Universitat Pompeu Fabra Barcelona

PhD Thesis

Essays on Heterogeneity and Macroeconomic Dynamics

Giacomo Caracciolo
giacomo.caracciolo@upf.edu

Graduate Programme in Economics, Finance and Management
Universitat Pompeu Fabra, Barcelona
Academic Year 2018/2019

Advisors: Prof. Davide Debortoli, Prof. Jordi Galí

Contents

List of Figures

Preface

The Great Recession and the following European sovereign debt crisis have exposed the fragility of the global financial system and the need to reinvigorate the set of tools available to policy-makers to respond to adverse shocks and unfavourable economic conditions. Few years after these episodes, the economic performance appears mixed in the countries of the western world. In Europe, in particular, the recovery has been weak in most cases, the policy interest rates remain very low and the inflation rate is systematically lower than its medium-run target. Such outcome calls for a revaluation of the current design of monetary and fiscal policies and their interaction. This PhD thesis consists of a collection of three separate papers that discuss the problems relative to some demographic and economic phenomena that are currently in place, posing new challenges for policy-makers. The first two chapters focus on the design of public pension systems in a secular stagnation environment, characterized by increasing inequality, financial market segmentation, low interest rates and an ageing population. The third chapter, instead, explores the use of the inflation tax for public finance reasons, when the economy features an underground sector and firms have an incentive to evade taxes.

More in detail, the first chapter examines the role played by social security in presence of limited asset market participation on consumption and wealth inequality. The introduction of a financial friction limiting access to capital markets in a standard medium-scale overlapping generations model alters the standard conclusions on public unfunded pensions and their impact on inequality. The second chapter explores a link between social security reforms, aimed at restoring the financial sustainability of standard PAYG pension schemes following demographic shocks and the equilibrium real interest rates. Different reforms, all sharing the goal to keep the balance budget, may amplify or mitigate the impact of population ageing on the interest rate. Such connection is identified, qualitatively, in a simple 3-generation OLG model and then tested quantitatively in a larger theoretical framework. The results point to an effect that is of a secondary magnitude relative to other drivers of the secular decline in interest rates, however the differential impact of alternative pension reforms is important when we consider different scenarios regarding the evolution of productivity growth projected in the future. The third and final chapter focuses on the apparent correlation between the size of the underground sectors and the inflation targets around the world. A theoretical model featuring an endogenous shadow economy is set up to determine the Ramsey optimal combination of monetary and fiscal policy needed to finance an exogenous level of government expenditure. Heterogenous firms face different incentives to pay taxes and stay in the formal sector or evade the taxes and risk to get caught. In contrast with some results in the literature, the analysis suggests that even in presence of tiny underground economies the social planner finds it optimal to use the inflation tax to raise some seigniorage revenues. Such outcome is reinforced when the case of distortionary taxation is considered.

I wish to thank my advisors, Prof. Davide Debortoli and Prof. Jordi Galí, whose contribution and input to my research have been crucial and inspiring, Prof. Vladimir Asriyan, Prof. Edouard Schaal and Prof. Isaac Baley for their precious advice, Prof. Teresa Garcia-Milà for her kindness and support. Moreover, I would like to thank Carolin Nerlich for giving me the chance to work as a trainee at the European Central Bank and Jacopo Bonchi, who co-authored the second chapter of this thesis and proved to be a valuable friend and colleague.

Chapter 1

Social Security and Inequality in Segmented Financial Markets

Abstract
Limited asset market participation is a well-known stylized fact and a widespread phenomenon even in developed economies. While existing models have already examined the effects of social security and its reforms on welfare and inequality, little attention has been devoted to the role of public pensions in the context of limited asset market participation. I develop a quantitative overlapping generations general equilibrium model where heterogenous agents face a financial friction limiting access to capital markets. I examine how, in presence of the market imperfection, a public pay-as-you-go system affects consumption and wealth inequality and compare the results with a standard model that does not account for limited asset market participation. In a second exercise, I study the implications, in terms of inequality, of an increase in the retirement age in response to a population ageing shock. I find that limited asset participation is important for the analysis of the impact of social security on overall inequality and on inequality within age groups.

1.1 Introduction

Limited asset market participation is a well-known stylized fact and a widespread phenomenon even in developed economies. Although the transaction costs associated with investing in stocks are relatively small and the average market performance, as reported by the standard indices, has been positive, there is a large fraction of the population that is excluded from financial markets. The Survey of Consumer Finances data documents that in 2016 the stock market participation rate for the cohorts aged 41-60 was around 58% in the US, as shown in Figure 1. This number likely represents an upper bound for the overall participation rate, as younger individuals tend to accumulate wealth before entering the asset market and older age groups start dissaving after retirement. Interestingly, the literature on the secular stagnation hypothesis (Caballero & Farhi 2017 [6]) highlights a rise of the equity risk-premium over time and in particular following the outbreak of the global financial crisis in 2007 (Figure 2), while, in the same period, the stock market participation rate has experienced a halt in its positive trend. Such evidence suggests that the inequality between asset market participants and non-participants has increased.

In addition to this first empirical fact, over the last years many countries in the western world have reformed their public pension systems. While a prolonged period of low interest rates and low growth has dramatically reduced the returns of existing schemes, an inexorable population ageing process threatens their financial sustainability. Several economies have adopted counter measures such as increasing the retirement age to rebalance the dependency ratio between a shrinking working age population and an expanding mass of retirees or modifying the way the first pension is calculated for a given amount of contributions (Carone et al. 2016 [8]). Some governments, in the attempt of reducing the liabilities of increasingly more costly and underperforming pension systems, have introduced a multi-pillar structure containing a funded defined-contribution component. Although the literature on pensions has extensively discussed the costs and benefits of social security and its impact on inequality, little attention has been devoted to the role played by public pensions and their distributional consequences in presence of limited asset market participation.

This paper studies how accounting for limited asset market participation alters the standard analysis of social security in regard to inequality. Specifically, I introduce a market imperfection, referred to as *minimum investment requirement*, limiting access to capital markets in an otherwise standard medium-scale overlapping generations model. I find that standard models, by abstracting from limited asset market participation, may understate the welfare and distributional consequences of social security.

In the theoretical framework under analysis, heterogeneous wealth levels at birth and labor efficiencies generate an endogenous wealth distribution and determine the ability of different agents to satisfy the minimum investment required to participate to financial markets. The presence of such entry barrier implies that the design of social security involves a policy trade-off: the public pension scheme offers a means to transfer consumption over time at a return that is higher than the one available to those agents who cannot invest in capital, but lower than the asset market return. Hence social security has a direct impact on consumption and wealth inequality by improving the conditions of some households and making the others worse off.

I conduct two quantitative exercises. First, abstracting from social security, I compare the consumption and wealth stationary distributions of two economies, one characterized by full asset market participation and one where, due to the market imperfection, only a fraction of the population can access the capital market. In this context, I introduce a public pay-as-you-go scheme and examine its impact on overall inequality and inequality by age. While in absence of the minimum investment requirement social security has a limited impact on inequality, in presence of limited asset market participation the public pension scheme amplifies the effect of the friction on the wealth and consumption distributions. In particular, it mitigates the social costs of the financial friction for the poorest households but negatively affects the consumption and wealth accumulation of the middle classes, as it pushes them out of the capital market. Moreover, by crowding out capital, social security favors the wealthiest agents, who would anyways manage to satisfy the minimum investment requirement, as they gain from an increased return to capital. As a result, overall inequality increases, but inequality at the bottom of the wealth and consumption distributions declines. Second, I analyze the inequality implications of an increase in the mandatory retirement age in response to a population ageing shock, in the form of a permanent increase of life expectancy at birth, in a frictionless economy and in one featuring limited asset market participation. In absence of the entry barrier to the capital market, the policy reform simply offsets the effects of population ageing. However, when limited asset market participation is accounted for, an increase in the retirement age reduces inequality even in comparison to the state of the economy before the longevity shock. In particular, the reform produces a redistribution of wealth in favor of poor agents and at the expenses of wealthy households, which is in contrast with the conventional view that an increase in the retirement age is a regressive type of measure.

1.1.1 Related literature

The analysis proposed here is connected with several strands of literature. In particular, the paper relates to the line of research on the long-term determinants of the wealth distribution and consumption inequality over the life-cycle (Benhabib et al. (2011) [4], Benhabib et al. 2015 [3], Storesletten et al. (2004) [35], Huggett et al. (2011) [24], De Nardi & Fang (2014) [10], Gabaix et al. (2016) [18]) and the one on the costs and benefits of social security (Huggett & Ventura (1999) [23], Matsen & Thøgersen (2004) [29], Miles (2000) [31], Krueger & Kubler (2006) [27], Imrohoroglu et al. (2003) [25], Hairault & Langot (2008) [20], McKay (2013) [30]). The conjecture that higher wealth is associated with higher returns is crucial, in my model, to justify the optimality of a public pension system. Fagereng et al. (2016) [14] show that this assumption is a realistic one using a rich dataset compiled from Norway's administrative tax records. Figure 3 reports their findings on the positive correlation between wealth and its returns. Interestingly, this evidence is robust to controlling for the risk associated to the different assets that enter in the households' portfolios. Furthermore, Fagereng et al. (2017) [13] convey that both capital market participation and the share invested in risky equity display a hump-shaped pattern over the lifetime, which is an implication of assuming a minimum investment requirement.

The literature on social security examines the two central questions of fiscal policy, equity and efficiency. In Krueger & Kubler (2006), a pay-as-you-go public transfer can be optimal by promoting intergenerational risk-sharing when it offers an asset whose return is imperfectly correlated with the market returns. Nonetheless, in their analysis, when public pensions are included in a general equilibrium framework, the well-known crowding out effect of capital due to social security outweighs the gains from risk-sharing. Therefore, overall, pensions are not Pareto optimal unless a very high risk-aversion coefficient is considered. A. Imrohoroglu, S. Imrohoroglu & Joines (2003) builds a model where agents, due to time-inconsistent preferences, do not save enough for retirement. Then the government, moved by paternalism, forces households to contribute to the pension system in order to solve the problem of old-age consumption. Although the main focus of this paper is the impact of social security on inequality, the presence of the financial friction implies that a pension system has an effect on welfare. Differently from the contributions mentioned here, this project does not consider the role played by risk or preferences in the design of social security. Instead, it focuses on limited asset market participation and heterogenous returns to wealth as a rationale for having a pay-as-you-go pension system in place. In this sense, the closest paper in the literature is McKay (2013), which recognizes that heterogenous agents have different ability to invest their savings in the financial market. Thus, McKay's model accounts for limited asset market participation and the concentration of wealth at the top of the distribution. The main quantitative exercise examines the impact of partial privatization of the pension system on welfare. The paper concludes that modelling investment decisions as an activity requiring effort does not substantially alter the results of standard models on pensions.

Conversely, in the analysis proposed here, I explore the role of social security in shaping wealth and consumption inequality in presence of limited asset market participation. Huggett & Ventura (1999) and Hairault & Langot (2008) study the distributional implications of social security reforms. The former analyzes the potential impact of the introduction of minimum pension floor in the US system and finds that such type of measure would worsen the conditions of the median ability household, which represents the bulk of the US population. The latter questions the conventional view that pay-as-you-go schemes produce less inequality than funded systems and concludes that the results crucially depend on the strength of the general equilibrium effects induced by pension reforms. Neither of the two, however, takes into account limited asset market participation.

Finally, the dynamic general equilibrium model developed in this work follows the tradition of employing overlapping generations models for studying fiscal policy along the life-cycle pioneered by Auerbach & Kotlikoff (1987) [2]. The paper proceeds as follows: section 2 introduces the minimum investment requirement in a simple 2-period model; section 3 illustrates the key ingredients of the heterogenous agents OLG model adopted in this paper, section 4 presents the results of the quantitative exercises and section 5 concludes.

1.2 Minimum investment requirement

The purpose of this section is to illustrate the main market imperfection studied in this paper at work in a stylized 2-generation OLG model with three types of agents, low, middle and high-skilled. For the sake of simplicity, I will conduct a partial equilibrium analysis where the return on savings and wages are exogenously fixed. Endogenous production will be introduced in the next section, with the prices of labor and capital determined in a general equilibrium framework.

In this simplified world, a generation lives for two periods. In the first period t each agent is born with a specific level of skill $e(i)$ and no wealth. She will supply a unit of labor earning $(1-\tau)w_t e(i)$ and decide what portion of it to consume and what portion to save for the next period in a standard lifetime utility maximization problem. The amount $\tau w_t e(i)$ is paid as a contribution to the public pension system which consists of a pay-as-you-go (PAYG) arrangement. The part of first-period income that is not consumed or paid to social security can be invested in two different financial opportunities, both having an exogenous deterministic return. The first investment opportunity

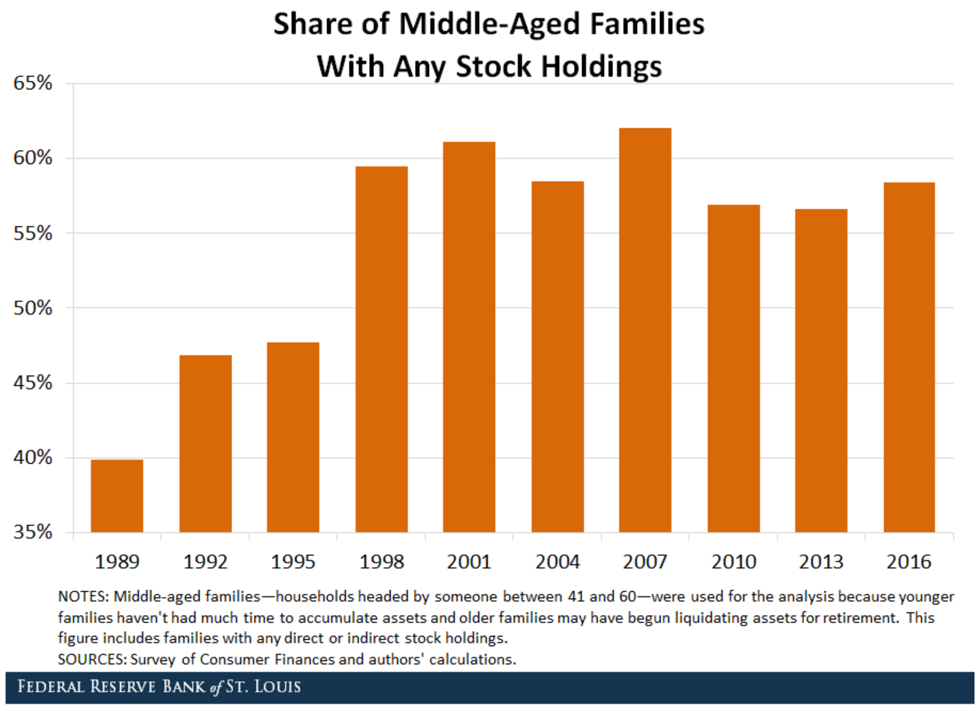


Figure 1.1: Stock market participation rate, from Ravikumar & Karson's "How Has Stock Ownership Trended in the Past Few Decades?"(2018)

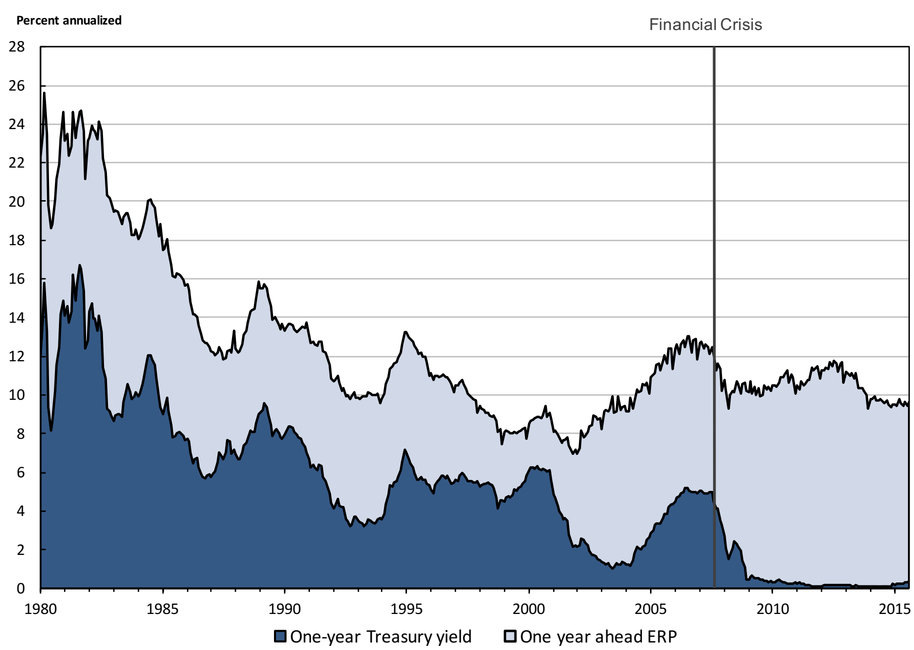


Figure 1.2: Equity Risk Premium over time, from Caballero & Farhi's "The Safety Trap" (2017)

(b) Risky and safe assets

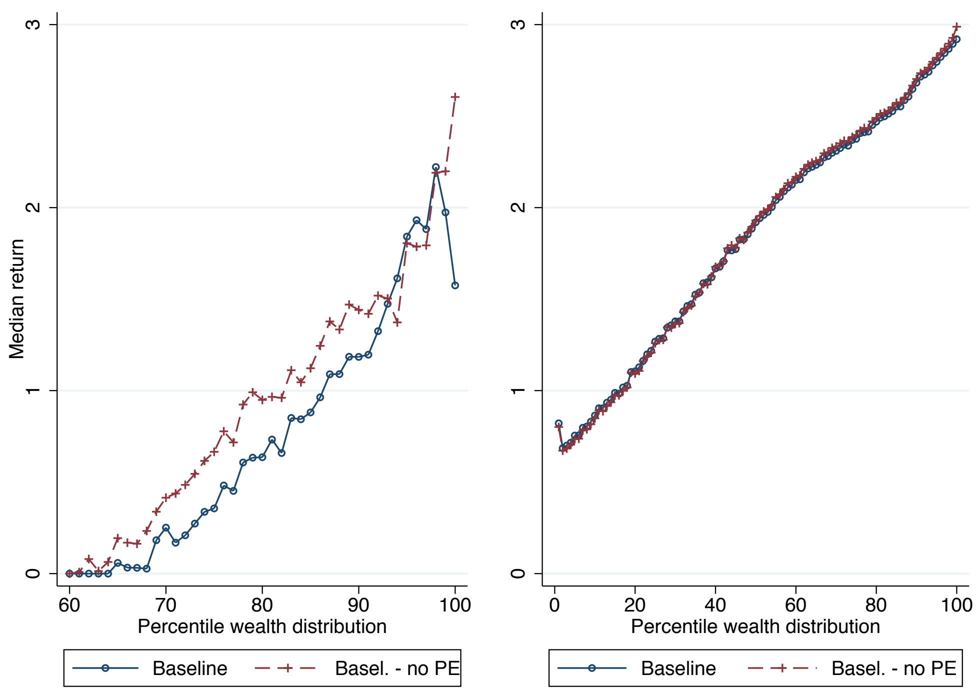


Figure 1.3: Heterogenous returns to wealth, from Fagereng, Guiso, Malacrino & Pistaferri's "Heterogeneity and Persistence in Returns to Wealth" (2016)

is a basic storage technology, with return $r^l = 0$; the second one is an asset with return $r^h > 0$. Furthermore, the latter can be purchased only with a minimum investment $a(i) \geq \tilde{a}$. This minimum investment requirement is a financial friction that hinders the access to the asset market. In the second and final period of her life $t+1$, each agent retires and consumes the return of her investment $(1+r)a(i)$ and her pension benefit. At the same time, a new cohort of agents is born with the same skill profile and size (population growth $n = 0$) as the one just retired. Thanks to an exogenous technological progress, wages grow over time at a constant rate g, so that $w_{t+1} = (1+g)w_t$. The newborn generation pays, through its contributions, the pension entitlements of the retired one. Therefore the pension system budget is:

$$\tau w_t e(i) = \phi(1-\tau)w_{t-1}e(i) \tag{1.1}$$

which states that overall contributions must equal, on a balanced budget, to the total amount of benefits, where ϕ is the replacement ratio. In practice, social security grants to the retirees a fraction ϕ of the pre-tax labor income $(1-\tau)w_t e(i)$ they earned when young. It follows that, once fixed ϕ and substituted $w_{t-1} = \frac{w_t}{1+g}$, the tax rate that keeps the budget balanced is:

$$\tau = \frac{\phi}{1+g+\phi} \tag{1.2}$$

Equation (1.1) reveals an implicit assumption of this stylized framework: the pension of the old retirees is paid by the young workers with the same skill level. Therefore we can interpret the three different skill levels, e^l, e^m and e^h as identifying three different professions, each one featuring a specific pension fund. Additionally, from equation (1.2) we can derive the replacement ratio as a function of the tax rate τ and the growth rate of wages g: $\phi = \frac{\tau}{1-\tau}(1+g)$. Then it can be easily obtained that the return of the pension funds is exactly g.

For the purposes of this section, it is assumed that $r^l < g < r^h$. Thus, social security offers an asset whose return is higher than the storing technology return, but lower than the return of the financial asset. As a consequence, those agents that cannot save enough to satisfy the minimum investment requirement in the asset market, will find it convenient to dispose of a pension fund that offers a return higher than the technology available to them. On the contrary, those agents that are in the position of investing some of their resources at the high return r^h, will suffer a loss if forced to contribute

to the pension fund. Such condition is at the core of the optimality of pension systems in the model proposed in this paper and represents the theoretical justification for studying the impact of social security on inequality.
Formally, the consumer problem is:

$$
\begin{aligned}
\max_{c_t^Y(i), c_{t+1}^O(i), a(i)} \quad & \log(c_t^Y(i)) + \beta \log(c_{t+1}^O(i)) \\
\text{s.t.} \quad & c_t^Y(i) + a(i) = (1-\tau) w_t e(i) \\
& c_{t+1}^O(i) = (1 + r(i)) a(i) + \tau (1+g) w_t e(i) \\
& r(i) = \begin{cases} r^h & \text{if } a(i) \geq \tilde{a} \\ r^l & \text{if } a(i) < \tilde{a} \end{cases} \\
& a(i) \geq 0
\end{aligned}
$$

Where the last inequality stands for a no-borrowing constraint. If we substitute $a(i)$ derived from the first period budget constraint into the second period one we can obtain the lifetime budget constraint:

$$
c_t^Y(i) + \frac{c_{t+1}^O(i)}{1 + r(i)} = (1-\tau) w_t e(i) + \tau w_t e(i) \frac{1+g}{1+r(i)} \tag{1.3}
$$

The right-hand side of equation (1.3) shows that social security can increase (or decrease) the lifetime income of agent i if $g > r(i)$ ($g < r(i)$). That crucially depends on the ability and willingness of the consumer to save enough in the first period in order to access the asset market in the second period, or, in other words, to pass the $\tilde{a}$ threshold. In the specific case in which $g = r(i)$, social security occurs to be neutral. The consideration that the private return to savings $r(i)$ may be different across heterogeneous agents constitutes the main deviation of this analysis from standard models on pensions. In fact, the conventional wisdom is to assume that agents earn the marginal return to capital r^K on their investment. Under such circumstance, social security is welfare improving and thus desirable only if its return, given by the sum of technological progress and population growth rates $g + n$ is higher than the return to capital r^K, i.e. if the economy is dynamically inefficient. In the proposed framework, different agents dispose of different asset returns depending on their wealth due to the presence of a financial friction that poses an entry barrier to the asset market. As a result, social security has a differential effect on heterogenous agents and brings about consequences in terms of inequality and welfare.
The solution of the optimization problem entails applying the standard Kuhn-Tucker

	$a(i) < \tilde{a}$	$a(i) = \tilde{a}$	$a(i) > \tilde{a}$
$c_t(i)^Y$	$\frac{w_t e(i)}{1+\beta}\left(1-\tau+\tau\frac{1+g}{1+r^l}\right)$	$(1-\tau)w_t e(i) - \tilde{a}$	$\frac{w_t e(i)}{1+\beta}\left(1-\tau+\tau\frac{1+g}{1+r^h}\right)$
$c_{t+1}(i)^O$	$\frac{\beta w_t e(i)}{1+\beta}\left((1-\tau)(1+r^l)+\tau(1+g)\right)$	$\tilde{a}(1+r^h)+\tau w_t e(i)(1+g)$	$\frac{\beta w_t e(i)}{1+\beta}\left((1-\tau)(1+r^h)+\tau(1+g)\right)$
$a_t(i)$	$\frac{w_t e(i)}{1+\beta}\left(\beta(1-\tau)-\tau\frac{1+g}{1+r^l}\right)$	$\tilde{a}$	$\frac{w_t e(i)}{1+\beta}\left(\beta(1-\tau)-\tau\frac{1+g}{1+r^h}\right)$

Table 1.1: Three different regions

conditions and identifying three different regions: $a(i) > \tilde{a}$, $a(i) = \tilde{a}$ and $a(i) < \tilde{a}$. Lifetime utility is:

1. $\log\left[\frac{w_t e(i)}{1+\beta}\left(1-\tau+\tau\frac{1+g}{1+r^l}\right)\right]+\beta\log\left[\frac{\beta w_t e(i)}{1+\beta}\left((1-\tau)(1+r^l)+\tau(1+g)\right)\right]$ when $a(i) < \tilde{a}$;
2. $\log\left[(1-\tau)w_t e(i)-\tilde{a}\right]+\beta\log\left[\tilde{a}(1+r^h)+\tau w_t e(i)(1+g)\right]$ when $a(i) = \tilde{a}$;
3. $\log\left[\frac{w_t e(i)}{1+\beta}\left(1-\tau+\tau\frac{1+g}{1+r^h}\right)\right]+\beta\log\left[\frac{\beta w_t e(i)}{1+\beta}\left((1-\tau)(1+r^h)+\tau(1+g)\right)\right]$ when $a(i) > \tilde{a}$.

The consumer, given her skill $e(i)$, the minimum investment requirement $\tilde{a}$ and the tax rate τ, will choose the amount of savings that produces the highest level of lifetime utility among the three just identified. Define with $\underline{e}$ the skill level that, given a specific tax rate τ, makes the lifetime utility in 1. and 2. equal and with $\bar{e}$ the one equalizing the lifetime utility in 2. and 3. Figure (1.4) displays how lifetime utility and savings vary with the level of skill $e(i)$, and the three different regions emerging from the optimization problem, under two scenarios: in black, when there is no social security and in red when $\tau > 0$. For levels of skill under $\underline{e}$ optimal savings are $a(i) < \tilde{a}$, for $\underline{e} < e(i) < \bar{e}$ it is optimal to choose exactly $a(i) = \tilde{a}$ and when $e(i) > \bar{e}$ the minimum investment requirement is not binding and $a(i) > \tilde{a}$. The introduction of the pension system shifts to the right the two threshold levels $\underline{e}$ and $\bar{e}$, as it reduces first period disposable income.

Consider, initially, the case in which there is no pension system, so that $\tau = 0$. In this simplified framework, I will assume that the high-skilled worker has a productivity $e^h > \bar{e}$, the middle-skilled $\underline{e} < e^m < \bar{e}$ and the low-skilled $e^l < \underline{e}$. For each household, the optimal consumption/savings decision depends on the size of the financial friction $\tilde{a}$ relative to first period disposable income $w_t e(i)$. For the high-skilled agent the minimum investment requirement $\tilde{a}$ is small relatively to her first period disposable income $w_t e^h$, implying that her consumption and investment decisions are not affected by its presence. Therefore, as it follows from log utility, she will invest a fraction $\frac{\beta}{1+\beta}$ of her initial endowment $w_t e^h$ entirely in the asset and benefit in period $t+1$ from the high market return r^h.

The middle-skilled household instead chooses to invest exactly $\tilde{a}$. The minimum investment requirement is binding and it is optimal for the household to sacrifice some first period consumption in order to be able to access the asset market. Finally, the low-skilled agent, given her labor efficiency e^l will only employ the low-yield storing technology because sacrificing the amount of first period consumption necessary to access the asset market in the following period is too costly in terms of utility or simply unfeasible due to the borrowing limit. Figure (1.5) summarizes these results. In this otherwise standard intertemporal decision problem, the minimum investment requirement in the asset market is graphically represented by a jump in the budget line, where the change in the slope is due to the different returns offered by the two investment opportunities. The horizontal distance between the point of discontinuity and the first period endowment $w_t e(i)$ measures $\tilde{a}$, the size of the friction.

At this point, let us study the effects of introducing social security. A positive τ reduces each household's first period disposable income and has a heterogenous effect on the lifetime income of different agents. As it emerged from equation (3), those agents whose private return to savings $r(i)$ is larger than g will suffer from a reduction of their lifetime income, while those that cannot access the high-yield asset return r^h will benefit from it. Figure 1.5 displays with the dashed lines how pensions alter the budget constraint and optimal consumption/savings decision of the three different types of agents. First of all, notice that a reduction in the first period disposable income implies that the minimum investment requirement becomes more stringent, i.e. the horizontal distance between the point where the jump in the budget line occurs, and the first period maximum feasible consumption increases. Such change is highlighted by the leftwards shift in the vertical dashed line. The low-skilled agent, who before the introduction of the pension system optimally decided not to save more than the requirement $\tilde{a}$, now can afford increasing both her first and second period consumption thanks to the return of the pension fund. In fact, her budget set expands. Moreover, the fall in the first period disposable income makes saving the amount $\tilde{a}$ not only suboptimal, but unfeasible due to the no-borrowing constraint. Graphically, the black dashed line moves to the left of the $c_t^Y = 0$ vertical axis. In absence of social security the middle-skilled agent chose to save exactly $\tilde{a}$. When $\tau > 0$ she does not find convenient anymore to sacrifice first period consumption in order to access high return in the second period, because of a reduction in her lifetime income. Then she ends up on a lower indifference curve, where she consumes more in the first period than she did before the introduction of the pension scheme, but less in the second period. Lastly, the high-skilled worker, whose intertemporal consumption decision was not affected by the presence of the minimum investment requirement, when forced to contribute to the pension system prefers to reduce her first period consumption to pass the threshold $\tilde{a}$ in the second period. Nonetheless she ends up, as the middle-skilled agent, on a lower indifference curve because social security reduces her lifetime income.

The bottom line of this theoretical exercise is to show that, in presence of a market imperfection that restricts access to financial markets and its returns, the introduction of social security improves the welfare of some agents and worsens the conditions of others. In other words, social security has an impact on consumption inequality. The literature on pensions has extensively discussed the costs and benefits of different ingredients shaping pension schemes and these aspects go beyond the scope of this analysis. Still, little attention has been devoted to the impact of restricted market participation on the design and size of social security. With this purpose in mind, the natural way to proceed is to abandon the simplified partial equilibrium framework examined in this section and move to a more realistic setting where social security does not only affect the lifetime income in the way described above, but also factor prices. In particular, generating a credible wealth distribution constitutes a necessary step in order to study how different agents, who find themselves in different parts of the wealth distribution, are affected by the minimum investment required to access the financial market and its higher returns. The next section introduces the overlapping generations model with heterogenous consumers which will be the workhorse model for our quantitative exercises.

