

Optimal Taxation of Wealthy Individuals*

Ali Shourideh

Wharton School, University of Pennsylvania

shouride@wharton.upenn.edu

August 31, 2014

Abstract

This paper studies the determinants of optimal taxes for wealthy individuals faced with capital income risk. I develop a model of optimal taxation of capital income in which wealth and income inequality is a result of capital income shocks together with frictions in financial markets. I use the model to study optimal taxation of various types of capital income: capital income from controlled businesses, outside the business as well as bequests. In presence of risk-return trade-offs, i.e., when more productive investments are riskier, I show that it is typically optimal to have progressive saving taxes. Furthermore, in an intergenerational context, I show that bequest taxes should be negative. Finally, I study the implications of the model on long run efficient distribution of wealth. I show that the long-run distribution of wealth has a fat-tail distribution and compare the efficient tail of the wealth distribution to the one resulting from an ad-hoc incomplete market model.

Keywords: Optimal taxation of capital, Capital income risk, Progressive taxes, Pareto Distribution

*This is a significantly revised version of the the first chapter of my PhD dissertation at the University of Minnesota, previously circulated as ‘Optimal Taxation of Capital Income: A Mirrleesian Approach to Capital Accumulation’. I am indebted to V. V. Chari and Larry Jones for valuable advice. This work has benefited from extensive comments and discussions from Stefania Albanesi, Florian Scheuer, and Ariel Zetlin-Jones. I would also like to thank Andy Abel, Laurence Ales, Alessandro Dovis, Andy Glover, Mike Golosov, Joao Gomes, Roozbeh Hosseini, Patrick Kehoe, Narayana Kocherlakota, Ellen McGrattan, Alessandro Pavan, Chris Phelan, Maryam Saeedi, Erick Sager, Chris Sleet, Alireza Tahbaz-Salehi, Maxim Troshkin, as well as seminar participants in Booth, Carnegie Mellon, Chicago, Columbia GSB, Northwestern, Princeton, Wharton, Yale, Minneapolis Fed, St Louis Fed, Ny Fed, 2012 AEA, NBER Macro Public Finance, NBER EFG meeting and Midwest Macro Conference for their comments. I acknowledge the support of University of Minnesota Doctoral Dissertation Fellowship. I thank Department of Economics at NYU for their hospitality.

1 Introduction

How should the income of the wealthy be taxed? When governments care about redistribution, how much wealth inequality should they allow? The answer to this question requires taking a stand on the process of wealth accumulation. Much of the optimal tax theory has tackled this question by using models where households are subject to idiosyncratic labor income risk and accumulate wealth as buffer against future income shocks. However, it has been documented that models with idiosyncratic labor income risk fail to generate a concentration of wealth at the top similar to that observed in the data. It has also been argued that models with households who are subject to capital income risk can generate a concentration of wealth similar to that in the data.¹ In this paper, motivated by this insight, I study optimal taxation of capital income and wealth.

I analyze optimal design of tax schedules by developing a model where households are subject to idiosyncratic capital income risk and private information. The productivity of investment projects stochastically evolves over time. In particular, investment productivity has two components, a component that is *known* by the households in advance and at the time of investment, and a *residual* component that is realized once investment is made.² The first component of productivity can be interpreted as *entrepreneurial ability*. I assume that productivity, investment and consumption are all private information to the entrepreneur. In such an environment, a planner would want to redistribute resources across households against productivity and income risk. These redistributive motives together with private information, leads to a trade-off between incentives to invest and redistribution as in [Mirrlees \(1971\)](#). So, the approach here can be thought of as a Mirrleesian approach to capital accumulation.

In this environment, I first analyze how taxes should be designed to achieve efficiency and redistribute across households. In particular, I study optimal taxes on different types of capital income. Second, I show how the developed model can generate long run distribution of wealth with Pareto tail. Using techniques from probability theory, I provide a method for calculation of the tail of the long run wealth distribution in the constrained efficient allocations. This would shed light on the question of what the wealth distribution should look like and how much wealth inequality should governments allow.

In order to discuss the implications of the model on optimal taxes, one should distin-

¹[Aiyagari \(1994\)](#)'s seminal paper is an example with idiosyncratic labor income risk that fails to capture the concentration of wealth among the wealthy. For successful models with capital income risk, see [Quadriani \(2000\)](#), [Cagetti and De Nardi \(2006\)](#), and [Benhabib et al. \(2011\)](#).

²This environment nests the models of entrepreneurship in [Evans and Jovanovic \(1989\)](#) and [Gentry and Hubbard \(2004\)](#).

guish between three types of capital income: 1. Outside saving: income from saving in a risk free asset outside of business, 2. Investment or inside saving: income from investing in the risky project, 3. Bequest: saving for descendants.³

First, as for taxes on outside saving, I show that risk and return trade-off leads to progressivity. Due to constant return to scale, the planner would like to have the most productive units produce. However, this comes at a cost above and beyond the resource cost of investment. In fact, due to unobservability of investment, full insurance is not achievable and the higher the level of investment, the higher the riskiness of consumption. Hence, as the desired investment level increases it becomes increasingly costly to provide incentive to invest. This convexity of cost of moral hazard implies that projects with higher ex-ante return should have higher investment. In other words, even though the model exhibits constant returns to scale at the individual level, the presence of frictions make it similar to a decreasing return to scale model.

Consequently, due to incentive reasons, more productive households are bearing higher degree of risk. Hence, absent taxes on outside saving, the households would like to self insure by investing in a risk-free bond and this demand is higher for more productive individuals. Therefore, the government should discourage households from investing outside the business and create higher distortions for more productive households. Hence, taxes on outside saving should be progressive.

Second, I characterize the determinants of optimal taxes on investment or inside saving. I provide a formula that relates these taxes to the investment rate and its dependence on productivities. There are two main forces: 1. Since more productive households are subject to higher risk, they should face regressive or decreasing marginal rates on inside saving. 2. Since more productive households are investing more, their average income is higher and redistributive motives leads to progressive taxes on inside saving. In most numerical simulations, the forces toward progressivity dominate, i.e., marginal tax rates on inside saving are increasing with investment productivity. The analysis here shows that with capital income risk, progressivity of the tax code is the name of the game. A feature that is to my knowledge novel to this model.

Third, when ideas do not persist across generations, i.e., productivities are stochastic, bequests should be subsidized. This result is driven by the way saving affects incentive to invest for future generations. In this model, since consumption is unobservable, higher transfers lowers the private cost of investment for the household. Hence, an increase in saving, increases investment by future generations, i.e., saving relaxes future incentive

³This outside saving can be interpreted as a risk-free bond or it could be interpreted as a an index of stocks of all the firms in the economy.

constraints. Therefore, it should be subsidized. This is in contrast with labor income risk models where saving tightens future incentives to work. That is absent saving taxes, households would like to save and work less hours and therefore saving should be taxed.

Finally, I provide a method to determine the properties of the tail of long-run distribution of wealth. I show that a simple modification of the model can deliver a stationary distribution for wealth and that this distribution is fat tailed. The main idea behind this result is the fact that consumption growth across generations is stochastic. This property together with appropriate borrowing limits for agents imply that a stationary distribution exists. I show that the tail of the wealth distribution can be calculated using a rather simple formula in the constrained efficient allocations. The methodology here, therefore, allows me to study the determinants of the efficient distribution of wealth using simple formulas.

This paper is a first attempt in the analysis of optimal capital taxation in a model that can capture reasonable properties of the wealth distribution. As I argue the model is consistent with a concentration of wealth at the top. In addition, the data on business ownership and entrepreneurship suggests that capital income risk is an important determinant of wealth inequality at the top. In particular, as noted by [Quadrini \(2000\)](#) and [Gentry and Hubbard \(2004\)](#), there is a high concentration of business owners at top of the wealth distribution. They establish that of the top %5 of the wealthiest Americans, around %70 are business owners. Furthermore, most of the wealthy households' assets is held in their business, approximately %41 and therefore subject to significant risk.

Despite this evidence, the literature on capital taxation has widely focused on labor income heterogeneity as the main driving factor of inequality.⁴ In this paper, I have investigated the other extreme where capital income together with frictions in financial markets is the only determinant of inequality. A developed economy like that of the United States, should be thought of as comprising of both households whose major risk in life is labor income risk and ones with significant capital income risk. The analysis in this paper shows that unlike models of optimal taxation so far, capital income risk cannot be ignored. It is, thereby, a first step toward better understanding of optimal taxes at the top of the wealth distribution.

Related Literature. This paper builds on the literature on optimal dynamic taxation

⁴An exception is [Piketty and Saez \(2012\)](#) who develop a model with accidental bequests and study its effect on capital taxation. While they argue that accidental bequests are consistent with micro data, they do not investigate the implications of their model about the distribution of wealth in the economy. One key difference between the two models is that the model developed in this paper is consistent with the observation that many people at the top of the wealth distribution hold a high fraction of their wealth in risky capital. Another exception is [Albanesi \(2011\)](#). However, she does not explore the implications of her model on wealth distribution.

(see [Golosov et al. \(2003\)](#), [Farhi and Werning \(2010a\)](#), [Golosov et al. \(2010\)](#) among others). This literature has mainly focused on environments with idiosyncratic labor income risk and their implications about dynamic taxation of various sources of income. In this paper, I study optimal taxation of various sources of income in a model with capital income risk and show that capital income risk overturns some of the main lessons from the literature, namely that the intertemporal wedge can be negative as well as progressivity.

My paper is also related to a growing literature on the effect of taxation on entrepreneurial behavior. [Cagetti and De Nardi \(2009\)](#) consider the effect of elimination of estate taxes on wealth accumulation. [Kitao \(2008\)](#) and [Panousi \(2009\)](#) study how changes in the capital income tax rate affects investment by entrepreneurs. However, none of these studies considers the optimal taxation of entrepreneurial income. In developing my model of entrepreneurs, I have relied on their benchmark models while abstracting from some details for higher tractability.

