag{3.22}$$

which is a necessary and sufficient condition for bubbles to exist in any rational bubble framework with overlapping generations.[19] With deterministic bubbles ($\tilde{\mu}_{t+1|t} = \mu_{t+1} = 0$), this condition is simply satisfied whenever the non-bubbly equilibrium is inefficient – i.e.,

[19] See for instance Samuelson (1958), Tirole (1985), Weil (1987), Galí (2014), and Galí (2017).

whenever $r^{NB} < n$ – since the bubbly steady state will coincide with the Golden Rule. Once uncertainty about the future value of bubbles is introduced, however, the inefficiency of the non-bubbly steady state becomes a necessary but insufficient condition for (3.22) to be satisfied. This is because uncertainty lowers the marginal expected return to assets, hence forcing agents to divert some resources back towards capital accumulation, which lowers r^B below the rate of population growth. The more risky the bubble is perceived as, the lower its marginal expected return, and the less likely it is that (3.22) will hold. In other words, in order for investors to believe it is worth for them to allocate any resources to asset purchases, it is necessary for them to believe that bubbles are not excessively risky. This leads to a first proposition:

> **Proposition 1.** A unique bubbly steady state exists if and only if bubbles are *perceived* not to be too risky, i.e., if and only if the perceived probability that the bubble will not burst is high enough:
>
> $$1 \geq 1 - \tilde{\mu}_{t+1|t} > \frac{1 + r^{NB}}{1 + n} \equiv M \tag{3.23}$$
>
> where M is the minimum probability below which bubbles cannot exist.

Let us rewrite condition (3.23) as $(1 - \tilde{\mu}_{t+1|t})(1 + n) > 1 + r^{NB}$. Hence, stochastic bubbles exist if and only if their expect return exceeds the return to saving all resources, taking into account the probability of burst. It follows that if condition (3.23) holds, (3.22) will as well, which is easy to see when considering equation (3.20) at the bubbly steady state. In fact, suppose $1 \geq M > 1 - \tilde{\mu}_{t+1|t}$. The expected return to saving all resources would then exceed the return to assets, as $(1 - \tilde{\mu}_{t+1|t})(1 + n) < 1 + r^{NB}$. Therefore, bubbles would not be able to exist. Similarly, if $M > 1$, then $r^{NB} > n$ and $M > 1 - \tilde{\mu}_{t+1|t}$, as $(1 - \tilde{\mu}_{t+1|t})$

must be below 1. This would imply that the non-bubbly steady state is Pareto efficient and it would be impossible to introduce bubbles without making some generations worse-off. Consequently, bubbles can only exist when $1 - \tilde{\mu}_{t+1|t} > M$.[20]

Whether or not agents will indeed perceive the bubble as not being too risky, in turn, depends on their degree of overconfidence. In fact, as explained in Section 3.3, agents form beliefs about μ_{t+1} based on their perceived signal precision, $\tilde{p}_\upsilon$, which in turn is determined by confidence κ. If agents receive a favourable signal ($Z_t < \bar{\mu}$), overconfidence makes them underestimate the probability of a bubble burst. If they receive an unfavourable signal ($Z_t > \bar{\mu}$), on the other hand, they overestimate it. As discussed above, Z_t and $\bar{\mu}$ define an interval of potential beliefs about μ_{t+1} at any point in time, such that $\tilde{\mu}_{t+1|t} \in \left[\underline{\delta}, \overline{\delta}\right]$.[21] In turn, κ determines where on that interval $\tilde{\mu}_{t+1|t}$ lies. This implies that if the threshold belief of a bubble burst $(1 - M)$ lies in the interval of currently attainable beliefs, overconfidence becomes a key determinant of bubble existence. In fact, in such cases, there exists a minimum level of confidence for bubbles to exist when agents receive optimistic news, and a maximum level of confidence when they receive pessimistic news. This is summed up in the following proposition:

> **Proposition 2.** Whenever Z_t is such that $(1 - M) \in \left[\underline{\delta}, \overline{\delta}\right]$, a bubbly steady state exists if and only if agents are overconfident enough about good news, or not excessively overconfident about bad news. In other words, for a given $p_\upsilon < +\infty$, a bubbly steady state exists if and only if $\kappa > \bar{\kappa}$ when $\left\{\underline{\delta}, \overline{\delta}\right\} = \{Z_t, \bar{\mu}\}$, and $\kappa < \bar{\kappa}$ when $\left\{\underline{\delta}, \overline{\delta}\right\} = \{\bar{\mu}, Z_t\}$. The threshold $\bar{\kappa} = \bar{\kappa}(r^{NB})$ is such that if

[20] Note that $(1 - \tilde{\mu}_{t+1|t})(1 + n) - 1 > 0$ whenever $(1 - \tilde{\mu}_{t+1|t}) > M$, as $M > \frac{1}{1+n}$, hence satisfying the bubbly b locus equation, (3.21).

[21] Intuitively, and consistent with the empirical evidence reported in Section 3.2, investors' cognitive biases manifest themselves in the form of excessive confidence in their ability to analyse the information they receive. Hence, investors' beliefs are bounded by their new information and their prior assumptions.

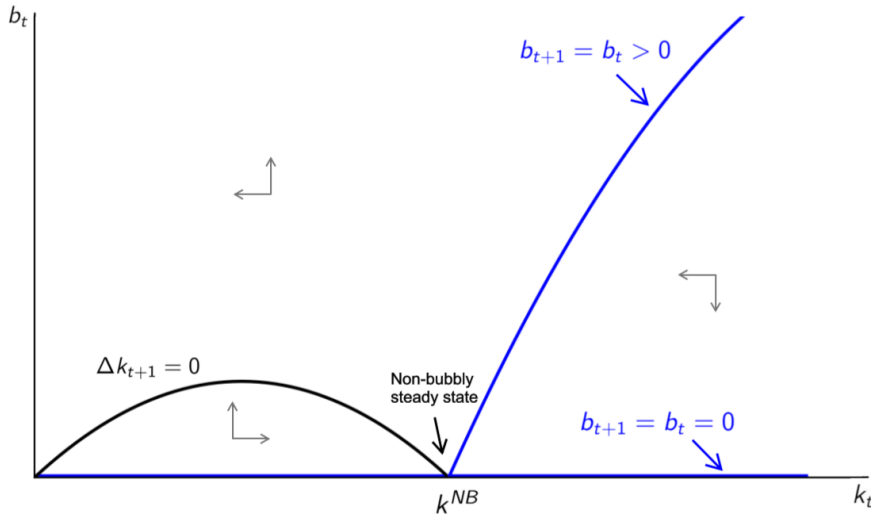
$\kappa = \bar{\kappa}$, then $\tilde{\mu}_{t+1|t} = M$ and $r^B = r^{NB}$, and can be expressed as:

$$\bar{\kappa} = \frac{p_{\bar{\mu}} \left[(1 - M) - \bar{\mu} \right]}{p_{\upsilon} \left[Z_t - (1 - M) \right]} \tag{3.24}$$

Intuitively, overconfidence about good (bad) news increases (lowers) the expected return to assets by lowering (increasing) the perceived risk of a bubble burst. Agents will therefore purchase the asset more (less) easily, hence relaxing (tightening) the condition for existence. As a consequence, condition (3.23) is more likely to be satisfied the more (less) overconfident agents are. In addition, overoptimism can allow bubbles to exist even when the true probability of a bubble burst does not satisfy existence condition (3.23), i.e., even when $1 - \mu_{t+1} < M$, as agents become overly optimistic about the likelihood of a burst. Conversely, bubbles may be kept from existence due to excessive confidence about bad news resulting in overpessimism, even when bubbles are actually not too risky.

Note that when both Z_t and $\bar{\mu}$ are above $(1 - M)$, the set of potential beliefs agents can hold is such that $(1 - \tilde{\mu}_{t+1|t}) < M$, i.e., such that the bubble is perceived as being too risky for all κ and a bubbly equilibrium cannot exist. When they are both below $(1 - M)$, then all attainable beliefs satisfy the condition for existence. Only when the value of Z_t drawn is such that $(1 - M)$ is a possible belief does overconfidence become crucial. This result is logical and consistent with the evidence laid out in Section 3.2. In fact, overconfidence is a deciding factor as to whether an asset is worth purchasing whenever the threshold risk level above which that asset becomes too risky lies in the range of risk levels investors can reasonably believe to be true. In turn, that range is determined by the set of knowledge they hold, which is comprised of their new information and their prior assumptions. In other words, if both the information investors receive and their prior assumptions indicate

Figure 3.6: Economy With $\kappa = \bar{\kappa}$



that bubbles are – or aren't – too risky, no degree of confidence can make them believe otherwise.

If the conditions from Propositions 1 and 2 are satisfied, then the model has two steady states, a bubbly one and a non-bubbly one, both of which are unique and stable, as shown in Figure 3.5. On the other hand, if the conditions are not satisfied, only the non-bubbly steady state will exist. The special case where $\kappa = \bar{\kappa}$ is represented graphically in Figure 3.6.

3.5.2 Comparative Statics

We now turn our attention to analysing the impact of overconfidence on the dynamics and steady states of the model. We start by looking at an unanticipated and permanent change in κ for a given degree of imperfect information, $p_v < +\infty$. As discussed above, an increase in κ results in agents assigning more weight to Z_t. With a favourable signal, this implies that agents underestimate the risk of a bubble burst, while an unfavourable one makes them overestimate it. Hence, in the case of good (bad) news, as agents become more confident, their perceived risk from holding the asset falls (increases), and so does their risk premium. This has several implications for the growth rate of bubbles and the steady state

of the economy.