1.3 General Equilibrium Model

The natural framework for studying problems with pensions and ageing is overlapping generation models. In the related literature, two are the options usually considered. On the one hand, the Blanchard-Yaari-Gertler perpetual youth model offers a relatively tractable environment, where the problem of aggregation is easily solved using the law of large numbers. Nonetheless, the assumption of age-independent probabilities of retirement and death represents a limitation for exercises that quantitatively examine consumers' choice along the life-cycle and inequality. For this reason, our choice falls on models a-là Auerbach-Kotlikoff, characterized by a full life-cycle structure with mul-

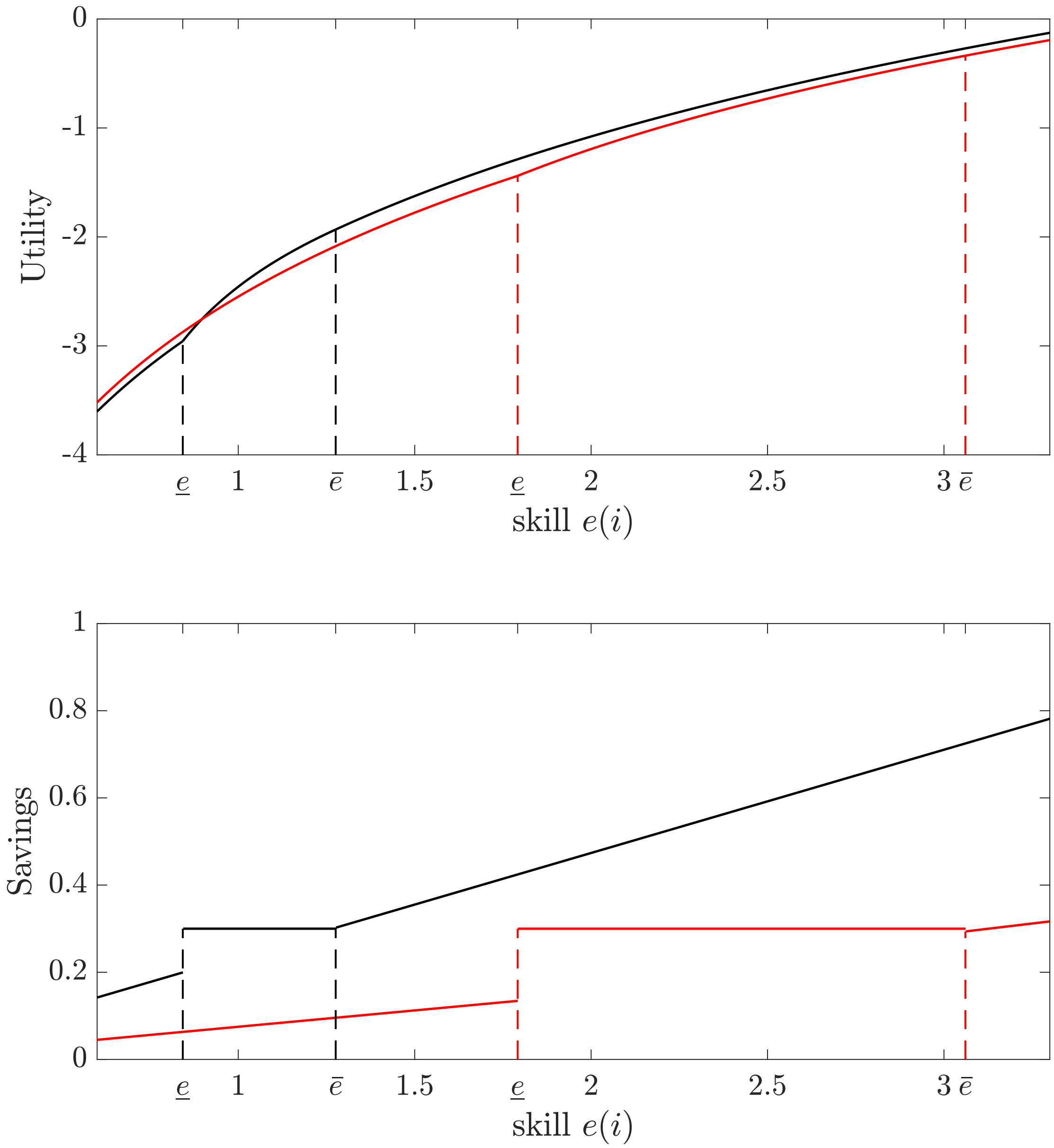


Figure 1.4: Utility and Savings for different levels of skill $e(i)$

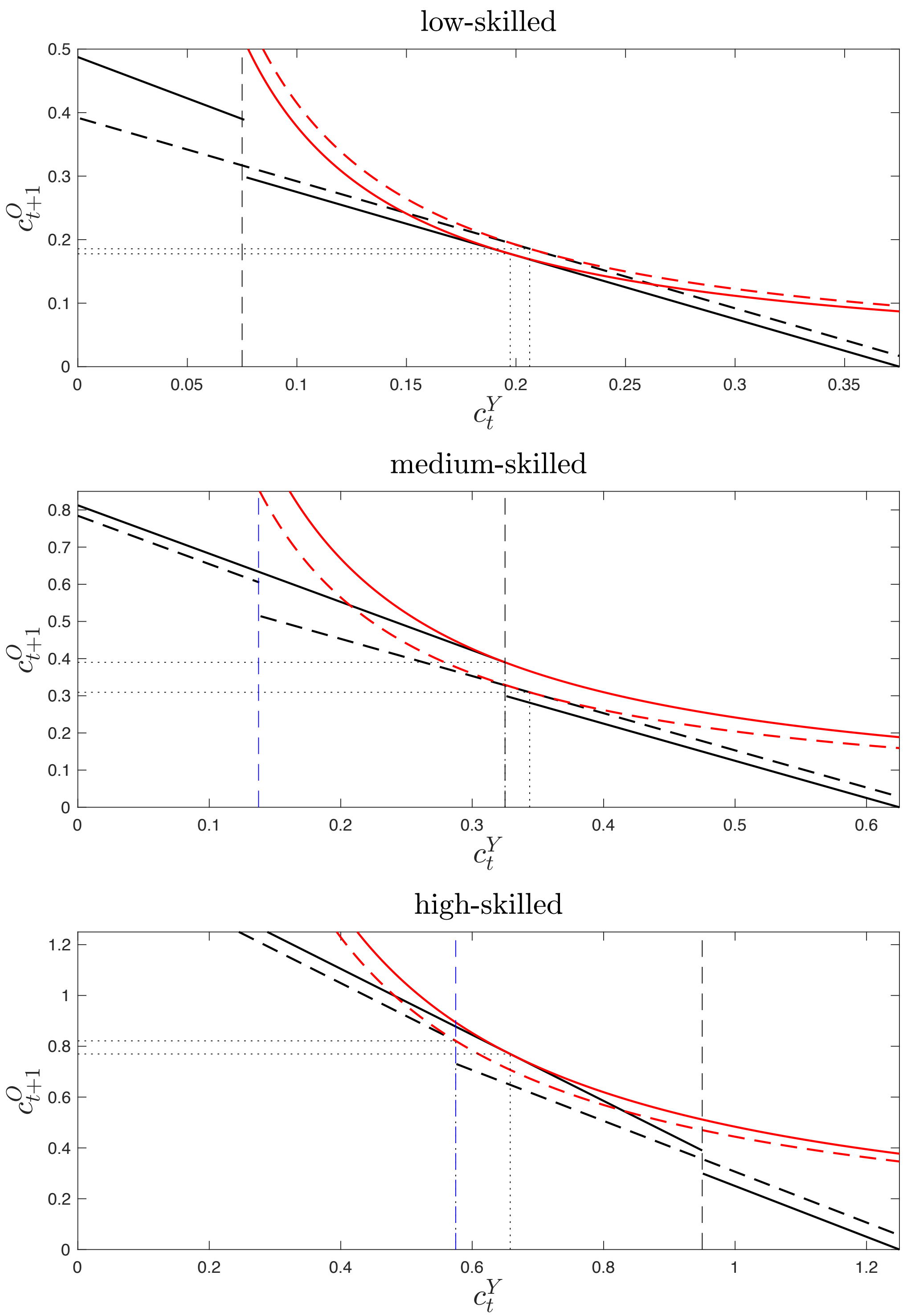


Figure 1.5: Consumption/Savings decision for different levels of skill $e(i)$

tiple generations co-exisiting at the same time and age-dependent survival probabilities. This class of models is more computationally demanding, but allows for more realistic demographic dynamics.

The overlapping generation model introduced in this section differs from the partial equilibrium framework described in the previous section in several aspects. The most relevant one concerns the determination of the interest rates. Specifically, while in the previously presented stylized economy both the storage technology return and the asset return were fixed parameters, here I will assume that only the former is exogenous and that the high market return is the return to capital, determined in equilibrium by demand and supply. This has a major implication: a change in the minimum investment requirement will affect the supply of capital, and, therefore, its return. Furthermore endogenous production replaces the exogenous endowment.

As the purpose of this paper is to determine the long-term inequality implications of social security, the analysis will focus on steady state exercises and leave aside transition dynamics.

1.3.1 Heterogenous households

Households have identical preferences, but differ in terms of wealth, ability and age. The life-cycle structure is as follows: households enter the economy at age 25 with an initial inherited stock of assets and an initial skill endowment, stay in the labor force until age 65 when they deterministically retire. After retirement, they live off of their asset returns and pension benefits until a maximum age of 100. The age of death is stochastic and the survival probabilities are age-dependent. Between 25 and 65, workers gain higher productivity through experience. Stochastic death and a bequest motive imply that households perish with positive assets which are automatically transferred, as inheritance, to the new generation of households replacing them in the next period. The stochastic age of death triggers the precautionary motive for saving. Such assumptions, together with the differential stock of wealth and skill at birth, will generate both a wealth and an income distribution. An endogenously generated wealth distribution will affect the ability of heterogenous agents to access the financial markets. The minimum investment requirement illustrated in the previous section, in combination with a borrowing limit, completes the set of constraints. The household entering the economy at time t and aged 25 faces the following lifetime utility maximization problem:

$$\max_{c_{t+k-1}(i),a_{t+k}(i)} E_t \sum_{k=1}^{75} \beta^{k-1} \left[\prod_{z=1}^{k} \big(1-\theta_{z-1}\big) \right] \left[U\big(c_{t+k-1}(i)\big) + \frac{1}{\beta}\theta_{t+k} W\big(a_{t+k+1}(i)\big) \right]$$

s.t.

- if young and employed:

$$c_{t+k}(i)+a_{t+k+1}(i) = a_{t+k}(i)\big(1+r_{t+k}-\delta\big)+(1-\tau_{t+k})w_{t+k}l_{t+k}e_{t+1}(i,k) \quad \forall \quad 0 \leq k \leq 39$$

- if old and retired:

$$c_{t+k}(i) + a_{t+k+1}(i) = a_{t+k}(i)\big(1 + r_{t+k} - \delta\big) + pen_{t+k}(i,k) \quad \forall k \geq 40$$

- borrowing constraint:

$$a_{t+k}(i) \geq 0 \quad \forall k$$

- minimum investment requirement:

$$r_{t+k} = \begin{cases} r^h_{t+k} & \text{if } a_{t+k}(i) \geq \tilde{a} \\ r^l_{t+k} & \text{if } a_{t+k}(i) < \tilde{a} \end{cases} \quad \forall k$$

- initial asset endowment:

$$a_t(i,0) = a^0(i)$$

The function U is a standard CRRA: $U\big(c_{t+k}(i)\big) = \frac{c_{t+k}(i)^{1-\sigma}}{1-\sigma}$ and expresses the utility coming from consumption. The function $W\big(a_{t+k+1}(i)\big) = \chi\frac{a_{t+k+1}(i)^{1-\eta}}{1-\eta}$ represents the utility derived from leaving bequests and takes as argument the amount of assets at the age of death. The period budget constraint at time $t+k$ depends on the state of the household i in the same period: employed or retired. It is assumed that, when young, each consumer inelastically supplies a unit of labor $l_{t+k} = 1$, augmented by the individual productivity or ability level $e_{t+k}(i,k)$. After the exogenously fixed retirement age of 65, i.e. 40 years after entering the labor markets, the households starts receiving a pension benefit $pen_{t+k}(i,k)$ until death. Uncertainty stems from one source only and only at an individual level: life duration. The stochastic process describing it is characterized by the sequence $\theta_0, ..., \theta_{75}$ of age-dependent probabilities of dying, with $\theta_0 = 0$ and $\theta_{75} = 1$.
At this point, we can rewrite the household problem more conveniently using two sets

of Bellman equations, one valid as long as the household is aged between 25 and 65 and one valid after retirement.
For $1 \leq k \leq 40$:

$$V\big(a_t(i); z_t(i,k); k\big) = \max_{c_t(i); a_{t+1}(i)} U(c_t(i)) + \theta_k W\big(a_{t+1}(i)\big) + \beta(1-\theta_k) V\big(a_{t+1}(i); e_{t+1}(i,k+1); k+1\big)$$

subject to

$$c_t(i) + a_{t+1}(i) = a_t(i)\big(1 + r_t - \delta\big) + (1 - \tau_t) w_t e_t(i,k)$$

For $k > 40$:

$$V\big(a_t(i); k\big) = \max_{c_t(i); a_{t+1}(i)} U(c_t(i)) + \theta_k W\big(a_{t+1}(i)\big) + \beta(1-\theta_k) V\big(a_{t+1}(i); k+1\big)$$

subject to

$$c_t(i) + a_{t+1}(i) = a_t(i)\big(1 + r_t - \delta\big) + pen_t(i,k)$$

The minimum investment requirement and borrowing constraint apply too.

1.3.2 Supply side

The production sector of the economy is standard. There is a representative firm hiring capital and labor in perfectly competitive markets and employing a constant returns to scale Cobb-Douglas production function $Y_t = A_t K_t^{\alpha} (E_t L_t)^{1-\alpha}$ where A_t represents an aggregate level of productivity growing at the exogenous and deterministic rate g and E_t the aggregate labor augmenting productivity which is assumed to be constant over time. As a consequence, production factors are paid their marginal productivities:

$$r_t^K = r_t^h = \alpha A_t \left(\frac{K_t}{EL}\right)^{\alpha-1}$$

$$w_t = (1-\alpha) A_t \left(\frac{K_t}{EL}\right)^{\alpha} Z$$

1.3.3 Government

The government collects a labor income tax, whose revenues are used to finance the pension transfers and runs a balanced budget in every period.

Pay-as-you-go

The PAYG scheme is a simple transfer from the working households to the ones retired. In practice, a certain replacement ratio ϕ_t is ensured, so that pensioners receive a fraction of the average wage they earned during their working age. It is, therefore, a defined benefit scheme. The tax rate on labor income is adjusted to preserve the financial sustainability of social security in each period.

$$\underbrace{\sum_{k=1}^{40}\sum_{i} N(i,k)\tau_t^{PAYG} w_t e_t(i,k)}_{\text{total revenues}} = \underbrace{\sum_{k=41}^{75} N(i,k) pen_t(i,k)^{PAYG}}_{\text{total expenditures}} \tag{1.4}$$

where

$$pen_t(i,k)^{PAYG} = \phi_t \frac{1}{40}\sum_{j=1}^{40}(1-\tau_{t-k+j}^{PAYG}) w_{t-k+j} e_{t-k+j}(i,k-40+j) \tag{1.5}$$

Equation 1.4 represents the government balanced budget, where total labor income tax revenues equate total expenditures for pensions. $N(i,k)$ is the mass of individual i at age k and $e_t(i,k)$ is her ability level at time t. The term $\sum_{k=39}^{74} N(i,k)pen_t(i,k)^{PAYG}$ constitutes total expenditure for pension benefits, paid to all retired households at time t, and crucially depends on the size of each cohort born at time $t-k$ and still alive at time t, $N_{i,k}$. Finally, equation 1.5 shows how an individual PAYG pension is calculated: it is a fraction $0 < \phi_t < 1$ of the average after tax labor income of agent i during her working age (25-65).

After fixing the replacement ratios $\phi_t = \phi$ the tax rate τ_t^{PAYG} is adjusted so that equations 1.4 - 1.5 are satisfied:

$$\tau_t^{PAYG} = \frac{\frac{\phi}{40}\sum_{k=41}^{75} N(i,k)\sum_{j=1}^{40}(1-\tau_{t-k+j}^{PAYG}) w_{t-k+j} e_{t-k+j}(i,k-40+j)}{\sum_{k=1}^{40}\sum_i N(i,k) w_t z_t(i,k)} \tag{1.6}$$

Which states that, in a PAYG system, the tax rate is increasing in the defined benefits due to the pensioners, determined by the retired cohorts' size $\{N(i,k)\}_{k=41}^{75}$, the average after tax wage they received during working age $\frac{1}{40}\sum_{j=1}^{40}(1-\tau_{t-k+j}^{PAYG})w_{t-k+j}e_{t-k+j}(i,k-40+j)$, the replacement ratio ϕ, and decreasing in aggregate labor income. As the last equation shows, PAYG schemes are exposed to some risks. In particular, population ageing, by reducing the size of young cohorts relative to the old ones, puts into question the sustainability of this arrangement. As population gets older, higher tax rates are required to ensure a balanced budget, if the government aims at maintaining the desired replacement ratio ϕ. This demographic effect is countered by wage growth: if the economy grows the upward pressure on the tax rate is relieved. Since, as in the previous section, the assumed scheme does not redistribute contributions across retirees of different labor earning classes, the return of social security is the return of the economy and benefits are proportional to contributions.

1.3.4 Competitive equilibrium

The competitive equilibrium of this economy is defined as:

- a series for the individual productivities $\{e_t(i,k)\} \quad \forall i,k$;
- a series of allocations $\{c_t(i), a_{t+1}(i)\}_{t=0}^{\infty} \quad \forall i$;
- a series of factor prices $\{w_t, r_t^K\}_{t=0}^{\infty}$;
- a series of tax rates $\{\tau_t\}_{t=0}^{\infty}$.

such that firms maximize their profits, households maximize their lifetime utility and the government runs a balanced budget in every period, ensuring the desired replacement ratios.

Although the objective of this paper is not to study optimal pensions in the presence of a minimum investment requirement, for the quantitative exercises I need to choose a reference size of the the pension system. Rather than calibrating the parameters to the real US social security system, given the stylized nature of the arrangement assumed in the model, and for the sake of simplicity, I choose the replacement ratio solving the Ramsey social planner problem. In practice, I pick the replacement ratio maximizing a utilitarian social welfare function such that the first order conditions describing the optimal behaviour of firms and households are satisfied and that the government budget is balanced.

$$\max_{\tau} \sum_{i=1}^{50} N(i) \sum_{j=1}^{75} \beta^{j-1} \left[\prod_{z=1}^{j} \left(1-\theta_{z-1}\right) \right] \left[U\left(c_{t+j-1}(i,j)\right) + \frac{1}{\beta}\theta_j W\left(a_{t+j}(i,j+1)\right) \right]$$

with $\theta_{k-1} = 0$ and where $N(i)$ indicates the mass of agent i at age k. The utilitarian social welfare function is the average of all existing households' expected lifetime utilities weighted by their size relative to the overall population and their life expectancy. The solution of the Ramsey social planner's problem entails dealing with a large system of non-linear equations that cannot be solved analytically. Hence, I rely on standard numerical methods to solve the problem.

How is the optimal level of social security determined? In order to answer this question, it is necessary to analyze the costs and benefits of pensions in relation with the assumptions made in this model. Obviously, the preference over the tax rate depends on the state, defined as a combination of age, productivity and financial position, featured by the individual agent.

On the side of costs, a higher tax rate means, for workers, a lower disposable income and therefore a crowding out effect on private accumulation of savings. Agents who cannot access the financial markets will sacrifice a greater portion of their after-tax labor income for savings, in the attempt of satisfying the minimum investment requirement in the following periods. Secondly, the pension system can have a lower return than the private return to savings, namely for those households being able to access the capital market. The return of the PAYG scheme is the growth rate of the economy. Forcing agents to contribute to a PAYG system when the return of the other available asset options is higher than the growth rate of the economy implies a reduction of the agent's lifetime income. In the model, the size of this crowding out effect is increasing in the difference $r^K - (n+g)$ and in the tax rate τ. However, the quantitative importance of these distortions caused by financing social security crucially depends, in this model, on the barrier limiting access to the return to capital. The example presented in the previous section showed that a household that is unable, independently from social security, to accumulate enough savings to satisfy the minimum investment requirement over her lifetime will not be affected by the crowding out effect. As a matter of fact, if her private savings return is r^l, the basic storage technology return and the return of the pension asset is higher than that, a positive tax rate will increase her lifetime income. This suggests that in this model a pension system represents an opportunity for some households and an impediment for others: the individual current and future expected asset position, relative to the minimum investment requirement, determines the desirability of social security as a means for transferring value into the future and smooth consumption.

In addition, a pension arrangement serves another purpose: it acts as an insurance policy against demographic risk. By pooling together the mandatory savings of all agents, social security ensures old-age consumption independently of the individual lifetime duration. In a PAYG scheme a defined benefit is granted even when the total amount of individual benefits paid by the government exceeds the individual contributions. However, there will be some pensioners who will never receive the total proceedings of their contributions because they perish early.

The presence of a pension system crowds out private wealth. This means that, as agents die stochastically, the size of the scheme directly affects the amount and distribution of bequests left to the future generations. As a consequence, the scope of social security has an impact of the ability of newborn generations to satisfy the minimum investment requirement to access the financial markets.

As a final note, the assumption of exogenous labor supply means that taxes on labor income will not distort the typical consumption/leisure trade-off. Thus, the assumed PAYG scheme promotes intergenerational risk-sharing only within labor earnings classes, does not have distortionary effects on the consumption/leisure decision and offers partial insurance against demographic risk, in absence of the annuities markets.

1.3.5 Calibration

Table 1.2 summarizes the calibration of most parameters. Most of it is due to the calibration/estimation performed in Benhabib, Bisin & Luo (2015). Although the model laid out in this project diverges from theirs, especially in relation with the presence of the minimum investment requirement and heterogeneous agents within each cohort, this analysis closely follows their approach to derive a realistic wealth distribution. An important result of their exercise is that $\sigma > \eta$, i.e. the utility from consumption displays higher curvature than the one obtained from bequests. This translates into increasing saving rates with respect to wealth. As explained by the authors, this is one of the determinants of the concentration in wealth observed in reality. A second crucial factor is labor earnings risk. In this regard, we refer to the work by Heathcote, Perri & Violante (2010) [21], which documents, based on PSID data, the evolution of the labor earnings distribution along the life-cycle. It uncovers the increasing dispersion of labor earnings over age and a typical hump-shaped profile. Table 1.3 presents their findings. Ten different labor earnings classes are identified featuring heterogenous initial labor proceedings and heterogenous labor income growth rates, which are accounted for in the model. As the data for earnings close to retirement age are unavailable, I assume that the observed fall in earnings between age 50 and 55 takes place in a larger time span, until 65. Moreover, as shown in the first column of Table 1.3, the bottom 10% of the earnings distribution features negative earnings, a symptom of debt positions.

Consistently with Benhabib, Bisin & Luo (2015), negative earnings are replaced by a fictitious tiny value as one of the model assumption is a no borrowing constraint. The presence of a bequests motive generates a distribution of wealth at birth. To keep track of inheritance and early life asset positions I distinguish 10 different classes at birth, based on the quintiles of the endogenous bequests distribution. However, to simplify the analysis, I will assume that labor earnings and initial asset positions groups are orthogonal, i.e. being born with a low or high initial wealth endowment does not influence the probability of being part of a specific labor earnings class. Therefore, the model does not account for human capital accumulation and education investment.

The data source for the age-dependent survival probabilities for the US is the 2014 Actuarial Life Table, provided by the US Social Security Administration. I assume that each period the size of the newborn cohort, entering the economy, is exactly 1. As both fertility and death rates are constant over time, population is also constant. Figure 6 shows the weight of each age cohort in the overall population. The area underneath the curve amounts to life expectancy at birth.

Finally, the benchmark model will feature 10 different households (each one accounting for 10% of the population) characterized by distinct labor earnings profiles and 10 different wealth positions at birth, determined by the quantiles of the endogenous bequests distribution. Thus, 100 heterogenous agents coexist in each cohort so that the model keeps track of 100x75 agents of different mass.

1.4 Results

1.4.1 The benchmark economy

The first experiment I conduct consists of examining how the equilibrium stationary distribution is affected by different levels of the minimum investment requirement in absence of social security. In particular, I will assess the impact of this friction on a set of variables, namely aggregate capital, the return to capital, the aggregate amount of savings of the economy, aggregate output, the Gini index for wealth inequality and for consumption inequality and the capital market participation rate through simple comparative statics exercises. The objective of this analysis to determine the scope of this market imperfection, or, in other words, whether the presence of even a small minimum investment requirement is quantitatively important for the equilibrium properties

Parameter	Value	Description
α	0.33	capital share in production
β	0.97	discount rate
σ	2	relative risk-aversion coefficient
δ	0.1	depreciation rate
η	1.186	curvature parameter for utility from bequests
χ	0.0312	weight of bequests in the utility function
r^l	0	storage technology return
A	1	total factor productivity
g	0	technological progress

Table 1.2: Calibration

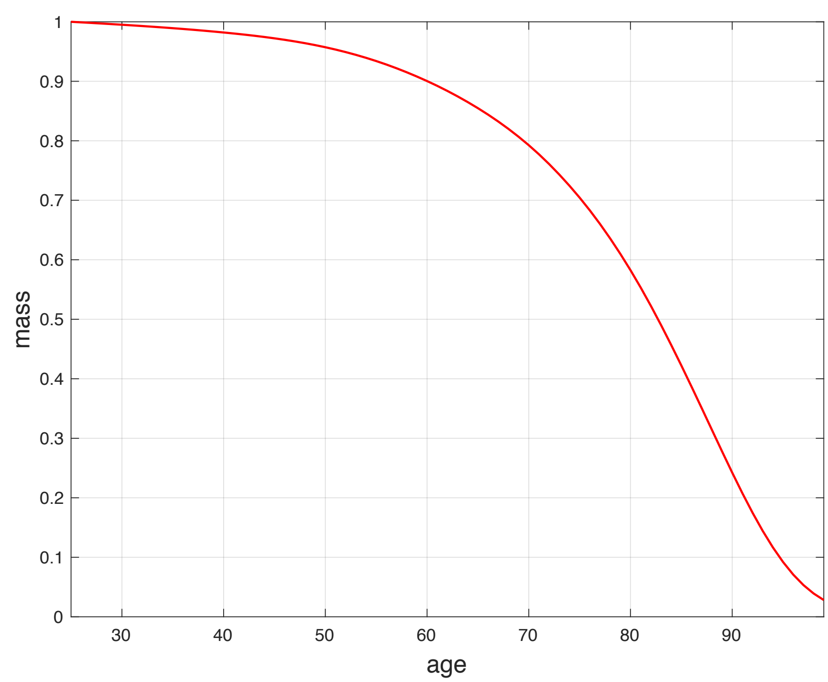


Figure 1.6: Population mass by cohort

Age range/%	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
25-30	-2.68	9.356	16.87	23.23	29.47	35.48	41.71	49.12	59.52	87.90
31-36	-1.68	12.90	21.88	29.78	37.10	44.21	52.06	61.69	75.01	123.5
37-42	-1.73	13.48	23.84	32.88	41.35	49.64	57.95	68.42	84.67	153.8
43-48	-2.73	13.59	24.54	33.73	42.76	51.46	60.73	72.46	90.04	165.5
49-54	-4.97	10.47	20.95	29.68	38.81	47.98	57.98	69.65	87.23	165.2
55-60	-8.22	10.47	11.31	19.63	28.21	37.60	47.20	59.23	77.07	156.5

Table 1.3: Labor earnings (in thousand dollars) profiles by age range from PSID and cleaned by Heathcote, Perri & Violante (2010)

of the economy under examination.

Table 1.4 summarizes the results of this first exercise, where the minimum investment requirement is increased from 0 to 0.1. A larger threshold $\tilde{a}$ implies that households will be able to access the capital market and its higher return only at later stage of their working life, after stockpiling enough savings. When the minimum investment requirement is high, the relatively low-skilled agents and those born with little wealth are completely excluded from the capital market and only the low-yield storing technology is available to them. When $\tilde{a}$ is increased from 0 to 0.1, sacrificing consumption to satisfy the investment requirement becomes too costly in terms of utility for many agents, so aggregate savings drop. As a consequence, the amount of capital supplied to firms falls and so do production and wages, according to the neo-classical production function. The equilibrium between demand and supply in the competitive capital market implies that return to capital r^K increases as $\tilde{a}$ rises, while the average interest rate or return to savings falls. The Gini indices for the wealth and consumption distributions both increase. A result of the assumed market imperfection is that limited asset market participation produces inequality. A minimum investment requirement $\tilde{a} = 0.1$ entails an asset market participation of 41.67%, which is close to what emerges in the Survey of Consumer Finances data. In absence of social security, such threshold is equivalent to, in equilibrium, 10.5 times the yearly working age average labor income of the median labor earnings class.

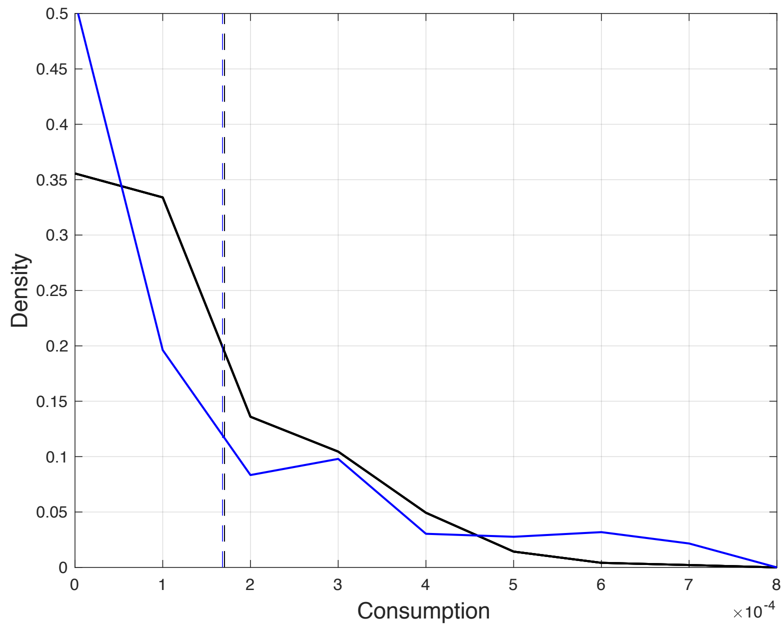
The impact of the minimum investment requirement on the consumption and wealth distributions is represented, in blue, in Figure 7. It shows that, relative to the case when $\tilde{a} = 0$, the financial friction has a marginal impact on the means of the distri-

Variable	$\tilde{a} = 0$	$\tilde{a} = 0.1$
w	0.0898	0.0880
r^K	0.1497	0.1567
K	0.2057	0.1924
Y	0.0933	0.0914
Gini - wealth	0.5195	0.7021
Gini - consumption	0.4140	0.6197
participation	100%	41.67%

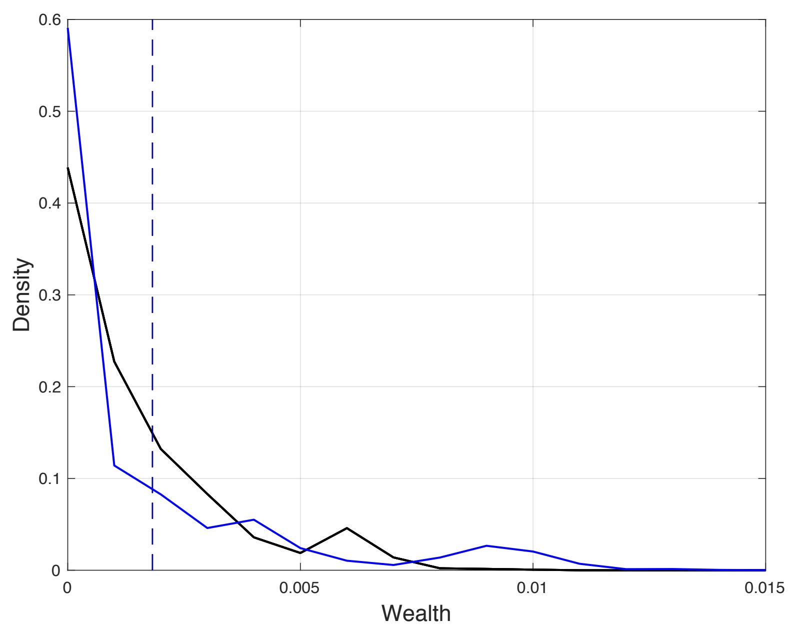
Table 1.4: Comparative statics with $\phi = 0$

butions, but significantly increases their dispersion, consistently with the information given by the Gini coefficients. The density of the population around the means falls, while there is a higher concentration of agents at the extremes of the distributions. This outcome can be explained as follows: poor agents, in presence of the friction, are now pushed out the capital market and can invest their savings only in the low-yield storing technology. Moreover, their wages are lowered. At the same time, agents in the middle of the labor earnings or wealth distribution at birth, who potentially can pass the threshold at some point of their life, will find it optimal to sacrifice part of their current consumption. This implies that their consumption levels, especially for the initial years of their working life, are similar to those of poorer agents who will never participate to the capital market. Finally, richer agents, for whom the investment requirement is not binding, enjoy higher levels of consumption and wealth relative to the case of a frictionless economy due to the increased return to capital r^K.

Figure 8 compares the results for the analysis of wealth and consumption inequality by age groups when $\tilde{a} = 0$ and when $\tilde{a} = 0.1$. Two measures of dispersion are taken into consideration: the Gini index and the interquartile range. The two criteria point to the same direction: the financial friction increases the dispersion of wealth and consumption at any age. In the case of full asset market participation, the model predicts that consumption inequality remains relative constant over the lifetime. This should not be surprising, given the model assumption that labor efficiency paths are determined at birth and not subject to idiosyncratic shocks. However, in presence of limited asset market participation, both measures report a more realistic increase of inequality along the life-cycle. Storesletten, Telmer & Yaron (2004) document an increase in consumption inequality with age and attribute this phenomenon to labor earnings risk. Huggett, Ventura & Yaron (2011) instead claim that the increasing consumption inequality over the lifetime can be mostly accounted for the initial conditions, defined in terms of asset position, skill level and learning ability. In a more sophisticated model, the existence of a minimum investment requirement entailing a restricted capital market participation, in combination with idiosyncratic shocks, may imply that small shocks can potentially have a large impact on savings and consumption decisions, and therefore, on inequality. The analysis of wealth heterogeneity conditional on age suggests that most of the increase in wealth inequality resulting from the introduction of the minimum investment requirement is due to the spread in the rate of wealth accumulation between rich and poor agents. In fact, whereas the interquartile range is constant and not dissimilar from its value in the frictionless economy until age 40, it steadily increases after, exactly when the agents in the middle of the initial wealth and labor earnings distributions manage to enter the financial market.

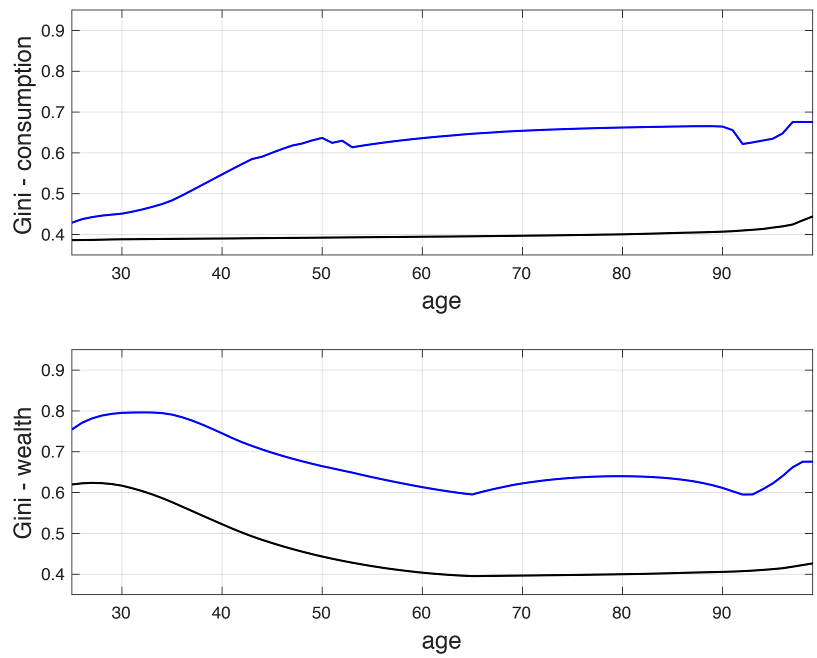


(a) Consumption distribution

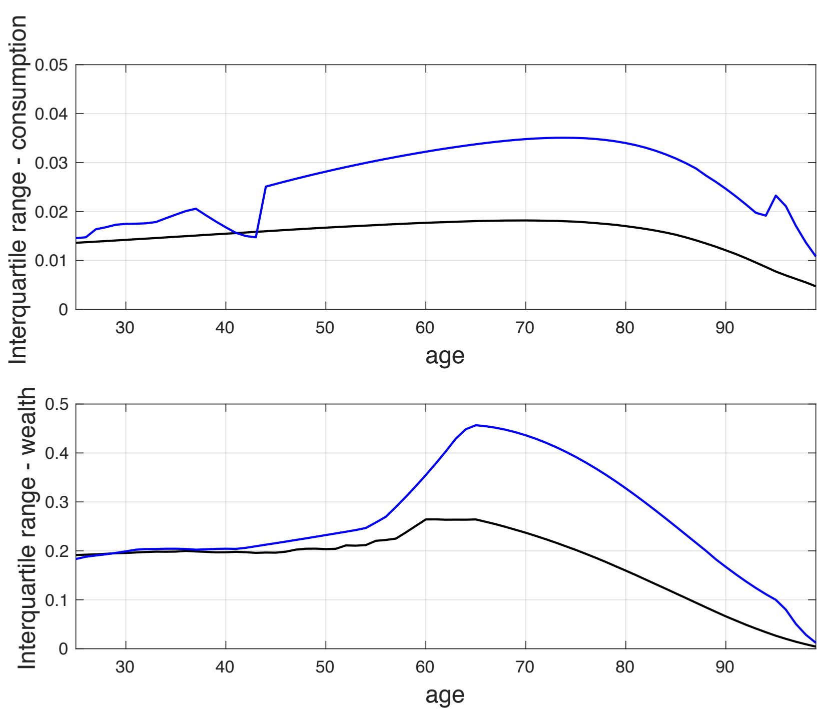


(b) Wealth distribution

Figure 1.7: In black $\tilde{a} = 0$, $\phi = 0$, in blue $\tilde{a} = 0.1$.