[Albanesi \(2011\)](#) and [Scheuer \(2010\)](#) are early attempts in studying optimal design of tax system for entrepreneurs and hence in spirit closer to this paper. [Scheuer \(2010\)](#) focuses on the decision of entry into entrepreneurship and its implication for differential treatment of entrepreneurs and workers. [Albanesi \(2011\)](#) is perhaps the closest study to this. She considers a two period model where the entrepreneurs are ex-ante identical and invest in risky projects. She shows that it is possible to have negative wedge on observable risky capital since it relaxes incentive constraint – while the wedge on risk-free asset is positive in contrast to the negative bequest tax in this paper. Furthermore, she considers different financing patterns by the firm and optimal taxes on various securities.

An important implication of my paper is the emergence of bequest subsidies when entrepreneurs are subject to capital income risk. This result is related to a large literature on optimal capital taxation including [Chamley \(1986\)](#), [Judd \(1985\)](#), [Kocherlakota \(2005\)](#), and [Conesa et al. \(2009\)](#), among others. In most of these studies the optimal tax rate on capital income/wealth is positive or zero.⁵ Exceptions are [Farhi and Werning \(2008\)](#) and [Farhi and Werning \(2010b\)](#) in which negative marginal tax rates emerge either as a result of a higher social discount factor or binding enforcement constraints in the future. In my model, however, bequest subsidies are optimal since they relax future incentive constraints.

This paper is also related to a body of research that studies power laws in economics,

⁵[Kocherlakota \(2005\)](#) actually shows that wealth taxes are zero in expectation and hence some time negative and some time positive. However, that result is specific to a particular implementation and there are other implementations for which capital income tax rate is equal to the investment wedge and hence positive; see [Werning \(2010\)](#).

Champernowne (1953), Simon (1955), Gabaix (1999) and Benhabib et al. (2011).⁶ It is specially close in its spirit to Benhabib et al. (2011). They show that in an intergenerational model with labor and capital income risk and warm-glow bequest motives, the long-run distribution of wealth is Pareto at the top with the tail being determined by the capital income risk only. In my paper, I achieve this via borrowing constraint and a more powerful technique introduced by Mirek (2011). Moreover, to my knowledge, this is the first study to characterize properties of the efficient distribution of wealth. Further, Benhabib et al. (2011)'s analysis relies on a particular market structure. The analysis in this paper is more powerful in that I can characterize the efficient distribution of wealth.⁷

Finally, from a technical point of view, parts of the model in this paper contains two main frictions, a hidden action problem and hidden type problem. In general, this makes the problem very hard to analyze. However, I use the first order approach, as in Pavan et al. (2009), to simplify the set of incentive constraints. Since there are two types of private information, this model shares the same structure as the model in Laffont and Tirole (1986) who study optimal regulation of a monopolist and more recently Garrett and Pavan (2010) and Fong (2009).

The paper is organized as follows: in section 2, I develop the main insights about progressive taxes via a two period model. Section 3, contains the dynamic extension of the two period model to an overlapping generations model and implications for bequest taxes. Section 4, discusses the properties of the long-run distribution of wealth. Section 5 concludes.

2 A Two Period Economy

In this section, I focus on a two-period economy in order to establish the main results of the paper regarding optimal taxation of two main sources of capital income – income from investments in controlled business and income from investments made outside of controlled business. Later, in section 3, I show that a fully dynamic extension of the two period model almost identically leads to the same qualitative results.

Consider a two period economy in which time is given by $t = 0, 1$. The economy is populated by a continuum of households.

Technology. Each household is the sole owner of an investment technology or project that is subject to idiosyncratic risk. In particular, entrepreneurs draw a productivity

⁶See Gabaix (2009) for an extensive review.

⁷See also Moll (2012) for an analysis of a model with capital income risk and stochastic death.

shock, $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$, at $t = 0$. I assume that θ is distributed according to the cumulative distribution function $F(\theta)$. I also assume that $F(\cdot)$ is differentiable over the interval $[\underline{\theta}, \bar{\theta}]$ and $f(\theta) = F'(\theta)$. The value of the shock, θ , determines the distribution of returns to individual investment. If a household with type θ invests k_1 in his private project, the project will yield an output of $y \in \mathbb{R}_+$ that is distributed according to the c.d.f. function $G(y|\theta k_1)$ (with p.d.f. given by $g(y|\theta k_1)$) where $G(\cdot|\cdot)$ is C^1 in all of its argument. Moreover, the mean value of y , given θ and k_1 is given by θk_1 , i.e., $\int_0^{\bar{y}} y g(y|\theta k_1) dy = \theta k_1$. That is, the production technology exhibits constant return to scale. Notice that this formulation can be a stand-in for a more general constant return to scale production function with various inputs (e.g., labor) in addition to capital. Furthermore, for each type θ and given an investment level k_1 , the entire distribution of outcomes $G(y|\theta k_1)$ will be realized at $t = 1$, i.e., a law of large number holds for all agents of type θ .⁸

To further simplify the analysis, we make the following assumption:

Assumption 1 *The distribution of output at $t = 1$ is Gamma, i.e., it satisfies⁹:*

$$y = \varepsilon(\theta k), \varepsilon \sim \Gamma(\eta, \eta^{-1}), \eta > 0$$

This implies that $E\varepsilon = 1$ and that its variance is given by η^{-1} . Moreover, the distribution of y satisfies

$$g(y|\theta k) = \frac{(\theta k)^{-1} \eta^\eta}{\Gamma(\eta)} (y(\theta k)^{-1})^{\eta-1} e^{-\eta y(\theta k)^{-1}}$$

I will refer to the p.d.f. of ε as $h(\varepsilon)$ while its c.d.f. is referred to as $H(\varepsilon)$. As will be shown later, the above assumption together with the log-utility assumption below, implies that optimal consumption in the second period should be linear in income, y . This will further simplify the analysis in characterizing saving taxes.

Note that the above assumption implies that $\frac{g_k(y|k, \theta)}{g(y|k, \theta)}$ or the likelihood ratio is increasing and linear (and hence it satisfies the conditions in see [Jewitt \(1988\)](#) and [Rogerson \(1985\)](#)).¹⁰,

$$\frac{g_k(y|k, \theta)}{g(y|k, \theta)} = \frac{1}{k} \eta \left[\frac{y}{\theta k} - 1 \right]$$

⁸One way to think about this is to assume that there is a double continuum of households, i.e., a continuum of types $\theta \in \Theta$ and there is a continuum of households for each type θ and the law of large number holds across all households of a particular type θ .

⁹Since gamma distributions are invariant to multiplications, the assumption that the mean value for ε is 1 is a normalization. Given any $\varepsilon \sim \Gamma(\eta, \chi)$, θ 's can be rescaled so that the mean values of ε 's is 1.

¹⁰Note that in general $\frac{g_k(\varepsilon \theta k|\theta, k)}{g(\varepsilon \theta k|\theta, k)} = -\frac{1}{k} (\varepsilon h'(\varepsilon) + h(\varepsilon))$.

Preferences. Households preferences are given by

$$\log(c_0) + \beta \log(c_1)$$

where c_0 and c_1 are consumption of the entrepreneur at each period. Households consume in each period and invest at $t = 0$. We assume for simplicity that each agent is endowed with e_0 units of consumption good at $t = 0$.

For this economy, an allocation is given by $\{c_0(\theta), c_1(\theta, \varepsilon), k_1(\theta)\}_{\theta \in \Theta, \varepsilon \in \mathbb{R}_+}$. An allocation is said to be *feasible* if it satisfies the following:

$$\int_{\underline{\theta}}^{\bar{\theta}} [c_0(\theta) + k_1(\theta)] dF(\theta) \leq e_0 \quad (1)$$

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_0^{\bar{y}} c_1(\theta, \varepsilon) dH(\varepsilon) dF(\theta) \leq \int_{\underline{\theta}}^{\bar{\theta}} \theta k_1(\theta) dF(\theta) \quad (2)$$

Throughout the paper, I will focus on a utilitarian planner whose objective is to maximize the sum of utilities of the households. In the context of the two-period model, this redistributive planner's objective is given by

$$\int_{\Theta} U(\theta) dF(\theta)$$

where $U(\theta)$ is the utility from allocations $\{c_0(\theta), c_1(\theta, \varepsilon)\}_{\varepsilon \in \mathbb{R}_+}$.

Friction-less Economy. Before characterizing properties of efficient allocations and their implications about optimal taxes, it is useful to start by characterizing the properties of optimal allocations in a frictionless economy, i.e., where there is no private information about investment, consumption and productivity. Not surprisingly, in absence of private information, it is efficient to allocate investment to the most productive households, i.e., households with $\theta = \bar{\theta}$. Note that in the second period, ε shocks cancel each other in the aggregate and total output in the second period is given by $\bar{\theta}k$ where k is the amount of capital invested at $t = 0$. Since the social planner has utilitarian objective, resources at each period are redistributed among everyone equally, i.e., perfect redistribution across θ 's at $t = 0$ and perfect insurance with respect to ε 's at $t = 1$. Due to log-preferences investment is a constant fraction of total resources in the first period and hence, the

following allocations are optimal without frictions:

$$\begin{aligned} k(\bar{\theta}) &= \frac{\beta}{1+\beta} e_0, k(\theta) = 0, \theta < \bar{\theta} \\ c_0(\theta) &= \frac{1}{1+\beta} e_0, \forall \theta \in \Theta \\ c_1(\theta, \varepsilon) &= \theta k(\bar{\theta}), \forall \theta \in \Theta, \varepsilon \in \mathbb{R}_+ \end{aligned}$$

It is needless to say that the above allocations are not incentive compatible: 1. Households with highest productivity, $\theta = \bar{\theta}$, have no incentives to invest – their consumption is not dependent on how much they invest, 2. Households with $\theta < \bar{\theta}$ would like to pretend to be of the highest productivity type, receive the transfer $k(\bar{\theta}) + c_0(\bar{\theta})$ and invest nothing.