First of all, a reduction in perceived risk following an increase in confidence for a given favourable signal makes assets more attractive to households at the bubbly steady state, hence raising the demand for them. This leads agents to free up resources by reducing their savings, which decreases capital accumulation. The reverse is true for the case of an unfavourable signal. This can be shown more formally by taking the derivatives of steady state bubbles and capital with respect to confidence:

$$\frac{db^B}{d\kappa} = \left[\frac{\beta}{1+\beta}\alpha(1-\alpha)(k^B)^{\alpha-1} - (1+n)\right]\frac{dk^B}{d\kappa} \tag{3.25}$$

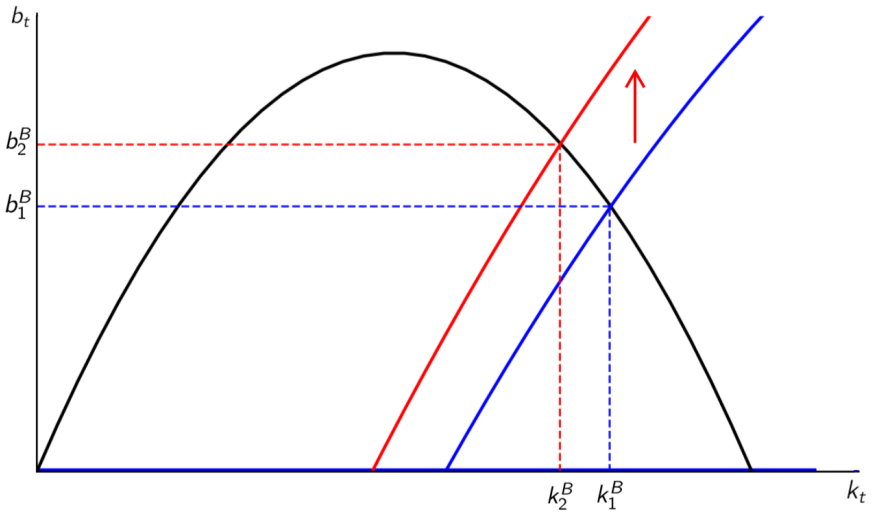
$$\frac{dk^B}{d\kappa} = \frac{1}{1-\alpha}\left[\frac{\alpha}{(1-\tilde{\mu}_{t+1|t})(1+n)-1}\right]^{\frac{\alpha}{1-\alpha}} \frac{\alpha(1+n)}{\left[(1-\tilde{\mu}_{t+1|t})(1+n)-1\right]^2}\frac{d\tilde{\mu}_{t+1|t}}{d\kappa} \tag{3.26}$$

where $\frac{\partial\tilde{\mu}_{t+1|t}}{\partial\kappa}$ is positive if $Z_t < \bar{\mu}$ and negative if $Z_t > \bar{\mu}$, and the term in brackets on the right-hand side of (3.25) is negative. Thus, if agents receive good news, higher confidence results in agents diverting more resources away from capital and toward bubbles at the bubbly steady state, i.e., $\frac{\partial b^B}{\partial\kappa} > 0$ and $\frac{\partial k^B}{\partial\kappa} < 0$. If they receive bad news, the opposite shift takes place.

Secondly, confidence has no impact on the non-bubbly steady state. This is because, at $(b_t, k_t) = (0, k^{NB})$, the aggregate bubble grows slower than the economy, resulting in an insignificantly small bubble per capita that becomes even smaller in each period. Consequently, any shift in it will be insignificant and will not affect capital per capita, i.e., $\frac{dk^{NB}}{d\kappa} = 0$. This result becomes evident when looking at equation (3.15) at the non-bubbly steady state.

Thirdly, it is interesting to point out that the effect of confidence on bubbles along the saddle path differs from its effect on the bubbly steady state. In fact, a stochastic bubble

Figure 3.7: Response of thee Economy to a Shift in κ When $Z_t < \bar{\mu}$



must grow at a faster rate than the rate of interest if it is to be held by risk-averse agents, by a factor equal to the risk premium, as was shown through equation (3.16). Higher confidence about bad news results in a higher risk premium which, ceteris paribus, implies that agents require a higher expected rate of return to the bubble in order to keep purchasing it. This return is equal to the expected rate of growth of the bubble, as it relates the price at which households expect to sell the asset tomorrow to the price they must pay today. Therefore, the more confident agents are about bad news, the faster the bubble will grow toward a smaller steady state value.

These results are summarized in the following proposition:

Proposition 3. If agents receive a favourable (unfavourable) signal, an increase in the degree of confidence κ leads to:

(i) a rise (fall) in the size of the bubble and a fall (rise) in capital accumulation at the bubbly steady state;

(ii) a deceleration (acceleration) of bubble growth along the saddle path;

(iii) no effect at the non-bubbly steady state.

Figure 3.8: Adjustment Path Following a Shift in κ When $Z_t < \bar{\mu}$

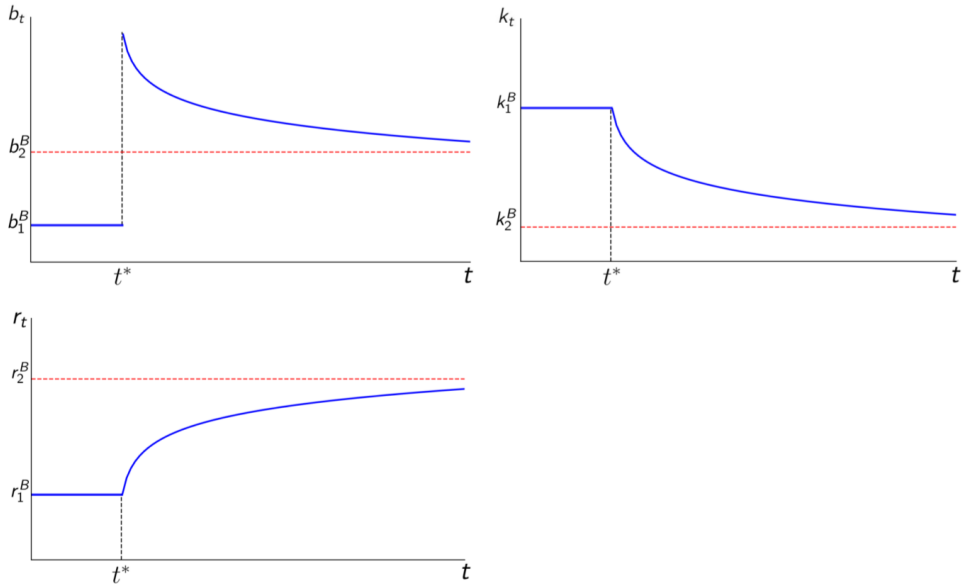
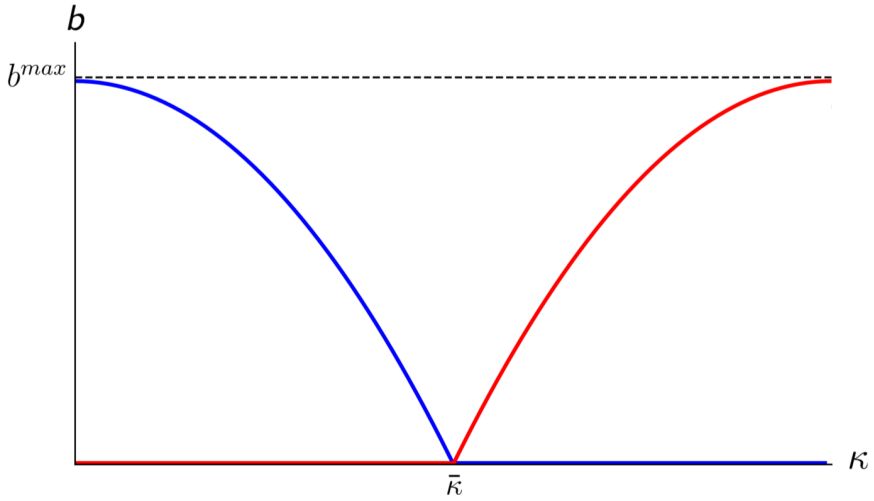


Figure 3.7 illustrates the impact of a permanent increase in κ from period t^* onward with $Z_t < \bar{\mu}$, such that $\kappa = \kappa_1$ for $t \in [0, t^*-1]$ and $\kappa = \kappa_2$ for $t \in [t^*, +\infty)$, with $\kappa_2 > \kappa_1$. Based on the loci defined in (3.19) and (3.21), this results in an upward-shift in the b locus. This is caused by a fall in the risk of holding the asset, which makes agents willing to purchase more assets for any given k_t. On the other hand, it has no impact on the k locus. In other words, $\frac{dL^b(k_t)}{d\kappa} > 0$ and $\frac{dL^k(k_t)}{d\kappa} = 0$ when $Z_t < \bar{\mu}$. As a consequence, b^B increases and k^B falls. Intuitively, when the shift in κ unexpectedly hits in period t^*, capital does not change as it is predetermined. Meanwhile, agents increase their asset purchases by an excessive amount in order to place themselves on the new saddle path and smooth their consumption over time. To do this, they reduce the amount they allocate to savings, resulting in a fall in capital accumulation in the following period. From period $t^* + 1$ onward, they decrease both their asset purchases and their savings until they eventually reach the new steady state in period t^{**}. Conversely, the risk-free interest rate slowly increases before eventually reaching its new steady state value. This adjustment path is shown in Figure 3.8.

Figure 3.9 puts together Propositions 2 and 3 to fully capture the relationship between

Figure 3.9: Confidence and Steady State Bubbles



Note: The red and blue curves capture the cases where $Z_t < \bar{\mu}$ and $Z_t > \bar{\mu}$, respectively. In addition, b^{max} is the steady state bubble achieved when $\tilde{\mu}_{t+1|t} = \underline{\delta}$. This figure holds whenever $(1 - M) \in [\underline{\delta}, \overline{\delta}]$.

bubbles and confidence whenever $(1 - M) \in [\underline{\delta}, \overline{\delta}]$. If agents receive good news, bubbles cannot exist for $\kappa < \bar{\kappa}$. Once confidence crosses the threshold, the size of the bubble starts increasing progressively, until it reaches $\hat{b}$ – the steady state bubble achieved when $\tilde{\mu}_{t+1|t}$ is minimised. This occurs when κ goes to infinity. On the other hand, if $Z_t > \bar{\mu}$, bubbles cannot exist when κ is above the threshold. They increase as κ falls below $\bar{\kappa}$, and they are maximised for $\tilde{\mu}_{t+1|t} = \bar{\mu}$ when $\kappa = 0$.

The channels through which overconfidence affects steady state bubbles become evident at this point. If agents overestimate the accuracy of their information, their perception of the risk they face when purchasing an asset becomes biased. In good times, they underestimate risk, while in bad times they overestimate it, resulting in a downward- and upward-biased risk premium, respectively. As a consequence, they overreact to the news they receive by over- or under-investing. This leads to the mispricing of the asset, with prices being excessively high in good times and excessively low in bad times. Hence, the channels of risk miscalculation, overreaction, and mispricing discussed in Section 3.2 all play a role in the

transmission of the cognitive bias to asset bubbles.

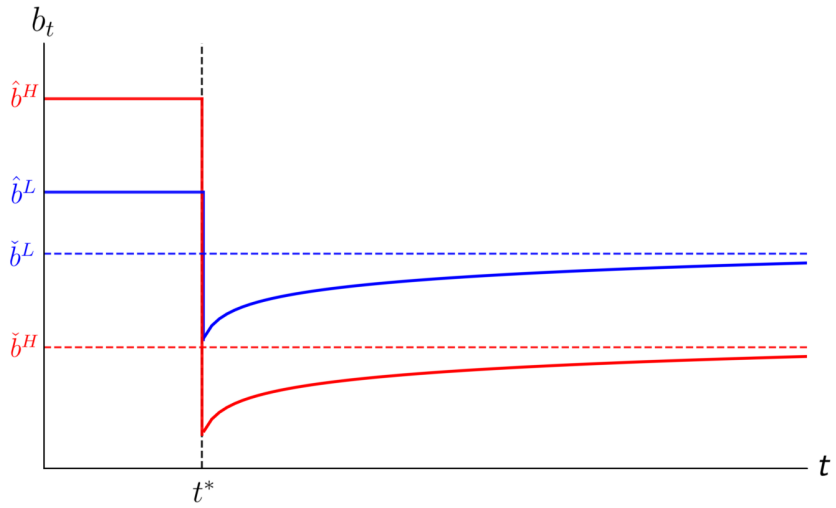
Another useful exercise that allows to better understand the full scope of the impact overconfidence can have on bubbles consists in analysing the effect of a change in the signal Z_t agents receive, for a given degree of overconfidence. Specifically, we look at the role played by a given $\kappa > 1$ when agents go from receiving good news to suddenly receiving bad news in period t^*. In other words, agents receive $Z_t = \hat{Z}_t$ until (t^*-1) and $Z_t = \check{Z}_t$ from t^* onward, where $\hat{Z}_t$ and $\check{Z}_t$ are favourable and unfavourable signals, such that $\hat{Z}_t < \bar{\mu} < \check{Z}_t$.