(a) Gini coefficient



(b) Interquartile range

Figure 1.8: Measures of dispersion, in black $\tilde{a} = 0$, $\phi = 0$, in blue $\tilde{a} = 0.1$.

1.4.2 Pay-as-you-go

This subsection studies the role played by a PAYG pension system in the benchmark economy described above. Firstly, I obtain the replacement ratio maximizing the social welfare function when $\tilde{a}$ is 0.1. Secondly, I compare the consumption and wealth distributions obtained with the optimal replacement ratio $\phi = 0.2$ and a minimum investment threshold $\tilde{a} = 0.1$ with their counterparts in absence of the friction, to examine how limited asset market participation alters the impact of social security on consumption and wealth inequality.

The main features of this type of pension arrangement are the way it is funded and the redistribution of the collected contributions it entails. In this model, each labor earnings class contributes to a specific pension fund, as if each class could be associated with a profession. Within a profession, each cohort pays in the fund the same proportion of her labor income. For simplicity, I assume that the replacement ratio of the system is the same across labor earnings classes, which, still, may imply different tax rates. In practice, agents of different age and different initial wealth, but with on the same labor productivity path or profession, pool their contributions to be entitled to a pension benefit from the fund once retired. The scheme allows households to partly insure against demographic risk. Therefore, this specific form of social security promotes risk-sharing both across households belonging to different initial wealth groups and across different generations, but still within each labor earning class.

In presence of a minimum investment requirement in the capital market, the return of a PAYG scheme is higher than the low-yield storing technology return. This assumption drives the result of the optimality of social security in presence of the market imperfection. At the same time, a positive tax rate τ will crowd out private savings and make the $\tilde{a}$ threshold harder to reach. As a result, fewer households will participate to the asset market. We can conclude that social security favours some households and endangers others. In particular, it benefits the agents at the top of the wealth distribution, because they are anyways able to pass the threshold and enjoy a even higher capital return r^K due to general equilibrium effects. But also the less skilled and poor agents who would not be able to invest in capital even in absence of the pension scheme, because they can at least enjoy a return to their contributions higher than the private return to their savings. However, social security will make the agents in the middle of the initial wealth and labor earnings distributions worse off, as they are hurt by a lower disposable income and therefore face a greater obstacle to access the best market return. Table 5 and 6 report the equilibrium summary statistics for different replacement ratios, respectively when $\tilde{a} = 0$ and when $\tilde{a} = 0.1$. The associated optimal replacement ratio ϕ is around 20%.

From the comparison between Table 5 and 6, a clear result emerges: the introduction of the pension system, in presence of the minimum investment requirement, implies a much bigger crowding out effect of capital. This translates into a bigger fall in aggregate output and wages, as well as a larger increase in the return to capital r^K. Moreover, while with full asset market participation the pension system only marginally affects the wealth and consumption Gini coefficients, when $\tilde{a} = 0.1$ social security generates substantially more inequality.

Figure 9, where the black lines indicate the distributions of an economy without any social security and the red line the distributions of the economy featuring the modelled pension arrangement, highlights a similar outcome. In absence of the market imperfection, social security has a little effect on the consumption and private wealth distributions. In particular, social security slightly increases mean consumption and slightly reduces (mechanically) mean private wealth, but leaves substantially unaffected the dispersion of the two distributions. On the contrary, when the capital market is only restricted to wealthy agents, social security not only has a larger impact on the means of the two distributions, but significantly increases their variance. Specifically, the density of agents around the means is negatively affected by the pension scheme and the density of the agents at the extremes of the distributions grows. Overall, social security seems to exacerbates the effect of the financial friction on inequality identified in the previous quantitative exercise.

Similarly, when we to turn to the analysis of wealth and consumption inequality conditional on age, based on the investigation of Figure 10, we find that social security has a very limited impact on the two chosen measures of dispersion, in the economy with full asset market participation. The little noticeable effects are due to the impact of social security on the wealth distribution at birth, through bequests: the pension scheme reduces the overall amount of inheritances and increases the heterogeneity in initial wealth. However, when we take into account limited asset market participation, we notice that social security has a stronger effect on inequality, and that the two measures of dispersion point in two different directions. The Gini index for consumption suggests that social security shifts the increase in inequality during the lifetime. It can be interpreted as the fact that paying contributions reduces disposable income and forces even the richest agents to postpone the moment of entry in the capital market. The interquartile range instead documents a fall of inequality at all ages. This is not surprising given that social security, by making the threshold $\tilde{a}$ harder to reach, pushes

Variable	$\phi = 0$	$\phi = 0.2$
w	0.0898	0.0881
r^K	0.1497	0.1562
K	0.2057	0.1935
Y	0.0933	0.0915
Gini - wealth	0.5195	0.5412
Gini - consumption	0.4140	0.4227
participation	100%	100%

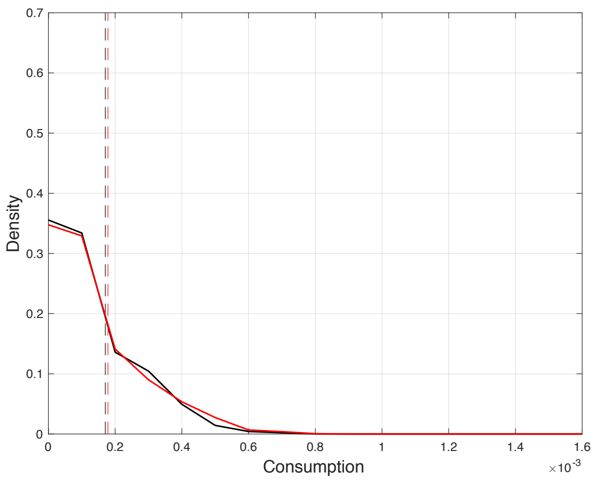
Table 1.5: Comparative statics with $\tilde{a} = 0$ and PAYG pension system

even households around the third quartile out of the capital market and fosters a concentration of agents at the bottom of the consumption distribution.

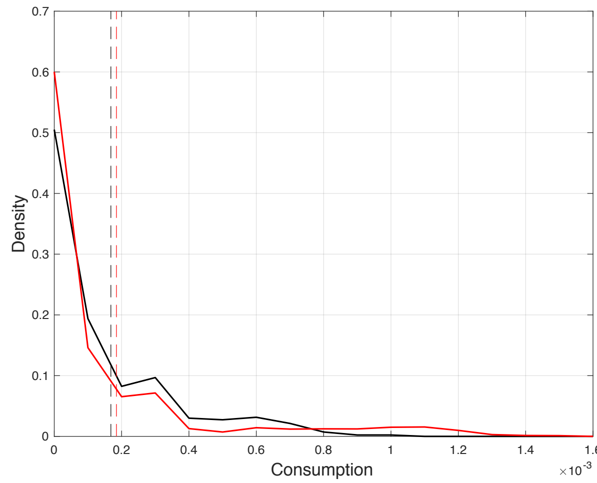
We can conclude that, in presence of the financial friction, the pension arrangement improves the conditions of the poorest and richest agents, but negatively hits the consumption and wealth accumulation of middle classes, who lose the opportunity to save enough to pass the threshold imposed by the minimum investment requirement. The agents at the top of the wealth distribution accumulate at faster rate, thanks to the increased capital return r^K; the agents at the bottom increase their consumption because they dispose of an asset, the pension scheme, which offers them a return higher than the one available to them (r^l); the agents in the middle of the wealth distribution that, in absence of social security, manage to access the capital market, suffer from a drop in their wealth accumulation and therefore in consumption levels. The inequality in the low part of the distribution shrinks, the difference between bottom and top expands. The presence of a PAYG scheme reduces, overall, the costs imposed by the minimum investment requirement and improves welfare. However, it has distributional consequences that may be important for policy-makers.

1.4.3 Population ageing

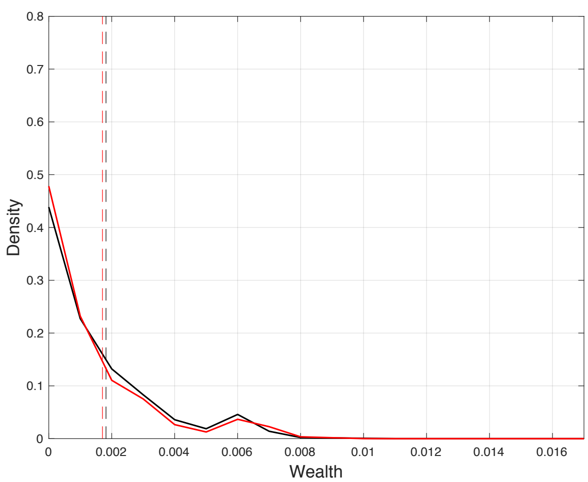
As a final experiment with the model, I study the effects of an increase in the regulated retirement age in response to a permanent longevity shock on the endogenous



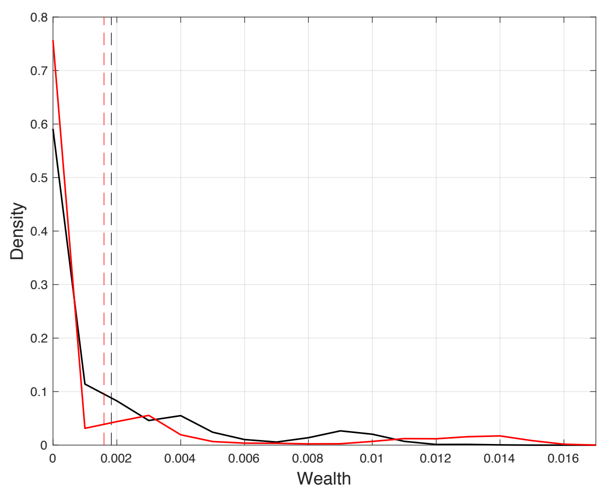
(a) $\tilde{a} = 0$



(b) $\tilde{a} = 0.1$

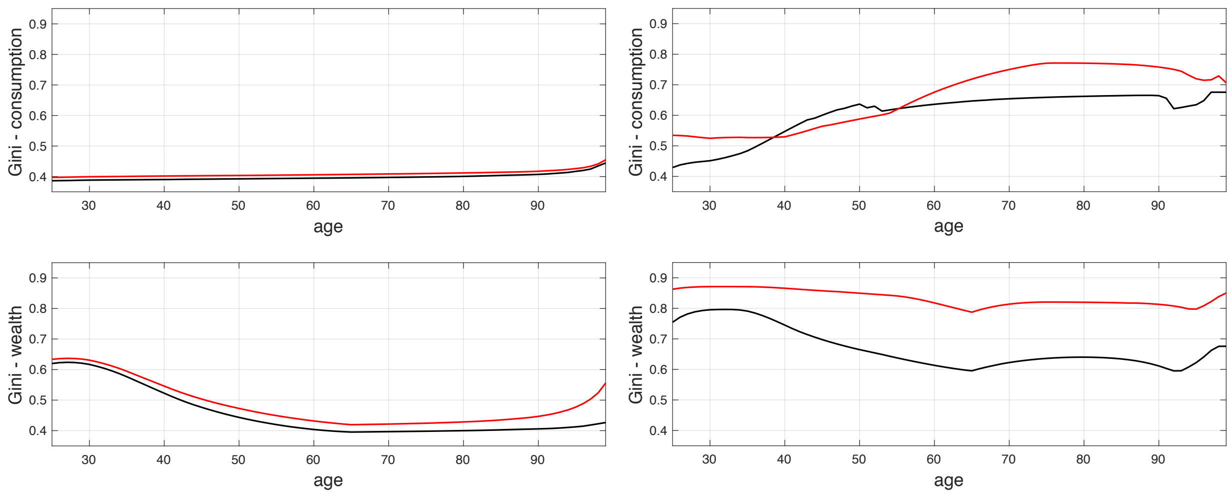


(c) $\tilde{a} = 0$



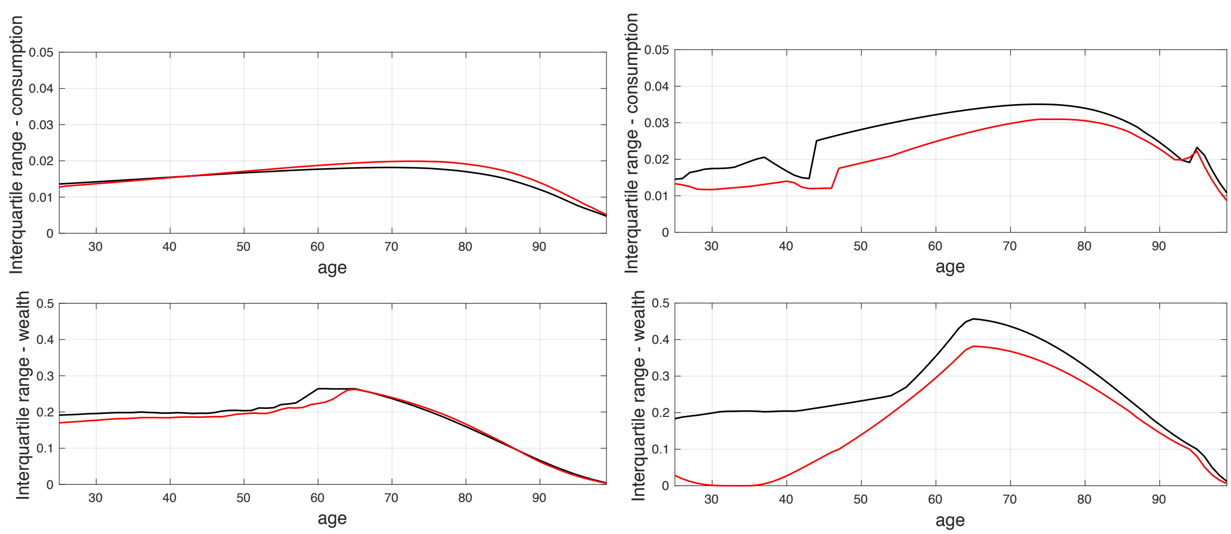
(d) $\tilde{a} = 0.1$

Figure 1.9: The effect of a PAYG pension system on the consumption and wealth distributions, in black $\phi = 0$, in red $\phi = 0.2$



(a) Gini index when $\tilde{a} = 0$

(b) Gini index when $\tilde{a} = 0.1$



(c) Interquartile range when $\tilde{a} = 0$

(d) Interquartile range when $\tilde{a} = 0.1$

Figure 1.10: The effect of a PAYG pension system on the consumption and wealth inequality by age groups, in black $\phi = 0$, in red $\phi = 0.2$

Variable	$\phi = 0$	$\phi = 0.2$
w	0.0880	0.0855
r^K	0.1567	0.1663
K	0.1924	0.1762
Y	0.0914	0.0888
Gini-wealth	0.7021	0.8502
Gini-consumption	0.6197	0.6822
participation	41.67%	24.76%

Table 1.6: Comparative statics with $\tilde{a} = 0.1$ and PAYG pension system

wealth and consumption distributions. Increasing the retirement age is a measure that has been adopted by many countries in the attempt to address the budget imbalances generated by demographics. Still, it is controversial from an economic standpoint and generally unpopular. In fact, it is often argued that it is a regressive type of intervention because poor individuals tend to have shorter lives compared to rich. Additionally, it may worsen the conditions of those workers that are employed in physically demanding activities and benefit who, on the contrary, is occupied in less arduous activities. At the current stage of the paper, the analysis abstracts from the role played by an endogenous labor supply and heterogeneous mortality rates across labor earnings classes, which may substantially alter the quantitative results.

I simulate the effect of an increase in life expectancy at birth from around 81 to 84 years. Such change is achieved through a horizontal shift of the curve representing the mass of each cohort, as reported in Figure 11. More specifically, there is a decline in the age-specific probabilities of dying, especially at old age. Such change brings about an increase in the discount factors. The simulated population ageing shock implies, in absence of any pension reform concerning the retirement age, a surge of the dependency ratio $\frac{\text{people aged 66-100}}{\text{people aged 25-65}}$ from 43.71% to 51.72%. I assume that the government keeps the initial replacement ratio ϕ at 0.2, which, as a consequence of the shock, induces an increase in the tax rates $\tau(i)$ for all the different labor earning classes. Importantly, here the outcome of the shock is not only determined by the movements in aggregate capital, as the increase in longevity also implies a reduction of aggregate labor which, in absence of the pension reform, drops from 0.6958 to 0.6591.

The pension reform consists of an increase in retirement age from 65 to 67, which brings the dependency ratio back to 45.09% and moderates the fall in total labor, which hovers at 0.6892. In order to calibrate the labor earnings of the additional two years of working age, I assume that the individual efficiencies grow at the same rate they experienced in the last 5 years before retirement. For most of the labor earnings classes, this means that labor income falls with age. Tables 7 and 8 report the results of the simulated shock and subsequent reform in an economy characterized by the presence of a PAYG scheme with replacement ratio $\phi = 0.2$, respectively when $\tilde{a} = 0$ and when $\tilde{a} = 0.1$.

The shock produces two opposite forces: on the one hand, as agents live longer, it pushes them to save more for old-age consumption; on the other hand, the government needs to raise the tax rates to maintain the same replacement ratio, which implies a reduction in disposable income and, therefore, in savings. Higher tax rates reduce private savings and crowd out not only capital, but also bequests. As documented in Tables 7 and 8, aggregate capital drops in the aftermath of the population ageing shock. Nonetheless, while the return to capital in the economy with full asset market participation falls, the economy featuring the financial friction experiences an increase in r^K. Although aggregate savings and capital fall, when $\tilde{a} = 0.1$, asset market participation, which is an extensive margin measure, slightly increases. Finally, in both economies the longevity shock produces a reduction of the Gini index for private wealth, but has heterogenous on consumption inequality: the Gini for consumption falls in the frictionless economy, but rises in the economy with limited asset market participation.

The third column of Tables 7 and 8 reports the results of the policy reform in response to the shock. Overall, in the economy with full asset market participation, the measure seems to mitigate all the above mentioned effects. On the contrary, in the economy with the minimum investment requirement to access the capital market, an increase in the retirement age sometimes mitigates the effects of the ageing shock on the aggregate statistics, sometimes amplifies them. In particular, the Gini coefficient for private wealth drops further, whereas the Gini coefficient for consumption, which in the aftermath of the shock had increased, falls to a level below the initial one. This is a sign of the heterogenous effects of the policy on the different agents populating the economy.

Figure 12 examines the impact of population ageing and the policy reform on the wealth and consumption distributions. If we focus on the first column, which refers to the frictionless economy, we can observe that the distributions implied by the policy reform, in red, are extremely close to the ones before the population shock, marked in black.

This indicates that the increase of retirement age contains the effects of the longevity shock on inequality. If we shift our attention to the second column, investigating the economy with limited asset market participation, we notice that the policy reform alters the distributions means, as well as their variance. Both mean consumption and wealth are negatively affected by ageing, but the policy reform moderates their decline. If we focus on the comparison between the blue and red curves, it emerges that: 1) the densities for low levels of consumption and wealth seem unaffected by the reform; 2) the densities of agents around the distributions means are increased by the policy; 3) the densities for high levels of consumption and wealth are lowered by the policy.
Finally, Figure 13 reports the effects of ageing and the subsequent policy reform on the two measures of inequality considered in the paper, the Gini coefficient and the interquartile range, along the life-cycle. Once again, in the case of full asset market participation, the inequality implications of the social security reform are not sizeable. When $\tilde{a} = 0.1$ the policy reform reduces the Gini coefficients for consumption and wealth at any age, even relative to the benchmark economy highlighted in black. However, when we consider the interquartile range, the demographic shock reduces consumption and wealth inequality for all age groups while the increase in retirement age increases it. We can conclude that this type of policy brings about distributional consequences that appear to benefit the poorest agents and to make the richest worse off. Such outcome is at odds with the conventional view that an increase in retirement age induces a shift of wealth from poor to rich agents. On the contrary, in this exercise, where the presence of a minimum investment requirement determines the ability of households to access the capital market, this type of policy intervention has progressive effects.

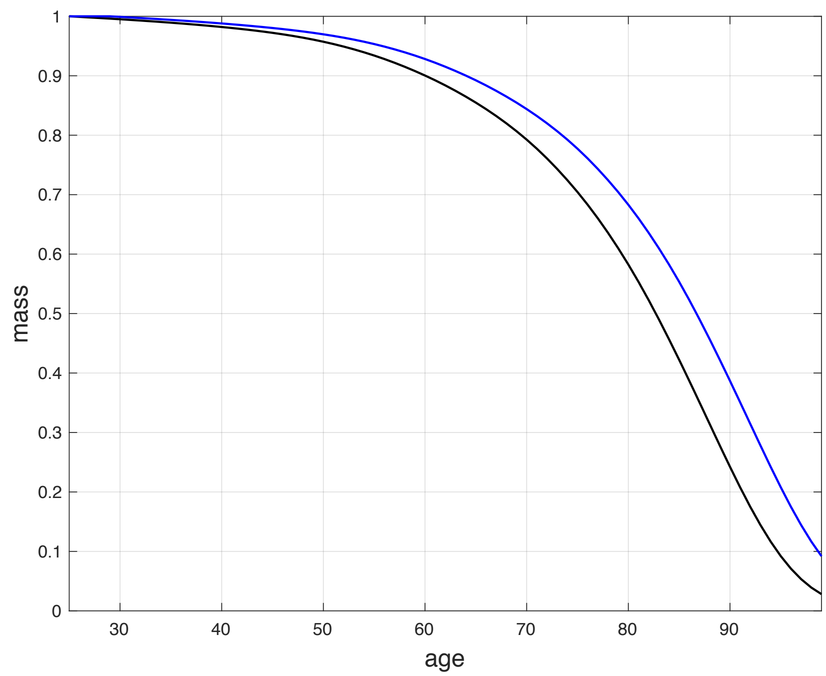
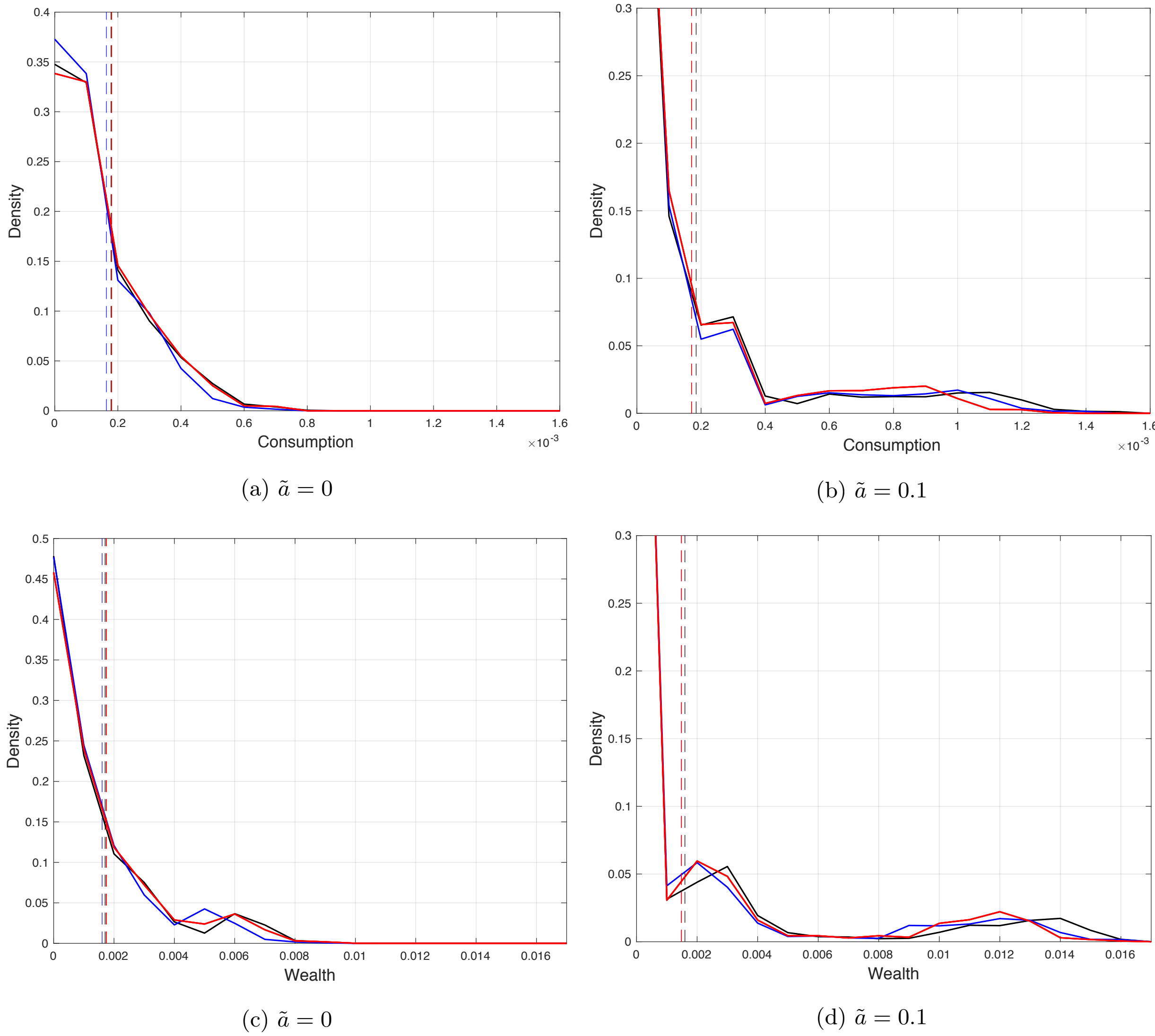


Figure 1.11: In blue, population mass by cohort after the longevity shock

Variable	before shock	after shock	after reform
w	0.0881	0.0837	0.0873
r^K	0.1562	0.1553	0.1560
K	0.1935	0.1750	0.1901
Y	0.0915	0.0823	0.0898
Gini-wealth	0.5412	0.5242	0.5288
Gini-consumption	0.4227	0.4129	0.4164
participation	100%	100%	100%

Table 1.7: Comparative statics with $\tilde{a} = 0$ and $\phi = 0.2$, before/after the demographic shock and after the increase in retirement age



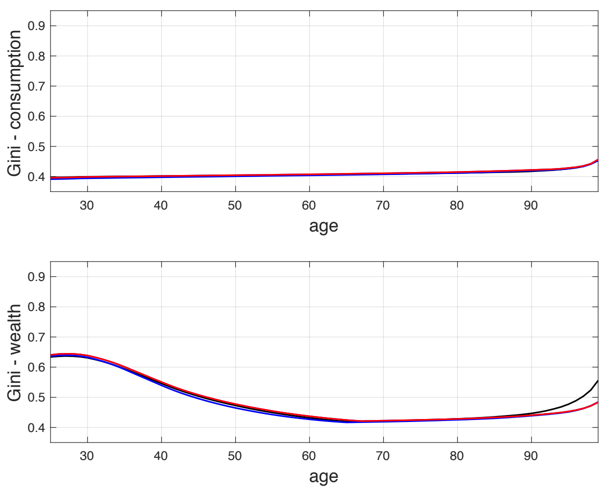
(a) $\tilde{a} = 0$

(b) $\tilde{a} = 0.1$

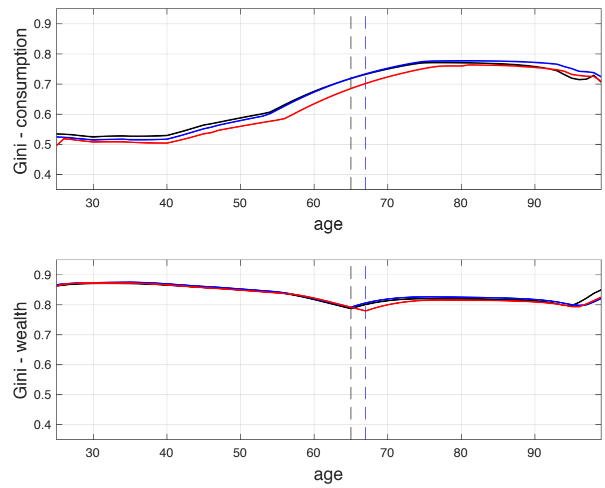
(c) $\tilde{a} = 0$

(d) $\tilde{a} = 0.1$

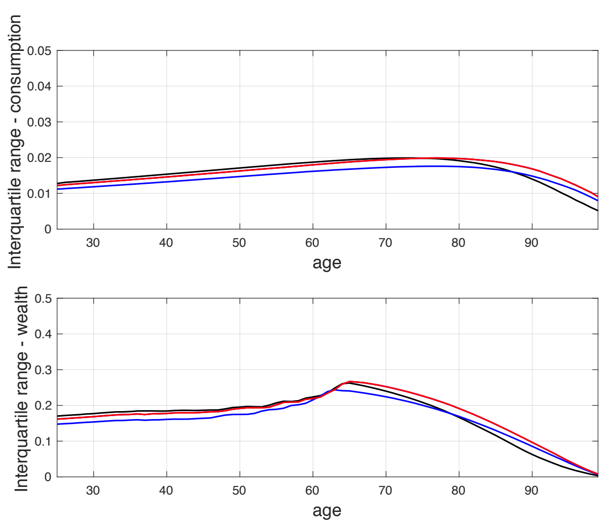
Figure 1.12: The effect of the policy reform on the consumption and wealth distributions, in black before the shock, in blue after the shock, in red after the reform.



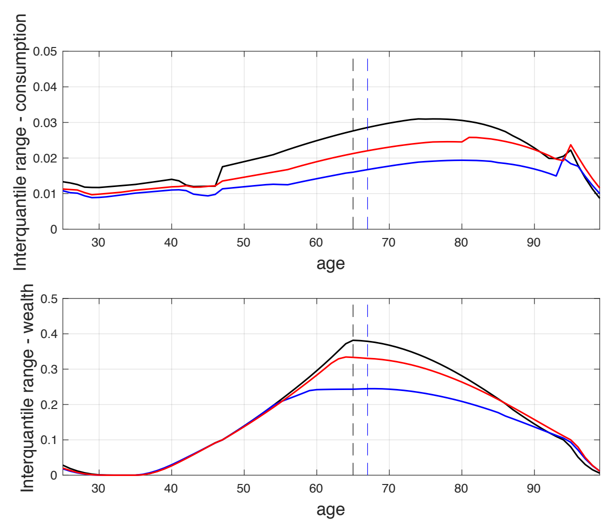
(a) Gini index with $\tilde{a} = 0$



(b) Gini index with $\tilde{a} = 0.1$



(c) Interquartile range with $\tilde{a} = 0$



(d) Interquartile range with $\tilde{a} = 0.1$

Figure 1.13: The effect of the policy reform on the consumption and wealth inequality by age groups, in black before the shock, in blue after the shock, in red after the policy reform

Variable	before shock	after shock	after reform
w	0.0855	0.0808	0.0836
r^K	0.1663	0.1667	0.1704
K	0.1762	0.1574	0.1666
Y	0.0888	0.0795	0.0860
Gini-wealth	0.8502	0.8499	0.8461
Gini-consumption	0.6822	0.6847.	0.6598
participation	24.76%	24.89%	24.87%

Table 1.8: Comparative statics with $\tilde{a} = 0.1$ and $\phi = 0.2$, before/after the demographic shock and after the increase in retirement age

1.5 Conclusions

This project investigates the role played by a minimum investment requirement to access the capital market in the design of the optimal pension system. It develops an overlapping generation model, in the tradition of Auerbach & Kotlikoff, populated by heterogenous households to study distributional issues. In order to derive wealth and income distributions aiming at realistically mimic the US data, the model features a bequest motive and heterogenous labor earnings as proposed by Benhabib, Bisin & Luo (2015). Agents, at birth, differ along two different dimensions: labor earnings path, exogenously fixed, and initial wealth, endogenously determined by bequests. To avoid further complications (and in absence of a reliable calibration) it is assumed that the skill level and the evolution of labor income over the life-cycle are independent of the initial stock of assets. Therefore, each of the different wealth classes at birth faces the same probability to end up in one of the ten labor earnings groups. The existence of an entry barrier to financial markets pushes some agents to adopt a low-yield storing technology to transfer value over time, smooth consumption and save against demographic risk. Social security offers an opportunity to insure agents against longevity uncertainty and to limit the social cost of the minimum investment requirement, but crowds out aggregate capital. Two are the main results of this analysis: the introduction of a PAYG pension system increases the long-term welfare but has distributional implications. Specifically, the determination of the optimal size of social security hinges on the relative densities of the consumers accumulating an amount wealth just below and above the minimum investment requirement. Those households that even in absence of the pension arrangement are not able to satisfy the minimum investment requirement, benefit from the introduction of the scheme. The rich agents that would, in any case, manage to invest their savings in the capital market benefit too, thanks to the endogenous response of the capital return r^K. However, those households who are pushed out of the capital markets because of the presence of social security are worse off.

A second exercise is conducted to evaluate the impact of an increase in retirement age, motivated by population ageing, on wealth and consumption inequality. The outcome challenges the conventional view that this type of measure is regressive. In fact, in presence of the financial friction, the policy reform improves the welfare of the poorest households at the expense of the richest.

Future developments of this line of research will involve going beyond the limitations imposed by the reduced-from assumption of a minimum investment requirement limiting asset market participation. In reality, there is not only heterogeneity in the returns to private savings among those that participate and those who do not, but also within the group of people having access to financial markets, reflecting the role of information acquisition and financial literacy. However some objectives seem more urgent at this stage. Firstly, the model lacks a realistic correlation between asset positions at birth and labor earnings evolution over the life-cycle. So far, this correlation is assumed to be zero. Taking into account the covariance between the distribution of skills at birth and wealth may produce a even more realistic wealth distribution and alter the results on inequality. Secondly, labor supply is exogenous in the model, which limits the external validity of the results that have emerged in the analysis of the simulated longevity shock.