2.1 Distortions and Policies

It is now useful to make precise the type of policies that are the focus of this paper. Although, the approach of this paper is to derive optimal policies from properties of optimal allocations¹¹, it is important to establish how allocations translate into tax functions. The market structure that is the basis of the analysis in this paper focuses on two types of assets or two types of saving: a risk-free bond and an equity in household's own project. I will refer to these as *Outside Saving* and *Inside Saving*; Risk free bond is what the households invest in other household's businesses (*outside* of their own) and equity is what households invest *inside* their business.

Given this asset structure, the budget constraints for each household is given by

$$\begin{aligned} c_0 + k_1 + b_1 &\leq e_0 \\ c_1 &\leq q^{-1} b_1 + \varepsilon \theta k_1 - T(q^{-1} b, \varepsilon \theta k_1) \end{aligned}$$

where T is the tax function that jointly depends on capital income outside and inside the business and q^{-1} is the risk free rate – q is the price of a risk-free bond. Note that the partial derivatives of T can be thought of as marginal tax rates. Given this, we define

¹¹This is in line with the New Dynamic Public Finance literature, see [Kocherlakota \(2010\)](#). [Diamond and Saez \(2011\)](#) refer to this as the *mechanism design approach to optimal taxation*.

the following marginal tax rates or wedges:

$$\tau_S(\theta) : u'(c_0(\theta)) = \beta q^{-1} (1 - \tau_S(\theta)) \int_{\mathbb{R}_+} u'(c_1(\theta, \varepsilon)) dH(\varepsilon) \quad (3)$$

$$\tau_K(\theta) : u'(c_0(\theta)) = \beta (1 - \tau_K(\theta)) \int_{\mathbb{R}_+} u'(c_1(\theta, \varepsilon)) \theta \varepsilon dH(\varepsilon) \quad (4)$$

I will refer to τ_S as outside saving wedge and τ_K as inside saving wedge. If one writes Euler equation for a consumer facing the above budget constraints, it can be seen that $\tau_S(\theta)$ and $\tau_K(\theta)$ are related to the partial derivatives of T . In fact, $\tau_S(\theta)$, is a weighted average of T_1 , weighted by marginal utilities. A similar property holds for $\tau_K(\theta)$. However, as I will show later, with log-preferences and $\varepsilon \sim \Gamma$, $c_1(\theta, \varepsilon)$ will be a linear function of ε and hence, it can be shown that $(1 - \tau_K(\theta)) \theta \varepsilon$ must coincide with the slope of $c_1(\theta, \varepsilon)$ with respect to ε . This implies that $\tau_K(\theta) = T_2(q^{-1}b(\theta), \varepsilon \theta k_1(\theta))$, i.e., the marginal tax rate on equity faced by individual θ .

Note that as the above analysis shows, the wedges defined above are not exactly the marginal tax rates faced by the individuals. One can think of them as distortions faced by the individuals. Nevertheless, throughout the next sections, I refer to an increasing pattern in $\tau_S(\theta)(\tau_K(\theta))$ as progressive taxes on outside (inside) saving. Hence, throughout the paper, progressive taxes are equivalent to *'more productive households face higher distortions'*.

Remark 1: In this implementation, I have focused on simple securities, i.e., debt and equity. One can think of an arbitrary security with dividend process d and define marginal tax rate for this security as above. Intuitively, securities are taxed based on their correlations with the idiosyncratic risk faced by the individuals. In particular, a security with payoffs independent of ε , are qualitatively taxed similar to a risk-free bond.

Remark 2: An underlying assumption in this implementation is that markets for insurance against shocks to θ 's are absent. Hence the tax code is partially insuring against the θ shocks as well as ε shocks. Furthermore, the existence of risk-free borrowing and lending and that different households can take different position implies that there is partial insurance coming from that market. Finally, note that in this model the need for taxation is due to redistribution.

2.2 Progressive (Outside) Saving Taxes

This section contains one of the main insights of the paper: that optimal taxes on outside saving are progressive. To show this, I first establish that there is a trade-off between return and risk: household with higher rate of return have riskier consumption sched-

ules. I then show that this property of the optimal allocations translates into progressive taxes on outside saving. The functional form assumptions, log-preferences and gamma distribution for shocks, simplify this analysis greatly. In the online appendix, I show that many of the results presented here hold for more general utility functions and distributions. Note that, these theoretical results are shown when θ is publicly observed. In section 2.4, I allow for private ex-ante types and discuss the implications of the model.

Information. Throughout the paper, I assume that consumption and investment decision of the individual as well as his ex-post productivity shock, ε , is privately by the individual only. An outside can observe income $y = \varepsilon\theta k_1$. Throughout the paper, I discuss cases where θ is known publicly as well as privately. I also assume that, funds allocated to investment and consumption are interchangeable at rate ρ . That is the funds that are allocated to investment by the mechanism designer, can be diverted to individual consumption at rate ρ . Given this information structure and under the assumption that θ is publicly observed, an allocation is said to be *incentive compatible* if¹²

$$\begin{aligned} \log(c_0(\theta)) + \beta \int \log(c_1(\theta, \varepsilon)) h(\varepsilon) d\varepsilon \geq \\ \max_{\hat{k}} \log\left(c_0(\theta) + \rho(k_1(\theta) - \hat{k})\right) + \beta \int \log\left(c_1\left(\theta, \frac{\varepsilon \hat{k}}{k_1(\theta)}\right)\right) h(\varepsilon) d\varepsilon \end{aligned} \quad (5)$$

The above constraint is worth interpreting. Note that in this model, the planner makes a transfer, $c_0(\theta) + k_1(\theta)$, in the first period together with a promise of delivering certain schedule of consumption as a function of ε . The household receives the transfers in the first period and choose an investment level \hat{k} resulting into consumption of $c_0(\theta) + \rho(k_1(\theta) - \hat{k})$. The planner on the presumption that the agent has chosen the suggested investment level and based on observing $\varepsilon \hat{k}$, imputes a value of $\varepsilon, \frac{\varepsilon \hat{k}}{k_1(\theta)}$, and provides the associated transfer. This determines the deviation value for the household.

It should be noted that, this specification of the friction, i.e., moral hazard, satisfies the conditions for the first order approach described in [Jewitt \(1988\)](#) and [Rogerson \(1985\)](#). The thrust of the first order approach is that the above incentive compatibility constraint

¹²A more familiar version of incentive compatibility can be written when allocations in the second period depend on income, i.e., $\hat{c}_1(\theta, y)$, as opposed to ε . In this case, incentive compatibility is given by

$$\begin{aligned} \log(c_0(\theta)) + \beta \int \log(c_1(\theta, y)) dG(y|\theta, k_1(\theta)) \geq \\ \max_{\hat{k}} \log\left(c_0(\theta) + k_1(\theta) - \hat{k}\right) + \beta \int \log(c_1(\theta, y)) dG(y|\theta, \hat{k}) \end{aligned}$$

It can be easily shown that the two formulations are equivalent. For illustrative reasons, I use the formulation in the text.

can be viewed as a maximization problem and hence it can be replaced by its first order condition with respect to \hat{k} .¹³ Hence, we can replace the above maximization with its first order condition with respect to \hat{k} evaluated at $k_1(\theta)$:

$$\frac{\rho}{c_0(\theta)} = \beta \int \log(c_1(\theta, \varepsilon)) \frac{1}{k_1(\theta)} (-h(\varepsilon) - \varepsilon h'(\varepsilon)) d\varepsilon \quad (6)$$

Note that when $\rho = 0$, the optimal allocation in the frictionless economy satisfies the above. Hence, it is incentive compatible – $\rho = 0$ means that the households cannot divert funds from investment to consumption and hence there is no friction.

Given this simplification, the planning problem is given by

$$\max_{c_0(\theta), c_1(\theta, y), k_1(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} \left[\log(c_0(\theta)) + \beta \int_0^{\bar{y}} \log(c_1(\theta, \varepsilon)) h(\varepsilon) d\varepsilon \right] dF(\theta) \quad (7)$$

subject to (6), (1), and (2). In what follows, I characterize the solution of the above planning problem and its implications about taxes.

I start the characterization of optimal allocation by studying how provision of efficient incentives at the individual level are balanced with redistribution. The following lemma is the first step in this analysis:

Lemma 1 *The solution to (7) satisfies the following:*

1. Consumption at $t = 1$ is linear in ε , i.e., $c_1(\theta, \varepsilon) = \gamma(\theta) + \zeta(\theta)(\varepsilon - 1)$, for some $\phi(\theta), \zeta(\theta) > 0$,
2. Average consumption at $t = 1$ is equated across types, i.e., $\int_0^{\infty} c_1(\theta, \varepsilon) h(\varepsilon) d\varepsilon = \gamma(\theta) = \beta/\lambda_1, \forall \theta \in \Theta$ where λ_1 is the multiplier associated with (2).

The proof can be found in the appendix.

The first part of the above lemma, establishes how incentives are provided. In fact, in order for households to undertake any investment, consumption schedule should be increasing in ε . That this schedule is linear is a direct consequence of log-preferences and gamma distribution. In general, as noted by [Hölmstrom \(1979\)](#), the inverse of marginal utility is linear in likelihood ratio $\frac{gk}{g} = -1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)}$. When $\varepsilon \sim \Gamma$, the likelihood ratio is linear and inverse of marginal utility is the identity function and hence, the above result

¹³As shown by [Jewitt \(1988\)](#), the first order approach is sufficient in moral hazard problems when: 1) $\int_0^y G(\hat{y}|\theta k) d\hat{y}$ convex and decreasing in k , 2) $\int y dG(y|\theta k)$ is concave and increasing in k , 3) hazard ratio $\frac{gk}{g}$ is increasing and concave in y , 4) $u((u')^{-1}(1/z))$ is concave, all of which are satisfied in this environment.

follows. The second part of lemma 7, illustrates how redistribution is done with observable θ . In fact the planner cannot equate consumption across all states and households but average consumption across all types can be equated. Note that a higher value of $\zeta(\theta)$ is associated with more risk in consumption and less insurance.