Based on Proposition 3, we know that prior to the news shift, overconfidence will lead households to overreact to optimistic news, resulting in an excessively large bubble at the steady state compared to an economy with lower overconfidence. In other words, $\hat{b}^H > \hat{b}^L$, where $\hat{b}^H$ and $\hat{b}^L$ capture the size of the bubble at the steady state under good news with, respectively, high and low overconfidence. Following the news shift, on the other hand, overconfidence will make agents excessively pessimistic, resulting in an even smaller bubble, such that $\check{b}^H < \check{b}^L$, where $\check{b}^H$ and $\check{b}^L$ are the steady state values of the bubble following bad news with high and low overconfidence, respectively. Consequently, the size of the shift at t^* is amplified on both ends, resulting in a much heftier drop in the bubble. This is illustrated in Figure 3.10 and summed up in the following corollary:

> **Corollary 1.** Suppose $Z_t = \hat{Z}_t$ for $t \in [0, t^* - 1]$ and $Z_t = \check{Z}_t$ for $t \in [t^*, +\infty)$, where $\hat{Z}_t < \bar{\mu} < \check{Z}_t$. The higher κ is, the larger the resulting drop in the size of the steady state bubble, i.e., $\hat{b}^H - \check{b}^H > \hat{b}^L - \check{b}^L$.

Moreover, overconfidence makes it much more likely that such a shift will result in a sudden and premature collapse of the bubble, i.e., prior to the exogenous bubble burst so far discussed. In fact, if $\check{Z}_t$ is such that $(1 - M) \in [\bar{\mu}, \check{Z}_t]$, then the higher κ is, the more likely it is that the bubble will suddenly fall to 0 at t^* as a result of the news shift. Similarly,

Figure 3.10: News Shift With and Without Overconfidence



Note: The size of the Z_t shift is the same in both cases. Shown here for $\bar{\mu} - \hat{Z}_t = \check{Z}_t - \bar{\mu}$.

let's suppose that news progressively became more and more negative, with Z_t falling each period. The higher overconfidence is, the sooner the economy would reach the non-bubbly steady state.

This model is therefore able to reproduce a number of phenomena and anomalies historically observed on stock markets. In particular, it is able to explain excessively high asset prices during boom times and excessively low ones during periods of market pessimism. It also helps understand the extreme market shifts often observed following important negative news announcements, as investors go from underestimating the risk of a crash to overestimating it. This also coincides with the reasoning behind the beauty-contest equivalent to this model, where participants also overreact to such news shift due to higher-order beliefs, compared to a model where they only exhibit first- or second-order beliefs.

3.6 Information, Dynamic Efficiency, and Welfare

The above results can now be used to develop a more comprehensive understanding of the role imperfect information and overconfidence play in determining the dynamic effi-

ciency of the pre-crash economy, and how the various information sets relate to each other. In addition, we are able to analyse the net impact of overconfidence on agents' welfare, once the costs of a potential bubble burst have been taken into account. We thus determine whether overconfidence, despite it being an unconscious bias, can actually be beneficial to investors.

3.6.1 Overconfidence and Pre-Crash Efficiency

First, recall that an efficient economy is defined as one where the risk-free rate is equal to the rate of growth of the economy, which in this model is equal to n. When this condition is satisfied, agents are able to achieve the highest consumption per unit of labour possible and we say that the economy is at the Golden Rule steady state, with $r^{GR} = n$. As discussed above, bubbles can only exist when the non-bubbly steady state is inefficient, such the economy over-accumulates capital and $r^{NB} < n$. This is because the introduction of an alternative store of value diverts resources away from capital, which is Pareto efficient when $r^{NB} < n$. In the particular case of deterministic bubbles, the bubbly equilibrium coincides with the Golden Rule, such that $r^D = r^{GR} = n$, where r^D is the rate of interest at the bubbly steady state with deterministic bubbles. In addition, as demonstrated by Tirole (1985), the maximum feasible bubble the economy can reach is the one achieved in the deterministic case, $b^D = b^{GR}$.

Incorporating uncertainty in the form of a positive and unobservable probability of future bubble bursts ($\tilde{\mu}_{t+1|t} \neq \mu_{t+1} > 0$ and $\kappa = 1$) adds risk to the purchase of assets. This leads to a shift of resources away from assets compared to the deterministic case, which brings the bubbly equilibrium into inefficient territory. We refer to this as the standard imperfect information (II) case, which is such that $r^{II} < r^D = n$.

If in addition agents are overconfident about the precision of the noisy information they receive ($\kappa > 1$), they will put higher weight on their signal when forming beliefs. In turn, if this signal suggests a lower-than-expected probability of burst, then this cognitive bias makes the pre-crash economy more efficient by getting it closer to the Golden Rule. If on the other hand the signal turns out to be more pessimistic than agents' prior, overconfidence makes the economy less efficient and brings it closer to the non-bubbly steady state. In particular, if $Z_t < \bar{\mu}$, the overconfident steady state lies between the standard imperfect information and the Golden Rule steady states, while if $Z_t > \bar{\mu}$, it lies below both. It also follows that overconfidence is able to make the bubbly equilibrium efficient ($r^{OC} = r^{GR}$) if it is such that $\tilde{\mu}_{t+1|t} = 0$, even when $\mu_{t+1} > 0$.

The above findings can be summarized in the following proposition:

Proposition 4. *(i)* Any uncertainty about the future worthiness of bubbles makes the bubbly equilibrium inefficient.

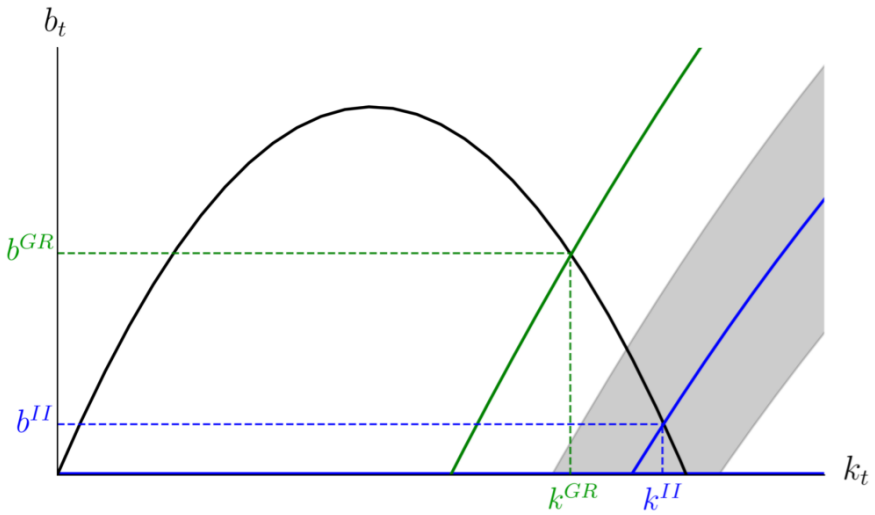
(ii) The maximum feasible bubble is the bubble achieved when $\tilde{\mu}_{t+1|t} = 0$ under imperfect information and when $\mu_{t+1} = 0$ under perfect information.

(iii) Overconfidence makes the pre-crash economy relatively more (less) efficient when agents receive good (bad) news. Specifically, $r^{OC} \in (r^{II}, r^{GR}]$ if $Z_t < \bar{\mu}$ and $r^{OC} < r^{II}$ if $Z_t > \bar{\mu}$.

(iv) Overconfidence can make stochastic bubbles efficient ($r^{OC} = r^{GR}$), even when $\mu_{t+1} > 0$, if it is such that $\tilde{\mu}_{t+1|t} = \underline{\delta}$ when $\underline{\delta} = 0$.

Figure 3.11 illustrates the above results graphically. It shows the relation between the standard imperfect information and overconfident cases, and how they relate to the Golden Rule. The green b locus captures the efficient economy, which can be achieved with either deterministic bubbles or with stochastic bubbles when $\tilde{\mu}_{t+1|t} = 0$. At the steady state

Figure 3.11: Information and Efficiency



Note: The shaded area captures the range of loci achievable under overconfidence. It is shown here for $(1-M) \in [\underline{\delta}, \overline{\delta}]$*, so that the case where* $1 - \tilde{\mu}_{t+1|t} = M$ *is included in the shaded area.*

reached at this locus, the maximum feasible bubble b^{GR} is achieved. The blue locus, on the other hand, shows the standard imperfect information case ($\kappa = 1$), which lies to the right of the Golden Rule, hence making it inefficient. The shaded area above this locus captures the range of potential steady states that can be achieved under overconfidence when $Z_t = \hat{Z}_t$, and is bounded from above by the belief held when $\tilde{\mu}_{t+1|t} = \hat{Z}_t$. Similarly, the range of potential steady states that overconfidence can lead to when $Z_t = \check{Z}_t$ are captured by the shaded area below the standard imperfect information locus, and is bounded from below by $\tilde{\mu}_{t+1|t} = \check{Z}_t$. Note that whenever the threshold belief $(1 - M)$ lies between $\underline{\delta}$ and $\overline{\delta}$, the non-bubbly steady state is included in the shaded area (as shown in Figure 3.11). The overconfident steady state and the non-bubbly steady state then coincide ($r^{OC} = r^{NB}$) whenever $\kappa = \bar{\kappa}$.

3.6.2 Welfare Analysis

The above discussion established that an increase in confidence about good news brings the pre-crash economy closer to the Golden Rule, hence making the bubbly steady state more

efficient by incentivizing agents to lower capital accumulation and purchase more assets. The reverse is true about bad news. Nevertheless, it would be incorrect to conclude that agents are necessarily better off being more over-optimistic or less over-pessimistic based on this. In fact, this is only true for pre-crash generations that are not impacted by the bubble burst. If agents are more overconfident about a favourable signal, this also implies a larger loss if the bubble happens to burst in the second period of their life. Hence, the question of whether a social planner should make agents more or less overconfident remains.