Nonetheless, the outlined model suggests that accounting for portfolio choice, limited asset market participation and heterogenous returns to wealth may be important for the analysis of social security and its impact on inequality.

Bibliography

[1] S Rao Aiyagari. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, 109184(3):659–684, 1994.

[2] Alan Auerbach and Laurence Kotlikoff. *Dynamic Fiscal Policy*. Cambridge University Press, 1987.

[3] Jess Benhabib, Alberto Bisin, and Mi Luo. Wealth distribution and social mobility in the US: A quantitative approach. *NBER Working Paper Series*, 21721:1–32, 2015.

[4] Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents. *Econometrica*, 79(1):123–157, 2011.

[5] Neil Bruce and Stephen J. Turnovsky. Social security, growth, and welfare in overlapping generations economies with or without annuities. *Journal of Public Economics*, 101(1):12–24, 2013.

[6] Ricardo J. Caballero and Emmanuel Farhi. The Safety Trap. *The Review of Economic Studies*, (2016), 2017.

[7] Frank N. Caliendo, Nick L. Guo, and Roozbeh Hosseini. Social security is NOT a substitute for annuity markets. *Review of Economic Dynamics*, 17(4):739–755, 2014.

[8] Giuseppe Carone, Per Eckefeldt, Luigi Giamboni, Veli Laine, and Stephanie Pamies Sumner. Pension Reforms in the EU since the Early 2000's: Achievements and Challenges Ahead. *European Economy Discussion Papers*, 8022(December), 2016.

[9] Juan C. Conesa and Dirk Krueger. Social Security Reform with Heterogeneous Agents. *Review of Economic Dynamics*, 2(4):757–795, 1999.

[10] Mariacristina De Nardi and Fang Yang. Bequests and heterogeneity in retirement wealth. *European Economic Review*, 72:182–196, 2014.

[11] Gauti B. Eggertsson and Neil R. Mehrotra. A Model of Secular Stagnation. 2014.

[12] Gauti B. Eggertsson, Neil R. Mehrotra, and Jacob A. Robbins. A model of secular stagnation: Theory and quantitative evaluation. *American Economic Journal: Macroeconomics*, 11(1):1–48, 2019.

[13] Andreas Fagereng, Charles Gottlieb, and Luigi Guiso. Asset Market Participation and Portfolio Choice over the Life-Cycle. *Journal of Finance*, 72(2):705–750, 2017.

[14] Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity and Persistence in Returns to Wealth. *National Bureau of Economic Research*, 2016.

[15] Hans Fehr, Christian Habermann, and Fabian Kindermann. Social security with rational and hyperbolic consumers. *Review of Economic Dynamics*, 11(4):884–903, 2008.

[16] Martin Flodén. The effectiveness of government debt and transfers as insurance. *Journal of Monetary Economics*, 48(1):81–108, 2001.

[17] Martin Flodén and Jesper Lindé. Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance? *Review of Economic Dynamics*, 4(2):406–437, 2001.

[18] Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. The Dynamics of Inequality. *Econometrica*, 84(6):2071–2011, 2016.

[19] Christian Geppert, Alexander Ludwig, and Raphael Abiry. Secular Stagnation? Growth, Asset Returns and Welfare in the Next Decades. *SAFE Working Paper*, 145:1–38, 2016.

[20] Jean Olivier Hairault and François Langot. Inequality and Social Security Reforms. *Journal of Economic Dynamics and Control*, 32(2):386–410, 2008.

[21] Jonathan Heathcote, Fabrizio Perri, and Giovanni L Violante. Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006. *Review of Economic Dynamics*, 13(1):15–51, 2010.

[22] Robert Hubbard and Kenneth Judd. Social Security and Individual Welfare: Precautionary Savings, Borrowing Constraints and the Payroll Tax. *American Economic Review*, 77(4):630–646, 1987.

[23] Mark Huggett and Gustavo Ventura. On the Distributional Effects of Social Reforms. *Review of Economic Dynamics*, (2):498–531, 1999.

[24] Mark Huggett, Gustavo Ventura, and Amir Yaron. Sources of Lifetime Inequality. *American Economic Review*, 101(December):2923–2954, 2011.

[25] Ayse Imrohoroglu, Selahattin Imrohoroglu, and Douglas H Joines. Time-Inconsistent Preferences and Social Security. *Quarterly Journal of Economics*, 118(2):745–784, 2003.

[26] Fabian Kindermann and Dirk Krueger. High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk. *SSRN Electronic Journal*, 2014.

[27] Dirk Krueger and Felix Kubler. Pareto Improving Social Security Reform when Financial Markets are Incomplete. *The American Economic Review*, pages 737–755, 2006.

[28] Dirk Krueger and Alexander Ludwig. On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. *Journal of Monetary Economics*, 77:72–98, 2016.

[29] Egil Matsen and Øystein Thøgersen. Designing social security - A portfolio choice approach. *European Economic Review*, 48(4):883–904, 2004.

[30] Alisdair McKay. Search for financial returns and social security privatization. *Review of Economic Dynamics*, 16(2):253–270, 2013.

[31] David K. Miles. Funded and Unfunded Pension Schemes: Risk, Return and Welfare. *CESifo Working Paper, No. 239*, 2000.

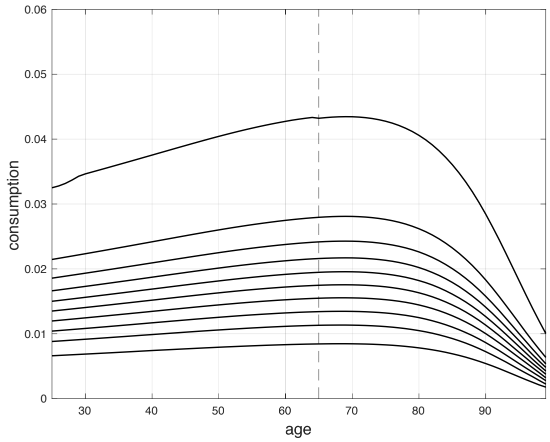
[32] Shinichi Nishiyama and Kent Smetters. Does social security privatization produce efficiency gains? *Quarterly Journal of Economics*, 122(4):1677–1719, 2007.

[33] Shinichi Nishiyama and Kent Smetters. The Optimal Design of Social Security Benefits. *Michigan Retirement Research Center Research*, (Paper No. 2008-197):153–164, 2008.

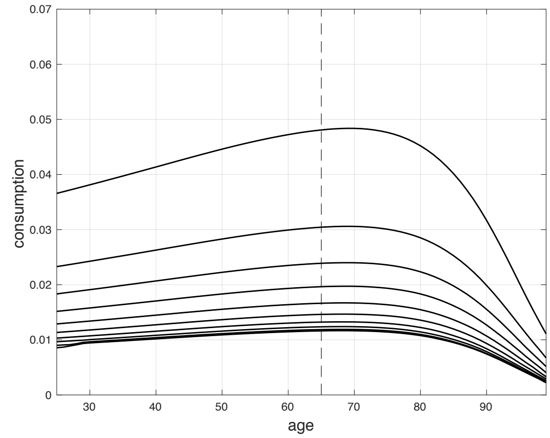
[34] Giorgio E. Primiceri and Thijs van Rens. Heterogeneous life-cycle profiles, income risk and consumption inequality. *Journal of Monetary Economics*, 56(1):20–39, 2009.

[35] Kjetil Storesletten, Christopher I. Telmer, and Amir Yaron. Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, 51(3):609–633, 2004.

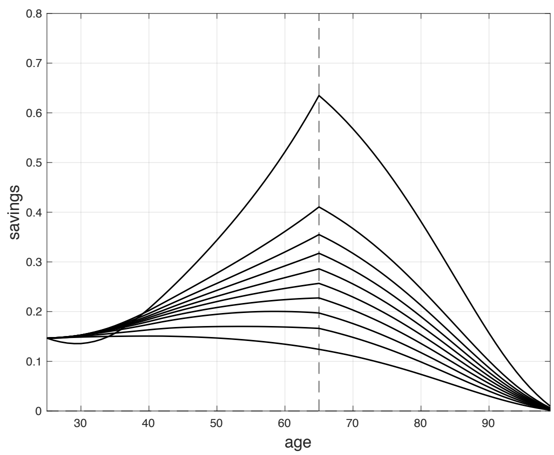
Appendices



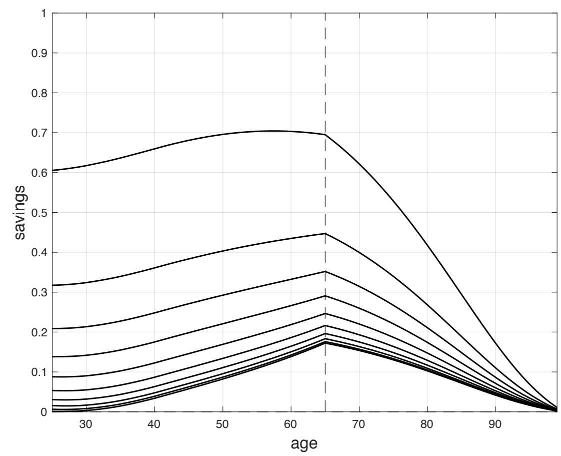
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes



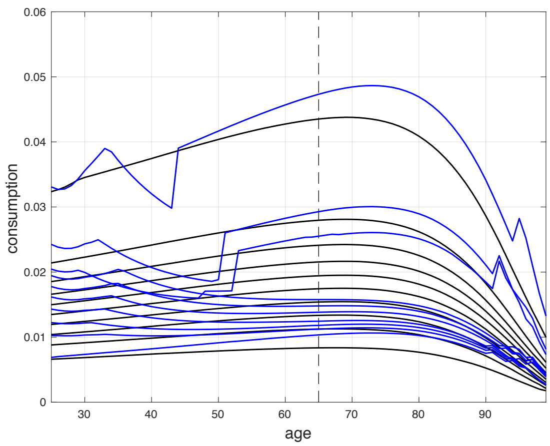
(c) Asset profiles, aggregating for the different initial wealth levels



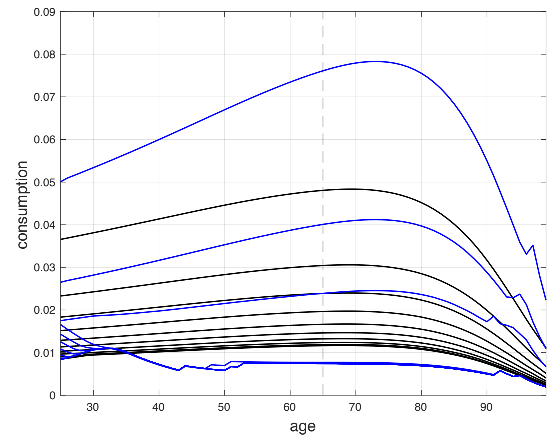
(d) Asset profiles, aggregating for the different labor earning classes

Figure 14: Consumption and asset profiles in absence of social security and of the financial friction.

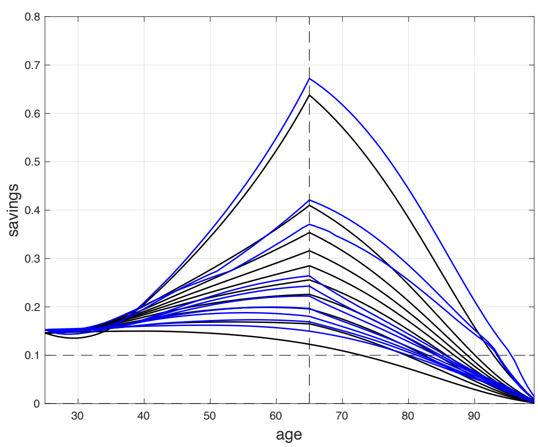
These appendices display the graphs highlighting the effects of introducing the financial friction, social security, population ageing and associated policy reforms on the consumption and asset life-cycle profiles of the different agents examined in the proposed model. For each graph table, the two columns show the role of each source of heterogeneity in determining consumption and wealth inequality along the life-cycle. In practice, in the first column results are reported aggregating the 10 different wealth-at-birth classes, so that the only source of inequality is labor income, while in the second column the 10 different labor earnings classes are aggregated in order to show the importance of initial wealth as a driver of consumption and wealth inequality.



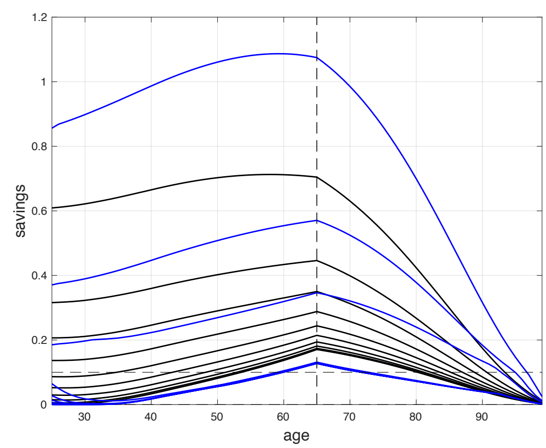
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes

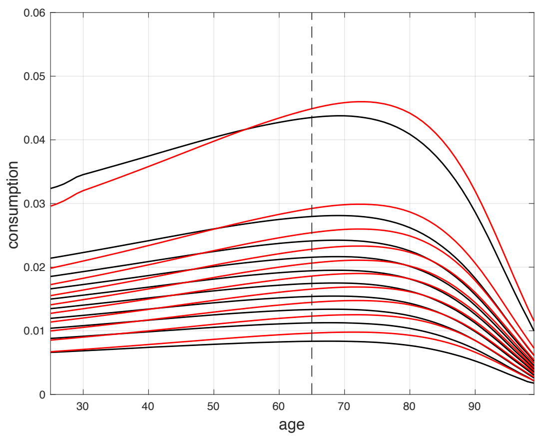


(c) Asset profiles, aggregating for the different initial wealth levels

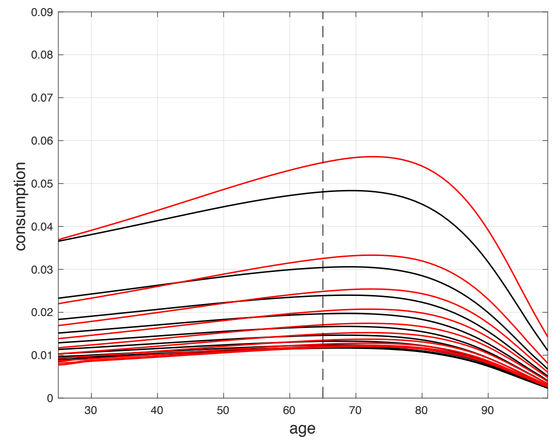


(d) Asset profiles, aggregating for the different labor earning classes

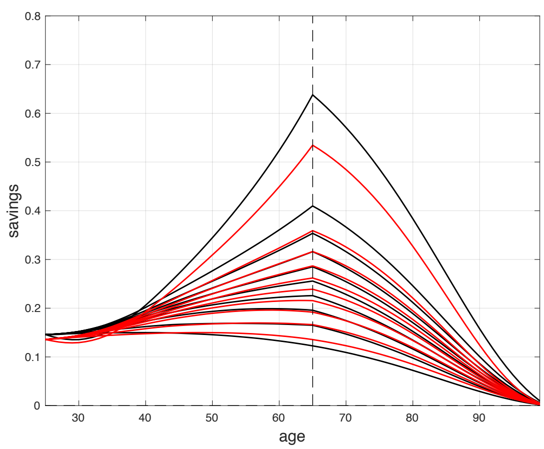
Figure 15: Consumption and asset profiles in absence of social security, in black when $\tilde{a} = 0$, in blue when $\tilde{a} = 0.1$.



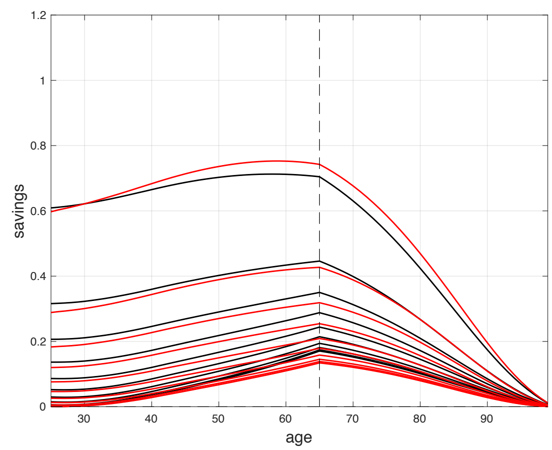
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes

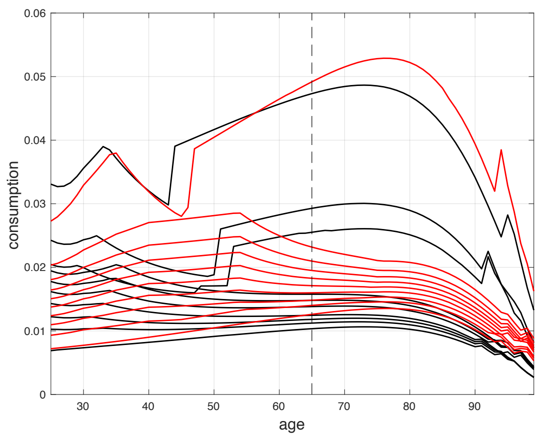


(c) Asset profiles, aggregating for the different initial wealth levels

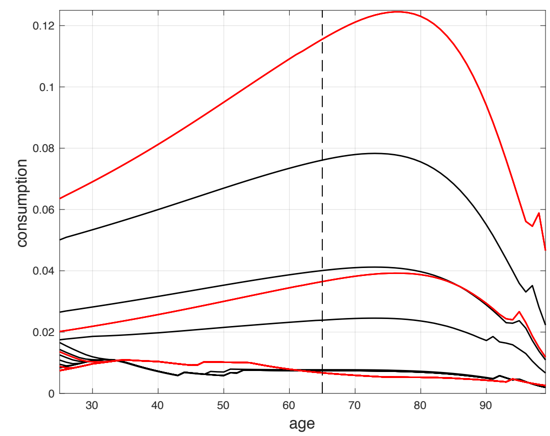


(d) Asset profiles, aggregating for the different labor earning classes

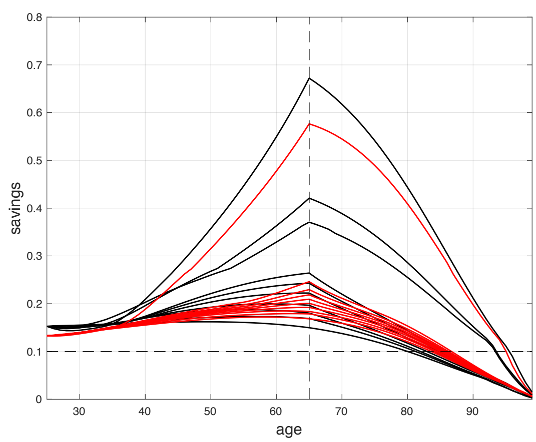
Figure 16: Consumption and asset profiles in absence of the financial friction ($\tilde{a} = 0$), in black when $\phi = 0$, in red when $\phi = 0.2$.



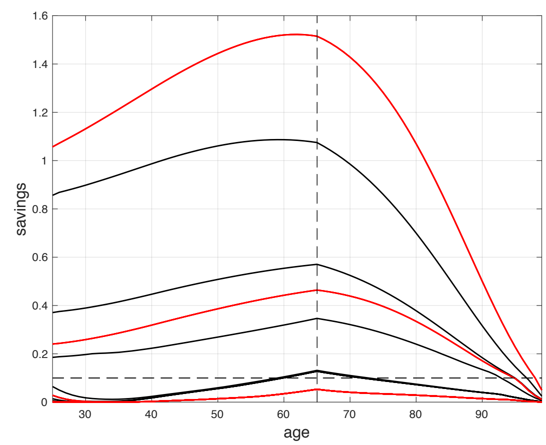
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes

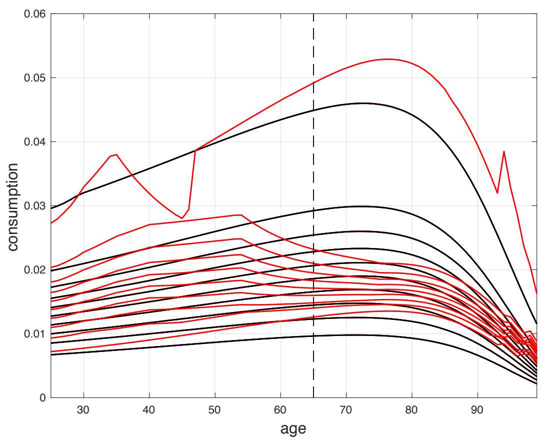


(c) Asset profiles, aggregating for the different initial wealth levels

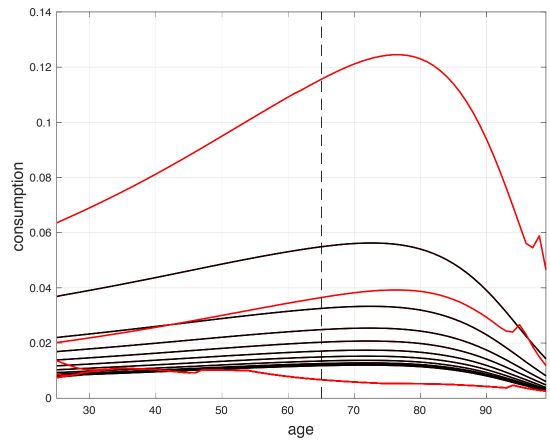


(d) Asset profiles, aggregating for the different labor earning classes

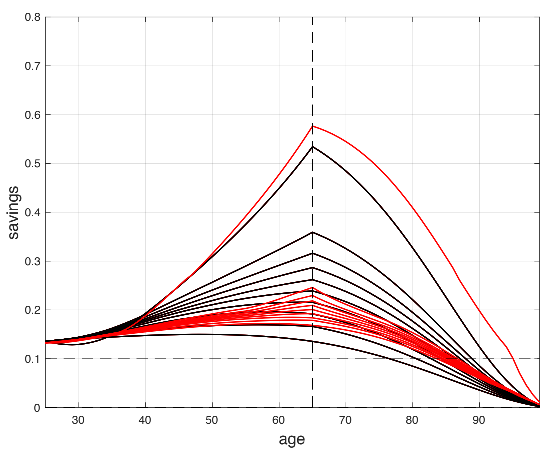
Figure 17: Consumption and asset profiles in presence of the financial friction ($\tilde{a} = 0.1$), in black when $\phi = 0$, in red when $\phi = 0.2$.



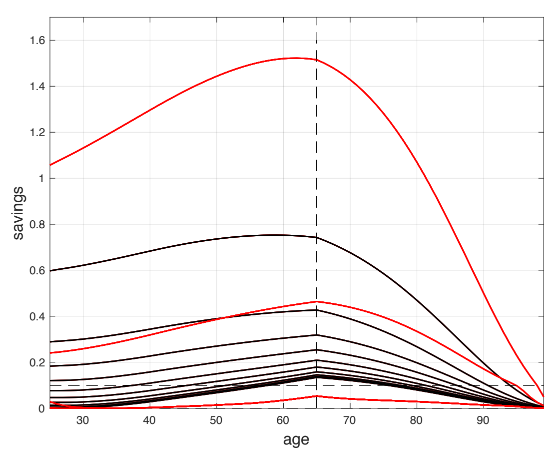
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes

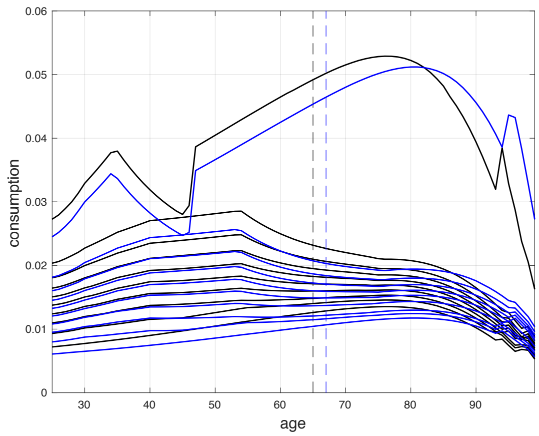


(c) Asset profiles, aggregating for the different initial wealth levels

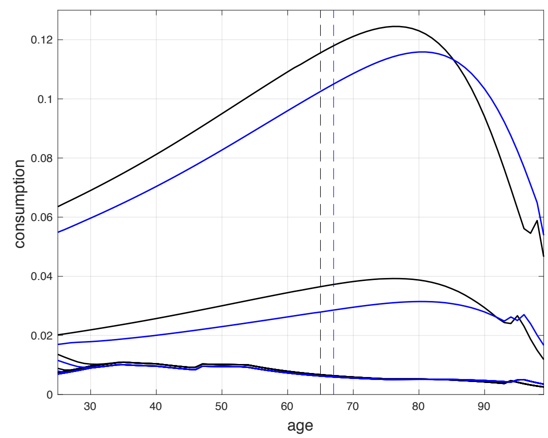


(d) Asset profiles, aggregating for the different labor earning classes

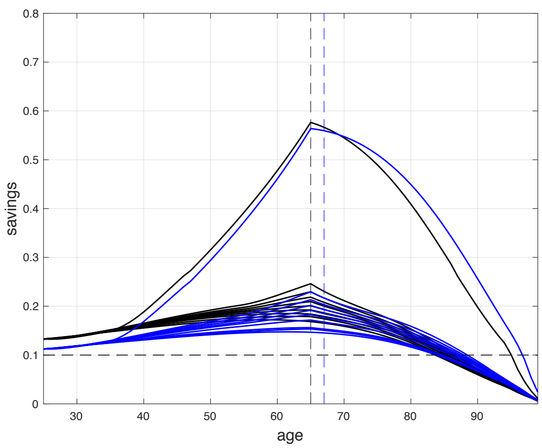
Figure 18: Consumption and asset profiles in presence of social security ($\phi = 0.2$), in black when $\tilde{a} = 0$ and $\phi =$, in red when $\tilde{a} = 0.1$.



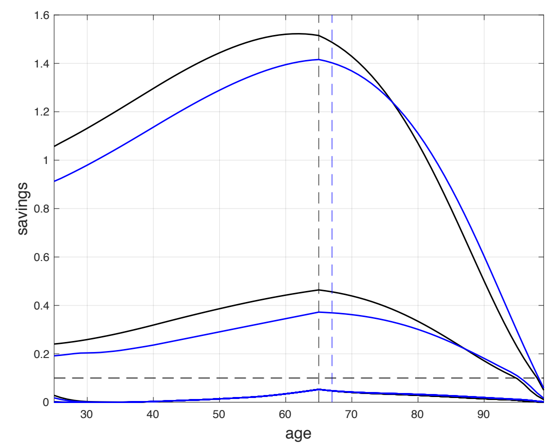
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes

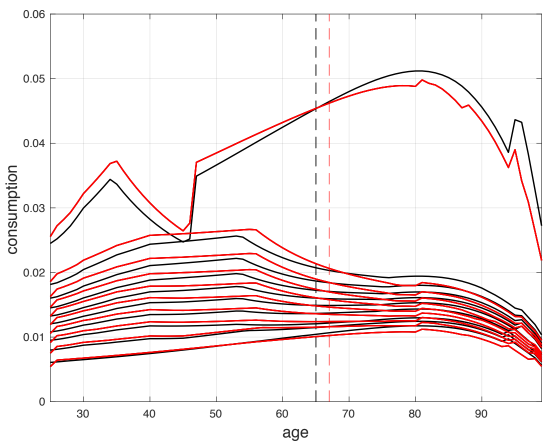


(c) Asset profiles, aggregating for the different initial wealth levels

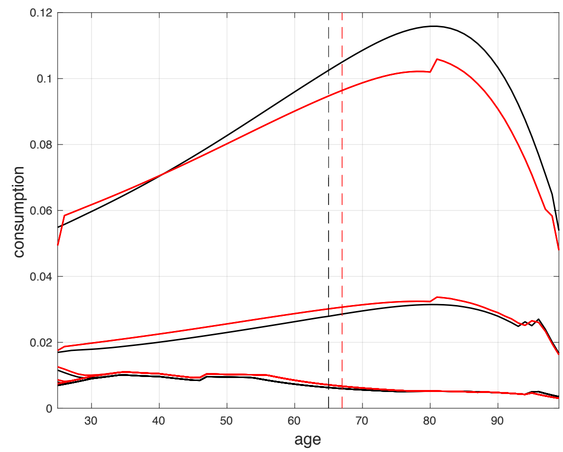


(d) Asset profiles, aggregating for the different labor earning classes

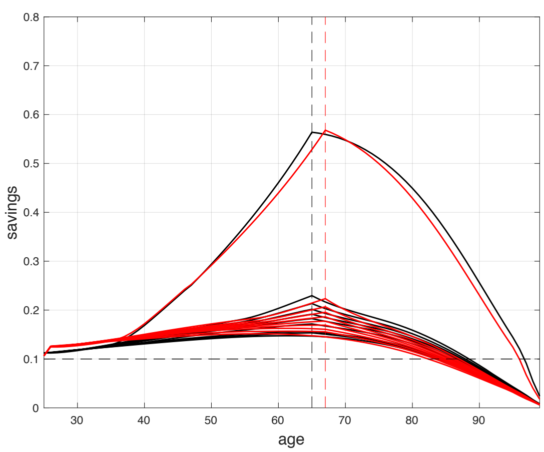
Figure 19: Consumption and asset profiles in presence of social security ($\phi = 0.2$) and the financial friction ($\tilde{a} = 0.1$), in black before the demographic shock, in blue after the demographic shock.



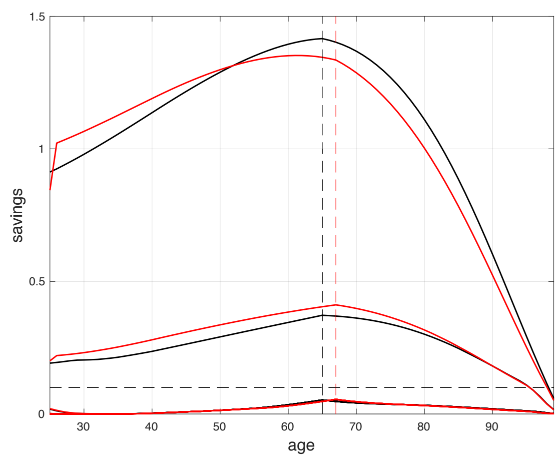
(a) Consumption profiles, aggregating for the different initial wealth levels



(b) Consumption profiles, aggregating for the different labor earning classes



(c) Asset profiles, aggregating for the different initial wealth levels



(d) Asset profiles, aggregating for the different labor earning classes

Figure 20: Consumption and asset profiles in presence of social security ($\phi = 0.2$) and the financial friction ($\tilde{a} = 0.1$), in black after the demographic shock, in red after the policy reform.

Chapter 2

Demographics and Low Interest Rates: the Role Played by Pension Reforms

joint with Jacopo Bonchi

Abstract

Demographic trends in the western world point to a lower population growth rate and a higher life expectancy. On the one hand, these trends increase the weight of middle-aged cohorts with a high stock of savings in the population pushing interest rates down (Baldwin and Teulings, 2014; Gottfries and Teulings, 2015; Lu and Teulings, 2016).[1] On the other hand, they undermine the financial sustainability of the PAYG pension schemes forcing governments to raise the age of retirement and support complementary private pension schemes (e.g. Italy, France and Germany). The aim of this work is to study how these reforms affect the equilibrium real interest rate. In particular, we want to investigate the existence and the quantitative relevance of an additional effect of demographics on interest rates mediated by the pension system. To fully understand the mechanism through which pension reforms affect interest rates, we first build an OLG model with three generations and population growth. Then, we measure the impact of pension reforms on the real interest rate through a quantitative model along the lines suggested by Eggertsson, Mehrotra and Robbins (2019).

[1] They also increase the weight of the old cohorts with lower propensity to save, anyway this would not significantly reduce aggregate savings (Backus et al., 2013; Carvalho et al., 2016).

2.1 Introduction

One of the topics at the center of academic research and policy work is what determines the equilibrium real interest rate and its historical evolution. This matter is of cardinal importance for policy-makers, who design monetary and fiscal policies depending on the response of the economy, in terms of consumption and savings decisions, predicted by the theoretical models they employ. In the aftermath of the 2007 Global Financial Crisis, based on the evidence of a sluggish recovery, inflation rates under target and policy rates at the zero lower bound, the idea of a secular stagnation, initially formulated by Alvin Hansen in 1938 and then proposed again by Larry Summers in 2014 has gained momentum. According to this view, the equilibrium real interest rate compatible with full employment has experienced a secular decline, at times masked by the emergence of speculative bubbles in the financial markets, which culminated with the Global Financial Crisis. Several candidates have been put forward as potential drivers of such phenomenon: population ageing, increases in wealth inequality, a productivity slowdown, a decline in investment goods relative prices, an increase in monopolistic power.

The decline in mortality and fertility rates appears to be an important determinant. Interestingly, demographic trends are also central for the design of unfunded (pay-as-you-go) pension systems, which are the prevailing ones in the western world. An increasingly larger retired population implies that social security, in order to keep its budget balanced, must be adjusted. Whenever keeping the same pension replacement rate through an increase in tax contributions has proven to be unfeasible, different reforms have been implemented: on the one hand parametrical reforms as a reduction in the replacement ratio, on the other hand more structural reforms as increases in the mandatory retirement age or partial privatizations. Such reforms took place in many developed countries and were demanded following the sovereign debt crisis, which exposed the vulnerability of the existing pension arrangements.

The objective of this paper is to study whether and to what extent the pension reforms aimed at restoring the sustainability of the existing systems have mitigated or amplified the impact of demographic trends on the equilibrium real interest rate. We are not aware of other papers that tackle this issue, which represents the main contribution of our analysis. We develop a stylized OLG model with 3 generations to qualitatively identify the mechanisms associated with the alternative pension reforms that can be implemented to restore the financial sustainability when the economy is hit by demographic shocks. We find that the policy to keep the same replacement ratio moderates the effect of ageing on the interest rate, while the one that adjusts the replacement ratio, maintaining the same contribution rate, amplifies the impact of an increase in longevity or drop in fertility on the interest rate. Lastly, if the retirement age is raised to rebalance the ratio between young and old cohorts, the adjustment of the pension scheme neither mitigates nor amplifies the effect of demographics on the interest rate whenever the extended working age duration does not affect lifetime income. Otherwise, the latter reform counteracts the decline in the interest rate produced by the increase in life expectancy or fall in the population growth rate.