Required Risk. The above characterization together with the incentive compatibility constraint (6) allows me to define the concept of *required risk*. This is the level of riskiness in consumption, ζ , that is required to sustain a certain level of $t = 0$ consumption, c_0 , and investment, k_1 . Note that the incentive constraint (6) can be written as

$$\rho \frac{k_1(\theta)}{c_0(\theta)} = \beta \int \log \left(\beta \lambda_1^{-1} + \zeta(\theta)(\varepsilon - 1) \right) (-h(\varepsilon) - \varepsilon h'(\varepsilon)) d\varepsilon$$

Given a value for λ_1 , we can write the above as

$$\zeta(\theta) = \Psi \left(\rho \frac{k_1(\theta)}{c_0(\theta)} \right) \quad (8)$$

where Ψ is an increasing function; a higher value of investment requires a higher degree of riskiness and a higher value of consumption implies a lower marginal cost of investment for household and hence a lower value of required risk.

The existence of required risk implies that there is an extra benefit to an infinitesimal increase in c_0 . This comes from the fact that an increase in c_0 decreases the required risk and hence increases expected utility in the second period. Equivalently, required risk induces an extra cost to an infinitesimal increase in k_1 . Figure 1, describes the determination of c_0 and k_1 . The curve A-B represents the relationship between c_0 and k_1 associated with optimal choice of c_0 — first order condition associated with c_0 . This relationship is independent of θ . The curve O-C represents the optimality condition for k_1 . Note that an increase in θ shifts this curve downward due to the increase in rate of return on investment. This figure establishes that both consumption and investment are increasing in θ . Intuitively, the planner would like to have investment to be done completely by the more productive type. However, this comes at a cost of riskiness in consumption in the second period. Equating the marginal cost of an increase in required risk to the rate of return on investment implies that investment should be higher for higher types.

The following proposition establishes that in addition to $c_0(\theta)$, $k_1(\theta)$, the required risk, $\zeta(\theta)$, is also increasing:

Theorem 1 Suppose that q is the intertemporal price of consumption. Furthermore, suppose that

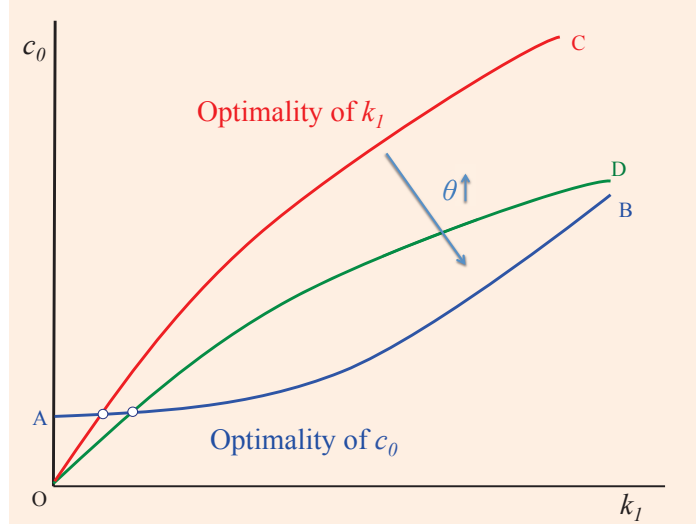


Figure 1: Determination of consumption and investment at $t = 1$

$q\bar{\theta} - 1 < \bar{x}$ where $\bar{x} = \rho \min_{0 \leq x \leq \frac{1}{\eta}} \left[\frac{1}{x} + \beta \eta e^{\frac{1}{x}-1} \left(\frac{1}{x} - \eta \right)^\eta \Gamma \left(-\eta, \frac{1}{x} - \eta \right) \right]^{-1}$, where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function. Then $c_0(\theta)$, $k_1(\theta)$, $\zeta(\theta)$, as well as investment rate, $\frac{k_1(\theta)}{c_0(\theta)}$, are increasing functions of θ . Furthermore, $k(\theta) = 0$ for all $\theta \in [\underline{\theta}, 1/q]$.

Proof can be found in the appendix.

The above proposition, establishes the risk-return trade-off: households with higher average productivity have to bear more risk in their second period consumption. This risk-return trade-off is endogenously determined by the extent of the moral hazard problem. In other words, although at the household level, the model satisfies constant returns to scale, due to the presence of moral hazard, it behaves similar to a model with decreasing returns to scale, e.g., Lucas (1978)'s span of control model. In fact, in this model, span of control by the household is *endogenously* derived from the underlying friction – that investment is unobservable by outsiders.

It is almost immediate from Theorem 1 that outside saving wedge should increase with θ . To see this, note that definition (3) for log-preferences can be written as

$$\frac{1}{1 - \tau_s(\theta)} = \beta q^{-1} c_0(\theta) \int_0^\infty \frac{h(\varepsilon)}{\lambda_1 + \zeta(\theta)(\varepsilon - 1)} d\varepsilon$$

As shown by above formula, $\tau_s(\theta)$ depends positively on first period consumption and expected marginal utility. From Theorem 1, $c_0(\theta)$, $\zeta(\theta)$ are increasing in θ . This implies that the expected marginal utility is increasing in θ – it is another measure of riskiness in

household's portfolio and hence it increases with risk, ζ . Hence, τ_S must be increasing in θ . We summarize this in the following Corollary. A formal proof is left for appendix.

Corollary 1 *Suppose that $\bar{\theta}$ satisfies the condition in Theorem 1. Then $\tau_S(\theta)$ as defined in (3) must be increasing in θ .*

The mechanism leading to progressive taxes on outside saving is worth further discussion. Intuitively, since more productive types are required to have a riskier consumption, absent progressive taxes on outside saving, they would like to decrease their investment and self-insure using outside risk free saving. Progressive taxes ensure that this does not happen and efficient level of investment is achieved.

Remark 3: The above intuition suggests that a version of this result for a more general class of utility functions and distribution of shocks must hold. In the online appendix, I show that in fact this is true. Hence it is natural to say that risk-return trade-offs in this model leads to progressive taxes on saving.

Remark 4: So far, I have assumed that θ is publicly observable. While I discuss the case with privately known θ in section 2.4, it is perhaps worth discussing the incentives for truth-full reporting given the above efficient allocation. Note that in this model, average continuation utility is decreasing in θ – due to higher risk in consumption. However, first period consumption as well as investment are increasing in θ . Hence, when θ 's are private and households face the above mechanism, it is not clear whether households would like to lie upward or downward. However, we can characterize the incentives for optimal reporting strategy. When an agent of type θ pretends to be $\hat{\theta}$ the utility attained by that agent is given by

$$U(\hat{\theta}, \theta) = \max \log \left(c_0(\hat{\theta}) + \rho \left(k_1(\hat{\theta}) - \hat{k} \right) \right) + \beta \int \log \left(c_1 \left(\hat{\theta}, \varepsilon \frac{\theta \hat{k}}{\hat{\theta} k_1(\hat{\theta})} \right) \right) dH(\varepsilon)$$

Using the envelope theorem and after some manipulations, we can show that

$$\frac{\partial}{\partial \hat{\theta}} U(\theta, \theta) = \frac{d}{d\theta} \left(\frac{k_1(\theta)}{c_0(\theta)} \right) + \frac{\partial}{\partial \hat{\theta}} \log \left(c_0(\hat{\theta}) + k_1(\hat{\theta}) \left(1 - \frac{\hat{\theta}}{\theta} \right) \right) \Big|_{\hat{\theta}=\theta}$$

The above derivative represents the household's local incentive for lying. The first term above, the investment rate, is increasing in θ , following proposition 1. Hence, the sign of the above is positive, if $c_0(\hat{\theta}) + \rho k_1(\hat{\theta}) \left(1 - \frac{\hat{\theta}}{\theta} \right)$ is increasing, i.e., when lying upward, utility in the first period increases. It turns out that in most simulations, this term is positive. Hence, facing the above mechanism, households would like to pretend to be

of higher productivity. Through this strategy, they would receive a higher transfer for consumption and investment at $t = 0$. Given that investment is privately known, they can reduce investment in order to reduce the riskiness of consumption in the second period.

2.3 Taxes on Inside Saving

In this section, I discuss the implications of the model on optimal taxation of inside capital or equity. I provide a formula that points to major determinants of taxes on inside capital and discuss the role of its ingredients.

To do so, first note that in the analysis above, since $c_1(\theta, \varepsilon)$ is linear with slope $\zeta(\theta)$:

$$\tau_K(\theta) = 1 - \frac{\zeta(\theta)}{\theta k_1(\theta)}$$

As the above definition shows, how the inside wedge changes with θ depends on how $\zeta(\theta)$ and $k_1(\theta)$ change as a function of θ . After further simplification, the following holds:

Lemma 2 *The solution to the problem (7) satisfies the following:*

$$\tau_K(\theta) = 1 - \frac{1}{\eta} \frac{\beta}{q} \frac{q\theta - 1}{\theta} \frac{c_0(\theta)}{\rho k_1(\theta)} \quad (9)$$

Proof can found in the appendix.

As it can be seen, similar to the saving taxes, the major determinant of the shape of $\tau_K(\theta)$ is the investment rate $\frac{k_1(\theta)}{c_0(\theta)}$ and its dependence on θ . In particular, when $\frac{k_1(\theta)}{c_0(\theta)}$ is increasing in θ , the above formula points toward two effects. The term $\frac{q\theta-1}{\theta}$ is increasing which cause a decrease in $\tau_K(\theta)$ while the rest of the terms are decreasing, which causes an increase in $\tau_K(\theta)$. Hence, the exact sign of $\frac{d}{d\theta}\tau_K(\theta)$ depends on which one of these effects dominates. Intuitively, higher investment rates by higher types implies that they should bear more risk and hence keep more of their income. At the same time, higher investment implies that the average income for them is higher and hence redistribution leads to higher taxes. For most simulations the latter dominates, hence taxes are progressive.

2.4 Unobservable Ex-ante Productivity

Here, I briefly discuss the implications of the model when θ is only privately observed by the households. In this case, in general it is hard to characterize properties of allocations. However, I provide some partial characterization of optimal allocations.