To answer this, we must first define the social planner's problem. One way to do this would be to assume that the social planner knows exactly when the bubble will burst and thus which specific generations will be affected. It would then be easy to see that only one generation would be negatively impacted by overconfidence, and the planner would choose to set κ as high (low) as possible for the young generations at $t \in [0, t^* - 2]$ and as low (high) as possible at $t = t^* - 1$ given good (bad) news. After that, $Q_t = 0$ and is not expected to rise again, so overconfidence would no longer have an impact on post-crash generations. By following this strategy, the social planner would ensure that the welfare of all generations is optimised.

However, a much more informative exercise consists in assuming that the social planner can perfectly observe the true probability of a crash, but does not know when it will actually occur. Overconfidence being an unconscious bias, households make decisions based on their beliefs, as before, which the social planner takes as given. The planner's problem hence consists in choosing the optimal κ for *any* pre-crash young generation in order to maximise their expected welfare based on the true crash probability, taking households' optimal consumption, savings, and investment choices under uncertainty as given.[22] Note

[22]Note that it would make little sense to compare the social planner's problem to that of an agent choosing κ for themselves. This is because overconfidence, by definition, is an unconscious bias originating from im-

that the social planner will now choose the same optimal κ for all pre-crash generations given Z_t, as she cannot know which particular cohort will suffer the consequences of the bubble crash. We can therefore write the problem as:

$$\max_{\kappa} \; W_t = (1-\mu_{t+1})W_t^+ + \mu_{t+1}W_t^0 \tag{3.27}$$

subject to households' optimal decisions $c_{1,t}^*$, s_t^*, and x_t^*, and the inequality constraints $\kappa \leq +\infty$ and $\kappa \geq 1$, where $W_t^0 = log(c_{1,t}) + \beta log(c_{2,t+1}^{0,*})$ and $W_t^+ = log(c_{1,t}^*) + \beta log(c_{2,t+1}^{+,*})$ are the generation's ex-post lifetime welfare in the event that the bubble does and does not collapse, respectively. In addition, $c_{2,t+1}^{+,*}$ and $c_{2,t+1}^{0,*}$ are consumption of the old such that $(b_{t+1}, k_{t+1}) = (b_{t+1}^*, k_{t+1}^*)$ and $(b_{t+1}, k_{t+1}) = (0, k_{t+1}^*)$, respectively. Note that both W_t^+ and W_t^0 depend on agents' beliefs $\tilde{\mu}_{t+1|t}$. We can re-express the welfare criterion as:

$$W_t = log(c_{1,t}^*) + (1-\mu_{t+1})log(c_{2,t+1}^{+,*}) + \mu_{t+1}log(c_{2,t+1}^{0,*}) \tag{3.28}$$

Since agents are risk averse, the social planner faces a trade-off. On the one hand, making households more (less) overconfident about optimistic (pessimistic) news increases welfare by diverting resources towards asset purchases, hence reducing capital overaccumulation and getting the economy closer to the Golden Rule. On the other hand, if the bubble does burst, over-optimistic agents will have purchased more assets, resulting in a larger loss when old.

Solving this optimization problem with inequality constraints leads to corner solutions, which are summarized in Proposition 5:

perfect information. In order to observe κ, one would need perfect information about the true signal precision as well. In addition, the concept of overconfidence as a cognitive bias would no longer make sense if agents could consciously select it.

Proposition 5. Let $\hat{\kappa}^*$ and $\check{\kappa}^*$ be the optimal overconfidence levels under good and bad news, respectively. A social planner seeking to maximise the expected lifetime utility of any pre-crash generation will choose to make agents as overconfident as possible about good news, i.e., set $\hat{\kappa}^*$ as high as possible, and to minimise their overconfidence about bad news, i.e., set $\check{\kappa}^* = 1$.

The social planner should therefore make agents fully rely on their signal when $Z_t < \bar{\mu}$, and put as much weight on it as they would under standard imperfect information when $Z_t > \bar{\mu}$. This ensures that agents are always as optimistic as possible about the likelihood of a bubble crash, i.e., it minimises $\tilde{\mu}_{t+1|t}$. Relating this result to Figure 3.11, we find that agents' expected welfare is maximised at the upper bound of the range of potential equilibria that overconfidence can achieve, i.e., when $\tilde{\mu}_{t+1|t} = \underline{\delta}$. As discussed above, when news is pessimistic, this upper bound is given by the standard imperfect information case. When it is optimistic, it is determined by $\tilde{\mu}_{t+1|t} = Z_t$.

What this result implies is that the potential benefits households would get from a reduction in capital overaccumulation in the event that the bubble remained valuable in the second period of their life outweigh the potential losses incurred if the bubble happened to burst. Going back to the evidence discussed in Section 3.2, the fact that risk-averse agents have a tendency to take on excessive risk during stock market booms due to their excessive confidence actually improves their expected welfare as they value their potential gains more than their potential losses. On the other hand, their inclination toward excessive cautiousness during periods of pessimism appears to hurt them.

3.7 Numerical Analysis

The framework discussed above has produced a number of key results regarding the relationship between stock market bubbles and investors' overconfidence. In this section, we extend our study by calibrating our model and presenting a numerical analysis. We start by motivating our calibration choices and outlining our calculations for parameters not previously covered by the literature. The second subsection then describes the simulation setup and illustrates its functioning by quantifying the impact of overconfidence on bubbles. In the third subsection, we pin down an empirically plausible degree of overconfidence and discuss its implications. We also test the empirical validity of our results by checking whether our calibrated framework produces realistic fluctuations and by conducting a robustness check.

3.7.1 Calibration

Our calibration follows the case of the United States, where a wider range of data is available. Standard parameters were calibrated based on previous studies on the U.S. economy. We therefore set $\{\beta, \alpha\} = \{0.99, 0.33\}$ based on Garriga (2017). Kolasa et al. (2009) calibrate the annual rate of population growth to 1.5%. To ensure the consistency of our calibration across all parameters, we choose a semi-annual frequency – as explained in more detail in sub-section 3.7.3 – resulting in n being set to 0.75%.

On the other hand, parameters relating to our information structure had to be calculated, due to the lack of coverage in the literature. Engle and Ruan (2019) develop a measure of the probability of market crashes for 23 different countries by building a model of systemic risk based on financial firms' under-capitalisation. We use their measure for the U.S. to calculate the mean and standard deviation of the probability of crashes over 6-month intervals,

Table 3.2: Calibrated Parameters

Parameter	Value	Description	Source
β	0.99	Discount factor	Garriga (2017)
α	0.33	Share of capital	Garriga (2017)
n	0.0075	Rate of population growth	Kolasa et al. (2009)
$\bar{\mu}$	0.177	Prior	See section 3.7.1
$\bar{\sigma}_\mu$	0.206	Standard deviation of crash probability	See Section 3.7.1
$\bar{\sigma}_\upsilon$	0.421	Standard deviation of noise	See Section 3.7.1

resulting in a calibration for $\{\bar{\mu}, \bar{\sigma}_\mu\}$ of $\{0.177, 0.206\}$. The volatility of noise $\bar{\sigma}_\upsilon$ has only been estimated, as far as we know, by Forni et al. (2017). However, rather than the volatility of noise itself, what matters here is its relation with $\bar{\sigma}_\mu$, as it will define the relative weights investors put on the signal and the prior in order to form beliefs. We therefore use Forni et al. (2017) to calibrate $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_\upsilon}$ to 0.489.[23] Plugging in the value of $\bar{\sigma}_\mu$ calculated from Engle and Ruan (2019), this implies a value for $\bar{\sigma}_\upsilon$ of 0.421.

The set of parameters used in the simulations presented in the next subsection are summarised in Table 3.2. In subsection 3.7.3, in order to test the empirical validity of our model, we will seek to pin down a specific value for the confidence parameter κ. To do so, we will need to calibrate an additional term, the volatility of agents' beliefs about the probability of crashes. This will be done using Yale Business School's Stock Market Confidence Indices for the United States. We will also calculate a target range of values for the volatility of bubbly returns by combining Shiller (1981) with historical data on S&P 500 prices. To avoid any confusion at this stage, we will elaborate on this when discussing our empirical strategy in subsection 3.7.3.

[23] As a robustness check, we will test our results on the full range of values for this ratio found in the literature.

3.7.2 Overconfidence in a Simulated Economy

So far in this paper, we have assumed that once a signal Z_t is drawn, it remains unchanged unless a news shift occurs. This simplification made sense for a number of reasons. First, it allowed for the analysis of the transition dynamics to a given steady state. In fact, since Z_t determines the equilibrium the economy converges to, each time a new signal is drawn, the steady state shifts. Combined with transition dynamics, this would have made the derivation of comparative statics results about the relationship between overconfidence and bubbles unnecessarily more intricate. Second, since agents only make decisions in the first period of their life and only observe a single signal, this assumption did not fundamentally alter our results.

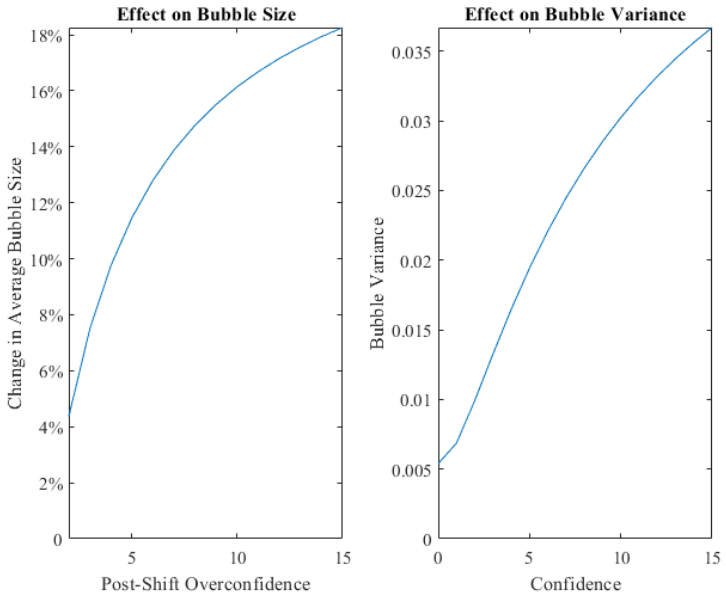
In this section, we are ultimately interested in comparing the volatility produced by our model to its empirical counterpart, as discussed in detail in the next subsection. In order to let bubbles fluctuate, we relax our assumption and let μ_{t+1} and v_t effectively be drawn each period from the truncated normal distributions described in Section 3.3. Furthermore, we assume that the economy starts at the steady state and we abstract from transition dynamics.[24] This way, all fluctuations in our economy originate from μ_{t+1} and v_t shifts.

The purpose of this subsection is to quantify the relationship between overconfidence and bubbles. In the next subsection, we will pin down an empirically plausible confidence level and analyse the economy at that particular equilibrium. We therefore start by reproducing the comparative statics exercises presented in Section 3.5, this time in our simulated economy.