Whether such mechanisms are quantitatively important is the second question we attempt to answer in this paper. To this end, we employ a larger OLG model with 81 generations featuring a more realistic population pyramid and a production economy. We calibrate the model to the US economy and we run several quantitative exercises. Firstly, we decompose the decline in the interest rate that took place between 1970 and 2015 and examine the role played by social security reforms. We obtain that on the one hand the increase in retirement age has mitigated the effect of higher longevity and lower fertility on the interest rate, on the other the drop in the replacement ratio, that can be interpreted as a gradual privatization of the pension system, has amplified the impact of demographics. Nonetheless, the importance of these pension system reforms for the determination of the real interest rate seems to be quantitatively dominated by other forces, as the direct effect of the decline in fertility and mortality rates or a productivity slowdown. Secondly, we simulate the demographic changes predicted by the UN for the US population between 2015 and 2060 and study the expected evolution of the equilibrium real interest rate implied by alternative social security reforms. While demographic trends are slow-moving and therefore easily predictable, the evolution of economic variables as productivity is object of speculation. As a consequence, we compare two different scenarios: one where the rate of productivity growth in 2060 remains at the 2015 low level, which we denominate "secular stagnation" scenario and one where it increases to 2%, defined as "normal growth" scenario. This distinction seems to matter for the choice of the pension reform, aimed at keeping the budget balanced, leading to higher average consumption and so higher social welfare. Whenever an increase in the retirement age is unfeasible, keeping a constant replacement ratio seems to be the better option relative to keeping a constant contribution rate in the secular stagnation scenario, while this conclusion is reversed in the normal growth scenario.

The paper is organized as follows: the next subsection on the related literature completes the introduction, section 2 shows the mechanisms produced by pension reforms at work in an ageing economy employing a basic 3-generation OLG model; section 3 develops a richer model used in section 4 to run our quantitative experiments and section 5 concludes.

2.1.1 Related literature

The main literature reference of this paper is Eggertsson, Mehrotra & Robbins (2019) [8], which sets up 2 OLG models to study the determinants of secular stagnation. The first one is a 3-generation model with the purpose of showing how different phenomena can potentially explain the secular decline in the interest rates. In particular, the authors illustrate how even in a simple life-cycle model, differently from its infinitely-lived agent counterpart, a shock to the debt limit, to the population growth rate or to income inequality can affect the equilibrium interest rate *permanently.* The second model they build is a quantitative OLG model, characterized by a more realistic population structure and a production economy, which aims at identifying, quantitatively, the importance of the different determinants that can account for the secular decline of the interest rate in the US since 1970.

Our work adopts exactly the same theoretical framework proposed by Eggertsson, Mehrotra & Robbins, but augments it with a realistic PAYG pension system, as the goal of this project is to explore the role of social security reforms for the determination of the equilibrium interest rate. As demographic trends are crucial for the design of a pension system, we neglect the other potential determinants of secular stagnation and focus on shocks to the fertility rate and life expectancy.

Another important paper in this literature is Carvalho, Ferrero & Nechio (2016) [7], that studies specifically the impact of demographics on savings decisions. They distinguish between the effects of a lower population growth and an increased longevity. They conclude that demographic trends produced a reduction in the equilibrium interest of at least 1.5% in developed economies, based on a life-cycle model along the lines of Gertler (1999) [11], Blanchard (1985) and Yaari (1965) [15]. Differently from their approach, we use an overlapping generation model à la Auerbach-Kotlikoff [2], where the probabilities of retiring and dying are age-dependent. Furthermore, while in Carvalho, Ferrero & Nechio (2016) social security is briefly considered in an extension of the paper, in this work we focus our attention on the pension system and the alternative reforms that can be adopted to restore its financial sustainability in the aftermath of demographic shocks. Therefore we model explicitly a social security budget, which is separated from the government budget.

2.2 Basic model

We have an endowment economy, in which people live for three periods. The size of generations is N_t^i with $i = y, m, o$. The growth rate of population is $1+n = \frac{N_t^y}{N_t^m}$, which also expresses the proportion between the middle and young generations. The middle cohort however, faces a probability $(1-s)$ to die before getting old, which implies that $\frac{N_t^m}{N_t^o} = \frac{1+n}{s}$. During youth, people borrow to finance their consumption and face an exogenous debt limit, D, which reflects the common view about the safe level of leverage. Each period middle-aged households receive an endowment Y, which is used to pay down debt, to consume and save for the retirement. There is just one asset serving as store of value: a one period riskless bonds, sold by the young generation. There is also a government, which runs a pay-as-you go pension system (hereafter PAYG). A lump-sum tax, T, is levied on the middle generation to pay the pension of elderly, θY, which is a fraction θ (replacement ratio) of their income. This implies a balanced budget constraint for the government, which is $N_t^m T = N_t^o \theta Y$ or alternatively $\theta Y = \frac{1+n}{s} T$. The representative household seeks to maximize:

$$\max_{C_t^y, C_{t+1}^m, C_{t+2}^o,} \ln C_t^y + \beta \ln C_{t+1}^m + \beta^2 s \ln C_{t+2}^o$$

s.t.

$$C_t^y = B_t^y \tag{2.1}$$

$$C_{t+1}^m = Y - (1+r_t)B_t^y - B_{t+1}^m - T \tag{2.2}$$

$$C_{t+2}^o = (1+r_{t+1})B_{t+1}^m + \frac{1+n}{s})T \tag{2.3}$$

$$(1+r_t)B_t^i \leq D \tag{2.4}$$

where C_t^i denotes the consumption of each generation and B_t^i the real value of risk-free bonds. Equation (2.4) is the exogenous borrowing limit, which is binding for young households:

$$B_t^y = \frac{D}{(1+r_t)} \tag{2.5}$$

The optimal condition for this problem is:

$$\frac{C_{t+1}^o}{C_t^m} = \beta(1+r_{t+1}) \tag{2.6}$$

The equilibrium in the loan market requires:

$$N_t^y B_t^y = N_t^o B_t^m$$

where the left-hand side is the amount of funds demanded by the young households and the right-hand side the amount supplied by the middle-aged ones. This equation can be rewritten as:

$$(1+n)B_t^y = B_t^m \tag{2.7}$$

Let me denote the loan demand with L_t^d and the loan supply with L_t^s. Using (2.5) to substitute for B_t^y, the loan demand can be expressed as:

$$L_t^d = \frac{1+n}{1+r_t} D \tag{2.8}$$

Combining (2.2), (2.3), (2.5) and (2.6), we obtain the loan supply:

$$L_t^s = B_t^m = \frac{\beta s}{1+\beta s}(Y - D - T) - \frac{1}{1+\beta s}\frac{1+n}{1+r_t}\frac{T}{s} \tag{2.9}$$

By equating loan demand and supply we derive the equilibrium real interest rate:

$$(1+r) = (1+n)\left[\frac{(1+\beta s)\,D + \frac{T}{s}}{\beta s\,(Y - D - T)}\right] \tag{2.10}$$

The effect of a PAYG pension scheme on the loan supply is twofold:

1. it reduces disposable income at middle age (see $-T$ in round brackets in equation 2.9)
2. it provides an income at the old age $(-\frac{1}{1+\beta s}\frac{1+n}{1+r_t}\frac{T}{s})$.

In any case, PAYG scheme decreases loan supply. As a consequence, we have a higher real interest rate than in the case of a fully funded pension scheme ($T = 0$).

2.2.1 Demographic Shock and Pension Reforms

In this subsection we simulate different demographic shocks and study the effects of three alternative pension reforms: a change in the replacement ratio holding contributions constant, a change in the contribution rate keeping the replacement ratio constant and finally a change in the retirement age, keeping both the replacement ratio and the tax rate constant. These parameter changes replicate the salient features of the sustainability-enhancing pension reforms in the Eurozone, which are a direct consequence of demographic trends.[2] In the model, two different types of demographic shocks can be examined: firstly, a longevity shock, captured by an in the probability of survival at middle age s; secondly, a permanent reduction of the population growth rate n.

Change in the replacement ratio

If tax contributions T are kept constant, the assumption of a balanced government budget implies that the replacement ratio θ necessarily falls if either s increases or n fall. In other words, given the same amount of contributions, an increase in longevity or a decrease in fertility negatively affect the return of the PAYG pension system. In our analysis, this represents the benchmark case.

$$\theta Y = \frac{1+n}{s}\overline{T} \tag{2.11}$$

2.2.2 Longevity shock

A surge in the survival probability s has two effects: first of all it increases the effective discount factor, implying that the agents attach greater importance to consumption at old age and save more when at middle age. This first effect occurs even in absence of a pension arrangement. Secondly, by reducing the return of the pension asset, it pushes agents to make up for the loss in terms of consumption at old age with further savings. As a consequence, the supply of loans increases and the interest rate r falls. This second effect is stronger the bigger is the ratio $\frac{1+n}{1+r}$, meaning that if the economy is far from the case of dynamic inefficiency, i.e. if $r >> n$, then the response of savings to the population ageing through this second channel will be quantitatively small. We can conclude that if the government stays inactive and simply accommodates the increase in longevity by reducing the replacement ratio θ, the pension system amplifies the effect of ageing on savings and on the equilibrium interest rate.

[2] "Longer and healthier lives are a remarkable achievement for our societies. Responsible policies are now needed to ensure that pension, healthcare and long-term care systems are financially sustainable and can provide adequate protection for all. There has been considerable progress in the reform of the European social protection models in the last decade, notably in the field of pension. A majority have adapted their systems to better withstand the demographic impact that will become apparent within the next decade. This not only includes general increases in retirement ages, but also restrictions on early retirement. These sustainability-enhancing pension reforms in most Member States can lead to new challenges. Generally, reforms went hand in hand with a streamlining of public pension schemes. To ensure that these reforms will enjoy lasting support and success, other flanking measures are likely to be necessary to maintain retirement incomes such as extending working lives and providing other means of retirement incomes through complementary pension savings. In parallel, Member States need to support the development of collective and individual pension plans to complement public pension schemes, including by removing obstacles at European level. Social partners may have an important role to play here, depending on national practices." (European Commission, Annual Growth Survey, 2016, p. 15)

$$\frac{\partial B^m}{\partial s} = \frac{\beta}{[1+\beta s]^2}(Y-D-T) + T\frac{1+n}{1+r}\left[\frac{1}{s^2} + \frac{\beta}{[1+\beta s]^2}\right] \tag{2.12}$$

Equation (2.12) further shows that increasing levels of taxes T alter the relative importance of the two different channels in favor of the second. Intuitively, a larger pension scheme reduces the resources available for private savings but also amplifies the impact of the pension transfer on lifetime income.

$$\frac{\partial(1+r)}{\partial s} = -(1+n)\left[\frac{D+\frac{T}{s}}{\beta s^2(Y-D-T)}\right] \tag{2.13}$$

Equation (2.13) shows the effect of an increase in longevity on the equilibrium interest rate. In absolute value, the effect is larger the larger is the fertility rate n and the larger are taxes T and debt D relative to income Y.

2.2.3 Fertility shock

Differently from the case of a positive longevity shock, a permanent reduction in n not only affects the supply of loans, but also the demand. As $L^d = \frac{1+n}{1+r}D$, a fall in n determines a lower demand of loans and the impact is stronger the larger is D. This is easily explained: a lower fertility rate implies a smaller young cohort relative to the middle one. On the side of the supply of loans:

$$\frac{\partial B^m}{\partial n} = -\frac{1}{1+\beta s}\frac{1}{1+r_t}\frac{T}{s} < 0 \tag{2.14}$$

Which reveals that a lower fertility rate n increases the supply of loans. Such effect would not be there in absence of the pension system (i.e. if $T = 0$). Therefore, the pension system again amplifies the effect of demographic dynamics on the decline in the equilibrium interest rate r:

$$\frac{\partial(1+r)}{\partial n} = \frac{\big(1+\beta s\big)D + \frac{T}{s}}{\beta s(Y-D-T)} > 0 \tag{2.15}$$

Change in tax contributions

If the replacement ratio θ is kept constant, the assumption of a balanced government budget implies that taxes T necessarily must increase if either s rises or n fall. In other words, given the same amount of social security expenditures, an increase in longevity or a decrease in fertility negatively affect the return of the PAYG pension system and therefore demands for a larger amount of tax contributions.

$$T = \frac{s}{1+n}\overline{\theta}Y \tag{2.16}$$

In order to study the effect of a positive longevity shock on the equilibrium interest rate through an increase in tax contributions, it is convenient to derive the supply of loans as a function of the replacement ratio θ, which is kept constant in this type of reform, using the government budget constraint:

$$L_t^s = B_t^m = \frac{\beta s}{1+\beta s}\left[Y - D\right] - \frac{1}{1+\beta s}\theta Y\left(\frac{1}{1+r} + \frac{s}{1+n}\beta s\right) \tag{2.17}$$

Combining equations (2.17) and (2.8), we obtain the equilibrium interest rate:

$$(1+r) = (1+n)\left[\frac{\left(1+\beta s\right)D + \frac{1}{1+n}\theta Y}{\beta s\left[Y\left(1-\theta\frac{s}{1+n}\right) - D\right]}\right] \tag{2.18}$$

2.2.4 Longevity shock

Equation (2.17) shows that when, in response to a rise in s and for a given interest rate r, the government keeps the replacement ratio constant by increasing taxes, the impact on savings and the loans supply is uncertain. In particular, while the increase in the discount factor pushes agents to save more (the term $\frac{\beta s}{1+\beta s}\left[Y - D\right]$ increases), the policy reaction may amplify or mitigate the effect of population ageing on the supply of loans, depending on the change in the term $-\frac{1}{1+\beta s}\theta Y\left(\frac{1}{1+r} + \frac{s}{1+n}\beta s\right)$. The intuition behind this is the following: due to the change in the effective discount rate βs, agents attach more importance now to old age consumption, which pushes them to save more. However, the increase in s also reduces the return of each unit invested in the pension system, causing a contraction of lifetime income. Hence, an increase in T may incentivize the agent to optimally reduce savings, given that the pension scheme pays the same benefit θY and the decline in disposable income $Y - D - T$ reduces the resources available for consumption at middle age. This second effect is stronger if taxes T are already high relative to middle age net income $Y - D$, and if the return of the pension system is relatively low (low n, for example).

We can compare equation (2.18) with the equilibrium interest rate emerging in our benchmark case, given by equation (2.10). If the government aims at keeping a constant replacement ratio and s rises to s', taxes must necessarily increase relative to the benchmark economy: $T' = \frac{s'}{1+n}\theta Y > T$. As a consequence:

$$\left[\frac{\left(1+\beta(s')\right)D+\frac{1}{1+n}\theta Y}{\beta(s')\left[Y\left(1-\theta\frac{s'}{1+n}\right)-D\right]}\right] > \left[\frac{\left(1+\beta(s')\right)D+\frac{T}{s'}}{\beta(s')\left(Y-D-T\right)}\right] \tag{2.19}$$

which implies that, in presence of the same longevity shock, the equilibrium interest emerging from a policy reform consisting of an increase in tax contributions is unambiguously higher than the one obtained if only the replacement ratio is adjusted to keep the government budget balanced. This is explained by the fact that when taxes are increased, the supply of loans increases less than in the benchmark case, or may even decrease.

2.2.5 Fertility shock

Again, a permanent fall in n reduces the demand of loans as $\frac{\partial L^d}{\partial n} = \frac{D}{1+r} > 0$. However, differently from the case of constant tax contribution, it also decreases the supply, given that:

$$\frac{\partial B^m}{\partial n} = \frac{\beta s^2 \theta Y}{[1+\beta s](1+n)^2} > 0 \tag{2.20}$$

This effect is due to the fact that a reduction in n implies a reduction in the pension return. But, since the pension benefit is kept constant at θY, taxes increase and shrink lifetime income. As a consequence the agent optimally reduces consumption at middle and old age, by reducing savings. A lower supply of loans, given the same drop in loans demand, following the same fertility shock, means that the equilibrium interest rate is higher than in the benchmark case, where tax contributions are constant.

Change in retirement age

We assume the government, after the positive longevity shock, or the negative fertility shock, wants to keep the replacement ratio constant ($\overline{\theta}$) without raising taxes ($\overline{T}$) and changing the proportion between middle-aged and old households (N^{mr} and N^{or} are the size of the middle and the old generation after the reform). This is a simple way to capture a change in the retirement age. In particular Δ agents remain middle-aged and they will receive pension in the next period instead of the current one. Before the demographic shock, the balanced budget constraint was:

$$\frac{N^m}{N^o}\overline{T} = \frac{1+n}{s}\overline{T} = \overline{\theta} Y$$

After the change in the retirement age, it is:

$$\frac{N^{mr}}{N^{or}}\overline{T} = \left(\frac{N^m + \Delta}{N^o - \Delta}\right)\overline{T} = \overline{\theta}Y \tag{2.21}$$

2.2.6 Longevity shock

Let us assume that s increases by δ, such that $s' = s + \delta$ and $\delta > 0$. Equation (2.21) imposes that:

$$\Delta = \frac{1}{2}\frac{\delta}{1+n}N^m = \nu N^m$$

where $\nu = \frac{1}{2}\frac{\delta}{1+n}$. The change in the size of middle-aged households affects the equilibrium in the loan market, which now requires:

$$N^y B^y = N^{mr} B^m = (N^m + \Delta)\, B^m = N^m(1+\nu)B^m$$

or equivalently:

$$\frac{1+n}{1+\nu}B^y = B^m$$

As a consequence, the loan demand becomes:

$$L^d = \frac{(1+n)}{(1+\nu)(1+r)}D$$

As $\nu > 0$, the new loan demand, resulting from the longevity shock and the increase in the retirement age, is lower than in our benchmark case, where $\nu = 0$.
In order to determine what happens to the loans supply, in the aftermath of an increase in the retirement age, two extreme cases should be examined:

1. the individual income Y, earned by all middle-aged agents, is not affected by the change in the relative size of the different cohorts;

2. the individual income Y is reduced, as the middle-aged cohort is larger, and aggregate resources remain the same.

In our benchmark economy, aggregate income $N^m Y$ is assumed to grow over time at the rate n. In the first scenario outlined above, after the pension reform, aggregate income $N^m Y$ starts growing at the rate $n+\nu$, so that each middle-aged individual keeps receiving the amount $Y' = Y$. Instead, in the second scenario, aggregate resources keep growing at the rate n. Then individual income must fall:

$$N^m Y = N^{mr} Y' = (N^m + \Delta)Y' = N^m(1+\nu)Y'$$

which implies that:

$$Y' = \frac{Y}{1+\nu}$$

The loan supply can be rewritten as follows:

$$L_t^s = B_t^m = \frac{\beta(s+\delta)}{1+\beta(s+\delta)}(Y' - D - T) - \frac{1}{1+\beta(s+\delta)}\frac{\theta Y}{1+r_t} \tag{2.22}$$

1. Under the first scenario, where $Y' = Y$ a positive $\delta > 0$ implies that the loan supply increases due to the increase in the discount factor, but not as much as in the benchmark economy, as in that case there was also an endogenous response of the pension system.

2. Under the second scenario, where $Y' = \frac{Y}{1+\nu}$ a positive $\delta > 0$ implies that the loan supply increases due to the increase in the discount factor, but not as much as under the first scenario. This follows from the fact that if the individual income falls to $Y' = \frac{Y}{1+\nu} = \frac{2(1+n)Y}{2(1+n)+\delta}$, a positive δ reduces lifetime income and savings too.

2.3 Quantitative model

This section develops a medium-scale life-cycle model to study the quantitative importance of demographic dynamics in explaining the fall in the equilibrium interest rates through the pension reform channel identified, qualitatively, in the previous section. Consistently with the analysis conducted above:

- two different demographic phenomena are examined: a decrease in the fertility rate and an increase in longevity;
- three different pension reforms are considered, all aimed at maintaining a balanced pension budget: a reduction of the replacement ratio keeping the tax rate constant, an increase in the tax rate keeping the same replacement ratio and, finally, an increase in the retirement age.

The theoretical framework we employ follows the tradition of Auerbach & Kotlikoff (1987) and is augmented with all the further ingredients introduced by Eggertsson, Mehrotra and Robbins (2019) in their quantitative evaluation of the Secular Stagnation hypothesis. However, here we explicitly model a public pay-as-you-go pension system as its reforms constitute the main focus of our analysis. This implies that the government disposes of two separate budgets: a first one for its spending, which is financed by issuing bonds and levying taxes on labor income, and a second one for paying the pension benefits to retirees, again through labor income taxation. The following subsections outline in detail the structure of the model.

2.3.1 Demographics

Households are subject to a life-cycle. They enter the economy at age 26, have kids at the same age, and participate to the labor market until age RA, when they retire and start receiving their pension benefits until death. The maximum age that households can reach is J. The demographic process which determines the ratio between workers and retirees is defined by a fertility rate and a series of age-specific survival probabilities $\{s_t(j)\}_{j=26}^{J}$ with $s_t(J) = 0$. Therefore, if $N_t(j)$ is the number of households aged j at time t, at time $t+1$ there will be $N_{t+1}(j+1) = s_t(j)N_t(j)$ households aged $j+1$. Associated to the fertility rate at each point in time there is a growth rate n_t which determines the size of the cohort aged 26 at time $t+1$ relative to the one aged 26 at time t: $N_{t+1}(26) = (1+n_t)N_t(26)$. In a stationary equilibrium the fertility rate and the survival probabilities do not depend on time. As a consequence, if $N(j)$ is the mass of agents aged j, in a stationary equilibrium:

$$N(j+1) = \frac{s(j)}{1+n}N(j)$$

for $j[25, J-1]$ and a given $N(26)$. We normalize $N(26)$ such that total population $N = \sum_{j=26}^{J} N(j)$ equals 1, meaning that the mass of each cohort is also the fraction relative to overall population $\frac{N(j)}{N}$.

2.3.2 Households

There is a representative household for each cohort. The problem faced by a newborn household at time t is the following:

$$\max_{c_{t+j-1}(j), a_{t+j}(j), x_{t+J-1}(J)} U = \sum_{j=26}^{J} s^j \beta^j u\big(c_{t+j-1}(j)\big) + s^J \beta^J \mu v\big(x_{t+J-1}(J)\big)$$

such that, when young and working ($j \leq RA - 26 + 1$):

$$c_t(j) + a_{t+1}(j+1) + TFR_{t-j+26}(26)x_t(j) =$$
$$(1 - \tau_t^b - \tau_t^p)w_t hc(j) + \Pi_t(j) + [r^k + \epsilon_t(1-\delta)]\Big(a_t(j) + q_t(j) + \frac{1-s(j)}{s(j)}a_t(j)\Big)$$

and when old and retired ($RA - 26 + 2 \leq j \leq J$):

$$c_t(j) + a_{t+1}(j+1) + TFR_{t-j+26}(26)x_t(j) =$$
$$pension_t(j) + [r^k + \epsilon_t(1-\delta)]\Big(a_t(j) + q_t(j) + \frac{1-s(j)}{s(j)}a_t(j)\Big)$$

where RA is retirement age and $a_t(26) = 0$. Utility from consumption is CRRA: $u(c_t(j)) = \frac{c_t(j)^{1-\frac{1}{\rho}}}{1-\frac{1}{\rho}}$ and so is utility from bequests: $v(x_t(J)) = \frac{x_t(J)^{1-\frac{1}{\rho}}}{1-\frac{1}{\rho}}$. Households leave bequests $x_t(J)$ only when they reach the maximum age J and receive inheritances $q_t(j)$ at age 57, which corresponds to model age 32. Moreover, households are assumed to participate to annuity markets so that involuntary bequests are shared among the surviving members of the same cohort. In this way, they insure themselves against the idiosyncratic age-dependent risk of death.

When young, agents earn firms' profits, purchase assets and supply exogenously their labor endowment. Their labor income depends on the age-dependent labor efficiency level $hc(j)$, the wage w_t and the tax rates τ^b and τ^p. The former tax is paid to finance government consumption, while the latter is the pension system contribution. After retirement, the only source of income is the proceedings from their investment decisions and the pension benefits. Finally, households face a borrowing limit:

$$a_t(j) \geq \frac{D_t}{1+r_t}$$

2.3.3 Firms

Three types of firms populate the supply side of the economy: final goods firms, intermediate goods firms and capital goods firms.

Final goods firms

Final goods firms purchase intermediate goods, transform them in differentiated final goods and operate in a regime of monopolistic competition. The different varieties are combined with a CES aggregator:

$$Y_t = \left[\int_0^1 y_t^f(i)^{\frac{\theta_t - 1}{\theta_t}} di \right]^{\frac{\theta_t}{\theta - 1}}$$

These firms can set the price in each period and face the following demand curve:

$$y_t^f(i) = Y_t \left(\frac{p_t(i)}{P_t} \right)^{-\theta_t}$$

The problem of the final good producer is the following:

$$\max_{p_t(i), y_t^f(i)} \frac{p_t(i)}{P_t} y_t^f(i) - \frac{p_t^{int}}{P_t} y_t^m$$

such that:

$$y_t^f(i) = y_t^m$$

$$y_t^f(i) = Y_t \left(\frac{p_t(i)}{P_t} \right)^{-\theta_t}$$

where the first constraint indicates that these firms employ a simple linear production technology of intermediate goods to produce their output. The solution of the problem above implies that the price is set charging a mark-up over the marginal cost:

$$\frac{p_t(i)}{P_t} = \frac{\theta_t}{\theta_t - 1} \frac{p_t^{int}}{P_t}$$

The intermediate good is homogeneous, therefore all final goods producers make the same pricing decision, so that $p_t(i) = P_t$. It follows that:

$$\frac{p_t^{int}}{P_t} = \frac{\theta_t - 1}{\theta_t}$$

Hence, aggregate profits are $\Pi_t = \frac{Y_t}{\theta_t}$. These profits are assumed to be distributed among households in proportion to their labor income. Thus:

$$\frac{Y_t}{\theta_t} = \sum_{j=26}^{J} N_{j,t} \Pi_{j,t}$$

Intermediate goods firms

The intermediate goods sector is perfectly competitive, hires labor and capital and transforms them into an intermediate good sold to final goods producers through a CES production function. The problem faced by the intermediate producer is:

$$\max_{L_t, K_t} \quad \frac{p_t^{int}}{P_t} Y_t - w_t L_t - r_t^K K_t$$

such that:

$$Y_t = \left[\alpha (A_{k,t} K_t)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(A_{l,t} L_t)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

It follows that intermediate goods factors of production are paid their marginal products:

$$w_t = \frac{p_t^{int}}{P_t} (1-\alpha) A_{l,t} \left(\frac{Y_t}{L_t} \right)^{\frac{1}{\sigma}}$$

$$r_t^K = \frac{p_t^{int}}{P_t} \alpha A_{k,t} \left(\frac{Y_t}{K_t} \right)^{\frac{1}{\sigma}}$$

For a no-arbitrage condition, the expected return of capital investment and risk-free bonds is the same:

$$1 + r_t = \frac{r_t^K + (1-\delta)\xi_t}{\xi_{t-1}}$$

Capital goods firms

In a perfectly competitive investment-specific production sector, the composite final good is converted into capital goods, using a linear production function:

$$\max_{Y_t^K} \quad \xi_t K_t - Y_t^K$$

such that:

$$K_t = z_t Y_t^K$$

where z_t is the productivity in the investment-specific production sector. The capital stock evolves over time according to the following law of motion:

$$K_{t+1} = (1-\delta) K_t + \frac{I_t}{\xi_t}$$

2.3.4 Government

The government manages two separate budgets. The first one is used to finance an exogenously given level of government expenditure G_t either through debt B_t or taxing labor income T_t. The tax rate τ^b is set in order to keep the budget balanced:

$$B_{t+1} = (1 + r_t)B_t + G_t - T_t$$

where $T_t = \tau_t^b w_t L_t$ and L_t is aggregate labor. The second budget is for social security, which consists of a standard pay-as-you-go system transferring resources from working to retired generations:

$$\tau_t^p w_t \sum_{j=26}^{RA} N_t(j) hc(j) = \nu \sum_{j=RA+1}^{J} N_t(j) \frac{1}{RA - 26 + 1} \sum_{i=1}^{RA-26+1} (1-\tau_{t-j+i}^p - \tau_{t-j+i}^b) w_{t-j+i} hc(25+i)$$

The left hand side of the above equation corresponds to the total amount of pension contributions paid by working households to social security at time t. The right hand side is the total amount of pensions benefits paid by the government to retirees. In particular, each pensioner receives a fraction, the replacement ratio ν, of the average labor income after contributions and taxes earned during the working age. Given ν and the past tax rates $\{\tau_j^p, \tau_j^b\}_{j=t-J+1}^{t-1}$, the government sets the current period pension contributions tax rate τ_t^p to keep also the second budget balanced.

2.3.5 Calibration

The introduction of a realistic pay-as-you-go system in the quantitative model set up in Eggertsson, Mehrotra and Robbins (2019) implies that the model needs to be recalibrated in order to perform the quantitative exercises in the next section. As in the original model, some parameters are estimated using the available data directly, some others are taken from the literature which has already estimated or calibrated them, while the values of the others are obtained minimizing a loss function having as arguments different data moments that we aim to match. We calibrate the model to the US economy in 1970 and in 2015, as one of the objectives of the analysis conducted here is to establish the quantitative importance of the different factors that may have influenced the secular decline in the real interest rate since 1970. Table 3.1 summarizes the results of the calibration. One of the focus of this paper is to study how the pension reforms that took place since 1970 in response to ageing and fertility shocks may have affected the savings decisions of consumers during their lifetime. As a consequence, a special attention is given to how the retirement age has changed since 1970, as well as to changes in the replacement ratio.

The US Social Security regulation establishes a full retirement age, which depends on the year of birth, but also allows agents to anticipate of a few years their exit from the labor markets at the cost of a reduction in the pension benefits paid to the retirees. In 2015, full retirement age for a person in her 60s was around 66 years and the early retirement age was 62. Data on 1970 is fuzzier as there was a large fraction of women who did not take part to the official labor force, while US Census Bureau data suggests that the average retirement age for men was around 65. However average statistics on retirement age can be misleading as the distribution is left-skewed, so a more telling indicator would be the median. Still, data lacks information on those that retire really late and is significantly affected by the decision to retire early of those who suffer from health issues and by the so-called survivorship bias. Given these considerations, we calibrate the effective retirement age in the model at 63 in 1970 and 65 in 2015. Such increase in the retirement age between 1970 and 2015 is modest and leads to a relatively conservative estimate of the contribution of this pension reform in the overall decline in the real interest rate.

2.4 Quantitative exercises

The basic model laid down in the second section of this paper examined, qualitatively, whether different pension reforms, aimed at maintaining a the pension budget balanced, produce an amplification or mitigation mechanism of the effect of demographic trends on the equilibrium interest rates. It concluded that:

1. a reform that keeps the same contribution rate by reducing the replacement ratio amplifies the impact of demographics on aggregate savings and the equilibrium interest rate;

2. a reform that keeps the replacement ratio by increasing the contribution rate mitigates the impact of demographics on aggregate savings and the equilibrium interest rate;

3. a reform that keeps the same contribution rate and replacement ratio by increasing the retirement age neither amplifies nor mitigates the effect of the demographic transition on savings and the interest rate, if lifetime income is unaffected by the policy, or mitigates if lifetime income falls as a result of the reform.

Parameter	Symbol	1970 value	2015 value	
Parameters estimated directly from the data				Source
Mortality profile	s_j			US mortality tables, CDC
Income profile	hc_j			Gourinchas and Parker (2002)
Total fertility rate	n	2.8	1.8	UN fertility data
Productivity growth	g	2.02%	0.65%	Fernald (2012)
Government debt (percent of GDP)	G	21.3%	20.75%	CEA
Public debt	b	42%	90%	Flow of Funds
Retirement age	RA	63	65	US Census Bureau
Replacement ratio	ν	45%	40%	US Social Security
Parameters taken from the literature				Source
Elasticity of intertemporal substitution	ρ	0.75	0.75	Gourinchas and Parker (2002)
Capital/labor elasticity of substitution	σ	0.6	0.6	Antras (2004)
Depreciation rate	δ	12%	12%	Jorgenson (1996)
Parameters calibrated matching data moments				
Rate of time preference	β	0.98	0.99	
Borrowing limit (percent of annual labor income)	D	12.6%	26.9%	
Bequests parameter	μ	21.62	13.25	
Retailer elasticity of substitution	θ	8.6	4.89	
Capital share parameter	α	0.19	0.24	
Targets				Source
Natural rate of interest		-2.62%	-1.47%	FED
Investment-to output ratio		16.8%	15.9%	NIPA
Consumer-debt-to-output ratio		4.2%	6.3%	Flow of Funds
Labor share		72.4%	66%	Elsby, Hobijin and Sahin (2013)
Bequests-to-output ratio		3%	3%	Hendricks (2001)

Table 2.1: Calibration

The current section employs the medium-scale life-cycle model outlined in section 4 to study, quantitatively, the importance of the amplification or mitigation effects generated by the alternative pension reforms following changes in life expectancy and in the fertility rate. The analysis consists of comparative statics between long-run steady-state equilibria. Three quantitative experiments follow. The first one is a positive exercise, where we try to determine the importance of the mechanisms under examination for the secular decline in interest rates that occurred since 1970 in a similar fashion to Eggertsson, Mehrotra and Robbins (2019). Differently from their paper, we account for social security and its reforms.

Then we conduct two normative exercises. Firstly, we simulate the response of the economy to the demographic trends expected for the next 30 years in the US, distinguishing between the effect of a marked increase in longevity and a modest increase in fertility. We focus our attention on how alternative pension reforms affect the long-term equilibrium. Secondly, we compare a 2060 steady-state characterized by a "pessimistic" prediction of the US productivity growth rate, which is assumed to stay at the 2015 level of 0.65% per year, with a "normal" one, where the productivity growth rate is assumed to be 2%. For both scenarios, we account for the expected evolution of demographic variables, featuring an increase in life expectancy and in the fertility rate taking place at the same time. The purpose of this test is to verify the impact of alternative pension reforms when, in addition to the expected permanent demographic shocks, we account for a stagnating GDP or a more optimistic economic outlook.

Throughout the normative exercises, the results are compared with a fictitious stationary equilibrium of the economy, where the demographic shocks have taken place but the government remains completely inactive on pensions (no change in the contribution rate, replacement ratio or mandatory retirement age) and accumulates an unfunded debt, allowing for a pension budget deficit in every period.