When θ is privately observed by the individuals, incentive compatible allocations are given by

$$\begin{aligned} \log(c_0(\theta)) + \beta \int_0^\infty \log(c_1(\theta, \varepsilon)) h(\varepsilon) d\varepsilon \\ \geq \max_{\hat{\theta}, \hat{k}} \log \left(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_0^\infty \log \left(c_1 \left(\hat{\theta}, \frac{\varepsilon \theta \hat{k}}{\hat{\theta} k_1(\hat{\theta})} \right) \right) h(\varepsilon) d\varepsilon \end{aligned} \quad (10)$$

The RHS of the above inequality is the utility that a type θ receives when he reports $\hat{\theta}$ and invests \hat{k} . The first order approach can be employed to simplify the planning problem.

When θ is private information, progressivity of saving taxes becomes harder to establish. However, it is possible to establish results that are similar to the previous results. In particular, the following lemma holds regarding the extent of redistribution and provision of incentives:

Lemma 3 *The solution to the planning problem with private θ satisfies the following:*

1. *Consumption at $t = 1$ is linear in ε , i.e., $c_1(\theta, \varepsilon) = \gamma(\theta) + \zeta(\theta)(\varepsilon - 1)$, for some $\gamma(\theta), \zeta(\theta) > 0$,*
2. *Average consumption at $t = 1$ is in general NOT equated across types, i.e., $\int c_1(\theta, \varepsilon) dH(\varepsilon) = \gamma(\theta)$ is not constant.*

Note that as before incentives for efficient investment are provided using linear consumption schedule. However, since θ is unobservable, the extent of redistribution is lower, i.e., average consumption is not equated across types. In fact, in most simulations, average consumption is decreasing in order to prevent the low types from pretending to be of high type. This property of efficient allocations under private information, makes it hard to show progressivity. However, it is possible to provide sufficient conditions so that progressivity holds. In fact, as long as investment ratio, $\frac{k_1(\theta)}{c_0(\theta)}$, is increasing in θ , it can be shown that saving taxes are increasing:

Proposition 1 *Suppose that in the solution to the planning problem with private θ , investment-consumption ratio, $\frac{k_1(\theta)}{c_0(\theta)}$, is increasing in θ . Then the implied outside saving wedge, $\tau_S^{PI}(\theta)$, is increasing in θ .*

Proof can be found in the appendix.

We can provide further characterization of the saving wedge when θ is private information. In fact, as it is common in models with hidden types, local incentive constraint are not binding for extreme values of θ . This implies that the distortion to saving is equal to saving distortions when θ is publicly observable. So we have the following corollary:

Corollary 2 *Consider the solution to the planning problem with private θ and its implied saving wedge, $\tau_S^{PI}(\theta)$. Then $\tau_S^{PI}(\theta) = \tau_S^{FI}(\theta)$ for $\theta = \underline{\theta}, \bar{\theta}$, where $\tau_S^{FI}(\theta)$ is the implied saving wedge with observable θ . Furthermore, $\tau_S^{PI}(\bar{\theta}) > \tau_S^{PI}(\underline{\theta})$.*

2.4.1 Numerical Example

In order to shed more light on mechanisms at play as well as the shape of tax schedule on inside saving, I provide a numerical example.¹⁴ Figure 2, shows the saving wedge as a function of θ as well as its relation to the saving wedge with full information about θ . Note that the saving wedge with private information is almost the same as the one with observable θ . Obviously this depends on the distribution of θ 's. However, it points to an important observation that in the model with observable θ 's, households incentive for lying is not that strong and hence the differences between the two models are small.

Finally, Figure 3 shows $\tau_K(\theta)$. This figure shows that the income tax schedule with respect to income from the business is progressive, i.e., increasing with respect to average productivity or average income. Note that this is true despite the fact that $\zeta(\theta)$ is increasing in θ . The figure also establishes that perhaps a progressive tax on total capital income – the sum of income from business and outside business, almost implements the optimal allocation since both tax $\tau_S(\theta)$ and $\tau_K(\theta)$ are increasing in θ and are close in numbers.

3 A Dynamic Extension

The two period model, although informative about optimal taxes on capital income, is short of a full analysis of optimal capital taxes. In particular, any relevant model of optimal capital taxes should be consistent with dynamics of wealth and its distribution in the economy. In this section, I develop a fully dynamic model of optimal capital

¹⁴I solve a cost minimization problem instead of utility maximization problem. Furthermore, we assume that θ is uniformly distributed so that the average return on investment varies from 3% to 50%. The other parameter values are $q = \beta = 0.95$, $\rho = 1$, and λ – the multiplier associated with promise keeping constraint, is set to 1. Note that the problem is homogenous in λ or the promise utility w . I will discuss this in more detail in section 3.

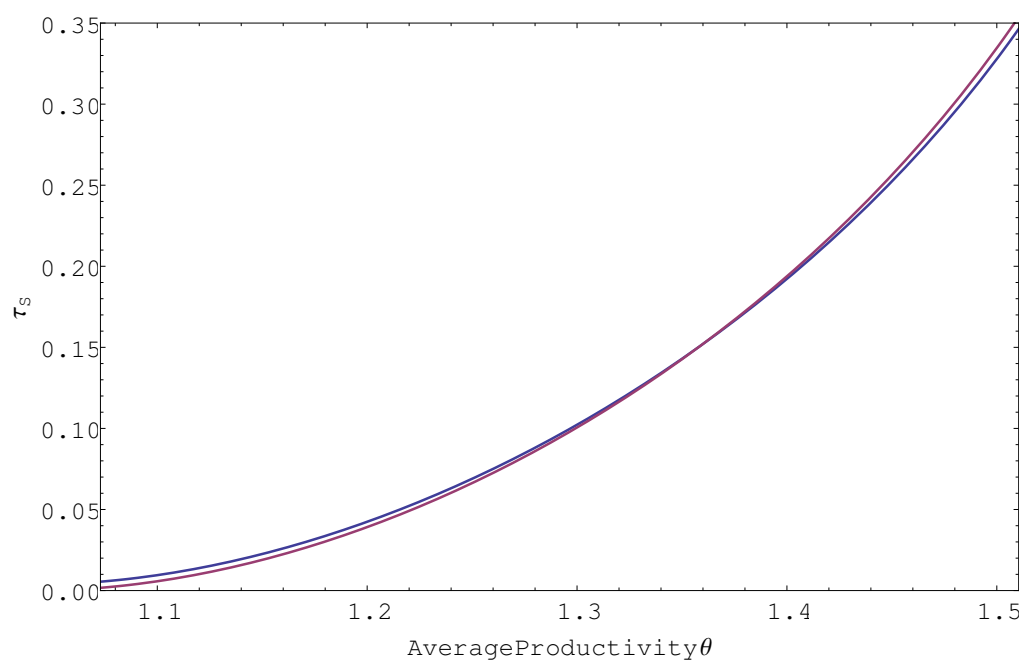


Figure 2: Saving Wedge with Public and Private Productivity

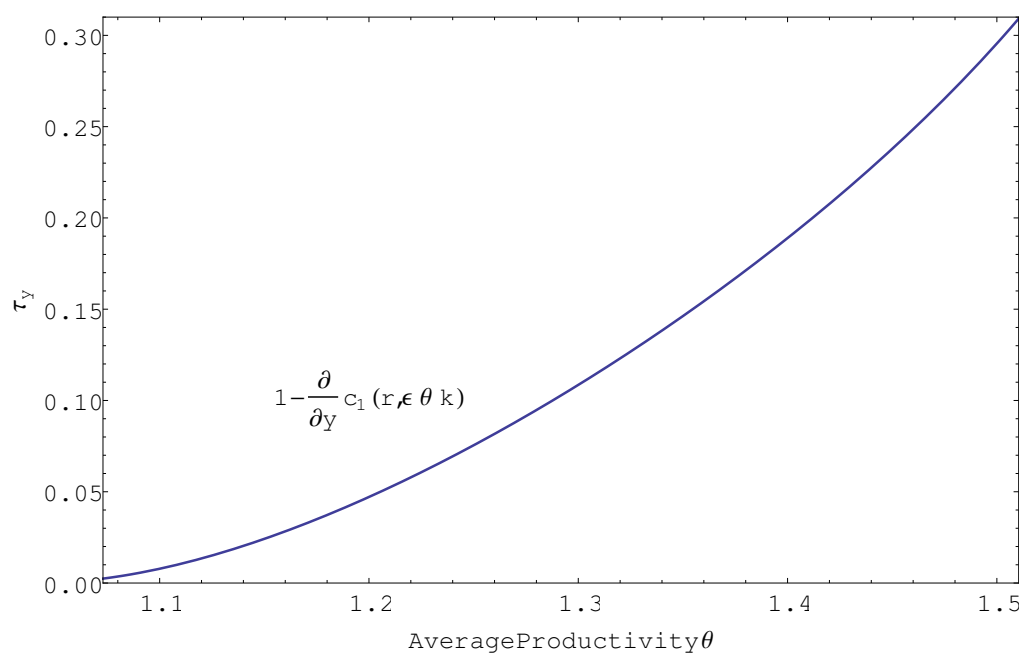


Figure 3: Marginal tax rate on business income, τ_y

taxes based on the two period economy described above. The model developed, satisfies diminishing returns in the aggregate, constant returns at the individual level as well as bequest motives via altruism. I will discuss the models implication on optimal estate taxes. Later in section 4, I discuss the implication of the model on optimal distribution of wealth.

The basic model is an overlapping generations extension of the two period model. Time is discrete, $t = 0, 1, \dots$. At each period t , a continuum of households are born, indexed by $i \in [0, 1]$. They live for two periods, t and $t + 1$. At t they are endowed with a unit of labor in each period, which they supply inelastically to perfectly competitive markets.