Let us first analyse the impact of κ on the size and volatility of bubbles. To that end, we

[24]This assumption allows to ignore transition dynamics between various steady states, which would create additional volatility not rooted in empirics. In fact, the frequency of signal draws would not allow for such transitions to actually occur. Instead, the economy would consist of the collection of first-period reactions following signal shifts each period, which are often amplified by overshooting, as shown in Figure 3.8.

Figure 3.12: Effect of Confidence on Average Bubble Size and Variance



Note: Panel A plots the percentage increase in the average size of the bubble following a shift in κ from 1 to the values plotted on the x-axis, which range from 2 to 15, in the case of good news. Panel B plots the variance of the bubble for various degrees of confidence.

simulate the economy for T periods, with $\kappa = 1$ (i.e., the standard imperfect information case) for $t = [0, \frac{T}{2} - 1]$, and $\kappa = \kappa^H$ for $t = [\frac{T}{2}, T]$, with $\kappa^H > 1$.[25] We then compare the average size of the bubble in the periods prior to the shift to its corresponding value in the presence of overconfidence. This allows us to quantify the degree to which overconfidence amplifies the bubbly component of asset prices. Panel A in Figure 3.12 illustrates this amplification effect by plotting the percentage change in the average size of the bubble between the first set of periods, $[0, \frac{T}{2} - 1]$, and the second, $[\frac{T}{2}, T]$, for $\kappa^H \in [2, 15]$ and $Z_t \in [0, \bar{\mu}]$.

Consistent with the results laid out in Proposition 3, confidence amplifies the average size of the bubble in the presence of good news, as investors shift their resources toward the risky asset. In fact, the bubbly component of asset prices is 11.47% and 16.15% higher when $\kappa^H = 5$ and $\kappa^H = 10$, relative to the standard imperfect information case.[26] In the

[25] T was set to $10,000$ throughout the analysis to guarantee the representativeness of results.

[26] Recall that we interpret overconfidence as agents perceiving their information as being κ times more precise than it truly is.

case of bad news, overconfidence results in excessively low prices, with similar magnitudes. A second effect of overconfidence is its impact on bubble volatility. Panel B in Figure 3.12 plots the variance of the bubble against the degree of confidence for $Z_t \in [0, 1]$. We find that the higher confidence is, the more volatility is amplified.[27] This is because more confident investors put higher weight on new, shifting information. For instance, letting $\kappa = 5$ amplifies the variance of the bubble nearly threefold, compared to the standard imperfect information case.

A similar exercise can be undertaken for a scenario akin to the one laid out in Corollary 1, whereby we let $Z_t \in [0, \bar{\mu}]$ for $t = [0, \frac{T}{2} - 1]$ and $Z_t \in [\bar{\mu}, 2 \cdot \bar{\mu}]$ for $t = [\frac{T}{2}, T]$.[28] Consistent with Corollary 1, as news shifts from positive to negative for any given κ, investors suddenly go from being over-optimistic to over-pessimistic, resulting in a drop in the size of the bubble. In order to quantify the impact of overconfidence on this shift, Figure 3.13 plots the magnitude of the drop in the average size of the bubble following the news shift as a function of confidence. In the standard imperfect information case, the news shift would result in a fall in the average size of the bubble of 5.68%. This drop is quickly amplified as confidence increases. It nearly doubles if $\kappa = 2$, and rises to 17.55% and 23.31% for $\kappa = 5$ and $\kappa = 10$, respectively. Overconfidence therefore has the potential to drastically magnify the size of market shifts following changes in news sentiment.

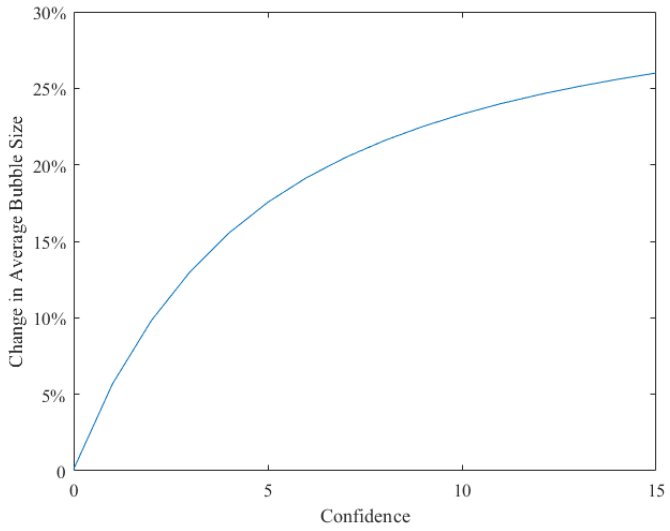
3.7.3 Empirical Strategy and Results

Having quantified the impact of various degrees of overconfidence on stock market bubbles, we are now interested in pinning down an empirically plausible value for κ. Measuring

[27] This result holds regardless of the range Z_t is drawn from.

[28] We restrict the range of possible values for Z_t following the shift to $[\bar{\mu}, 2 \cdot \bar{\mu}]$ to ensure that it is of same size as $[0, \bar{\mu}]$. If we let $Z_t \in [\bar{\mu}, 1]$, the economy would automatically be more volatile in the bad news state, since $\bar{\mu} < 0.5$.

Figure 3.13: Drop in Average Bubble Size Following A News Shock



overconfidence in practice is a rather challenging task due to the fact that it cannot be directly observed. In fact, while absolute confidence can easily be quantified – for instance, by asking survey respondents to rate their level of certainty about future market outcomes – overconfidence is a relative term, as it measures agents' level of confidence relative to an unobservable scenario where they are neither over- nor under-confident. Trying to measure it directly is therefore likely to lead to a very imprecise estimation of its true value, which would be problematic in the context of the present exercise. Fortunately, our framework offers an alternative approach. Let us express equation (3.4) in variance terms, and solve for κ. This yields the following expression:

$$\kappa = \frac{p_\mu}{p_\upsilon \left(\sqrt{\frac{var(Z_t)}{var(\tilde{\mu}_{t+1|t})}} - 1 \right)} \tag{3.29}$$

where $var(Z_t) = \bar{\sigma}_\mu^2 + \bar{\sigma}_\upsilon^2$. It is therefore possible to obtain an empirically-based measure of overconfidence by calibrating $\bar{\sigma}_\mu$, $\bar{\sigma}_\upsilon$, and $var(\tilde{\mu}_{t+1|t})$ – the variance of the perceived probability of a market crash. The first two have already been calibrated. A measure of the

latter can be obtained from Yale Business School's U.S. Crash Confidence Index from the United States Stock Market Confidence Indices database. From 1989 to this day, this survey has asked respondents – both individual investors and wealthy individual investors – to rate their perceived likelihood of a catastrophic market crash – similar to that of October 28, 1929 - to occur in the following 6 months. Since such a forward-looking measure would be hard to re-express over a different frequency, we have instead calibrated the rest of the model on a semi-annual basis for consistency.[29] Based on this survey, we calculate $\widehat{var(\tilde{\mu}_{t+1|t})} = 0.058$ – where $\widehat{var(\tilde{\mu}_{t+1|t})}$ is the calibrated value of $var(\tilde{\mu}_{t+1|t})$ – averaged across the two types of investors.[30] Therefore, we estimate the empirically-relevant value of κ to be:

$$\kappa^{emp} = \frac{p_\mu}{p_\upsilon \left(\sqrt{\frac{var(Z_t)}{\widehat{var(\tilde{\mu}_{t+1|t})}}} - 1 \right)} \tag{3.30}$$

resulting in $\kappa^{emp} = 4.427$. Notice that this value is larger than 1, seemingly confirming the claim made in the behavioural finance literature that investors are on average overconfident. In fact, this result suggests that investors are more than four times more confident than they should be. Put differently, they believe their information to be nearly four and a half times more precise than it truly is.

Before we can make inferences about the impact of our estimated degree of overconfidence on stock market bubbles, we need to verify that this calibration – and indeed this framework as a whole – produces empirically plausible results. One way to do this is to compare the second moment of the bubbly component of asset prices implied by our model when $\kappa^{emp} = 4.427$ to its empirical counterpart. To ensure consistency of the units of

[29] We have also averaged the measure obtained from Yale Business School over 6-month intervals to remain consistent with our previous calculations.

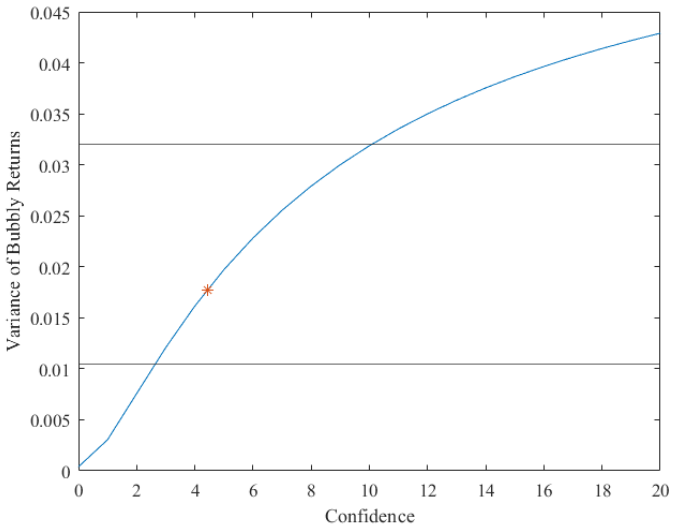
[30] Comparing this to $\bar{\sigma}^2_\mu = 0.042$, we note that the perceived probability of a market crash is more volatile than its (estimated) true counterpart – as would be expected.

measurement between these two values, let us focus on the returns to the risky asset, as opposed to its absolute value. In order to find an empirical estimate of the bubbly part of the volatility of returns to stocks, we must first decompose asset returns into their bubbly and fundamental parts. To do so, we use Shiller (1981), who calculates the share of stock price volatility attributable to fundamentals by looking at the "sample standard deviations of dividends around their long-term exponential growth path". We can then compare it to the total volatility of stock prices to infer the share attributable to bubbles. His findings can be used to derive a ratio between bubble volatility and market volatility, suggesting that bubbles explain between 82% and 92.5% of stock price variations. We apply these ratios to variations in semi-annual stock market returns, calculated using monthly data on S&P 500 real prices taken from the Nasdaq's online database from 1989 to 2021. We find that the volatility of 'bubbly returns' over this period lied between 0.0105 and 0.032. This constitutes the target range our model will need to match.

We are now in a position to test our framework's empirical validity. We do this by determining the variance of returns to the risky asset implied by our model when $Z_t \in [0, 1]$, given the empirical value of κ derived above.[31] We find that for $\kappa^{emp} = 4.427$, the variance is equal to 0.0177, hence falling well within our target range. Figure 3.14 illustrates this by plotting the range of volatility obtained for $\kappa = [0, 20]$. The upper and lower bounds of our target range is delimited by horizontal lines, and the level of variance corresponding to our calibrated degree of confidence is represented by a star. Our model is therefore able to produce realistic fluctuation magnitudes, and it does so for a value of κ above 1. What this seems to imply is that a large portion of the volatility of stock market bubbles in the real

[31] We let μ_{t+1} and v_t be drawn such that $Z_t \in [0, 1]$, given that we do not differentiate between optimistic and pessimistic periods when looking at the empirical data. Note that Z_t still follows a truncated normal distribution, with $\bar{\mu}$, $\bar{\sigma}^2_{\mu}$, and $\bar{\sigma}^2_{v}$ as calibrated above.