2.4.1 Positive exercise

Similarly to the analysis conducted in Eggertsson, Mehrotra and Robbins (2019), we perform of a decomposition of the different factors that contributed to the decline in the interest rate observed since 1970, with a specific focus on the role played by pension reforms. In order to quantify the impact of the different drivers, we take the 2015 stationary distribution and we shock, one at a time, the exogenous parameters associated with the different phenomena that potentially explain the fall in the interest rate, keeping all the other parameters constant. Concerning the endogenous response of

		Our results	Eggertsson, Mehrotra & Robbins (2019)	
Forcing variable	Δ in r	Percent of total Δ	Δ in r	Percent of total Δ
Total interest rate change	-4.09%	100	-4.02%	100
Mortality rate	-1.06%	26	-1.82%	43
Total fertility rate	-1.02%	25	-1.84%	43
Productivity growth	-1.22%	30	-1.90%	44
Government debt (percent of GDP)	+0.95%	-23	+2.11%	-49
Labor share	-0.56%	14	-0.52%	12
Relative price of investment goods	-0.41%	10	-0.44%	10
Debt limit	+0.14%	-3	+0.13%	-3
Replacement ratio	-0.20%	5	-	-
Retirement age	+0.38%	-9	-	-

Table 2.2: Interest rate fall decomposition

the pension system, the main changes that took place in the US consist of an increase in both the mandatory and the average retirement age as well as a decrease in the replacement ratio. In the model representation, such adjustments translate into an increase in the retirement age from 63 years in 1970 to 65 years in 2015, and a decline of the replacement ratio from 45% of the working age after taxes average labor income in 1970 to 40% in 2015. Table 2.2 compares the results of our decomposition with the ones obtained in Eggertsson, Mehrotra & Robbins (2019), where social security and its reforms between 1970 and 2015 are not accounted for. What emerges from this comparison is that factors such as the decline in the mortality and fertility rates are significant drivers of the secular decline in interest rates but their quantitative importance, once we include a realistic pay-as-you-go pension system in the model, is reduced. Furthermore, the increase in retirement age and replacement ratio mitigate the impact of ageing and the productivity slowdown on aggregate savings, together with the increase in government debt, but the role they play is quantitatively small.

2.4.2 Normative exercises

Longevity shock

The positive longevity shock considered here consists of an increase in the old-age survival probabilities $s(j)$ such that life expectancy increases from 78.7 years in 2015 to 85.3 years, the level predicted for 2060 by the UN. Figure 1 (a) shows how the size of each cohort, relative to the overall population, is affected by population ageing, while Table 2.3 reports some summary statistics of the economy, under the different scenarios. The outcome of this exercise is consistent with the predictions of the basic model. Compared to the stationary equilibrium before the longevity shock, if no pension reform is adopted, the real interest rate falls by 87 basis points, while aggregate savings increase by 3.97%. When, instead, the government decides to maintain a balanced pension budget by reducing the replacement ratio, the real interest rate falls even more, by 114 basis points and aggregate savings increase by 5.29%. Alternatively, if the government keeps constant the replacement ratio by increasing the pension contribution rate, aggregate savings increase only by 3.74% and the equilibrium interest rate falls only by 82 basis points. Finally, if the mandatory retirement age is raised to 66 (so that agents work one year more), aggregate savings increase by 2.91% and the interest rate falls by 64 basis points.

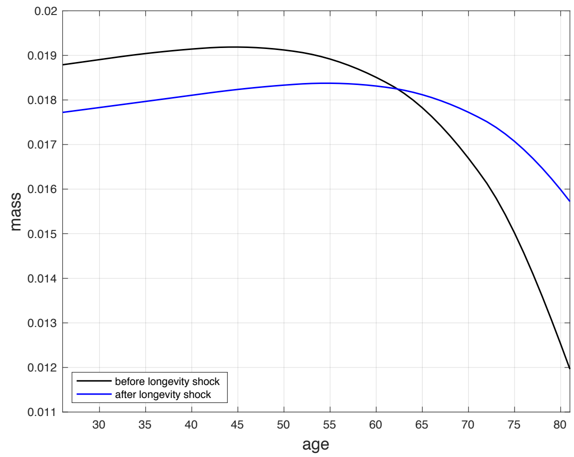
Figure 1 (c) and (e) display the impact of the alternative pension reforms, in response to the ageing shock, on the consumption and asset profiles over the life-cycle. The pension reform that brings about a decline in the replacement ratio, that can be interpreted as a partial privatization of the pension system, features an equilibrium allocation with higher consumption and savings for young cohorts and lower consumption for old cohorts. In other words, it redistributes resources from old to young generations. The policy keeping a constant replacement ratio, and therefore implying an increase in the pension contribution rate, reduces consumption at all ages with respect to the case in which the government does not implement any reform and runs a pension budget deficit. Finally, the measure entailing an increase in the mandatory retirement age of one year allows agents to moderate the increase in savings at younger age and increase consumption when old.

Fertility shock

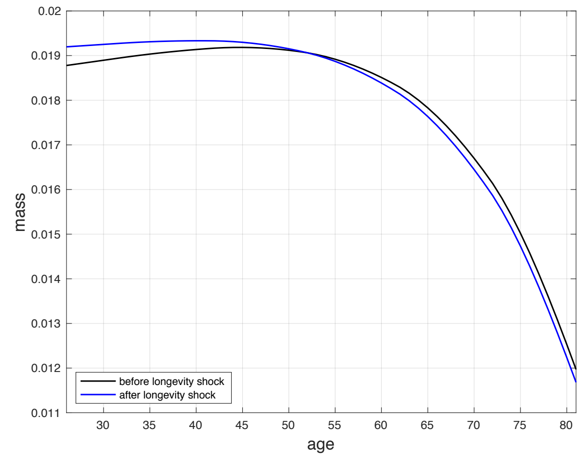
The UN predicts that the US total fertility rate will increase between 2015 and 2060, going from 1.875 to 1.913 children per woman. Such increase in the fertility rate, despite being modest, counteracts the secular decline in the interest rate. Table 2.4 displays the

Variable	no reform	constant replacement ratio	constant tax rate	increase retirement age
w	0.76	0.76	0.77	0.76
r	-2.34%	-2.29%	-2.61%	-2.12%
K	1.50	1.49	1.53	1.47
S	2.53	2.52	2.56	2.50
Y	1.15	1.15	1.15	1.15
social welfare	-178.56	-179.29	-180.30	-177.35
τ^b	27.55%	27.63%	27.14%	27.90%
τ^p	6.99%	8.16%	6.99%	7.49%
replacement ratio	40%	40%	33%	40%
retirement age	40	40	40	41
dependency ratio	37.80%	37.80%	37.80%	34.45%
labor share	66.43%	66.40%	66.57%	66.32%
investment-output ratio	16.71%	16.66%	16.98%	16.50%

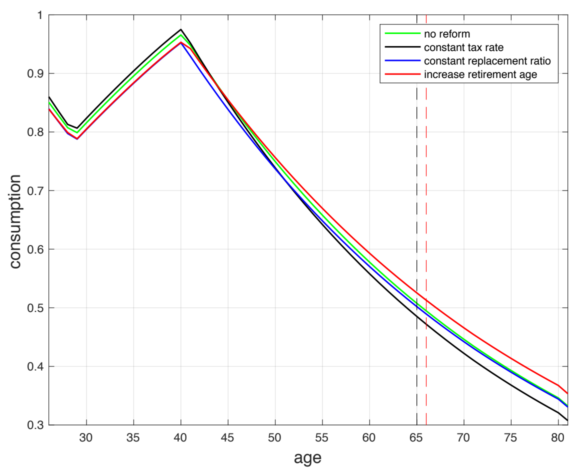
Table 2.3: Longevity shock



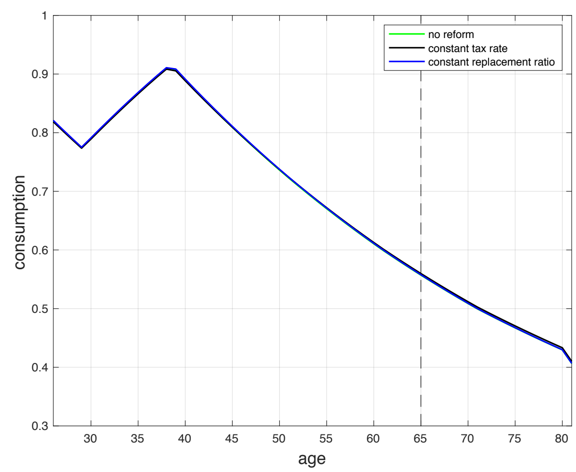
(a) Population mass by cohort - longevity shock



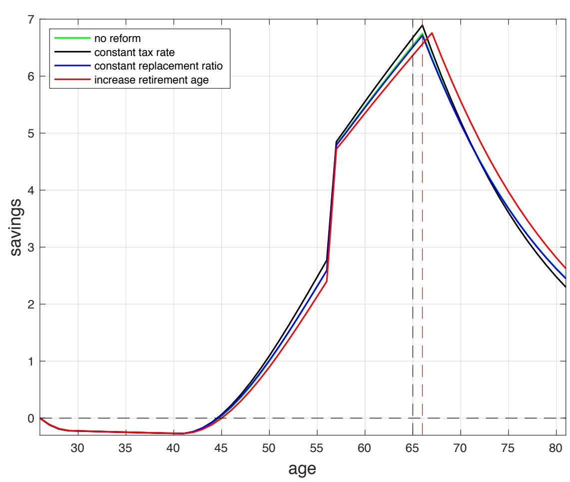
(b) Population mass by cohort - fertility shock



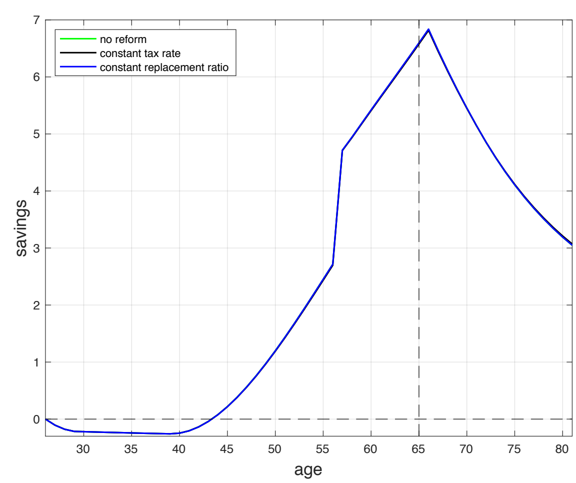
(c) Consumption profiles - longevity shock



(d) Consumption profiles - fertility shock



(e) Asset profiles - longevity shock



(f) Asset profiles - fertility shock

Figure 2.1: Results

summary statistics of this exercise, reporting again the effects of the positive fertility shock under the different pension reforms aimed at maintaining a balanced social security budget. In the fictitious case in which the government does not act, the equilibrium interest rate increases by 6 basis points while aggregate savings drop by around 0.3%. If, instead, the tax rate for the individual contributions is kept constant, allowing for an increase in the replacement ratio, the interest rate increases by 9 points and aggregate savings drop by 0.6%. In the scenario in which the government decides to maintain the same replacement ratio, the necessary adjustment in the tax rate τ^p is so little that the real interest rate increases only by 6 basis points, similarly to the case where no reform is adopted.

Figure 1 (d) and (f) show that impact of the alternative pension reforms, in response to the positive fertility shock, on the consumption and asset profiles over the life-cycle is very modest, which should not come as a surprise, because the shock itself is of little magnitude.

Secular stagnation vs. "normal" growth

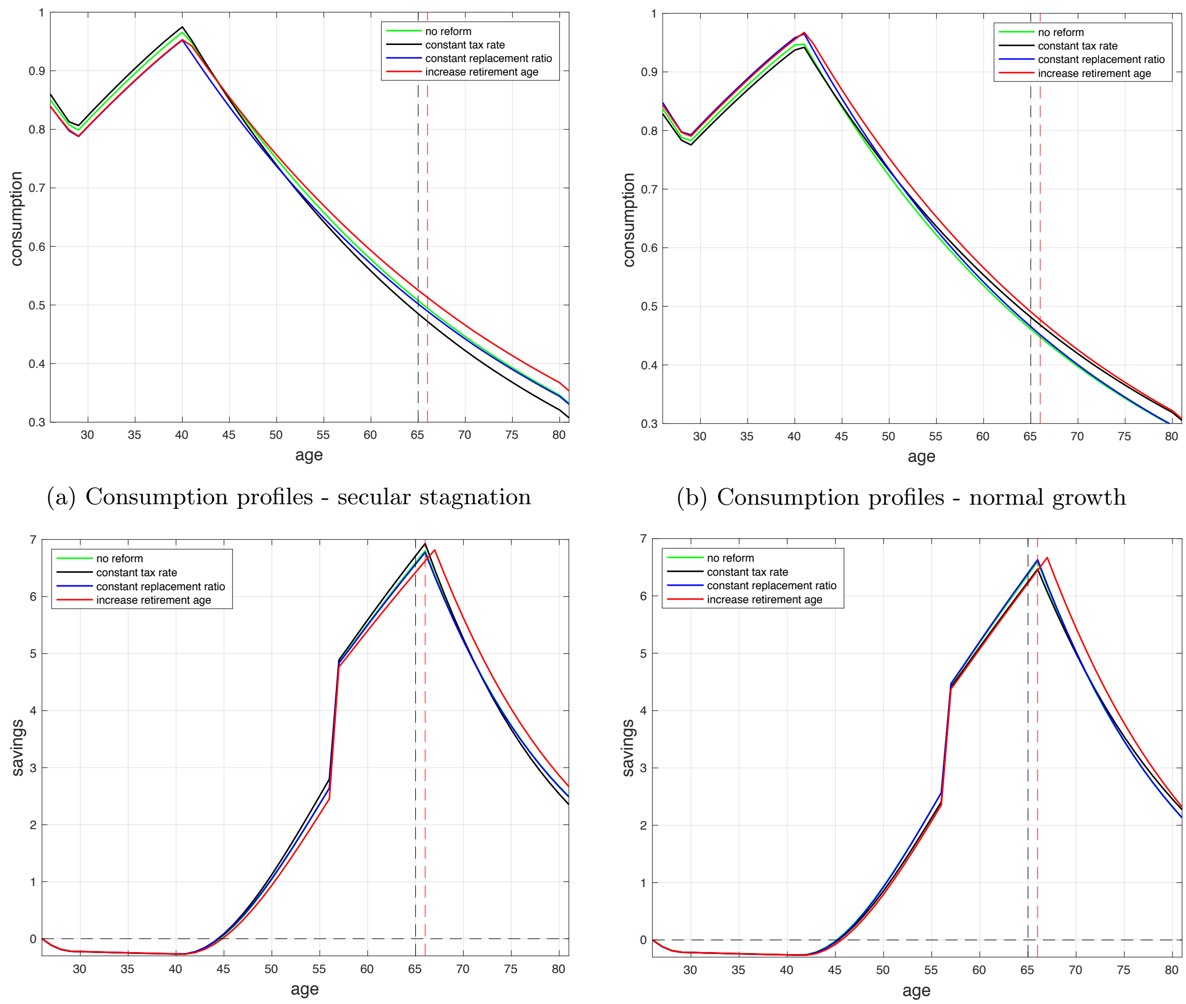
The final exercise of this paper attempts to examine the impact of different pension reforms on aggregate savings and the equilibrium interest rate in response to the two demographic shocks studied in the previous experiment, combined together. Moreover, two cases are considered. In the first one, the productivity growth rate stays at the 2015 level of 0.65%, which we denote as a "secular stagnation" or pessimistic growth rate. In the second one, it is raised to 2%, denominated as "normal". Table 2.5 shows the results of this last exercise.

What Table 6 reveals is that, relative to the 2015 real interest rate of -1.47%, the different assumptions regarding the 2060 productivity growth rate imply movements of the equilibrium interest rate in opposite directions. When, in the pessimistic scenario, the productivity growth rate remains low, the increase in longevity and fertility produce a further drop of the equilibrium interest rate. On the contrary, in the optimistic scenario characterized by a productivity growth rate of 2% per year such exogenous shock more than counters the effect of ageing, so that the equilibrium interest rate actually increases. Concerning the different reforms that the government can apply to maintain a budget social security budget in response to the demographic shocks, we obtain that the reform that dominates the others in terms of social welfare is the increase in retirement age, independently of the productivity growth rate.

Interestingly, the ranking between the other two options, the reform that keeps the

Variable	no reform	constant replacement ratio	constant tax rate
w	0.75	0.75	0.75
r	-1.41%	-1.41%	-1.38%
K	1.40	1.40	1.40
S	2.43	2.43	2.42
Y	1.14	1.14	1.14
social welfare	-162.33	-162.25	-162.18
τ^b	28.91%	28.90%	28.96%
τ^p	6.99%	6.86%	6.99%
replacement ratio	40%	40%	40.89%
retirement age	40	40	40
dependency ratio	31.63%	31.63%	31.63%
labor share	65.96%	65.96%	65.94%
investment-output ratio	15.95%	15.95%	15.92%

Table 2.4: Fertility shock



(a) Consumption profiles - secular stagnation

(b) Consumption profiles - normal growth

(c) Asset profiles - secular stagnation

(d) Asset profiles - normal growth

Figure 2.2: Results

Variable	no reform pessimistic	no reform normal	constant replacement ratio pessimistic	constant replacement ratio normal	constant tax rate pessimistic	constant tax rate normal	increase retirement age pessimistic	increase retirement age normal
w	0.76	0.75	0.76	0.75	0.77	0.74	0.76	0.74
r	-2.28%	-1.11%	-2.24%	-1.16%	-2.51%	-0.84%	-2.06%	-0.93%
K	1.49	1.38	1.48	1.38	1.51	1.35	1.47	1.36
S	2.52	2.40	2.52	2.40	2.55	2.37	2.50	2.38
Y	1.15	1.13	1.15	1.13	1.15	1.13	1.14	1.13
social welfare	-178.59	-160.64	-179.22	-160.07	-180.10	-159.27	-177.29	-158.22
τ^b	27.55%	27.59%	27.61%	27.51%	27.18%	28.01%	27.88%	27.88%
τ^p	6.99%	6.99%	8.01%	5.90%	6.99%	6.99%	7.34%	5.39%
replacement ratio	40%	40%	40%	40%	34.15%	48.5%	40%	40%
retirement age	40	40	40	40	40	40	41	41
dependency ratio	36.94%	36.94%	36.94%	36.94%	36.94%	36.94%	33.67%	33.67%
labor share	66.40%	65.81%	66.38%	65.84%	66.52%	65.68%	66.29%	65.72%

Table 2.5: Secular stagnation vs. "normal" growth

replacement ratio constant, and the one that keeps the same contribution rate τ^p depends on the realized productivity growth rate. In fact, in the secular stagnation case keeping the same replacement ratio and increasing taxes seems preferable, while in the case of "normal" growth keeping the same and so increasing the replacement ratio leads to higher utility, at steady-state. Such outcome can be rationalized taking into consideration that, in both scenarios, the growth rate of the economy, given by the sum of fertility rate and the productivity growth rate $n + g$ is higher than the corresponding equilibrium interest rate r, which makes the economy dynamically inefficient. Therefore, in a secular stagnation scenario, the representative household prefers to invest more resources in the pension system despite the fall in the rate of return of the pay-as-you-go scheme, due to the increase in longevity, because each unit contributed to social security pays off more than each unit invested into capital. On the contrary, in the scenario of normal productivity growth, the representative household prefers to keep contributing the same amount and therefore enjoying a higher replacement ratio as the net effect of the increase in longevity and increase in productivity on the pension system return is positive. As a consequence, the reform consisting of maintaining the same replacement ratio would imply a smaller amount of resources invested in social

security (as the drop in the tax rate τ^p from 6.99% to 5.9% shows) which is not preferable as the economy is dynamically inefficient.
These results are confirmed by the asset and consumption profiles over the lifetime reported in Figure 2. An increase in the retirement age shifts resources from young to old age in both scenarios. Keeping the same tax rate favours consumption at young age in the pessimistic outlook and consumption at old age in the optimistic outlook. The situation is reversed for the reform that adjusts the tax contributions to maintain the same replacement ratio.

2.5 Conclusions

Slow-moving demographic dynamics, reinforced by unfavorable economic conditions as the sovereign debt crisis, have put a strain on the existing pay-as-you-go public pension systems. Many governments in the western world have adopted different reforms with the goal of restoring their financial sustainability. This paper has attempted to study whether and how such reforms have impacted aggregate savings and the equilibrium real interest rate. A first result of our analysis suggests that the alternative reforms considered here, sharing the common objective of rebalancing the social security budget, may amplify, mitigate or remain neutral with respect to the impact of demographic trends on the interest rate. With reference to the evolution of the real interest rate in the US since 1970, the increase in the statutory full retirement age seems to have moderated the effects of a higher longevity and lower the fertility rate, whereas the reduction in the replacement ratio granted by social security to the retirees could have amplified it. However, the quantitative exercises reveal that these changes were of secondary significance for the determination of the interest rate in the US, when compared to other drivers. Future research could focus on European economies like Italy, Germany and France, whose populations have experienced, over the last 60 years, a drop in fertility and an increase in life expectancy even more pronounced than in the US.
In a second quantitative exercise we have explored how different pension reforms, adopted in response to the demographic trends predicted by the UN for the US population over the next 30 years, may affect the consumption and savings decisions of households over their lifetime. In particular, we have compared two alternative scenarios, one where the rate of growth for productivity stays at a level really close to 0, with one where productivity grows at 2% per year. In the first scenario, the expected changes in fertility and longevity push the equilibrium real interest rate further down,

while in the second the net effect of the changes in productivity and demographics on the interest rate is small but positive. In both scenarios, raising the retirement age by one year to restore the sustainability of the pension system leads to the highest level of average consumption and social welfare. However, the ranking between the other two possible options, i.e. keeping constant the replacement ratio vs. keeping constant the tax contribution rate, depends on the realization of the productivity growth rate. In the secular stagnation scenario, keeping a constant replacement ratio through an increase in pension contributions seems to be preferable over maintaining the same tax rate at the cost of a reduction in the replacement ratio. Such conclusion is reversed in the normal growth scenario. This appears as an interesting finding because the policy of increasing the retirement age could be not always feasible in practice and its costs are not fully accounted for in our workhorse model, as the labor supply decision is exogenous. We interpret this outcome as a consequence of dynamic inefficiency, which implies that the reform that produces the largest pension system increases lifetime income the most. This last result is at odds with the widespread view that a privatization of the pension systems would be welfare improving, a proposal that seemed to gather consensus in the political and economic debates over the last years.

Bibliography

[1] Orazio Attanasio, Sagiri Kitao, and Giovanni L. Violante. Global demographic trends and social security reform. *Journal of Monetary Economics*, 54(1):144–198, 2007.

[2] Alan Auerbach and Laurence Kotlikoff. *Dynamic Fiscal Policy*. Cambridge University Press, 1987.

[3] Richard Baldwin and Coen Teulings. Secular Stagnation: Facts, Causes and Cures. In *VoxEU, Ebook*, number August 2014, pages 1–164. 2014.

[4] Gianluca Benigno and Luca Fornaro. Stagnation Traps. *ECB Working Paper Series*, March(2038):1–51, 2017.

[5] Ricardo J. Caballero and Emmanuel Farhi. The Safety Trap. *The Review of Economic Studies*, (2016), 2017.

[6] Ricardo J Caballero, Emmanuel Farhi, and Pierre-Olivier Gourinchas. Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares. *Mimeo*, (617):1–12, 2017.

[7] Carlos Carvalho, Andrea Ferrero, and Fernanda Nechio. Demographics and real interest rates: Inspecting the mechanism. *European Economic Review*, 88:208–226, 2016.

[8] Gauti B. Eggertsson, Neil R. Mehrotra, and Jacob A. Robbins. A model of secular stagnation: Theory and quantitative evaluation. *American Economic Journal: Macroeconomics*, 11(1):1–48, 2019.

[9] Gauti B. Eggertsson, Neil R. Mehrotra, Sanjay R. Singh, and Lawrence H. Summers. A contagious malady? Open economy dimensions of secular stagnation. *IMF Economic Review*, 64(4):581–634, 2016.

[10] Christian Geppert, Alexander Ludwig, and Raphael Abiry. Secular Stagnation? Growth, Asset Returns and Welfare in the Next Decades. *SAFE Working Paper*, 145:1–38, 2016.

[11] Mark Gertler. Government debt and social security in a life-cycle economy. *Carnegie-Rochester Conference Series on Public Policy*, 50:61–117, 1999.

[12] Robert J Gordon. The Economics of Secure Stagnation: A Supply-Side View. *American Economic Review: Papers & Proceedings*, 105(5):54–59, 2015.

[13] Alvin H. Hansen. Economic Progress and Declining Population Growth. *American Economic Journal: Macroeconomics*, 29(March):1–15, 1939.

[14] Lawrence H. Summers. U.S. economic prospects: Secular stagnation, hysteresis, and the zero lower bound. *Business Economics*, 49(2):65–73, 2014.

[15] Menahem E. Yaari. Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies*, 32(2):137–150, 1965.

Chapter 3

Optimal Monetary and Fiscal Policy in a Model with an Underground Sector

Abstract
This paper investigates the relation between the size of the underground sector of developing economies and the optimal monetary and fiscal policy mix necessary to fund a given stream of government expenditure. Due to the limited ability to measure the scope and size of informality, the vast majority of research in this literature assumes that the size of the underground sector is exogenous and invariant to policy decisions. In this project I model the decision of heterogenous firms to operate in the formal sector and pay the sales tax or to conceal their activities and evade taxes. Producing informally involves running the risk of getting caught by the authorities and therefore losing profits. By assumption, the probability of being sanctioned is an increasing function of the capital-labor ratio employed in production. As a consequence, informal firms adopt a sub-optimal capital-labor ratio to reduce the risk of getting caught. The government faces a trade-off when choosing the optimal mix of profits and inflation tax to finance its expenditure. In fact, a higher tax rate pushes more firms to evade taxes and adopt inefficient technologies, whereas a higher inflation tax increases the transaction cost associated with consumption. The policy mix that solves the Ramsey problem crucially depends on the firms' productivity distribution.

3.1 Introduction

The area of economics related to the size, origins and scope of the so-called underground economy is surrounded by an aura of mystery and a large degree of uncertainty. The main reason is the great difficulty of measuring the size of this phenomenon due to its own nature, which, by definition, involves productive activities that are concealed from statistical offices and tax reports. A second issue is directly linked: the estimation of the shadow economy depends on the assumed definition. Unfortunately, the literature does not agree on a unique definition, but proposes several ones, each capturing a particular aspect of the underground economy. Therefore many estimation methods have been developed, which sometimes do not converge to the same results.

Studying the informal sector might be of scarce interest if the consequences of its existence were relatively unimportant for the overall economy. Experts in this field stress that possessing an accurate information on the shadow activities is crucial for guiding the hand of policy makers, who, otherwise, would take decisions based on only partially representative data (Schneider & Enste [24], 2000). Secondly, the analysis of the interaction between the official and the underground economy allows us to evaluate different policies aimed at improving the economic performance. As an example, if an expanding shadow economy was due to the perception by tax-payers of an overwhelming and unjustified tax burden, measures of austerity, as the ones undertaken in many countries during the great recession, would potentially generate a vicious circle of increasing budget deficit and further expansion of the underground economy.

In this regard, whether the existence of an occult economy is conducive or not for economic growth remains controversial in the literature: some scholars support the idea that the presence of underground economy, caused by tax evasion, hinders the provision and worsens the quality of public goods (Loayza [13] 1996). Other researchers emphasize that the negative effects are overstated and actually, "underground economies increase overall economic efficiency by circumventing the inefficiencies brought about by taxation" (Feige [10] 1989). Busato & Chiarini [4] 2004 builds a model where the underground economy acts as a buffer during recessions, partly absorbing the increasing unemployment by allowing a reallocation of labor between formal and informal activities. Then, the "risk insurance" services offered to workers and firms weaken the motive for fighting the shadow economy as a policy objective. Asea [3] 1996 finds that an underground economy can foster competition, increase efficiency, contribute to create markets and provide financial resources unavailable in the official environment.

The objective of this project is to study the policy implications of the underground economy, in terms of the optimal combination of fiscal and monetary policy necessary to finance government spending. Or, from a different perspective, to determine the optimal size of the informal sector. Three different strands of literature are linked to this topic: the one focusing on the size, causes and consequences of the underground economy, the one exploring tax evasion and the one studying welfare maximizing monetary policy and the costs and benefits of inflation.

The empirical facts I address and try to explain concern the relationship between the different sizes of the underground economies across countries, and the corresponding inflation rates. Put in other words, the main goal is to examine whether the different degrees of tax evasion experienced by different countries can account for the heterogeneity in monetary policy targets set by central banks. Data show that developing countries tend to have larger shadow sectors than developed ones (check the appendices). Nevertheless, it would be wrong to think about the underground economy as a phenomenon confined to emerging countries, where the it constitutes a significant portion of overall activity, that disappears at later stages of economic development. OECD countries like Italy or Spain are examples of industrialized countries where it is firmly established and quite vast. Moreover, observed inflation targets set by the central banks of OECD countries concentrate in the 2% - 3% interval, while those of emerging economies appear generally higher. This evidence suggests that seignorage is an important revenue resource for the governments of developing countries and reinforces the public finance motive for inflation.

Figure 1 displays the relationship between observed inflation targets in 2015 and the estimated size of the informal sector as a percentage of official GDP (average 1999-2007), based on Schneider, Buehn & Montenegro [23] (2010) calculations, for more than 50 countries that have inflation targeting as the monetary policy regime. Economies being part of the European monetary union are excluded, since they do not individually set their policy objective. Even though the evidence emerging from the graph is not clear, we can still recognize a positive correlation between the two variables. What appears unambiguous is that the variance of the monetary policy target increases with the size of the underground economies. This fact should not be surprising, as the literature on optimal monetary policy offers less guidance for emerging countries.

Although the data points for the countries that have adopted the Euro are not reported in the figure, there is some heterogeneity in terms of shadow economy size among them. In particular, we can distinguish three groups: the central-northern Europe group, with countries like France, Germany, Luxembourg, Austria, Netherlands, where it ranges between 10% and 15%; the southern-western Europe group, with Italy, Spain, Portugal and Greece, historically characterized by high public debt and inflation rates, where the size of the informal sector ranges between 22% and 30%; and the eastern Europe group, i.e. Latvia, Lithuania, Estonia, Slovak Republic and Slovenia, that have joined the monetary union more recently and where the informal sector sometimes constitutes

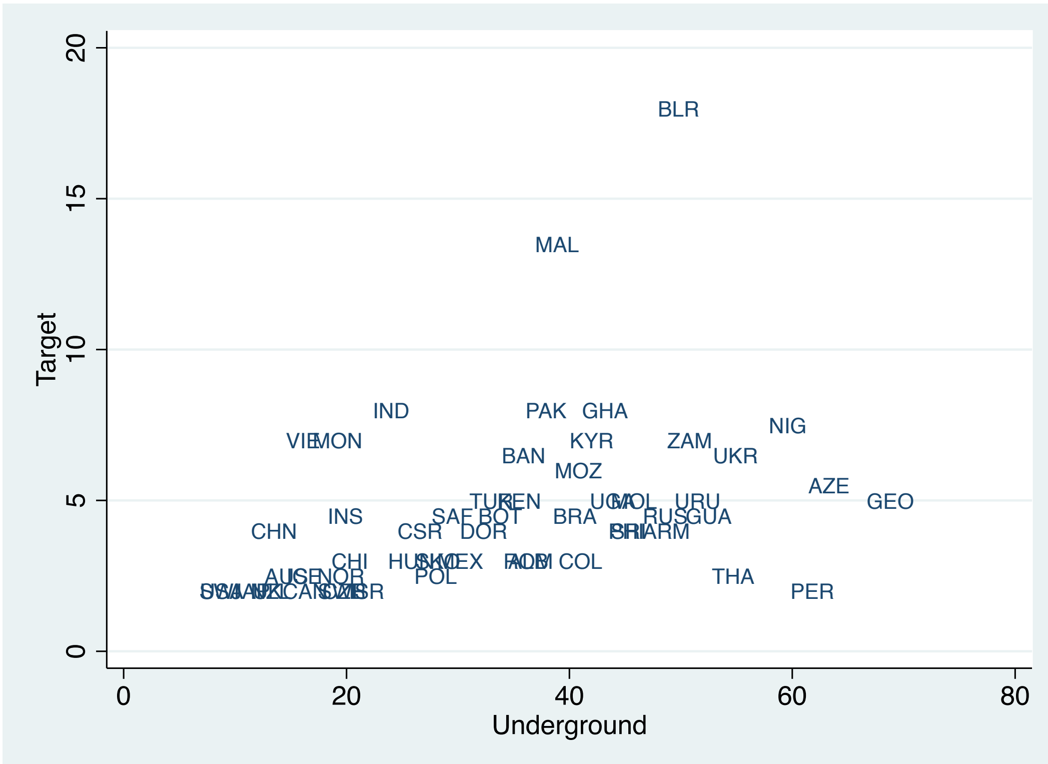


Figure 3.1: Underground Economy as % of GDP (average 1999-2007) vs. Inflation Target in % (2015)

40% of official GDP. Such heterogeneity raises some doubts on the optimality of the common inflation target of <2% set by the European Central Bank and calls for further investigation on the benefits and costs of joining a currency union.

Interestingly, there is a mismatch between theory results and empirical evidence: while models with money non-neutrality postulate that the optimal rate of inflation ranges between minus the real interest rate and zero, observed inflation targets are positive. Therefore, a second objective is to test whether taking into account the existence of an unofficial market helps filling this gap between theory and practice.

Differently from other papers tackling the same subject, in the model I propose the size of the informal sector is endogenously determined by optimizing decisions of firms. The government disposes of two policy instruments: the tax rate on firms' profits and the nominal interest rate. The use of the tax rate to finance the government budget encompasses a policy trade-off: a greater rate, on the one hand, increases the tax revenues for a given level of firms' profits, on the other hand pushes some firms from the formal to the tax evading informal sector, and by doing so decreases the tax base. Informal

firms employ a sub-optimal capital-labor ratio in the attempt of eluding tax authorities. The incomplete taxation and distortions implied by taxes create the incentive for the benevolent social planner to set a positive nominal interest rate and raise some seignorage revenues. However a positive nominal interest makes private consumption purchases costly. Having her space of manoeuvre constrained by multiple trade-offs, the benevolent Ramsey social planner engineers the right policy mix in order to maximize the welfare of the representative household.
Section 2 presents a review of the literature related to this topic. Section 3 illustrates the model. Section 4 presents the results and section 5 highlights the main weaknesses of the proposed framework and describes some potential future advancements of this project.