When young at period t , households are endowed with a technology that transforms period t -consumption and labor input into period $(t + 1)$ -consumption. That is, each household at date t is endowed with a production function

$$y_{it+1} = (A_{it+1}k_{it+1})^\alpha l_{it+1}^{1-\alpha}$$

where k_{it+1} is the amount of capital invested at t , l_{it+1} is the amount of labor hired at $t + 1$ and A_{it+1} is productivity. Similar to the two period model, $A_{it+1} = \theta_{it}\varepsilon_{it+1}$ where θ_{it} is known at t , when household i is deciding about how much to invest and ε_{it+1} will be realized at $t + 1$. The ex-ante productivity is drawn from a distribution $F(\theta)(\theta \in \Theta = [\underline{\theta}, \bar{\theta}])$ and ε_{it+1} is distributed according to a cumulative distribution function, $H(\varepsilon)$. Throughout the paper, I refer to the distribution of y_{t+1} induced by θ_t , k_{t+1} and l_{t+1} as $\hat{G}(y|\theta_t, k_{t+1}, l_{t+1})$. Both θ_t and ε_{t+1} are i.i.d. In order to produce, households hire labor from competitive labor market. At t , households decide how much to consume versus to invest in the technology.¹⁵

When old at period $t + 1$, households collect revenue from production less of wages paid to labor, consume and leave bequests. Hence a household born at t , consumes $c_{0,t}$ at t and $c_{1,t}$ at $t + 1$. It determines how much capital to purchase with wealth inherited from their parents, k_{t+1} , at t and determines how much labor to employ. A household born at t has the following utility function:

$$V_t = E_t [u(c_{0,t}) + \beta u(c_{1,t}) + \hat{\beta} V_{t+2}]$$

¹⁵Note that there are two ways to interpret this model. One is that capital fully depreciates across periods and y_{it} is the output produced at t . Another interpretation is that y_{it} is the value of output produced at t plus the depreciated value of capital. The basic assumption, under this interpretation, is that an outsider cannot distinguish changes in the value of physical capital from revenue generated from it.

where

$$u(c) = \begin{cases} \frac{c^{1-\sigma}}{1-\sigma} & \sigma \neq 1 \\ \log c & \sigma = 1 \end{cases}$$

and V_t is the utility of a household born at t .

Given this environment, an allocation is given by

$$\left\{ c_{0,t}(\theta^t, y^{t-1}), c_{1,t}(\theta^t, y^t), k_{t+1}(\theta^t, y^{t-1}), l_{t+1}(\theta^t, y^t) \right\}_{t \in \mathbb{N} \cup \{0\}}$$

where

$$\begin{aligned} \theta^t &= (\theta_t, \theta_{t-2}, \theta_{t-4}, \dots, \theta_{t \bmod 2}) \\ y^{t-1} &= (y_{t-1}, y_{t-3}, \dots, y_{(t-1) \bmod 2}) \end{aligned}$$

are the histories of shocks for a household born at t . Moreover, $\mu_{0,t}(\theta^t, y^{t-1})$ is the distribution of histories for the young generation at t and $\mu_{1,t-1}(\theta^{t-1}, y^t)$ is the distribution of histories for the old generation at t . We call an allocation feasible if

$$\int_{\Theta^{\lceil (t+1)/2 \rceil} \times \mathbb{R}_+^{\lceil t/2 \rceil}} \left[c_{0,t}(\theta^t, y^{t-1}) + k_{t+1}(\theta^t, \varepsilon^{t-1}) \right] d\mu_{0,t}(\theta^t, y^{t-1}) \quad (11)$$

$$\begin{aligned} &+ \int_{\Theta^{\lceil t/2 \rceil} \times \mathbb{R}_+^{\lceil (t-1)/2 \rceil}} c_{1,t-1}(\theta^{t-1}, y^t) d\mu_{1,t-1}(\theta^{t-1}, y^t) \\ &\leq \int_{\Theta^{\lceil t/2 \rceil} \times \mathbb{R}_+^{\lceil (t-1)/2 \rceil}} y_t d\mu_{1,t-1}(\theta^{t-1}, y^t) \end{aligned}$$

$$\int_{\Theta^{\lceil t/2 \rceil} \times \mathbb{R}_+^{\lceil (t-1)/2 \rceil}} l_t(\theta^{t-1}, y^{t-1}) d\mu_{1,t-1}(\theta^{t-1}, y^t) = 1 \quad (12)$$

where $\mu_{0,t}$ and $\mu_{1,t-1}$ are the measure of histories induced by the history of investment and hiring k_t, l_t .

Information. Throughout this section, I assume that the agent privately observes ε_{t+1} as well as investment and consumption. The planner can observe the output from the project, ex-ante productivity¹⁶ θ_t as well as labor input and total transfers $\{\tau_{0,t}(\theta^t, y^{t-1}), \tau_{1,t}(\theta^t, y^{t+1})\}$ – used for consumption and investment, to the households. Hence given a transfer scheme $\tau_t(\theta^{t-2}, y^{t-1}) = \{\tau_{\tau,\tau}(\theta^\tau, y^{\tau-1}), \tau_{\tau,\tau+1}(\theta^\tau, y^{\tau+1})\}_{\tau \geq t}$,

¹⁶For a discussion of results with θ_t private, I refer the readers to the online appendix.

each agent solves the following:

$$V_t \left(\tau_t \left(\theta^{t-2}, y^{t-1} \right) \right) = \max_{c_{0,t}, c_{1,t}, k_{t+1}, l_{t+1}} \int_{\Theta \times \mathbb{R}_+} \left[u(c_{0,t}) + \beta u(c_{1,t}) + \hat{\beta} V_{t+2} \left(\tau_{t+2} \left(\theta^{t-1}, \theta_t, y^{t+1} \right) \right) \right] dF(\theta_t) d\hat{G}(y_{t+1} | \theta_t, k_{t+1}, l_{t+1})$$

subject to

$$\begin{aligned} c_{0,t}(\theta_t) + k_{t+1}(\theta_t) &= \tau_{0,t}(\theta_t) \\ c_{1,t}(\theta_t, y_{t+1}) &= \tau_{1,t}(\theta_t, y_{t+1}) \end{aligned}$$

where in the above budget constraints, I have suppressed the history of shocks prior to t .

An allocation $\{c_{0,t}, c_{1,t}, k_{t+1}, l_{t+1}\}$ is said to be incentive compatible, if it is a solution to the above problem for appropriate transfer values.

I assume that a planner evaluates allocation according to the following welfare function:

$$V_0 + \delta V_1 \tag{13}$$

that is the planner cares about the welfare of the first two generations for some value of $\delta > 0$. This welfare criterion, although time-inconsistent, implies that the planning problem can be simplified significantly, as we show in what follows.

We call an allocation incentive efficient if it maximizes the above objective, it is feasible and incentive compatible.

Labor Demand. Before starting to characterize incentive efficient allocations, I discuss labor demand by firms. Note that since labor input is hired at the spot and is observable by the planner, we can write the output by the production unit as

$$y_t = \hat{y}_t l_t^{1-\alpha}$$

and write all the allocations in terms of histories of \hat{y}^t as opposed to y^t —one can think of \hat{y}_t as labor productivity. Using this modification, the return to l_t from planner's point of view is $(1 - \alpha) \hat{y}_t l_t^{-\alpha}$ for each individual production unit with labor productivity \hat{y}_t . Hence it should be equated to its cost which is the same for all entrepreneurs – it is equal to the ratio of the Lagrange multiplier on [12](#) to [11](#). Hence,

$$(1 - \alpha) \hat{y}_t l_t^{-\alpha} = p_t \tag{14}$$

where p_t is the same for all households. Replacing the labor demand in the production function, we will have

$$\begin{aligned} y_t &= \left(\frac{1-\alpha}{p_t} \right)^{\frac{1-\alpha}{\alpha}} \hat{y}_t^{\frac{1}{\alpha}} = \left(\frac{1-\alpha}{p_t} \right)^{\frac{1-\alpha}{\alpha}} A_t k_t \\ &= \kappa_t \varepsilon_t \theta_t k_t \end{aligned}$$

and hence production is linear in k . We denote the distribution of y_{t+1} induced by k_{t+1} and θ_t , $G(y_{t+1}|k_{t+1}, \theta_t)$.

Recursive Formulation and Incentive Compatibility. In our setup, since shocks are i.i.d., promise utility is a sufficient statistic to keep track of history for each individual. In particular, for any agent born at t , ex-ante utility is given by

$$w_t(\theta^{t-2}, y^{t-1}) = \int_{\Theta \times \mathbb{R}_+} \left[u(c_{0,t}) + \beta u(c_{1,t}) + \hat{\beta} w_{t+2}(\theta^t, y^{t+1}) \right] dF(\theta_t) dG(y_t|\theta_t, k_{t+1})$$

Given this definition, incentive compatibility can be written as the following

$$\begin{aligned} U_t(\theta^t, y^{t-1}) &= u(c_{0,t}(\theta_t)) + \int_{\mathbb{R}_+} [\beta u(c_{1,t}(\theta_t, y_{t+1})) + \hat{\beta} w_{t+1}(\theta_t, y_{t+1})] dG(y_t|\theta_t, k_{t+1}) \geq \\ &\max_{\hat{k}} u(c_{0,t}(\hat{\theta}) + k_{t+1}(\hat{\theta}) - \hat{k}) + \int_{\mathbb{R}_+} [\beta u(c_{1,t}(\hat{\theta}, y_{t+1})) + \hat{\beta} w_{t+1}(\hat{\theta}, y_{t+1})] dG(y_t|\theta_t, \hat{k}) \end{aligned}$$

where history before t is suppressed for easier notation. As before, we focus on local incentive constraints:

$$u'(c_{0,t}) = \int_{\mathbb{R}_+} [\beta u(c_{1,t}(\theta_t, y_{t+1})) + \hat{\beta} w_{t+1}(\theta_t, y_{t+1})] g_k(y_t|\theta_t, k_{t+1}) dy_t$$

Given the above simplification of the incentive compatibility constraint, instead of focusing on a welfare maximization problem, we focus on minimizing the cost of providing certain level of utility to households. We also focus on the component planning problem of providing a certain level of utility to an agent with a certain history. This component planning problem can be written as

$$\begin{aligned} P_t(w) &= \max_{c_0, c_1, k, w', U} \int_{\Theta} [\alpha q_t \kappa_{t+1} \theta k(\theta) - c_0(\theta) - k(\theta) + p_t \\ &\quad + q_t \int_0^\infty [-c_1(\theta, y) + q_{t+1} P_{t+2}(w'(\theta, y))] dG(y|\theta, k(\theta)) dF(\theta) \end{aligned} \quad (15)$$

subject to

$$\int U(\theta) dF(\theta) = w \quad (16)$$

$$u(c_0(\theta)) + \int [\beta u(c_1(\theta, y)) + \hat{\beta} w'(\theta, y)] dG(y|\theta, k(\theta)) = U(\theta) \quad (17)$$

$$\int_0^\infty [\beta u(c_1(\theta, y)) + \hat{\beta} w'(\theta, y)] g_k(y|k(\theta), \theta) dy = u'(c_0(\theta)) \quad (18)$$

where q_t is the intertemporal price of consumption.