Figure 3.14: Variance of Bubbly Returns



Note: The horizontal lines capture the upper and lower bounds of the range of target values. The variance of returns obtained when $\kappa^{emp} = 4.427$ is represented by the red star.

world originates from changes in information about their riskiness. It also suggests that we can make empirically-relevant inferences about the quantitative impact of overconfidence on bubbles.

We therefore return to the results laid out in the previous subsection and interpret them for our calibrated overconfidence level, κ^{emp}. Based on our model and calibration, our findings suggest that the bubbly component of asset prices is on average 10.48% larger due to overconfidence during optimistic times. Similarly, overconfidence amplifies bubble volatility by a factor of 2.58, compared to the standard imperfect information case. Following a news shock, as pointed out above, the average size of the bubble drops by 5.68% when $\kappa = 1$. Under κ^{emp}, this drop increases to 16.40%, i.e., nearly threefold. These results suggest that overconfidence plays a significant role in determining the size and volatility of bubbles, as well as the magnitude of market shifts following shocks. They also support findings from the behavioural finance literature regarding the presence and role of cognitive biases in financial markets, and confirm for the first time that these effects are corroborated

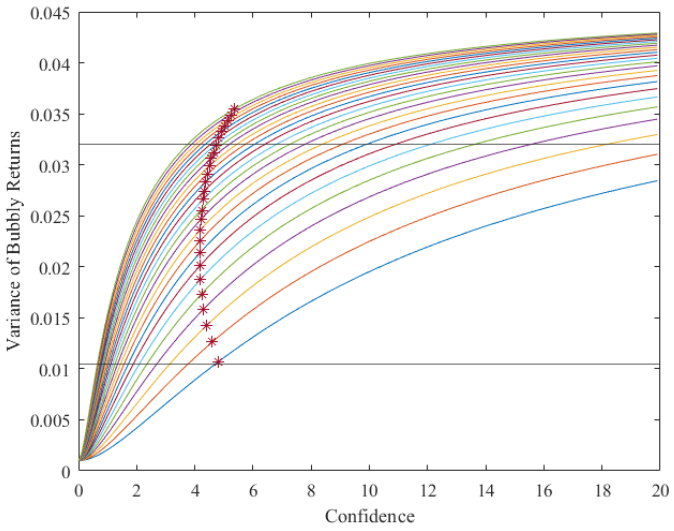
in a general equilibrium framework.

Recall that $\bar{\sigma}_\upsilon$ was chosen based on the estimation of $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_\upsilon}$ produced by Forni et al. (2017). We chose this calibration due to their paper representing the only attempt – as far as we know – at empirically estimating the ratio of the standard deviations of the components of a noisy signal. However, agents in their model actually receive a signal about the fundamental value of assets, not the probability of a market crash. As a way to verify the validity of our results, we perform a robustness check by testing our model on the full range of calibrated values of this ratio that we were able to find in the literature.[32] The lowest such value was calibrated by Daniel et al. (1998), which set $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_\upsilon}$ to 0.365. The highest was found in Peng (2018), with $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_\upsilon} = 1.118$. Letting $\bar{\sigma}_\mu = 0.206$, as before, we obtain a vast set of possible $\bar{\sigma}_\upsilon$ values, lying between 0.184 and 0.564. Figure 3.15 plots the relationship between the variance of returns and confidence for this entire range of $\bar{\sigma}_\upsilon$ values, as well as the empirically-relevant point at which $\kappa = \kappa^{emp}$ in each of these scenarios, denoted by a star. Each curve captures the relationship for a different $\bar{\sigma}_\upsilon$, with the closest to (furthest from) the x-axis corresponding to the highest (lowest) value. This is because the higher $\bar{\sigma}_\upsilon$ is, the less weight investors put on their signal, hence dampening fluctuations in the economy.

We note that κ remains above 1 across the entire range, with a minimum value of 4.165, hence providing further support for our finding that investors are on average overconfident. We also find that 86.52% of the empirically-relevant points derived for this $\bar{\sigma}_\upsilon$ interval produce a variance lying within the target range. In fact, only when $\bar{\sigma}_\upsilon \in [0.184, 0.203]$ does volatility exceed its target upper bound of 0.032. Recall that the calibration of $\frac{\bar{\sigma}_\mu}{\bar{\sigma}_\upsilon}$ derived from Forni et al. (2017) resulted in $\bar{\sigma}_\upsilon = 0.421$. Given that this is the only calibration in the

[32] In all of these papers, investors received a signal about the fundamental value of a risky asset. As far as we know, no study has previously calibrated or estimated the parameters of a noisy signal about the probability of a market crash.

Figure 3.15: Robustness Check



Note: The curves closest to and furthest from the x-axis correspond to $\bar{\sigma}_\upsilon = 0.564$ and $\bar{\sigma}_\upsilon = 0.184$, respectively. Thirty curves are plotted, each capturing a different value of $\bar{\sigma}_\upsilon$ in an equally-spaced manner.

entire interval that is empirically estimated, and that we would need noise to be less than half as high as this to produce magnitudes outside of the empirical target range, we feel confident that our results on the validity of the framework would remain unchanged for any empirically reasonable calibration of $\bar{\sigma}_\upsilon$.

That being said, future research should further test the robustness of our quantitative findings to the inclusion of Real Business Cycle or New Keynesian model features. This notwithstanding, the analysis above provides a solid starting point to quantify the impact of overconfidence on stock market bubbles.

3.8 Concluding Remarks

Motivated by the empirical evidence on the role of overconfidence in financial markets, we develop a stochastic bubble framework with imperfect information where investors overestimate the precision of the information they receive. We use this model to shed light on

the various ways in which overconfidence impacts stock market bubbles, and the channels through which this happens.

We find that agents are only willing to purchase an asset if they deem it not to be too risky, i.e., if their belief about the probability that it may suddenly lose its value is above a certain threshold. Whenever this threshold belief lies in the range of opinions that is reasonable for agents to hold based on their information and prior assumptions, overconfidence becomes a key determinant of bubble existence. In particular, investors must be overconfident enough about good news and not excessively overconfident about bad news.

We also find that when agents receive optimistic news, overconfidence makes them underestimate the risk they are taking, and leads them to misprice the asset and over-invest in it. By contrast, overconfident investors' excessive pessimism after receiving bad news results in excessively low prices. This result rationalises the seemingly irrational behaviour empirically observed on stock markets during boom times and recessionary periods.

In addition, overconfidence makes the pre-crash economy more efficient following positive news, while excessive pessimism exacerbates the problem of overaccumulation. Despite the potential risks associated with holding the asset, agents are better off maximising their overconfidence about good news and minimising it after receiving an unfavourable signal, as the potential benefits of greater efficiency outweigh the (large) costs of a potential crash.

Finally, we conduct a numerical analysis to quantify our results and verify the empirical validity of our model. Our calibration suggests that agents are indeed overconfident and believe their information to be more than four times more precise than it truly is. As a result, bubbles are on average 10.48% larger during optimistic times, and 2.58 times more volatile over the full sample. Overconfidence can also lead to much larger market swings following news shifts. We also find that this calibration is able to produce bubble volatility

of similar magnitude to what can be estimated using empirical data.

The novelty of this model lies in its ability to analyse the impact of overconfidence on bubbles in a general equilibrium model for the first time. The importance of this contribution is underscored by the empirical evidence on overconfidence constituting one of the biggest contributors to the size and volatility of stock market bubbles (Jlassi et al., 2014). In addition, based on recent findings by Huo and Pedroni (2020), the information structure we develop is equivalent to having an asymmetric information setup with a continuum of investors exhibiting higher-order beliefs. Therefore, the model we propose offers an easily extensible yet comprehensive theoretical framework that can serve as the basis for future analysis on the role of overconfidence. In fact, it allows to draw generalisable propositions on overconfidence and bubbles while isolating the macroeconomic channels through which these effects arise.

In particular, future research on the topic should try to understand the impact of this behavioural bias on the conduct of monetary policy and the extent to which central banks should factor it in their decisions. Such pursuit should not limit itself to the more traditional instruments available to central banks. In fact, this could have particularly strong implications for central banks' communication and forward guidance strategies, which constitute strong signals that policy-makers can send to market participants. The above framework lays some of the foundations from which this research can expand.

References

Alpert, M. and Raiffa, H. (2013), A progress report on the training of probability assessors, *in* D. Kahneman, P. Slovic and A. Tversky, eds, 'Judgement Under Uncertainty: Heuristics and Biases', Cambridge University Press, chapter 21.

Angeletos, G.-M. and La'O, J. (2009), 'Noisy Business Cycles', *National Bureau of Economic Research, Working Paper No.* **14982**.

Asriyan, V., Fornaro, L., Martin, A. and Ventura, J. (2016), 'Monetary Policy for a Bubbly World', *National Bureau of Economic Research No. w22639* .

Barber, B. M. and Odean, T. (2001), 'Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment', *The Quarterly Journal of Economics* **116**(1), 261–292.

Barlevy, G., Gale, D. and Allen, F. (2017), 'On Interest Rate Policy and Asset Bubbles', *Society for Economic Dynamics* .

Ben-David, I., Graham, J. R. and Harvey, C. R. (2007), 'Managerial Overconfidence and Corporate Policies', *National Bureau of Economic Research, Working Paper* **13711**.

Benhabib, J., Wang, P. and Wen, Y. (2015), 'Sentiments and Aggregate Demand Fluctuations', *Econometrica* **83**(2), 549–585.

Benos, A. V. (1998), 'Aggressiveness and survival of overconfident traders', *Journal of Financial Markets* **1**(3-4), 353–383.

Bernardo, A. E. and Welch, I. (2004), 'On the Evolution of Overconfidence and Entrepreneurs', *Journal of Economics & Management Strategy* **10**(3), 301–330.

Biais, B., Hilton, D., Mazurier, K. and Pouget, S. (2005), 'Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market', *The Review of Economic Studies* .

Binswanger, M. (2005), 'Bubbles in Stochastic Economies: Can They Cure Overaccumulation of Capital?', *Journal of Economics* **84**(2).

Blanchard, O. J. (1979), 'Speculative bubbles, crashes and rational expectations', *Economic Letters* **3**(4), 387–389.

Blanchard, O. J. and Watson, M. W. (1982), 'Bubbles, Rational Expectations and Financial Markets', *National Bureau of Economic Research, Working Paper No.* **0945**.

Bondt, W. F. M. D. and Thaler, R. (1985), 'Does the Stock Market Overreact?', *The Journal of Finance* **40**(3), 793–805.