3.2 Literature Review

A standard result in monetary economics is the optimality of the Friedman rule [11] (1969) which prescribes a zero nominal interest rate and implies a negative inflation rate equal to the real interest rate. Such outcome is justified by the fact that the social cost of printing money is negligible while the private (opportunity) cost of holding money is the interest rate paid on bonds. Then, in order to minimize the welfare loss due to holding cash, it is optimal for monetary authorities to set a zero nominal interest such that the private and social costs of money are equated.
However, in the literature on the costs and benefits of inflation, it is often argued that a positive inflation rate allows the government to finance its spending through seignorage whenever other forms of taxation are not available or do not offer a complete coverage of all sources of income. Phelps [17] (1973) obtains that a positive inflation rate is optimal when the government can only use distortionary taxes as inflation acts as a tax on consumption. Nevertheless, many papers present classical monetary models in which the Friedman rule continues to be optimal also in a second-best world where lump-sum taxes are unavailable (Chari, Christiano & Kehoe [6] 1996, Correia & Teles [7] 1999). Schmitt-Grohé & Uribe [22] (2011) examines the opportunity of setting a zero nominal interest rate in a variety of contexts. The aim of this paper is to analyze the apparent divergence between policy prescriptions derived in theoretical models and the actual decisions of central banks all around the world. Inflation-targeting monetary authorities have a policy objective of price stability around an inflation rate ranging between 2% and 3% in many countries, which is systematically higher than the opti-

mal inflation rate implied by leading monetary models. The Friedman rule is tested, on the one hand, in a transaction costs model with demand for fiat money, and, on the other hand, in a model where money is not neutral due to the presence of sticky prices. Schmitt-Grohé & Uribe (2011) shows that a positive nominal interest rate is optimal when firms manage to evade taxes on the value of production in a transaction costs model. The way the underground sector is formalized is relatively simple: an amount u of the representative firm total production is concealed from tax authorities which levy a tax on total income. Two polar assumptions on the functional form of u are confronted. When u is a degree 1 homogenous function of the amount of labor hired by the representative firm (e.g. the level of activity as $Y_t = N_t$) the underground economy is proportional to the above-ground economy and a proportional tax on the above-ground output is also a proportional tax on total output. Then the Friedman rule is still optimal. When, instead, u does not depend on the level of activity then a positive interest rate is optimal. The latter case is examined in a simulation of the model: as a result, only the existence of unrealistically large underground sectors can justify optimal inflation rates close to the levels targeted by central banks around the world.

Other papers have already examined quantitatively the implications for optimal monetary policy of tax evasion in an underground sector. Nicolini [15] (1998) builds a model where the set of available goods is partitioned between those that are traded in the informal sector, evading taxes, and those traded in the formal sector, where consumers pay a consumption tax. In both sectors, an exogenously given portion of goods is purchased by cash, while the rest can be bought using credit. Inflation acts as a tax on cash goods: on those traded in the formal sector which regularly pay the consumption tax and on those traded in the informal sector. The results of the quantitative exercise, taking as given the size of the informal and formal sectors and the consumption tax rate, point to a small effect of tax evasion on interest rates and on optimal monetary policy, even for a country with an estimated large underground economy as Peru.

De V. Cavalcanti & Villamil [9] (2003) perform a quantitative test on the welfare gains of increasing or decreasing inflation with respect to a baseline observed level. The analysis is based on a shopping time model where agents can evade labor income taxes operating in the underground labor market. The representative firm hires workers in both the informal and the informal labor markets and the production function shows a constant elasticity of substitution between the two types of labor. The theoretical result is that, when a shadow labor market exists, the Friedman rule is no longer optimal. The numerical experiments suggest that the optimal inflation tax can significantly change

among different economies depending on the dimension of the informal sector.
Koreshkova [12] (2006) introduces different production technologies between the formal and the informal sector: the former employs a linear production function, while the latter disposes of a decreasing returns to scale technology. Nevertheless, technologies are symmetric across goods inside the two sectors. Then perfect competition implies that every firm operating in any of the two sectors produces the same quantity of output at the same price so production can be easily aggregated. Different exogenous productivity levels in the two sectors are assumed in order to calibrate a realistic underground sector size in the quantitative exercise. Transactions can be performed with cash or using a costly credit and workers decide the allocation of the time worked in the formal and in the informal sector, where labor income is taxed. As a results of the quantitative analysis, optimal inflation rates are much higher than the ones obtained by Nicolini (1998) and are also much more sensitive to the size of the shadow economy. In this framework, the productivity difference between the two sectors plays a key role.
The idea of this project comes from the observation that all these models share a common feature: the size of the informal sector is exogenously given and optimal monetary and fiscal policies depend on that. But this means that an increase in the level of taxes has no effect on the size of the informal sector. Therefore, it is interesting to study what are the determinants of the informal sector and how it can emerge as a consequence of agents' choices and policy decisions. Let's then turn to the model.

3.3 Model

The model I propose consists of a closed economy featuring three types of agents: heterogeneous firms, the representative household and the government, which sets the tax rate and the nominal interest rate. The next subsections focus on each of them in detail.

3.3.1 Firms

The production sector of the economy is populated by a continuum of heterogeneous firms, indexed by $i \in [0, 1]$, characterized by the same constant returns to scale Cobb-Douglas production function $Y_{i,t} = A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}$ which employs both capital and labor, that are hired in perfectly competitive markets, and where A_t represents an exogenous aggregate level of productivity. Firms produce a differentiated good and differ in terms of the parameter α_i, characterized by the density function $g(\alpha_i)$, defined in the interval $[0, 1]$ and time invariant. In particular, the production sector features a regime of monopolistic competition among firms where each one can freely reset its price in every period. This means that they would naturally use different capital-labor ratios, set different prices and sell different quantities. Total production is obtained via the Dixit-Stiglitz aggregator $Y_t = \left(\int_0^1 Y_{i,t}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}$ where parameter $\epsilon > 1$ denotes the intratemporal elasticity of substitution across different varieties. The implied set of demands for each differentiated good is:

$$Y_{i,t} = \left(\frac{P_{i,t}}{P_t} \right)^{-\epsilon} Y_t$$

where the price index is $P_t = \left(\int_0^1 P_{i,t}^{1-\epsilon} di \right)^{\frac{1}{1-\epsilon}}$.

Firms face a further decision: they can operate either in the formal sector, where a tax on profits must be paid to the government or in the informal sector evading taxes, but exposing themselves to the risk of getting caught by the tax authorities, and then seeing their sales confiscated. The novelty introduced by this model is that the probability of detection in the informal sector is endogenous and an increasing function of the capital-labor ratio chosen by the firm. The rationale of such design relies on the idea that labor can be easily concealed from tax authorities while the use capital, intended as machinery, equipment or even plants, makes informal firms more detectable. This assumption is consistent with the empirical evidence concerning the type of businesses that are usually carried out in the shadow economy: as Andreoni, Erard & Feinstein [2] (1998) illustrates, typical underground activities are low-skilled labor-intensive.

Two are the main consequences: firstly, the same firm employs a lower capital-labor ratio if it produces in the informal sector than in the formal sector due to the risk of getting caught; secondly, the size of the informal sector is endogenously determined by the tax rate, the distribution of parameter α_i and the characteristics of the function describing the probability of detection, given the factors' prices.

Formally, the problem of an individual firm consists of maximizing the profit function with respect to price and capital-labor ratio subject to the technological constraint given by the production function and the demand for its output. Once determined the optimal price and capital-labor ratio, each firm compares the profits that it can obtain in the formal and in the informal sector and chooses one of the two accordingly. The next subsections illustrate the details.

The formal firm

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t}} \quad (1-\tau_t)\left[\frac{P_{i,t}}{P_t}Y_{i,t} - w_t N_{i,t} - R_t K_{i,t}\right]$$

$$\text{s.t.} \qquad Y_{i,t} = \left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t$$

$$Y_{i,t} = A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}$$

Where τ represents the tax rate, the ratio $\frac{P_{i,t}}{P_t}$ the relative firm's price, w_t and r_t respectively the prices of labor and capital in real terms. The maximization involves two constraints: the Cobb-Douglas production function and the demand faced by each firm. The problem can be solved setting up a Lagrangean:

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t},\lambda_{i,t}} \quad (1-\tau_t)\left[\left(\frac{P_{i,t}}{P_t}\right)^{1-\epsilon} Y_t - w_t N_{i,t} - R_t K_{i,t}\right] - \lambda_{i,t}\left[\left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t - A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}\right]$$

The first order conditions with respect to the price, capital and labor imply:

$$\left(\frac{P_{i,t}}{P_t}\right)^*_{\text{for}} = \frac{\epsilon}{\epsilon-1}\lambda_{i,t}^{\text{for}} \tag{3.1}$$

$$w_t = \lambda_{i,t}^{\text{for}}(1-\alpha_i)A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i}_{\text{for}} \tag{3.2}$$

$$R_t = \lambda_{i,t}^{\text{for}}\alpha_i A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i-1}_{\text{for}} \tag{3.3}$$

Equation (1) states that firms set the price as a mark-up over the marginal cost and that a greater tax rate force them to set a higher price. Equations (2) and (3) suggest that the firm hires respectively labor and capital to the point that their marginal product times the marginal cost correspond to their price. Dividing equation (2) by equation (3) allows us to compute the optimal capital-labor ratio for the firm:

$$\frac{w_t}{R_t} = \frac{1-\alpha_i}{\alpha_i}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{for}} \tag{3.4}$$

$$\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{for}} = \frac{w_t}{R_t}\frac{\alpha_i}{1-\alpha_i} \tag{3.5}$$

Substituting (5) in either (2) or (3) lets us recover the marginal cost for the formal firm as a function of parameters and prices:

$$\lambda^{\text{for}}_{i,t} = \frac{1}{A_t}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i}\left(\frac{w_t}{1-\alpha_i}\right)^{1-\alpha_i} \tag{3.6}$$

The informal firm

Analogously:

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t}} \left[1 - m\left(\frac{K_{i,t}}{N_{i,t}}\right)\right]\frac{P_{i,t}}{P_t}Y_{i,t} - w_t N_{i,t} - R_t K_{i,t}$$

$$\text{s.t.} \qquad Y_{i,t} = \left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t$$

$$Y_{i,t} = A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}$$

The Lagrangean is:

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t},\lambda_{i,t}} \left[1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)\right]\left(\frac{P_{i,t}}{P_t}\right)^{1-\epsilon} Y_t - w_t N_{i,t} - R_t K_{i,t} - \lambda_{i,t}\left[\left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t - A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}\right]$$

The first order conditions with respect to the price, capital and labor imply:

$$\left(\frac{P_{i,t}}{P_t}\right)^*_{\text{inf}} = \frac{\epsilon}{\epsilon-1}\frac{\lambda^{\text{inf}}_{i,t}}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}} \tag{3.7}$$

$$w_t = \lambda^{\text{inf}}_{i,t}(1-\alpha_i)A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i}_{\text{inf}}\left[1 + \frac{\epsilon}{\epsilon-1}\frac{1}{1-\alpha_i}\frac{m'(\quad)}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}\right] \tag{3.8}$$

$$R_t = \lambda^{\text{inf}}_{i,t}\alpha_i A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i-1}_{\text{inf}}\left[1 - \frac{\epsilon}{\epsilon-1}\frac{1}{\alpha_i}\frac{m'(\quad)}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}\right] \tag{3.9}$$

Again, equation (7) states that firms set the price as a mark-up over the marginal cost and that a greater probability of being caught by tax authorities $m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}$ forces them to set a higher price. Equations (8) and (9) show that firms' decision to hire capital and labor is distorted by the presence of the endogenous probability of detection which, by assumption, is increasing in the capital-labor ratio ($m'(\quad) > 0$). Dividing equation

(8) by equation (9):

$$\frac{w_t}{R_t} = \left[\frac{1-\alpha_i + \frac{\epsilon}{\epsilon-1}\frac{m'(\quad)}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}{\alpha_i - \frac{\epsilon}{\epsilon-1}\frac{m'(\quad)}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\right]\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}} \tag{3.10}$$

By comparing equation (10) with equation (4) we can recognize one of the first results of this setup: the same firm, identified by a specific parameter α_i, will hire a higher capital-labor ratio in the formal sector than in the informal sector. In practice, the term between square brackets in equation (10) is bigger than $\frac{1-\alpha_i}{\alpha_i}$.
Let's define $f\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}} = \alpha_i - \frac{\epsilon}{\epsilon-1}\frac{m'(\quad)}{1-m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}}\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}} \leq \alpha_i$. Then the optimal capital-labor ratio for the informal firm can be written as:

$$\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}} = \frac{w_t}{R_t}\frac{f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}{1-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}} \tag{3.11}$$

$$\lambda^{\text{inf}}_{i,t} = \frac{1}{A_t}\left(\frac{R_t}{f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}\right)^{\alpha_i}\left(\frac{w_t}{1-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}\right)^{1-\alpha_i} \tag{3.12}$$

In order to determine it explicitly we need to specify a functional form for the probability of detection. Let us assume that the latter is a linear function of the capital-labor ratio: $m\left(\frac{K_{i,t}}{N_{i,t}}\right) = \psi\frac{K_{i,t}}{N_{i,t}}$ where $\psi > 0$. Then:

$$\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}} = \frac{-\left[(1-\alpha_i) + \frac{w_t}{R_t}\psi(\alpha_i + \frac{\epsilon}{\epsilon-1})\right] + \sqrt{\left[(1-\alpha_i) + \frac{w_t}{R_t}\psi(\alpha_i + \frac{\epsilon}{\epsilon-1})\right]^2 + 4\frac{w_t}{R_t}\psi\alpha_i(\frac{\epsilon}{\epsilon-1} - (1-\alpha_i))}}{2\psi(\frac{\epsilon}{\epsilon-1} - (1-\alpha_i))} \tag{3.13}$$

The size of the underground economy

The size of the underground economy depends on the optimizing decisions of firms as each firm freely chooses the sector that grants higher profits. Given parameter ψ and the distribution of parameter α_i among firms, there are three possible cases:

1. all firms are formal: the tax rate set by the government is low enough that all firms find convenient to stay formal and pay taxes.

2. all firms are informal: the tax rate is high enough to push all firms to evade taxes.

3. some firms decide to be formal, some other informal.

In the latter case, which is the most interesting one, there will be one (or more?) marginal firm $\bar{\imath}$, identified by $\bar{\alpha}$, that is indifferent between the formal and the informal sector as both offer the same profits. As firms are ordered according to α_i, the position of firm $\bar{\imath}$ with respect to the whole distribution determines the size of the shadow economy.

$$\big(\text{Profits}\big)^{\text{for}}_{\bar{\imath},t} = \big(\text{Profits}\big)^{\text{inf}}_{\bar{\imath},t} \tag{3.14}$$

We need then to solve equation (14) for $\bar{\alpha}$, given a set of factors prices and the tax rate. Unfortunately it is transcendental equation which cannot be solved in closed form:

$$\left\{\left[\frac{f\left(\frac{K_{\bar{\imath},t}}{N_{\bar{\imath},t}}\right)^{*}_{\text{inf}}}{\bar{\alpha}}\right]^{\bar{\alpha}}\left[\frac{1-f\left(\frac{K_{\bar{\imath},t}}{N_{\bar{\imath},t}}\right)^{*}_{\text{inf}}}{1-\bar{\alpha}}\right]^{1-\bar{\alpha}}\right\}\left[1-m\left(\frac{K_{\bar{\imath},t}}{N_{\bar{\imath},t}}\right)^{*}_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}} = \Big[1-\tau\Big]^{\frac{1}{\epsilon-1}} \tag{3.15}$$

Nevertheless, as proven in the appendices, one can show that the left hand side of equation (15) takes value 1 when α_i is 0 and is always decreasing in α_i. Therefore, as the right hand side does not depend on α_i and is constant for a given tax rate τ, there can be either one or no $0 < \bar{\alpha} < 1$ that satisfies the equation. If the solution exists, all firms with $\alpha_i > \bar{\alpha}$ will be formal. Otherwise, all firms will stay in the informal sector.

3.3.2 Households

The part of the model regarding the behaviour of households corresponds exactly to the one set up in Schmitt-Grohé and Uribe (2011) to study the optimality of the Friedman rule in classical monetary models. Peculiar is the way a motive for holding money is introduced: the representative household faces some transaction costs, proportional to consumption, that are increasing in consumption spending and decreasing in real money holdings. Defining velocity:

$$v_t = \frac{P_t C_t}{M_t}$$

The transaction cost function $s(v_t)$ presents the following features:

1. it is nonnegative and twice continuously differentiable;
2. there exists a satiation level of velocity $\underline{v} > 0$ such that $s(\underline{v}) = s'(\underline{v}) = 0$;
3. $(v - \underline{v})s'(v) > 0$ for $v \neq \underline{v}$;
4. $2s'(v) + vs''(v) > 0$ for all $v \geq \underline{v}$.

This sets of assumption ensures that a zero nominal interest rate makes the transaction cost zero and that it is not associated with an infinite demand for money. Moreover, the transaction cost is increasing in the nominal interest rate.

The problem of the infinitely-lived representative household consists of maximizing the utility function, such that its budget constraint is satisfied:

$$\max_{C_t,N_t,I_t,v_t,B_t,M_t} \sum_{t=0}^{\infty} \beta^t U(C_t, N_t)$$

$$\text{s.t.} \quad C_t[1+s(v_t)] + I_t + \frac{M_t}{P_t} + Q_t\frac{B_t}{P_t} = \frac{M_{t-1}}{P_t} + \frac{B_{t-1}}{P_t} + w_t N_t + R_t K_t + \text{Profits}_t \quad (3.16)$$

The period utility function U is increasing in consumption C_t and decreasing in the amount of hours worked N_t. C_t is a Dixit-Stiglitz index of the goods consumed: $C_t = \left(\int_0^1 C_{i,t}^{\frac{\epsilon-1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}$. The discount factor β takes a value between 0 and 1. The household disposes of 1-period government bonds whose price is $Q_t \leq 1$, the inverse of the gross nominal interest rate set by the central bank and pay a consumption unit in the next period. The variable M_t indicates nominal money balances, demanded to alleviate the transaction costs associated with consumption purchases. The household budget is the sum of money holdings inherited from the previous period, the interest paid by government bonds purchased in the last period, labor and capital income earned in the perfectly competitive labor and capital markets, and firms' profits. The representative household can increase the capital stock through investment:

$$K_{t+1} = K_t(1-\delta) + I_t \quad (3.17)$$

where $0 < \delta < 1$ is the period depreciation rate. Finally, the household cannot roll on its debt forever and is subject to a no-Ponzi game condition:

$$\lim_{j\to\infty}(M_{t+j} + B_{t+j}) \geq 0$$

Once we substitute equation (17) into the budget constraint (16) we can calculate the first order conditions of the problem with respect to consumption, hours worked, next period capital, bonds and nominal money balances, which allows us to derive the optimal conditions describing the household behaviour:

$$-\frac{U_{n,t}}{U_{c,t}} = \frac{w_t}{1+s(v_t)+v_t s'(v_t)} \quad (3.18)$$

$$v_t^2 s'(v_t) = 1 - Q_t \quad (3.19)$$

$$\frac{U_{c,t}}{1+s(v_t)+v_t s'(v_t)}\frac{1}{\beta}\pi_{t+1}Q_t = \frac{U_{c,t+1}}{1+s(v_{t+1})+v_{t+1}s'(v_{t+1})} \tag{3.20}$$

$$R_{t+1} = \frac{1}{\Pi_{t+1}Q_t} \tag{3.21}$$

Equation (18) shows the intratemporal consumption/leisure trade-off while the demand for money is represented by equation (19): a positive nominal interest rate ($Q_t < 1$) implies that the transaction cost is not zero, creating a wedge $1+s(v_t)+v_t s'(v_t)$ between the marginal rate of substitution of consumption for leisure and the real wage. Equation (20) is the Euler equation. Finally equation (21) is the typical Fisher equation relating the real interest rate, the nominal interest rate and the inflation rate.

3.3.3 Government

The government runs a balanced budget in every period and finances its (exogenously given) expenditure issuing nominal 1-period bonds, printing money and levying taxes on formal firms' sales:

$$G_t + \frac{B_{t-1}}{P_t} + \frac{M_{t-1}}{P_t} = T_t + Q_t\frac{B_t}{P_t} + \frac{M_t}{P_t} \tag{3.22}$$

where:

$$T_t = \tau_t \int_{\bar{i}}^{1} \left(\frac{P_{i,t}^{\text{for}}}{P_t}Y_{i,t}^{\text{for}} - w_t N_t - R_t K_t\right) g(\alpha_i)di + \int_0^{\bar{i}} m\left(\frac{K_{i,t}}{N_{i,t}}\right)_{\text{inf}}^{*} \frac{P_{i,t}^{\text{inf}}}{P_t}Y_{i,t}^{\text{inf}} g(\alpha_i)di \tag{3.23}$$

The latter expression highlights that an increase in the tax rate τ has two opposite effects on the total amount of taxes collected by the government: on the one hand it increases the amount of taxed profits collected by the government from formal firms, on the other hand, by pushing some firms away from the formal towards the informal sector it effectively decreases the tax base, but increases the amount obtained from sanctioning the informal firms that are caught evading taxes. The first term in equation (23) exactly identifies the total sum of taxes paid by formal firms, while the second one gives the total amount of resources collected by the tax authorities as sanctions to the underground economy.

3.3.4 Competitive equilibrium

Given the fiscal and monetary policy tools set by the government and the central bank (the tax rate τ_t and the price of bonds Q_t) and an exogenously given stream of government spending G, a competitive equilibrium is characterized by a sequence from $t = 0$ to ∞ of real variables $\{C_t, N_t, K_t, \frac{M_t}{P_t}, \frac{B_t}{P_t}, v_t, Y_t, w_t, r_t, \bar{\alpha}\}$ satisfying:

$$-\frac{U_{n,t}}{U_{c,t}} = \frac{w_t}{1 + s(v_t) + v_t s'(v_t)}$$

$$v_t^2 s'(v_t) = 1 - Q_t$$

$$\frac{U_{c,t}}{1 + s(v_t) + v_t s'(v_t)} = \beta R_{t+1} \frac{U_{c,t+1}}{1 + s(v_{t+1}) + v_{t+1} s'(v_{t+1})}$$

$$C_t[1 + s(v_t)] + K_{t+1} - (1-\delta)K_t + \frac{M_t}{P_t} + Q_t \frac{B_t}{P_t} = \frac{M_{t-1}}{P_t} + \frac{B_{t-1}}{P_t} + w_t N_t + R_t K_t + \text{Profits}_t$$

$$G_t + \frac{B_{t-1}}{P_t} + \frac{M_{t-1}}{P_t} = T_t + Q_t \frac{B_t}{P_t} + \frac{M_t}{P_t}$$

$$\left\{ \left[\frac{f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{\bar{\alpha}} \right]^{\bar{\alpha}} \left[\frac{1 - f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{1-\bar{\alpha}} \right]^{1-\bar{\alpha}} \right\} \left[1 - m\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}} \right]^{\frac{\epsilon}{\epsilon-1}} = \left[1 - \tau \right]^{\frac{1}{\epsilon-1}}$$

$$\lim_{j\to\infty} (M_{t+j} + B_{t+j}) \geq 0$$

e.g. the no-Ponzi game condition, the government budget constraint and the optimal conditions derived from the household's and firms' problems. We need three more equations describing the clearing of goods, labor and capital markets:

$$Y_t = C_t[1 + s(v_t)] + K_{t+1} - (1-\delta)K_t + G_t \tag{3.24}$$

$$K_t = K_t^{\text{inf}} + K_t^{\text{for}} = \int_0^{\bar{i}} K_{i,t} g(\alpha_i) di + \int_{\bar{i}}^1 K_{i,t} g(\alpha_i) di = \int_0^1 K_{i,t} g(\alpha_i) di \tag{3.25}$$

$$N_t = N_t^{\text{inf}} + N_t^{\text{for}} = \int_0^{\bar{i}} N_{i,t} g(\alpha_i) di + \int_{\bar{i}}^1 N_{i,t} g(\alpha_i) di = \int_0^1 N_{i,t} g(\alpha_i) di \tag{3.26}$$

where equation (24) represents the resource constraint of the economy, and equations (25)-(26) state the equilibrium condition in the perfectly competitive capital and labor markets.

3.3.5 Calibration

As Schmitt-Grohé & Uribe (2011) constitutes the main literature reference of this project, the calibration I adopt follows theirs. Specifically, the demand side of the economy is identical and the parameters describing the preferences and the transactions cost associated with the purchase of consumption goods are exactly the same.

The utility function is logarithmic and additively separable in leisure: $U(C_t, N_t) = \log(C_t) + \theta \log(1 - N_t)$, where the parameter θ gives the weight of leisure in the utility function. The transaction cost is a function of $v_t = \frac{P_t C_t}{M_t}$: $s(v_t) = \phi_1 v_t + \frac{\phi_2}{v_t} - 2\sqrt{\phi_1 \phi_2}$
The main novelties introduced in this work pertain the production side of the economy. In particular, differently from Schmitt-Grohé & Uribe (2011), the model I develop features heterogeneous firms, rather then a representative one, a capital market and an endogenous underground economy. Therefore the main calibration efforts focus on quantifying the parameters that describe the heterogeneity in terms of technology across firms as well as the risk of getting caught evading taxes. Unfortunately, it is really hard to calibrate the technological characteristics of informal firms, as we have limited data information on them. Therefore in the analysis I conduct here I will study how different assumptions on the distribution of the parameter α_i and the probability of detection in the underground economy affect the results. Nonetheless I impose the restriction that the mean of the α_i distribution $\int_0^1 \alpha_i g(\alpha_i) di = 0.33$, which is a recurrent calibration of the capital share in the related literature with a representative firm, while different values for the dispersion of the distribution are tested. Finally, in combination with each assumed degree of dispersion in α_i, I calibrate aggregate productivity A such that the government expenditure $G = 0.04$ corresponds to about 20% of GDP, as in Schmitt-Grohé & Uribe, and the the parameter $\epsilon = 5$, which implies a mark-up $\frac{\epsilon}{\epsilon - 1} = 1.25$, compatible with the theoretical evidence.

3.4 Results

The objective of this paper is to study what is the optimal policy in this economy, intended as the combination of fiscal and monetary policy necessary to finance an exogenous public expenditure delivering the highest welfare for the representative household, that is compatible with the competitive equilibrium allocation. In other words, I solve the Ramsey problem. In order to explore the qualitative and quantitative properties of the proposed model, I conduct three comparative statics exercises. In the first one, I examine how different assumptions on the dispersion of the parameter α_i across firms, keeping the mean of the distribution constant, affect the optimality of the Friedman rule in this framework. In the second one, I keep the distribution of α_i constant and determine the optimal size of the underground economy under alternative assumptions on the parameters governing the risk of getting caught evading taxes. In the last exercise I depart from the assumption of profits taxation and I study the Ramsey problem

Parameter	Symbol	Value
Parameters taken from Schmitt-Grohé & Uribe (2011)		
Discount factor	β	$\frac{1}{1.04}$
Weight of leisure in the utility function	θ	2.9
Transaction cost parameter 1	ϕ_1	0.0111
Transaction cost parameter 2	ϕ_2	0.07524
Government expenditure	G_t	0.04
Public debt	B_t	0
Parameters taken from the literature		
Depreciation rate	δ	12%
Mean of capital share parameter	$\int_0^1 \alpha_i g(\alpha_i) di$	0.33
Parameters calibrated matching data moments		
Productivity	A	1.61
Elasticity of substitution across varieties	ϵ	5

Table 3.1: Calibration

when the only conventional fiscal tool is a sales tax.

3.4.1 The Ramsey problem

The Ramsey problem consists of choosing the mix of monetary and fiscal policy that delivers the highest welfare for the representative household. Solving it implies also determining what is the optimal size of the informal sector, given the parameters describing firms' technology and the probability of detection in the underground economy. Formally:

$$\max_{C_t,N_t,K_{t+1},v_t,\frac{B_t}{P_t},\frac{M_t}{P_t},\bar{\alpha},w_t,r_t,\tau,\pi_t,Q_t} \sum_{t=0}^{\infty} \beta^t U(C_t, N_t)$$

subject to:

$$-\frac{U_{n,t}}{U_{c,t}} = \frac{w_t}{1 + s(v_t) + v_t s'(v_t)}$$

$$v_t^2 s'(v_t) = 1 - Q_t$$

$$\frac{U_{c,t}}{1 + s(v_t) + v_t s'(v_t)} \frac{1}{\beta} \Pi_{t+1} Q_t = \frac{U_{c,t+1}}{1 + s(v_{t+1}) + v_{t+1} s'(v_{t+1})}$$

$$R_{t+1} = \frac{1}{\Pi_{t+1} Q_t}$$

$$K_t = K_t^{\text{inf}} + K_t^{\text{for}}$$

$$N_t = N_t^{\text{inf}} + N_t^{\text{for}}$$

$$C_t[1 + s(v_t)] + K_{t+1} - (1 - \delta)K_t + G_t = Y_t$$

$$G_t + \frac{B_{t-1}}{P_t} + \frac{M_{t-1}}{P_t} = T_t + Q_t \frac{B_t}{P_t} + \frac{M_t}{P_t}$$

$$Q_t \leq 1$$

$$0 < \tau < 1$$

where:

$$T_t = \tau \int_{\bar{i}}^{1} \left(\frac{P_{i,t}^{\text{for}}}{P_t} Y_{i,t}^{\text{for}} - w_t N_t - R_t K_t \right) g(\alpha_i) di + \int_0^{\bar{i}} m \left(\frac{K_{i,t}}{N_{i,t}} \right)^*_{\text{inf}} \frac{P_{i,t}^{\text{inf}}}{P_t} Y_{i,t}^{\text{inf}} g(\alpha_i) di$$

$$K_t^{\text{inf}} = Y_t A_t^{\epsilon-1} \left(\frac{\epsilon - 1}{\epsilon} \right)^{\epsilon} \int_0^{\bar{i}} (1 - m(\alpha_i))^{\epsilon} \left(\frac{R_t}{f(\alpha_i)} \right)^{\alpha_i(1-\epsilon)-1} \left(\frac{w_t}{1 - f(\alpha_i)} \right)^{(1-\alpha_i)(1-\epsilon)} g(\alpha_i) di$$

$$K_t^{\text{for}} = Y_t A_t^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\int_{\bar{i}}^{1}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i(1-\epsilon)-1}\left(\frac{w_t}{1-\alpha_i}\right)^{(1-\alpha_i)(1-\epsilon)} g(\alpha_i)di$$

$$N_{i,t}^{\text{inf}} = Y_t A_t^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\int_{0}^{\bar{i}}(1-m(\alpha_i))^{\epsilon}\left(\frac{R_t}{f(\alpha_i)}\right)^{\alpha_i(1-\epsilon)}\left(\frac{w_t}{1-f(\alpha_i)}\right)^{(1-\alpha_i)(1-\epsilon)-1} g(\alpha_i)di$$

$$N_{i,t}^{\text{for}} = Y_t A_t^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\int_{\bar{i}}^{1}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i(1-\epsilon)}\left(\frac{w_t}{1-\alpha_i}\right)^{(1-\alpha_i)(1-\epsilon)-1} g(\alpha_i)di$$

$$v_t = \frac{P_t C_t}{M_t}$$

In the framework set up above, the government nominal debt B_t is actually undetermined. For the sake of simplicity, I will assume it is zero for all t. Obviously, it is a good idea to reduce the size of the problem, by eliminating some variables through the constraints. Furthermore, as I am interested in determining the optimal inflation tax in the long-run, I restrict my attention to the steady state(s) of this economy. Thus, I impose the conditions $K_{t+1} = K_t = K$ and $C_{t+1} = C_t = C$ and simplify the system of equations:

$$\max_{Q,\tau} U(C,N)$$

subject to:

$$-\frac{U_n}{U_c} = \frac{w}{1+s(v)+vs'(v)}$$

$$v^2 s'(v) = 1-Q$$

$$R = \frac{1}{\beta}$$

$$\Pi = [RQ]^{-1}$$

$$K = K^{\text{inf}} + K^{\text{for}}$$

$$N = N^{\text{inf}} + N^{\text{for}}$$

$$C[1+s(v)] - \delta K + G = Y$$

$$G = T + \frac{\Delta M}{P}$$

$$Q \leq 1$$

$$0 < \tau < 1$$
$$v = \frac{C}{\frac{M}{P}}$$
$$\frac{\Delta M}{P} = \frac{C}{v}\left(1 - \frac{1}{\Pi}\right)$$

What is the policy trade-off?

Firstly, it is important to identify the policy trade-off and the multiple sources of distortion present in this framework. On the demand side, consumers face a transaction cost associated to consumption purchases. Specifically, they need money balances to perform transactions, but holding money is costly, whenever the nominal interest rate is positive. In fact, a Q below 1 creates a wedge between the marginal rate of substitution of consumption for leisure and the real wage, pushing agents to consume and work less. On the production side, firms are responsible for two forms of inefficiency. First of all, they operate in a regime of monopolistic competition. Thus, thanks to some market power, they act as price-makers and extract profits. Secondly, firms face an incentive to evade taxes at the cost of risking to get caught by the authorities. Therefore, those firms which decide to operate informally adopt a suboptimal capital-labor ratio in order to conceal their activities, as, by assumption, the probability of being sanctioned is an increasing function of the capital-labor ratio employed.
The sources of inefficiency just described shape the policy trade-off faced by the government when choosing the mix of monetary and fiscal policy necessary to finance its expenditure. A positive nominal interest rate lifts some tax burden from the shoulders of formal firms and favours the movement of some informal firms to the formal economy. However, it makes transactions more costly and depresses consumption and the labor supply. A higher tax rate reduces the reliance on seigniorage revenues but favours the development of the underground sector. In fact, increasing the tax rate to finance public expenditure has two opposite effects: on the one hand it raises the tax revenues obtained by taxing more the profits of formal firms, on the other it pushes some firms to evade taxes and therefore it reduces tax revenues. The net effect depends on the density of firms around the cutoff firm $\bar{i}$, the one which is indifferent between operating formally or informally. If there are relatively fewer (more) firms right below the threshold than the ones above $\bar{i}$, increasing the tax rate, for given factor prices, would increase (decrease) tax revenues. Moreover, due to the fact that the probability

of getting caught producing informally is, by assumption, increasing in the capital-ratio employed by informal firms, general equilibrium effects mitigate the incentives of formal firms to evade taxes following an increase in the tax rate. A rise in τ encourages some formal firms to produce in the underground economy and to adopt lower capital-labor ratios. As a consequence, a higher tax rate increases the demand for labor for a given real wage. It follows that the real wage goes up, making a movement from the formal to the informal sector less attractive and likely. The strength of this general equilibrium effect crucially depends on the elasticity of the risk of being detected evading taxes with respect to the capital-labor ratio employed and, again, on the mass of firms around the cutoff $\bar{i}$.