An allocation is incentive efficient if for a sequence $\{q_t, \kappa_t\}$ it is the solution to the above sequence of problems given a w_0 and w_1 as well as (14). Furthermore, q_t, κ_t as well as w_0 and w_1 are determined so that (11) and (12) are satisfied. Accordingly, *stationary allocation* is defined when q_t 's and κ_t 's are constant.

Characterization. In what follows, we characterize the solution to the component planning problem 15 and discuss the main properties of the distortions. The first point to be noted in characterizing the solution to 15 is that the problem is homogenous in w . This simplifies the analysis to a great extent and the following proposition holds:

Proposition 2 *Given $\{q_t, \kappa_t\}$, the value function and policy functions associated with (15) satisfy the following:*

1. When $\sigma = 1$,

$$\begin{aligned} P_t(w) &= -B_t e^{\frac{1-\hat{\beta}}{1+\hat{\beta}}w} + H_t \\ c_0(\theta, w) &= c_{0t}(\theta) e^{\frac{1-\hat{\beta}}{1+\hat{\beta}}w} \\ c_1(\theta, y, w) &= c_{1t}(\theta, y) e^{\frac{1-\hat{\beta}}{1+\hat{\beta}}w} \\ k(\theta, w) &= k_t(\theta) e^{\frac{1-\hat{\beta}}{1+\hat{\beta}}w} \\ w'(\theta, y, w) &= \hat{w}_t(\theta, y) + w \\ U(\theta, w) &= U_t(\theta) + w \end{aligned}$$

for some $B_t, c_{0t}, c_{1t}, k_t, \hat{w}_t$ and U_t . H_t is the present value of labor income:

$$H_t = p_t + \sum_{s=0}^{\infty} q_t \cdots q_{t+2s+1} p_{t+2s+2}$$

2. When $\sigma \neq 1$,

$$\begin{aligned}
P_t(w) &= -B_t((1-\sigma)w)^{\frac{1}{1-\sigma}} + H_t \\
c_0(\theta, w) &= c_{0t}(\theta)((1-\sigma)w)^{\frac{1}{1-\sigma}} \\
c_1(\theta, y, w) &= c_{1t}(\theta, y)((1-\sigma)w)^{\frac{1}{1-\sigma}} \\
k(\theta, w) &= k_t(\theta)((1-\sigma)w)^{\frac{1}{1-\sigma}} \\
w'(\theta, y, w) &= \hat{w}_t(\theta, y)(1-\sigma)w \\
U(\theta, w) &= U_t(\theta)(1-\sigma)w
\end{aligned}$$

for some $B_t, c_{0t}, c_{1t}, k_t, \hat{w}_t$ and U_t .

Proof can be found in the Appendix.

There are three main takeaways from the above proposition. First, that the present value $H_t - P_t(w)$ is a random variable with i.i.d. growth rate:

$$\frac{H_{t+2} - P_{t+2}(w_{t+2})}{H_t - P_t(w_t)} = \begin{cases} \frac{B_{t+2}}{B_t} e^{\frac{1-\hat{\beta}}{1+\hat{\beta}} \hat{w}_{t+2}(\theta_t, y_{t+1})} & \sigma = 1 \\ \frac{B_{t+2}}{B_t} ((1-\sigma) \hat{w}_{t+2}(\theta_t, y_{t+1}))^{\frac{1}{1-\sigma}} & \sigma \neq 1 \end{cases} \quad (19)$$

More specifically, for a stationary economy where B_t and $P_t(\cdot)$ are independent of time, the above growth rate is an i.i.d. process. This result is specially important since one way to think about $H_t - P_t(w_t)$ is the sum of financial wealth, $-P_t(w_t)$ and present value of future labor income. We use this result later to show that in a modified version of the above model, the long-run distribution of wealth has a Pareto tail.

Second, the above characterization suggests that in the stationary version of the problem, distortions or wedges are independent of history w . Hence, wedges or marginal tax rates inherit the correlation structure of the shocks, i.e., they are i.i.d.

Finally, when old, utility $u(c_{1,t})$ is always directly related to promised utility, w_{t+2} , in a manner that is independent of history. When $\sigma = 1$, $w_{t+2} = \frac{1+\beta}{1-\beta} \log c_{1,t} + \xi_t$ where ξ_t is some constant that depends on B_{t+2} and q_{t+1} . When $\sigma \neq 1$, $w_{t+2} = \xi_t \frac{c_{1,t}^{1-\sigma}}{1-\sigma}$ where ξ_t is a constant that depends on B_{t+2} and q_{t+1} . This implies that the component planning problem is identical to the two-period model where the discount factor is not β . Therefore, the results from the two-period economy would apply here and the taxes are qualitatively similar and we have the following proposition:

Proposition 3 Consider the solution to the component planning problem (15) and its implied wedges, $\tau_{K,t}$, and $\tau_{S,t}$. Then:

1. The wedges, $\tau_{K,t}$ and $\tau_{S,t}$ only depend on θ_t .
2. If $\bar{\theta} \leq \tilde{\theta}(\sigma)$, then $\tau_{S,t}$ is increasing in θ_t , i.e., outside saving taxes are progressive.

In what follows, for general exposition, I write $c_{0,t}(\theta, w) = c_{0,t}(\theta) \rho(w)$ where $\rho(w)$ is defined by

$$\rho(w) = \begin{cases} e^{\frac{1+\beta}{1-\beta}w} & \sigma = 1 \\ ((1-\sigma)w)^{\frac{1}{1-\sigma}} & \sigma \neq 1 \end{cases} \quad (20)$$

As it is clear from Proposition 2, a similar property holds for other allocations, i.e., c_1, k .

3.1 Optimal Estate Taxes

In this section, I discuss the implications of the model on taxation of bequests, the third type of savings discussed in the paper. I first start by proving a modified version of the Inverse Euler Equation for this environment in order to characterize how efficient incentive provision affects dynamics of consumption. I then show that optimal estate taxes are negative in contrast with taxes on outside saving that are positive.

I start the analysis by studying the intertemporal distortions between the young and the old. The following modified inverse Euler equations shed more light on these distortions:

Proposition 4 *In the solution to 15, optimal consumption must satisfy the following equation:*

$$\frac{\beta q_{t+1}}{\hat{\beta}} E_t \left[(\alpha q_{t+2} \kappa_{t+3} \theta_{t+2} - 1) k_{t+3} \frac{u''(c_{0,t+2})}{u'(c_{0,t+2})^2} \right] + \frac{\beta q_{t+1}}{\hat{\beta}} E_{t+1} \frac{1}{u'(c_{0,t+2})} = \frac{1}{u'(c_{1,t})} \quad (21)$$

$$- (\alpha q_t \kappa_{t+1} \theta_t - 1) k_{t+1} \frac{u''(c_{0,t})}{u'(c_{0,t})^2} + \frac{q_t}{\beta} E_t \frac{1}{u'(c_{1,t})} = \frac{1}{u'(c_{0,t})} \quad (22)$$

Proof can be found in the appendix.

The above equations show that saving distortions are different when young and old. In particular, when old, there are forces toward subsidization of saving while when young, saving should be taxed. The difference lies in the role of saving and its effect on the incentive constraint. In particular, for an old household, an extra unit of saving would relax their descendants incentive constraint. This is because when lying households are consuming more and hence have a lower marginal utility. A small increase in consumption in every state of the world, increases household's utility from telling the truth more than that of lying and hence relaxing future incentive constraint. Hence, bequests should be taxed negatively. When young, this effect is the opposite. An extra

unit of saving tightens the young household's incentive to invest in the project and hence saving should be taxed.

The above modified Inverse Euler Equations although informative, do not necessarily pin down the sign of the saving wedge for the old. The following theorem establishes the main result on bequest taxes:

Theorem 2 *Consider the solution to the component planning problem above. It must satisfy:*

$$\frac{\beta}{q_{t+1}} E_{t+1} u'(c_{0,t+2}) < u'(c_{1,t}).$$

That is bequests should be subsidized.

The proof is an illustration of the above intuition and is left for the appendix.

Note that this result is in contrast with the results from labor income risk models, e.g., [Goloso et al. \(2003\)](#) among others. In particular, in models with labor income risk and separable utility function, risk-free saving or bequests should always be taxed due to its perverse effect on labor supply. Absent bequest taxes, households would like to save and work less. More technically, the most attractive lying strategies are those in which consumption decreases upon lying. Due to concavity of the utility function, a unit of saving increases the value of lying by more than its impact on the value of telling the truth. Therefore, saving/bequest should be taxed. In this model, however, the opposite effect is satisfied. The most attractive lying strategy for the young is lying upward in which consumption increases. Hence a unit of bequest from the old relaxes future incentive for their descendants to invest and should be subsidized. In other words, saving by the old increases resources in the future and thereby increases investment by future youth and hence should be subsidized. On the other hand, in models with labor supply, saving decreases labor supply by future youth and hence should be taxed.