Brunnermeier, M. K. and Oehmke, M. (2012), 'Bubbles, Financial Crises, and Systemic Risk', *National Bureau of Economic Research, Working Paper* **18398**.

Caballé, J. and Sákovics, J. (2003), 'Speculating against an overconfident market', *Journal of Financial Markets* **6**(2), 199–225.

Cahuc, P. and Challe, E. (2009), 'Produce or Speculate? Asset Bubbles, Occupational Choice and Efficiency', *IZA DP No.* **4630**.

CFI (2022), 'The efficient market hypothesis'.
URL: *https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/efficient-markets-hypothesis/*

Choi, J., Laibson, D. and Metrick, A. (2002), 'How does the Internet Affect Trading? Evidence from Investor Behavior in 401(k) Plans', *Journal of Financial Economics* **64**(3), 397–421.

Chuang, W.-I. and Lee, B. (2006), 'An Empirical Evaluation of the Overconfidence Hypothesis', *Journal of Banking & Finance* **30**(9), 2489–2515.

Cochrane, J. H. (2002), 'Stocks as Money: Convenience Yield and the Tech-Stock Bubble', *National Bureau of Economic Research Working Paper* **8987**.

Contessi, S. and Kerdnunvong, U. (2015), 'Asset Bubbles: Detecting and Measuring Them Are Not Easy Tasks', *Federal Reserve Bank of St Louis* .

Cooper, A. C., Woo, C. Y. and Dunkelberg, W. C. (1988), 'Entrepreneurs' perceived chances for success', *Journal of Business Venturing* **3**(2), 97–108.

Dalbar (2020), '2020 QAIB Report', *Quantitative Analysis of Investor Behavior* .

Daniel, K. and Hirshleifer, D. (2016), 'Overconfident Investors, Predictable Returns, and Excessive Trading', *National Bureau of Economic Research Working Paper* **21945**.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998), 'Investor Psychology and Security Marker Under- and Overreactions', *Journal of Finance* **LIII**(6).

de Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990), 'Positive Feedback Investment Strategies and Destabilizing Rational Speculation', *The Journal of Finance* **45**(2).

Diamond, P. (1965), 'National debt in a neoclassical growth model', *American Economic Review* **55**(5), 1126–1150.

Engle, R. F. and Ruan, T. (2019), 'Measuring the Probability of a Financial Crisis', *Proceeding of the National Academy of Sciences* **116**(37).

Fama, E. F. (1970), 'Efficient Capital Markets: A Review of Theory and Empirical Work', *Journal of Finance* **25**(2), 383–417.

Forni, M., Gambetti, L., Lippi, M. and Sala, L. (2017), 'Noise Bubbles', *American Economic Journal* **9**(4), 122–152.

French, K. R. (2008), 'Presidential Address: The Cost of Active Investing', *The Journal of Finance* **63**(4).

Froot, K. A. and Obstfeld, M. (1991), 'Intrinsic Bubbles: The Case of Stock Prices', *The American Economic Review* **81**(5), 1189–1214.

Galí, J. (2014), 'Monetary Policy and Rational Asset Pricing Bubbles', *American Economic Review* **104**(3), 721–752.

Galí, J. (2017), 'Monetary Policy and Bubbles in a New Keynesian Model with Overlapping Generations', *CREI Working Paper* .

Garriga, C. (2017), 'Optimal Fiscal Policy in Overlapping Generations Models', *Public Finance Review* **47**(1), 3–31.

Gervais, S., Heaton, J. B. and Odean, T. (2011), 'Overconfidence, Compensation Contracts, and Capital Budgeting', *The Journal of Finance* **66**(5), 1735–1777.

Gervais, S. and Odean, T. (2001), 'Learning to Be Overconfident', *The Review of Financial Studies* **14**(1), 1–27.

Glaeser, M., Langer, T. and Weber, M. (2012), 'True Overconfidence in Interval Estimates: Evidence Based on a New Measure of Miscalibration', *Journal of Behavioral Decision Making* **26**(5), 405–417.

Grinblatt, M. and Keloharju, M. (2009), 'Sensation Seeking, Overconfidence, and Trading Activity', *The Journal of Finance* **64**(2), 549–578.

Grossman, S. J. (1976), 'On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information', *Journal of Finance* **31**(2), 573–585.

Guerron-Quintana, P. A., Hirano, T. and Jinnai, R. (2019), 'Recurrent Bubbles and Economic Growth', *Crepe Discussion Paper No.* **52**.

Guerron-Quintana, P. A., Hirano, T. and Jinnai, R. (2021), 'Bubbles, Crashes, Ups and Downs in Economic Growth Theory and Evidence', *CIGS Working Paper Series* .

Han, B., Hirshleifer, D. and Walden, J. (2022), 'Social Transmission Bias and Investor Behavior', *Journal of Financial and Quantitative Analysis* **57**(1).

Hashimoto, K.-I. and Im, R. (2019), 'Asset Bubbles, Labour Market Frictions and RD-Based Growth', *Canadian Journal of Economics* **52**(2), 822–846.

Herdegen, M. and Schweizer, M. (2015), 'Strong Bubbles and Strict Local Martingales', *Swiss Finance Institute* **15**(05).

Hirano, T. and Yanagawa, N. (2017), 'Asset Bubbles, Endogenous Growth, and Financial Frictions', *The Review of Economic Studies* **84**(1).

Hogarth, R. M. and Karelaia, N. (2012), 'Entrepreneurial Success and Failure: Confidence and Fallible Judgement', *Organization Science* **23**(6), 1733–1747.

Huo, Z. and Pedroni, M. (2020), 'A Single-Judge Solution to Beauty Contests', *American Economic Review* **110**(2), 526–568.

Ideka, D. (2017), 'Monetary Policy, Inflation and Rational Asset Price Bubbles', *mimeo, Bank of England* .

Jlassi, M., Naoui, K. and Mansour, W. (2014), 'Overconfidence Behavior and Dynamic Market Volatility: Evidence from International Data', *Procedia Economics and Finance* **13**, 128–142.

Kantorovitch, I. (2021), 'Exuberant and Uninformed: How Financial Markets (Mis-)Allocate Capital during Booms', *Job Market Paper, Universitat Popeu Fabra and Barcelona GSE* .

Karpoff, J. M. (1987), 'The Relation Between Price Changes and Trading Volume: A Survey', *The Journal of Financial and Quantitative Analysis* **22**(1), 109–126.

Kelley, E. and Tetlock, P. (2013), 'Why Do Investors Trade?', *Columbia Business School, Working Paper* .

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan.

Kim, Y. H. (2013), 'Self Attribution Bias of the CEO: Evidence from CEO Interviews on CNBC', *Journal of Banking and Finance* **37**(7), 2472–2489.

Kolasa, M., Rubaszek, M. and Skrzypczyński, P. (2009), 'Putting the New Keynesian DSG Model to the Real-Time Forecasting Test', *European Central Bank, Working Paper Series* **1110**.

Kyle, A. S. and Wang, F. A. (1997), 'Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?', *The Journal of Finance* **52**(5).

Lamont, O. A. and Thaler, R. H. (2001), 'Can the Market Add and Substract? Mispricing in Tech Stock Carve-Outs', *National Bureacu of Economic Research Working Paper* **8302**.

Lichtenstein, S., Fischhoff, B. and Phillips, L. D. (2013), Calibration of probabilities: The state of the art to 1980, *in* D. Kahneman, P. Slovic and A. Tversky, eds, 'Judgement Under Uncertainty: Heuristics and Biases', Cambridge University Press, chapter 22.

Malkiel, B. G. (2003), 'The Efficient Market Hypothesis and Its Critics', *Journal of Economic Perspectives* **17**(1), 59–82.

Malmendier, U. and Tate, G. (2005), 'CEO Overconfidence and Corporate Investment', *The Journal of Finance* **60**(6), 2661–2700.

Michailova, J. and Schmidt, U. (2016), 'Overconfidence and Bubbles in Experimental Asset Markets', *Journal of Behavioural Finance* **17**(3), 280–292.

Milgrom, P. and Stockey, N. (1982), 'Information, Trade and Common Knowledge', *Journal of Economic Theory* **26**(1), 17–27.

Montier, J. (2006), 'Behaving Badly', *Dresdner Kleinwort Wasserstein* .

Moore, D. A. and Healy, P. J. (2008), 'The Trouble With Overconfidence', *Psychological Review* **115**(2), 502–517.

Odean, T. (2002), 'Volume, Volatility, Price, and Profit When All Traders Are Above Average', *The Journal of Finance* **53**(6), 1887–1934.

Ofek, E. and Richardson, M. (2003), 'DotCom Mania: The Rise and Fall of Internet Stock Prices', *Journal of Finance* **58**.

Pathirawasam, C. (2011), 'The Relationship Between Trading Volume and Stock Returns', *Journal of Competitiveness* **3**.

Peng, Q. (2018), 'Noisy Rational Bubbles', *mimeo* .

Phillips, P. C. B., Shi, S.-P. and Yu, J. (2015), 'Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500', *International Economic Review* **56**(4).

Phillips, P. C. B., Wu, Y. and Yu, J. (2011), 'Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?', *International Economic Review* **52**(1), 201–226.

Rad, F. (2018), 'Detecting and Measuring Financial Market Bubbles', *The University of Manitoba* .

Samuelson, P. A. (1958), 'An Exact Consumption-loan Model of Interest with or without the Social Contrivance of Money', *The Journal of Political Economy* **66**(6), 467–482.

Scheinkman, J. A. and Xiong, W. (2003), 'Overconfidence and Speculative Bubbles', *Journal of Political Economy* **111**(6).

Shiller, R. J. (1981), 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?', *The American Economic Review* **71**(3), 421–436.

Shiller, R. J. (2003), 'Irrational Exuberance', *Princeton, N.J.* .

Smith, V. L., Suchanek, G. L. and Williams, A. W. (1988), 'Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets', *Econometrica* **56**(5), 1119–1151.

Statman, M., Thorley, S. and Vorkink, K. (2006), 'Investor Overconfidence and Trading Volume', *The Review of Financial Studies* **19**(4), 1531–1565.

Stevenson, O. (1981), 'Are We All Less Risky and More Skillful than our Fellow Drivers?', *Acta Psychologica* **47**(2), 143–148.

Stiglitz, J. E. (2003), *The Roaring Nineties: A New History of the World's Most Prosperous Decade*, New York: W.W. Norton.

Teulings, C. N. (2016), 'Secular Stagnation, Rational Bubbles, and Fiscal Policy', *Tinbergen Institute Discussion Paper* **16-081**.

Tirole, J. (1982), 'On the possibility of speculation under rational expectations', *Econometrica* **50**(2), 1163–1181.

Tirole, J. (1985), 'Asset Bubbles and Overlapping Generations', *Econometrica* **53**(6), 1071–1100.

van Gasteren, L. (2016), 'The Effect f Overconfidence on Stock Market Bubbles, Velocity and Volatility', *Radbound University Nijmegen* .