In this framework, the Ramsey social planner may find it convenient to raise some seignorage revenues through a positive nominal interest rate as the existence of an informal sector evading taxes limits the ability of the government to fund its expenditure through fiscal policy. This public finance motive for inflation is reinforced by the inefficiency implied by producing in the informal sector: as informal firms try to hide their activities from tax authorities, they employ a capital-labor ratio that is lower than the one they would choose in absence of taxes on profits, for the technology they possess. Therefore, in order to make-up for this distortion and to find the resources to cover the government expenditure, it may be optimal to set a lower tax rate and a higher inflation rate than those that result to be optimal in the Schmitt-Grohé and Uribe model.

Is a zero nominal interest policy optimal?

A second question that naturally arises is whether the Friedman rule, i.e. $Q = 1$ (zero nominal interest rate), can be optimal in this framework. When such policy is implemented, the optimal conditions derived from the household's problem become:

$$v = \underline{v}$$

$$-\frac{U_n}{U_c} = w$$

$$\Pi = \frac{1}{R}$$

Velocity v is at its satiation point and the wedge between the marginal rate of substitution of consumption for leisure and the real wage vanishes. At steady state, inflation would be negative and equal to minus the real interest rate. Is this optimal?

Such interest rate rule would cancel any distortion due to the consumption transaction costs, but the government would have to finance the exogenous public expenditure

(and the negative seigniorage outflow) via the tax on profits uniquely. It is worth distinguishing two alternative scenarios: if the density $g(\alpha_i)$ and the parameters governing the probability of being detected evading taxes were such that no firms had the incentive to operate informally even of the highest tax rates, the government would have the full ability to extract the tax revenues needed to balance its budget from the economy production side, without incurring in the inefficiencies associated with the underground sector. In this situation, the combination of the Friedman rule and a tax on profits would be the solution of the Ramsey problem, but only if the tax revenues were enough to cover the government expenditure and the negative seigniorage outflow implied by the calibration. In fact, the tax on profits acts as lump-sum, as it does not distort the optimal decisions of firms, but the maximum amount of its proceedings is constrained by the condition that firms' profits cannot be negative. Whenever such revenues are not enough to make up for the government expenses, the planner resorts to the lowest positive interest rate compatible with the government budget. This consideration aside, the Friedman rule would be optimal in this scenario as any positive interest rate would only imply transaction costs for the representative household.
A more interesting scenario appears when the coefficients governing the probability of detecting underground activities are such that some firms actually decide to operate informally for any tax rate τ. In this case the tax on profits would be distortionary, as it would provide some firms with the incentive to evade taxes and to employ sub-optimal capital-labor ratios causing an efficiency loss. Under these circumstances, the solution of the Ramsey optimal policy problem would involve a positive nominal interest rate as a tool to impose an indirect tax on the shadow economy.

3.4.2 The role of firms' heterogeneity

In this section I examine the role played by the distribution of the parameter α across firms in relation to the solution of the Ramsey problem. In the quantitative exercise I assume that the probability of being caught evading taxes $m\left(\frac{K_{i,t}}{N_{i,t}}\right) = \psi_1 + \psi_2\left(\frac{K_{i,t}}{N_{i,t}}\right)$ is characterized by the following calibration: $\psi_1 = 0.15$ and $\psi_2 = 19$ and that $A = 1.61$. Such values ensure that the size of underground is realistic and that a government expenditure of $G = 0.04$ amounts to around 20% of GDP, as in Shmitt-Grohé & Uribe. The experiment consists of studying how the optimal mix of inflation and tax rate is affected by the dispersion of the distribution $g(\alpha_i)$, holding the its mean $\int_0^1 \alpha_i g(\alpha_i) di$ constant and equal to 0.33. I pick three different values for the dispersion of $g(\alpha_i)$, referred to, in increasing order, as σ^l, σ^m and σ^h and compare the steady-state allocations obtained adopting the optimal policy.

Figure (3.2) plots the densities resulting from the three different values of dispersion and table (3.2) displays the results of the analysis. It emerges that higher degrees of dispersion in terms of technology are associated with higher optimal rates of inflation, larger underground sectors and lower tax rates. Such outcome can be interpreted observing that a higher degree of dispersion means that there many firms at the bottom of the distribution that may find it optimal to evade taxes. As a consequence, setting the tax rate τ too high may reduce the tax revenues considerably and favour the inefficiency associated with producing informally. Therefore, the government finds it optimal to set a higher nominal interest rate, a higher inflation rate to finance its expenditure and to contain the size of the underground economy. Given the chosen calibration, the cut-off value $\bar{\alpha}$ implied by the optimal policy mix is decreasing in the degree of dispersion. Interestingly, the real wage w and output Y resulting from implementing the optimal combination of monetary and fiscal policy are increasing in the value of dispersion. The latter outcome emerges also when we analyze the a similar economy where firms cannot evade taxes. Table (3.3) highlights the main steady state summary statistics under this assumption and reveals that in this case the optimal inflation rate would be insensitive to the degree of dispersion of α_i across firms and equal the one implied by the Friedman rule. By comparing table (3.2) and (3.3) we can observe the loss in terms of efficiency and welfare imposed by the presence of the underground sector, as well as how an incomplete tax system due to tax evasion creates an incentive for the planner to use inflation as an indirect tax on the informal economy.

3.4.3 The role of the probability of detection

The second quantitative exercise focuses on the parameters governing the probability of getting caught producing informally and evading taxes. For the sake of simplicity, I have assumed throughout the paper that such probability is a linear function of the capital-labor ratio employed: $m\left(\frac{K_{i,t}}{N_{i,t}}\right) = \psi_1 + \psi_2\left(\frac{K_{i,t}}{N_{i,t}}\right)$. The value of ψ_1 can be interpreted as the quality of the government law enforcement. In fact, it corresponds to the probability of being found producing informally when the firm employs no capital. The magnitude of the parameter ψ_2 has a double function: it gives a measure of how the firm's size makes the firm more likely to be caught evading taxes as well as the strength of the misallocation due to the underground sector. Here, I examine how different values of the coefficients ψ_1 and ψ_2 affect the optimal inflation rate π - tax rate τ mix. Table (3.4) reports the optimal combinations of fiscal and monetary policy when ψ_2

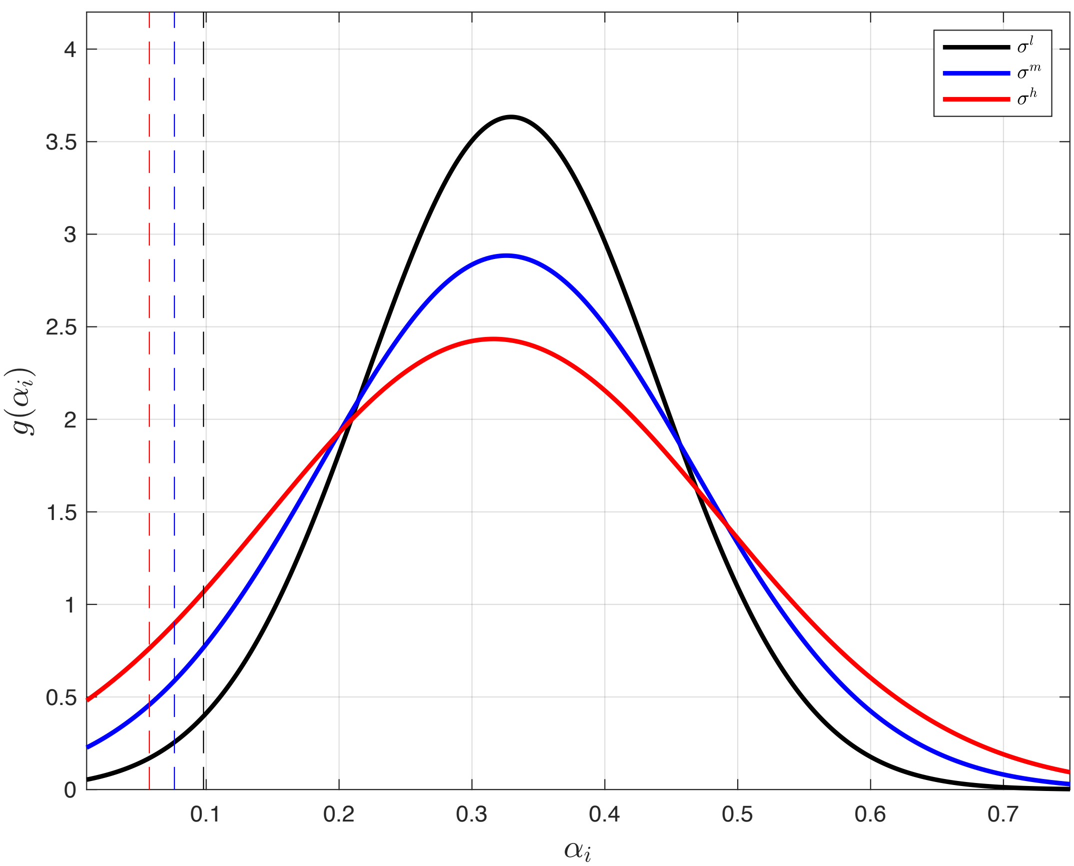


Figure 3.2: Density $g(\alpha_i)$ for different values of σ

Variable	σ^l	σ^m	σ^h
optimal π	13%	15%	18%
τ	85.98%	81.98%	77.61%
$\frac{\Delta M}{P}$	0.0041	0.0047	0.0055
T	0.0359	0.0353	0.0345
Y	0.2082	0.2143	0.2208
C	0.1616	0.1679	0.1744
w	0.6094	0.6376	0.6670
$\bar{\alpha}$	0.0980	0.0761	0.0572
$\frac{Y^{inf}}{Y}$	1.28%	3.11%	4.08%

Table 3.2: Optimal inflation rates for different values of σ

Variable	σ^l	σ^m	σ^h
optimal π	-3.85%	-3.85%	-3.85%
τ	98.56%	95.10%	92.29%
$\frac{\Delta M}{P}$	-0.0026	-0.0028	-0.0029
T	0.0426	0.0428	0.0429
Y	0.2163	0.2249	0.2324
C	0.1712	0.1802	0.1881
w	0.6252	0.6625	0.6941
$\bar{\alpha}$	0	0	0
$\frac{Y^{inf}}{Y}$	0%	0%	0%

Table 3.3: Optimal inflation rates for different values of σ, when firms have no incentive to evade taxes

is kept constant at 19 and ψ_1 takes three different values: $\psi_1^l = 0.1$, $\psi_1^m = 0.15$ and $\psi_1^h = 0.2$. Larger values of ψ_1 are associated with an increasingly weaker incentive to evade taxes. As a consequence, the Ramsey allocation features smaller underground sectors and lower optimal inflation rates.

Tables (3.5) instead compares three different sets of steady-state statistics obtained following optimal policy when ψ_1 is kept constant at 0.15 and $\psi_2^l = 12$, $\psi_2^m = 19$ and $\psi_2^h = 25$. A larger ψ_2 determines, for the same chosen capital-labor ratio, a higher probability of incurring into a sanction for evading taxes. The results of this test point in the same direction of the previous one: if evading taxes gets riskier, less firms decide to produce informally and the underground sector shrinks. This favours the use of the conventional tax on profits to finance government spending and setting a lower nominal interest rate results to be optimal.

The bottom line of this exercise is that differences in the effectiveness of the law enforcement system can account for the positive correlation between the size of the underground sectors and the inflation rates observed in across countries. Developing economies, where the function of the state to control tax evasion and detect informal activities faces greater difficulties, have a stronger incentive to use the inflation tax to meet public finance goals.

3.4.4 Distortionary taxation

In the framework analyzed so far taxation is distortionary as long as it provides an incentive to some firms to evade taxes. Those firms that decide to operate informally adopt lower capital-labor ratios than formal firms in the attempt of avoiding to get caught by the authorities, generating an efficiency loss. However, if law enforcement was so strong that no firm found it convenient to evade taxes, taxation would act as lump-sum and the optimal demand of labor and capital would not be distorted. In this section I study whether substituting the tax on profits with a tax on sales would affect the conclusions drawn so far. A tax on sales would be distortionary regardless of the incentive to evade taxes and the presence of the underground economy. In order to see

Variable	ψ_1^l	ψ_1^m	ψ_1^h
optimal π	40%	18%	10%
τ	74.44%	77.61%	79.74%
$\frac{\Delta M}{P}$	0.0082	0.0055	0.0038
T	0.0318	0.0345	0.0362
Y	0.2179	0.2208	0.2238
C	0.1701	0.1744	0.1782
w	0.6636	0.6670	0.6751
$\bar{\alpha}$	0.0711	0.0572	0.0395
$\frac{Y^{inf}}{Y}$	6.91%	4.08%	1.82%

Table 3.4: Optimal inflation rates for different values of ψ_1

Variable	ψ_2^l	ψ_2^m	ψ_2^h
optimal π	23%	18%	16%
τ	75.78%	77.61%	78.4%
$\frac{\Delta M}{P}$	0.0063	0.0055	0.0051
T	0.0337	0.0345	0.0349
Y	0.2202	0.2208	0.2212
C	0.1734	0.1744	0.1750
w	0.6667	0.6670	0.6677
$\bar{\alpha}$	0.0591	0.0572	0.0554
$\frac{Y^{inf}}{Y}$	4.52%	4.08%	3.77%

Table 3.5: Optimal inflation rates for different values of ψ_2

this, let us solve the problem of the formal firm under this new assumption:

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t}} \quad (1-\tau_t)\frac{P_{i,t}}{P_t}Y_{i,t} - w_t N_{i,t} - R_t K_{i,t}$$

$$\text{s.t.} \qquad Y_{i,t} = \left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t$$

$$Y_{i,t} = A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}$$

Where τ_t represents the sales' tax rate. Setting up a Lagrangean:

$$\max_{\frac{P_{i,t}}{P_t},K_{i,t},N_{i,t},\lambda_{i,t}} \quad (1-\tau_t)\left(\frac{P_{i,t}}{P_t}\right)^{1-\epsilon} Y_t - w_t N_{i,t} - R_t K_{i,t} - \lambda_{i,t}\left[\left(\frac{P_{i,t}}{P_t}\right)^{-\epsilon} Y_t - A_t K_{i,t}^{\alpha_i} N_{i,t}^{1-\alpha_i}\right]$$

The first order conditions with respect to the price, capital and labor imply:

$$\left(\frac{P_{i,t}}{P_t}\right)^{*}_{\text{for}} = \frac{\epsilon}{\epsilon-1}\frac{1}{1-\tau_t}\lambda_{i,t}^{\text{for}}$$

$$w_t = \lambda_{i,t}^{\text{for}}(1-\alpha_i)A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i}_{\text{for}}$$

$$R_t = \lambda_{i,t}^{\text{for}}\alpha_i A_t\left(\frac{K_{i,t}}{N_{i,t}}\right)^{\alpha_i-1}_{\text{for}}$$

After combining the three latter equations with the constraints we can derive the optimal demand for capital and labor and the price set by the formal firm:

$$\left(\frac{K_{i,t}}{N_{i,t}}\right)^{*}_{\text{for}} = \frac{w_t}{R_t}\frac{\alpha_i}{1-\alpha_i} \tag{3.27}$$

$$\left(\frac{P_{i,t}}{P_t}\right)^{*}_{\text{for}} = \frac{\epsilon}{\epsilon-1}\frac{1}{1-\tau_t}\frac{1}{A_t}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i}\left(\frac{w_t}{1-\alpha_i}\right)^{1-\alpha_i} \tag{3.28}$$

$$K_{i,t}^{\text{for}} = Y_t\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}(1-\tau_t)^{\epsilon}A_t^{\epsilon-1}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i(1-\epsilon)-1}\left(\frac{w_t}{1-\alpha_i}\right)^{(1-\alpha_i)(1-\epsilon)} \tag{3.29}$$

$$N_{i,t}^{\text{for}} = Y_t\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}(1-\tau_t)^{\epsilon}A_t^{\epsilon-1}\left(\frac{R_t}{\alpha_i}\right)^{\alpha_i(1-\epsilon)}\left(\frac{w_t}{1-\alpha_i}\right)^{(1-\alpha_i)(1-\epsilon)-1} \tag{3.30}$$

We can observe that, even though the capital-labor ratio is the same as the one chosen by the formal firm when taxes are on profits, the tax on sales pushes the formal firm to set a higher price and to produce less, by demanding less capital and labor. Furthermore, this distortion would arise also in absence of the possibility to evade taxes. What are

Variable	Full law enforcement	When firms can evade taxes
π	29%	105%
τ	20.89%	19%
$\frac{\Delta M}{P}$	0.0050	0.0084
T	0.0350	0.0316
Y	0.1673	0.1685
C	0.1224	0.1213
w	0.4580	0.4723
$\bar{\alpha}$	0	0.0229
$\frac{Y^{inf}}{Y}$	0%	0.4656%

Table 3.6: Optimal inflation rates with and without an underground sector, when sales are taxed

the implications for the optimal combination of fiscal and monetary policy?
I solve again for the Ramsey social planner problem, which consists of choosing the mix $\{Q_t, \tau_t\}$ necessary to fund the same exogenous amount of government spending G_t, that is compatible with the competitive equilibrium allocation and delivering the highest level of welfare. With the exception of the optimizing conditions for the formal firm, the problem is characterized by the same set of equations laid down in section 4.1. The calibration is analogous to the one performed for the previous quantitative exercises. In particular, I assume that $\sigma = \sigma^l$, $\psi_1 = 0.15$ and $\psi_2 = 19$. Table (3.6) reports the results for the comparison of the steady state allocations obtained following the Ramsey policy when firms cannot vs. when firm can evade taxes. The first column of the table shows that under the assumption of a tax on sales, the optimal inflation rate is very high, around 29%, even in absence of the inefficiency due to the shadow economy. This level is very far from the inflation rate consistent with the Friedman rule and can be accounted for by the distortionary nature of taxation. If we open up the possibility for firms to evade taxes, the optimal inflation rises much further, up to 105%: the support for using the inflation tax is strengthened by the fact that informal firms adopt a capital-labor ratio that is closer to the one that would emerge in absence of distortionary taxation. In other words, the existence of the underground

sector mitigates the distortion imposed by the sales' tax on the demand of labor and capital, but only on those firms evading taxes. As a result, a very high inflation rate is optimal even in presence of a tiny underground sector.

3.5 Some Remarks and Future Prospects

In this section some shortcomings and some potential advancements of this framework are discussed. The key feature of the proposed model lies in the assumption of an endogenous probability of getting caught while operating in the informal sector. Such probability is an increasing function of the capital-labor ratio employed by the firm. The rationale for such assumption is based on the hypothesis that the size of firms positively affects the likelihood to be audited by the tax authorities. This allows the size of the underground economy to be endogenous, which is innovative in the literature. Formal firms fully pay taxes on profits, informal firms fully evade them. Obviously, such outcome is not realistic: in the empirical literature on tax evasion, audits reveal that firms and workers tend to underreport only a part of their earnings. Secondly, in the proposed model the underground economy is represented as a set of firms hiding their production from the tax authorities. However, the phenomenon of tax avoidance and evasion seems to be especially relevant in the labor market, where employers sometimes hire workers illegally to avoid regulations and costs. Thirdly, in this framework the only driver for the existence of an underground economy is tax evasion, while, in reality, it is a much more complex phenomenon. But the main shortcoming of the analysis is that any possible simulation exercise to test the implications of the model requires calibrating from the data the parameters related to the probability of detection in the informal sector and those describing the technological features of the entire production sector of the economy. This does not appear an easy task, given the lack of information and data on the underground economy. However the proposed exercises show that even for uncalibrated but conservative values of such parameters, relatively high inflation rates can be optimal in economies featuring small underground sectors. Moreover, the Friedman rule fails to be optimal even though taxes are lump-sum, because of the incentive they provide to labor-intensive firms to evade them. These results go against the conclusions drawn in Schmitt-Grohé & Uribe (2011), where only unrealistically large underground economies can account for the observed inflation targets. Moreover, if we substitute the tax on profits with one on sales, e.g. we introduce distortionary taxation, the support for using the inflation tax is further reinforced, even in absence of the informal sector.

The public finance motive for inflation modelled here helps explaining the correlation between the size of the underground sectors and the observed inflation rates prevailing among developing economies. The framework, however, falls short in accounting for realistic combinations of inflation rates and underground economies. In fact, from the chosen calibration, the optimal inflation rates appear to be much higher than the observed ones. Such limitation can be overcome if we augment the theoretical setup with an incentive, for the policy-maker, to keep inflation low, as nominal rigidities. Sticky prices, for example, call for an optimal inflation rate equal to zero in standard frameworks like the new Keynesian model. Departing from the assumption of flexible prices may introduce a further degree of complexity to the model, but could help to find realistic inflation rates as a solution of the Ramsey problem.

[]

Bibliography

[1] Ari Aisen and Francisco José Veiga. The political economy of seigniorage. *Journal of Development Economics*, 87(1):29–50, 2008.

[2] James Andreoni, Brian Erard, and Jonathan Feinstein. Tax Compliance. *Journal of Economic Literature*, 36(2):818–860, 1998.

[3] Patrick K. Asea. The Informal Sector: Baby or Bath Water? *Carnegie-Rochester Conference Series on Public Policy*, 45(December):163–171, 1996.

[4] Francesco Busato and Bruno Chiarini. Market and Underground Activities in a Two-Sector Dynamic Equilibrium Model. *Economic Theory*, 23(4):831–861, 2004.

[5] Alina Carare and Mark R. Stone. Inflation Targeting Regimes. *IMF Working Paper*, January:1–36, 2003.

[6] V.V. Chari, Lawrence J. Christiano, and Patrick J. Kehoe. Optimality of the Friedman Rule in Economies with Distorting Taxes. *Journal of Monetary Economics*, 37:203–223, 1996.

[7] I. Correia and P. Teles. The Optimal Inflation Tax. *Review of Economic Dynamics*, 2(2):325–346, 1999.

[8] Isabel Correia, Juan Pablo Nicolini, and Pedro Teles. Optimal fiscal and monetary policy: Equivalence results. *Journal of Political Economy*, 116(1):141–170, 2008.

[9] Tiago V. de V. Cavalcanti and Anne P. Villamil. Optimal Inflation Tax and Structural Reform. *Macroeconomic Dynamics*, 7(03):333–362, 2003.

[10] Edgar L. Feige. *The Underground Economies: Tax Evasion and Information Distortion*. Cambridge University Press, 1989.

[11] Milton Friedman. The Optimum Quantity of Money. *The Optimum Quantity of Money and Other Essays*, 1969.

[12] Tatyana A. Koreshkova. A Quantitative Analysis of Inflation as a Tax on the Underground Economy. *Journal of Monetary Economics*, 53(4):773–796, 2006.

[13] Norman V. Loayza. The Economics of the Informal Sector: a Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy*, 45(December):129–162, 1996.

[14] N. Gregory Mankiw. The optimal collection of seigniorage. Theory and evidence. *Journal of Monetary Economics*, 20(2):327–341, 1987.

[15] Juan Pablo Nicolini. Tax Evasion and the Optimal Inflation Tax. *Journal of Development Economics*, 55:215–232, 1998.

[16] Renzo Orsi, Davide Raggi, and Francesco Turino. Size, trend, and policy implications of the underground economy. *Review of Economic Dynamics*, 17(3):417–436, 2014.

[17] Edmund S. Phelps. Inflation in the Theory of Public Finance. *The Swedish Journal of Economics*, 75:67–82, 1973.

[18] Paulina Restrepo-Echavarria. Macroeconomic volatility: The role of the informal economy. *European Economic Review*, 70:454–469, 2014.

[19] Stephanie Schmitt-Grohé and Martín Uribe. Optimal fiscal and monetary policy under imperfect competition. *Journal of Macroeconomics*, 26(2):183–209, 2004.

[20] Stephanie Schmitt-Grohé and Martín Uribe. Optimal fiscal and monetary policy with sticky wages and sticky prices. *Journal of Economic Theory*, 114:198–230, 2004.

[21] Stephanie Schmitt-Grohé and Martín Uribe. Optimal Simple and Implementable Monetary and Fiscal Rules. *Nber Working Paper Series 10253*, (January), 2004.

[22] Stephanie Schmitt-Grohé and Martín Uribe. The Optimal Rate of Inflation. *Handbook of Monetary Economics*, 3(1):653–722, 2011.

[23] Friedrich Schneider, Andreas Buehn, and Claudio E. Montenegro. New Estimates for the Shadow Economies all over the World. *International Economic Journal*, 24(4):443–461, 2010.

[24] Friedrich Schneider and Dominik H. Enste. Shadow Economies: Size, Causes and Consequences. *Journal of Economic Literature*, 38(1):77–114, 2000.

[25] Joel Slemrod. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48, 2007.

Appendices

I demonstrate here that equation (15) has a unique, if any, solution in the unknown variable $\bar{\alpha}$:

$$\left\{\left[\frac{f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{\bar{\alpha}}\right]^{\bar{\alpha}}\left[\frac{1-f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{1-\bar{\alpha}}\right]^{1-\bar{\alpha}}\right\}\left[1-m\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}}=\left[1-\tau\right]^{\frac{1}{\epsilon-1}}$$

The strategy of the proof involves the following steps:

1. show that $\left\{\left[\frac{f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{\bar{\alpha}}\right]^{\bar{\alpha}}\left[\frac{1-f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{1-\bar{\alpha}}\right]^{1-\bar{\alpha}}\right\}\left[1-m\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}}$ is equal to 1 when $\alpha_i=0$;
2. show that $\left[\frac{f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{\bar{\alpha}}\right]^{\bar{\alpha}}\left[\frac{1-f\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}}{1-\bar{\alpha}}\right]^{1-\bar{\alpha}}$ decreases as α_i increases;
3. show that $\left[1-m\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}}$ decreases as α_i increases.

First, let us evaluate the functions functions $m\left(\frac{K_{i,t}}{N_{i,t}}(\alpha_i)\right)$ and $f\left(\frac{K_{i,t}}{N_{i,t}}(\alpha_i)\right)$ at the optimal capital-labor ratio chosen by the informal firm (equation (13)), under the assumption that the probability of detection is linear:

$$m\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}=\frac{-\left[(1-\alpha_i)+\frac{w_t}{r_t}\psi(\alpha_i+\frac{\epsilon}{\epsilon-1})\right]+\sqrt{\left[(1-\alpha_i)+\frac{w_t}{r_t}\psi(\alpha_i+\frac{\epsilon}{\epsilon-1})\right]^2+4\frac{w_t}{r_t}\psi\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))}}{2(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))}$$

$$f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}=\frac{\alpha_i+\frac{\epsilon}{\epsilon-1}+\frac{r_t}{w_t\psi}(1+\alpha_i)-\sqrt{\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))}}{2(\frac{r_t}{w_t\psi}+1)}$$

Notice that both expressions take value 0 when $\alpha_i=0$. This implies that $\left[1-m\left(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\right)^*_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}}=1$ when $\alpha_i=0$. Focussing on the latter one, subtract α_i from both sides:

$$f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}-\alpha_i=\frac{-\alpha_i+\frac{\epsilon}{\epsilon-1}+\frac{r_t}{w_t\psi}(1-\alpha_i)-\sqrt{\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))}}{2(\frac{r_t}{w_t\psi}+1)} \tag{31}$$

Let us prove that it is always ≤ 0.

1.

As r_t, w_t and ψ are always positive, we can just focus on the sign of the numerator on the right hand-side, which must be negative:

$$-\alpha_i+\frac{\epsilon}{\epsilon-1}+\frac{r_t}{w_t\psi}(1-\alpha_i) \leq \sqrt{\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i\left(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i)\right)}$$

There are two possible cases: either $-\alpha_i+\frac{\epsilon}{\epsilon-1}+\frac{r_t}{w_t\psi}(1-\alpha_i)<0$ or not. In the first case, the proof would be already complete. In order to check the second case, let us square both sides. After few steps we obtain that the condition is:

$$\frac{r_t}{w_t\psi}+1 \geq 0$$

which is always true as r_t, w_t and ψ are always positive and this completes the proof.$\diamond$

Let us now turn to the following expression:

$$\left[\frac{f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}{\alpha_i}\right]^{\alpha_i}\left[\frac{1-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}{1-\alpha_i}\right]^{1-\alpha_i} = \exp\left\{\alpha_i \log\left[\frac{f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}{\alpha_i}\right]+(1-\alpha_i)\log\left[\frac{1-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}}{1-\alpha_i}\right]\right\}$$

Interestingly, the term in the curly brackets on the right-hand side can be interpreted as to what in thermodynamics is called the relative entropy between two Bernoulli probability distributions, characterized by the probabilities α_i and $f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}$. For its peculiar features, it decreases as the distance between α_i and $f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}$ increases. It is null when $f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}=\alpha_i$, which occurs only when $\alpha_i=0$. In such case, the exponential function is equal to 1.

At this point, multiply both sides of equation (27) for $-2(\frac{r_t}{w_t\psi}+1)$:

$$2\left(\frac{r_t}{w_t\psi}+1\right)\left(\alpha_i-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}\right)=$$
$$\alpha_i-\frac{\epsilon}{\epsilon-1}-\frac{r_t}{w_t\psi}(1-\alpha_i)+\sqrt{\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))}$$

We want to study how the distance between α_i and $f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}$ evolves as α_i increases. We will show that it grows.

2.

Take the first derivative of the right hand-side with respect to α_i. After some steps we obtain:

$$(\frac{r_t}{w_t\psi}+1)+(\frac{r_t}{w_t\psi}+1)\left(\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))\right)^{-\frac{1}{2}}\big[\alpha_i(\frac{r_t}{w_t\psi}+1)-\frac{r_t}{w_t\psi}+\frac{\epsilon}{\epsilon-1}$$

Let us check when this derivative is positive. As $\left(\frac{r_t}{w_t\psi}+1\right)>0$ we just need to focus on when:

$$\left(\left[\frac{r_t}{w_t\psi}(1-\alpha_i)+\alpha_i+\frac{\epsilon}{\epsilon-1}\right]^2+4\frac{r_t}{w_t\psi}\alpha_i(\frac{\epsilon}{\epsilon-1}-(1-\alpha_i))\right)^{\frac{1}{2}}+\big[\alpha_i(\frac{r_t}{w_t\psi}+1)-\frac{r_t}{w_t\psi}+\frac{\epsilon}{\epsilon-1}\big]\geq 0$$

Which occurs when:

$$\frac{\epsilon}{\epsilon-1}\big(\frac{r_t}{w_t\psi}+1\big)\geq 0 \tag{32}$$

Since $\epsilon>1$, by assumption, the last inequality is always satisfied. Then the difference $\alpha_i-f\big(\frac{K_{i,t}}{N_{i,t}}\big)^*_{\text{inf}}$ increases as α_i increases.$\diamond$

3.

By the definition of the function $f(\quad)$:

$$\alpha_i-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}=\frac{\epsilon}{\epsilon-1}\left(\frac{1}{1-m\big(\frac{K_{i,t}}{N_{i,t}}\big)^*_{\text{inf}}}-1\right)$$

Then, taking the first derivative with respect to α_i:

$$D_{\alpha_i}\left[\alpha_i-f\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}\right]=\frac{\epsilon}{\epsilon-1}\frac{1}{\big(1-m\big(\frac{K_{i,t}}{N_{i,t}}\big)^*_{\text{inf}}\big)^2}D_{\alpha_i}\left[m\left(\frac{K_{i,t}}{N_{i,t}}\right)^*_{\text{inf}}\right]$$

Which implies that also the function $m(\quad)$ is increasing in α_i. As a consequence, $\left[1-m\big(\frac{K_{\bar{i},t}}{N_{\bar{i},t}}\big)^*_{\text{inf}}\right]^{\frac{\epsilon}{\epsilon-1}}$ decreases as α_i increases.$\diamond$

The proof is complete.

Country	Inflation target (2015)	Target set by	Underground economy as % of GDP (Average 1999-2007)
Albania	3% ± 1		36.3%
Armenia	4% ± 1.5	G and CB	48.7%
Australia	2% - 3%	G and CB	14.6%
Azerbaijan	5% - 6%		63.3%
Bangladesh	6.5%		35.9%
Belarus	18% ± 2		49.8%
Botswana	3% - 6%		33.8%
Brazil	4.5% ± 2	G and CB	40.5%
Canada	2% ± 1	G and CB	16.3%
Chile	3% ± 1	CB	20.3%
China	4%		13.5%
Colombia	2% - 4%	CB	41%
Costa Rica	4% ± 1		26.6%
Czech Republic	2% ± 1	CB	19.8%
Dominican Republic	4% ± 1		32.3%
Georgia	5%		68.8%
Ghana	8% ± 2%	G and CB	43.2%
Guatemala	4.5% ± 1	CB	52.5%
Hungary	3%	CB	25.8%
Iceland	2.5%	G and CB	16.2%
India	8%		24%
Indonesia	4.5% ± 1	G and CB	19.9 %
Israel	1% - 3%	G and CB	21.8%
Japan	2%		11.4%
Kenya	5% ± 2.5		35.5%
Kyrgyzstan	7%		42%
Malawi	12% - 15%		38.9%
Mexico	3% ± 1	CB	30.2%
Moldova	5% ± 1.5		45.8%
Mongolia	7%		19.2%
Mozambique	6%		40.8%
New Zealand	1% - 3%	G and CB	13.2%
Nigeria	6% - 9%		59.6%
Norway	2.5%	G	19.5%

Country	Inflation target (2015)	Target set by	Underground economy as % of GDP (Average 1999-2007)
Pakistan	8%		37.9%
Peru	2% ± 1	CB	61.8%
Philippines	4% ± 1	G and CB	45.1%
Poland	2.5% ± 1	CB	28%
Romania	3% ± 1	G and CB	36.3%
Russia	4.5% ± 1.5		48.6%
Serbia	4% ± 1.5	G and CB	–
South Africa	3% - 6%	G	29.5%
South Korea	2.5% - 3.5%	G and CB	28.2%
Sri Lanka	3% - 5%		45.3%
Sweden	2%	CB	19.6%
Switzerland	$< 2\%$		8.6%
Thailand	2.5% ± 1.5	G and CB	54.7%
Turkey	5% ± 1.5	G and CB	33%
Uganda	5%		43.9%
Ukraine	4% - 9%		54.9%
United Kingdom	2%	G	12.9%
Uruguay	3% - 7%		51.5%
USA	2%		8.8%
Vietnam	7%		16.1%
Zambia	7%		50.8%