Wang, X., Zheng, L., Li, L., Zheng, Y., Sun, P., Zhou, F. A. and Guo, X. (2017), 'Immune to Situation: The Self-Serving Bias in Unambiguous Contexts', *Frontier in Psychology* **822**(8).

Weil, P. (1987), 'Confidence and the Real Value of Money in an Overlapping Generations Economy', *The Quarterly Journal of Economics* **102**(1), 1–22.

Appendix 3.A Truncated Normal Distribution

Let X follow a truncated normal distribution, such that it is drawn within the range (a, b) of a normal distribution with mean μ and variance σ^2, with $-\infty < a < b < +\infty$ Then, the first two moments of the doubly-truncated normal distribution are given by:

$$\mathbb{E}(X|a < X < b) = \mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}$$
$$Var(X|a < X < b) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}\right)^2\right]$$

where $\alpha = \frac{a-\mu}{\sigma}$, $\beta = \frac{b-\mu}{\sigma}$, $\phi(\cdot)$ is the Probability Density Function (PDF) of the standard normal distribution, and $\Phi(\cdot)$ is the Cumulative Distribution Function (CDF) of the standard normal distribution.

Appendix 3.B Proofs

Proof of Proposition 1. As pointed out in Section 3.5, the condition for bubble existence in any rational bubble framework with overlapping generations is that it is profitable for agents to divert some resources away from assets. In other words, the return to assets must be greater than the return to saving all resources:

$$r^B > r^{NB}$$

which corresponds to condition (3.22). Using equation (3.20) at the steady state, we can express r^B as:

$$1 + r^B = (1 - \tilde{\mu}_{t+1|t})(1 + n) \tag{3.31}$$

Hence, the return to savings at the bubbly steady state, which in equilibrium must be equal to the return to assets, is a function of agents' beliefs. Plugging (3.31) into (3.22) and rearranging it gives:

$$1 - \tilde{\mu}_{t+1|t} > \frac{1 + r^{NB}}{1 + n} \tag{3.32}$$

Adding to this the fact that, by definition, $(1 - \tilde{\mu}_{t+1|t}) \in (0, 1)$ allows to obtain condition (3.23). □

Proof of Proposition 2. Recall that $(1 - M)$ is the threshold probability of burst above which agents will deem the asset as being too risky to be purchased. In addition, recall that $\tilde{\mu}_{t+1|t}$ is bounded at any time by Z_t and $\bar{\mu}$.

Therefore, if both Z_t and $\bar{\mu}$ are above $(1 - M)$, all achievable values of $\tilde{\mu}_{t+1|t}$ will be above $(1 - M)$, so that assets will be perceived as too risky for any value of κ. On the other hand, if both Z_t and $\bar{\mu}$ are below $(1 - M)$, $\tilde{\mu}_{t+1|t}$ will lie below $(1 - M)$ for any value of κ, so that a bubbly equilibrium always exists.

Whenever the value of the signal drawn is such that $(1 - M) \in [Z_t, \bar{\mu}]$ or $(1 - M) \in [\bar{\mu}, Z_t]$, the threshold belief for bubble existence lies in the range of possible. In this case, whether a bubbly equilibrium exists will depend on the value of κ. In particular, there exists a threshold confidence level $\bar{\kappa}$ such that $\tilde{\mu}_{t+1|t} = (1 - M)$ when $\kappa = \bar{\kappa}$. In other words:

$$1 - M = \frac{\bar{\kappa} p_{\upsilon}}{\bar{\kappa} p_{\upsilon} + p_{\bar{\mu}}} Z_t + \frac{p_{\bar{\mu}}}{\bar{\kappa} p_{\upsilon} + p_{\bar{\mu}}} \bar{\mu} \tag{3.33}$$

which determines $\bar{\kappa}$ given Z_t. Solving for $\bar{\kappa}$ leads to condition (3.24). □

Proof of Proposition 3. *(i)* To obtain the value of the bubble at the bubbly steady state, let

us impose condition (3.18) on equation (3.19), which gives us:

$$b^B = \frac{\beta}{1+\beta}(1-\alpha)(k^B)^\alpha - (1+n)k^B \tag{3.34}$$

This captures the point on the k locus where the bubbly steady state is located. Imposing condition (3.17) on (3.21) yields:

$$b^B = \frac{\beta}{1+\beta}(1-\alpha)(k^B)^\alpha - (1+n)\frac{\alpha}{(1-\tilde{\mu}_{t+1|t})(1+n)-1}^{\frac{1}{1-\alpha}} \tag{3.35}$$

which corresponds to the point on the b locus where the bubbly steady state is located. Equating these two conditions and solving for k^B allows us to obtain capital per capita at the bubbly steady state:

$$k^B = \left[\frac{\alpha}{(1-\tilde{\mu}_{t+1|t})(1+n)-1}\right]^{\frac{1}{1-\alpha}} \tag{3.36}$$

Taking the derivative of b^B and k^B with respect to κ leads to equations (3.25) and (3.26).

(ii) This follows from arbitrage equation (3.16).

(iii) At the non-bubbly steady state, the value of the bubble per capita is zero. Hence, capital accumulation equation (3.15) at the steady state becomes:

$$k^{NB} = (1+n)^{-1}\left[\frac{\beta}{1+\beta}w\left(k^{NB}\right)\right] \tag{3.37}$$

which corresponds to steady state capital per capita in the standard Diamond (1965) model. Since at the non-bubbly steady state bubbles do not affect the economy, and since overconfidence enters the model through beliefs about the probability that the bubble may lose its

value, κ has no effect on k^{NB}. □

Proof of Corollary 1. Follows from Proposition 3 *(i)*. □

Proof of Proposition 4. *(i)* From Diamond (1965), we know that any steady state in an overlapping generations economy is efficient when the risk-free interest rate equals the rate of growth of the economy. In other words, the bubbly equilibrium in this model is only efficient when $r^B = n$. Looking at equation (3.31), it is clear that this can only be the case when $\tilde{\mu}_{t+1|t} = 0$. Whenever agents believe the bubble may burst in the following period ($\tilde{\mu}_{t+1|t} > 0$), the bubbly equilibrium exhibits over-accumulation as $r^B < n$.

(ii) We know from Tirole (1985) that the maximum feasible bubble is achieved at the Golden Rule steady state, whenever $r^B = n$. It follows from (i) that this can only be achieved when $\tilde{\mu}_{t+1|t} = 0$.

(iii) Based on Proposition 3 *(i)*, we know that when $Z_t < \bar{\mu}$, $\frac{dk^B}{d\kappa} < 0$ and, by extension, $\frac{dr^B}{d\kappa} > 0$. Hence, an economy with $\kappa > 1$ will have a lower steady state capital level than one with $\kappa = 1$, implying that $r^{OC} > r^{II}$. The former will be closer to the Golden Rule and will therefore be relatively more efficient than the latter, in the sense that it will suffer to a lesser extent from capital overaccumulation. Conversely, when $Z_t > \bar{\mu}$, $\frac{dk^B}{d\kappa} > 0$ and $\frac{dr^B}{d\kappa} < 0$, hence an economy with overconfidence will have a lower interest rate than one with standard imperfect information.

Furthermore, from *(ii)* we know that the maximum feasible bubble is reached when $\tilde{\mu}_{t+1|t} = 0$. Hence, the overconfident economy cannot go any further than the Golden Rule steady state, so that r^{OC} cannot be higher than r^{GR}, but can achieve it if it is such that $\tilde{\mu}_{t+1|t} = 0$.

(iv) Follows from *(ii)* and *(iii)*.

Proof of Proposition 5. Recall that $p_\upsilon \equiv \frac{1}{\bar{\sigma}_\upsilon^2}$ and $\tilde{p}_\upsilon \equiv \frac{1}{\tilde{\bar{\sigma}}_\upsilon^2}$. The perceived signal volatility can therefore be expressed as: $\tilde{\bar{\sigma}}_\upsilon^2 = \chi \bar{\sigma}_\upsilon^2$, where $\chi \equiv \frac{1}{\kappa}$. For ease of notation, let us re-express the social planner's problem in terms of χ:

$$\max_{\chi} W_t = log(c_{1,t}^*) + (1-\mu_{t+1}) log(c_{2,t+1}^{+,*}) + \mu_{t+1} log(c_{2,t+1}^{0,*})$$

$$\text{s.t. } c_{1,t}^* = w(k_t^*) - (1+n)k_{t+1}^* - b_t^*$$

$$c_{2,t+1}^{+,*} = \big[1 + r(k_{t+1}^*)\big](1+n)k_{t+1}^* + \frac{1}{1-\tilde{\mu}_{t+1|t}} \frac{1+r(k_{t+1}^*)}{1+n} b_t^*$$

$$c_{2,t+1}^{0,*} = \big[1 + r(k_{t+1}^*)\big](1+n)k_{t+1}^*$$

$$\chi \leq 1$$

$$\chi \geq 0$$

We obtain the following conditions:

$$\lambda_2 \left[\frac{1}{(1-\tilde{\mu}_{t+1|t})^2} \frac{1+r(k_{t+1}^*)}{1+n} \right] b_t^* \frac{d\tilde{\mu}_{t+1|t}}{d\chi} = \lambda_4 - \lambda_5 \tag{3.38}$$

$$\lambda_1 \Big[c_{1,t}^* - w(k_t^*) + (1+n)k_{t+1}^* + b_t^* \Big] = 0 \tag{3.39}$$

$$\lambda_2 \Big[c_{2,t+1}^{+,*} - \big[1 + r(k_{t+1}^*)\big](1+n)k_{t+1}^* - b_{t+1}^* \Big] = 0 \tag{3.40}$$

$$\lambda_3 \Big[c_{2,t+1}^{0,*} - \big[1 + r(k_{t+1}^*)\big](1+n)k_{t+1}^* \Big] = 0 \tag{3.41}$$

$$\lambda_4(\chi - 1) = 0 \tag{3.42}$$

$$\lambda_5\chi = 0 \tag{3.43}$$

where λ_1, λ_2, λ_3, λ_4, and λ_5 are the non-negative Lagrangian multipliers corresponding to the constraints listed above, in their order of appearance. Given agents' optimising behaviour, we know that λ_1, λ_2, and λ_3 must be positive. We therefore have to determine whether $\lambda_4 \geq 0$ and $\lambda_5 \geq 0$.

It is easy to see that when $Z_t < \bar{\mu}$, $\frac{d\tilde{\mu}_{t+1|t}}{d\chi} < 0$, so that $\chi < 1$ and $\chi = 0$ ($\lambda_4 = 0$ and $\lambda_5 > 0$) are the only solution to the social planner's problem. In other words, the social planner would choose to set κ as high as possible. On the other hand, when $Z_t > \bar{\mu}$, $\frac{d\tilde{\mu}_{t+1|t}}{d\chi} > 0$, so that the optimisation problem is solved for $\chi = 1$ and $\chi > 0$ ($\lambda_4 > 0$ and $\lambda_5 = 0$). This means that the social planner should minimise overconfidence, such that $\kappa^* = 1$. □