3.2 Steady State

Here we consider the effect of private information on steady state level of capital as well as properties of the stationary distribution of consumption, capital and promised utility. Note that in a steady state, q_t is constant as well as κ_t and B_t .

We define steady state as an allocation C_0, C_1, K, Y , and L such that

$$\begin{aligned} C_0 + C_1 + K &= Y \\ L &= 1 \end{aligned}$$

where C_0 is the aggregate consumption by the young, C_1 is the aggregate consumption by the old and K , Y , and L are capital stock, output, and aggregate hours respectively. These aggregates should be consistent with the solutions of the component planning problem above.

To characterize the steady state of this economy, we start by aggregation. Suppose that distribution of promised utility for the young in each period is given by $\Psi_{0,t}(w)$. Then aggregates in each period t are given by

$$\begin{aligned} C_{0,t} &= \int_{\Theta} c_0(\theta) dF(\theta) \int \rho(w) d\Psi_{0,t}(w) \\ C_{1,t} &= \int_{\Theta} c_1(\theta, y) dG(y|\theta, k(\theta)) dF(\theta) \int \rho(w) d\Psi_{0,t-1}(w) \\ K_{t+1} &= \int_{\Theta} k(\theta) dF(\theta) \int \rho(w) d\Psi_{0,t}(w) \\ Y_t &= \int_{\Theta} \kappa \theta k(\theta) dF(\theta) \int \rho(w) d\Psi_{0,t-1}(w) \end{aligned}$$

Furthermore, from labor demand, we must have

$$l(\theta, \varepsilon, w) = \kappa^{\frac{1}{1-\alpha}} \varepsilon \theta k(\theta, w)$$

and hence aggregate labor demand is given by

$$L_t = \kappa^{\frac{1}{1-\alpha}} \int \theta k(\theta) dF(\theta) \int \rho(w) d\Psi_{0,t-1}(w)$$

Then, feasibility implies that

$$\begin{aligned} C_{0,t} + C_{1,t} + K_{t+1} &= Y_t \\ L_t &= 1 \end{aligned}$$

Note that given the policy function for w , we must have

$$\int \rho(w) d\Psi_{0,t+2}(w) = \int \rho(\hat{w}(\theta, y)) dG(y|\theta, k(\theta)) dF(\theta) \int \rho(w) d\Psi_{0,t}(w)$$

The above analysis suggests that in order to have stationarity of the optimal allocation, it must be that

$$\int \rho(\hat{w}(\theta, y)) dG(y|\theta, k(\theta)) dF(\theta) = 1 \quad (23)$$

The above condition, together with the fact that $P'(w_t)$ has some martingale property, i.e., it is a supermartingale or submartingale, gives us some information about the sta-

tionary market-clearing value of q . In particular, we have the following:

$$E_t P' (w_{t+2}) = \frac{\hat{\beta}}{q^2} P' (w_t) \quad (24)$$

From Proposition 2, we have

$$\frac{P' (w_{t+2})}{P' (w_t)} = \begin{cases} ((1 - \sigma) \hat{w} (\theta_t, y_{t+1}))^{\frac{\sigma}{1-\sigma}} & \sigma \neq 1 \\ e^{\frac{1+\beta}{1-\beta} \hat{w} (\theta_t, y_{t+1})} & \sigma = 1 \end{cases}$$

So, (24) can be written as

$$\int_{\Theta \times \mathbb{R}} ((1 - \sigma) \hat{w} (\theta, y))^{\frac{\sigma}{1-\sigma}} dG (y|k (\theta), \theta) dF (\theta) = \frac{\hat{\beta}}{q^2}, \sigma \neq 1 \quad (25)$$

$$\int_{\Theta \times \mathbb{R}} e^{\frac{1+\beta}{1-\beta} \hat{w} (\theta, y)} dG (y|k (\theta), \theta) dF (\theta) = \frac{\hat{\beta}}{q^2}, \sigma = 1 \quad (26)$$

The above equations combined with (23) pin down the stationary interest rate, q . In particular, in case of log-preferences, since $\rho (w) = e^{\frac{1+\beta}{1-\beta} w}$, stationarity implies that $\hat{\beta} = q^2$. In other case, we can provide partial characterizations using Jensen's inequality. We have the following proposition:

Proposition 5 *In any steady state of the dynamic economy:*

- When $\sigma = 1$, $q = \sqrt{\hat{\beta}}$,
- When $\sigma < 1$, $q < \sqrt{\hat{\beta}}$,
- When $\sigma > 1$, $q > \sqrt{\hat{\beta}}$.

Proof can be found in the appendix.

The idea behind the above results is similar to Angeletos (2007). On the one hand, the existence of private information and exposure to risk means that households have high demand for a risk-free asset, i.e., a lower q . At the same time, a lower level of investment in risky assets, implies a lower level of capital stock and hence a higher rate of return. However, unlike Angeletos (2007), these two effects completely offset each other in the log-case. In addition, when $\sigma > 1$, the rate of return is lower than natural rate of discounting across periods and when $\sigma < 1$, the rate of return is higher.¹⁷

¹⁷A similar result holds in Farhi and Werning (2012).

3.2.1 Stationary Distribution

It can be show that, similar to most models of dynamic contracting with private information, the stationary distribution is trivial and almost all the agents will be at the lowest possible utility level, 0 in case of $\sigma < 1$ and $-\infty$ in case of $\sigma \leq 1$. To see this, note that when $\sigma \neq 1$, $w_{t+2} = w_t \cdot (1 - \sigma) \hat{w}(\theta_t, y_{t+1})$, i.e., w_t is a geometric random walk. Hence, either $E[\log((1 - \sigma) \hat{w}(\theta, y))] < 0$, in which case by law of large number w_t converges to its lower bound almost surely, or $E[\log((1 - \sigma) \hat{w}(\theta, y))] > 0$, in which case w_t converges to infinity almost surely.¹⁸ When $\sigma = 1$, w_{t+2} is a random walk with drift, $E[\hat{w}]$, and (26) implies that $E[\hat{w}] < 0$. Hence, w_t converges to $-\infty$ almost surely. Nevertheless, in all cases, variance of w_t grows without bound over time. We summarize this discussion in the following corollary:

Corollary 3 *In a stationary economy, promised utility w_t satisfies the following long-run properties:*

- When $\sigma \geq 1$, $w_t \rightarrow -\infty$, a.s.,
- When $1 > \sigma > 0$, $w_t \rightarrow 0$, a.s.,
- For all values of σ , $\text{Var}[w_t] \rightarrow \infty$.

The proof can be found in the appendix.

The above characterization, implies that although a stationary economy exists, promised utility, which can be thought of as a proxy for wealth, is not a stationary process. That is a version of immiseration holds in this economy – see [Thomas and Worrall \(1990\)](#), [Atkeson and Lucas \(1992\)](#), among many others. In the following section, we introduce a method to resolve this issue in order to achieve stationarity.

4 Long Run Distribution of Wealth

In this section, I discuss how the model developed above can be extended to generate a long-run distribution of wealth with a Pareto tail and what an optimal Pareto tail looks like. As Corollary 5 shows, a version of the infamous immiseration result holds in steady state of the model developed above. Note that this immiseration result can be thought of as ever increasing borrowing by each generation against their descendents labor and capital income. One natural assumption would be that a social planner can have the

¹⁸An application of Jensen's inequality and (25) implies that when $\sigma > \frac{1}{2}$, $E[\log((1 - \sigma) \hat{w})] < 0$.

ability to protect future generations in face of this ever-increasing borrowing. That is, it is possible to look at Pareto efficient allocations that do not only favor the initial two generations. This is the approach in this section.

Augmented Problem. I augment the contracting problem in (15) with a constraint of the following form:

$$w_t \geq \underline{w}$$

where $\underline{w} > \inf_{c>0} u(c)$ (which is either 0 or $-\infty$). One way to interpret the above constraint is that the social planner cares about future generations more than the initial ones, i.e., households born in periods 1 and 2. Hence, imposing the above constraint is equivalent to consider points on the Pareto frontier that put positive weight on the utility of households born in periods $t = 2, 3, \dots$ (Recall that the planning problem considered in section 3 only puts positive weight on the utility of households born in periods 1 and 2; see objective (13)). A similar approach for establishing stationarity in economies with private information is used in Atkeson and Lucas (1995) and Albanesi and Sleet (2006).¹⁹

Under the above augmentation, the component planning problem can be written as

$$P_c(w) = \max_{\Theta} \int_{\Theta} [\alpha q \kappa \theta k(\theta) - c_0(\theta) - k(\theta) + p + q \int_0^{\infty} [-c_1(\theta, y) + q P_c(w'(\theta, y))] dG(y|\theta, k(\theta))] dF(\theta) \quad (27)$$

subject to

$$\begin{aligned} \int U(\theta) dF(\theta) &= w \\ \log c_0(\theta) + \int [\beta \log(c_1(\theta, y)) + \hat{\beta} w'(\theta, y)] dG(y|\theta, k(\theta)) &= U(\theta) \\ \int_0^{\infty} [\beta \log(c_1(\theta, y)) + \hat{\beta} w'(\theta, y)] g_k(y|k(\theta), \theta) dy &= u'(c_0(\theta)) \\ w'(\theta, y) &\geq \underline{w}. \end{aligned}$$

Note that we have already assumed that the economy is stationary – both labor income, p , and risk free rate, q^{-1} , are constant.

Wealth and Taxes. The above planning problem, although informative about the behavior of allocations, does not necessarily have sharp predictions about wealth. Nevertheless, the object $P_c(w)$, being the difference between present value of income and

¹⁹This is in contrast with papers like Farhi and Werning (2007) or Sleet and Yeltekin (2006) where the constraints are of the form, $\int w_t d\mu_t \geq \underline{w}$. I suspect that the implications of the model are not that different under this assumption.

consumption for a generation with history represented by w , can tell us something about wealth. In fact, here, I construct a tax function that implies that $P(w)$ is total wealth for an individual. To do so, I consider a decentralization of the following form:

$$c_{0,t} + b_{t+1} + k_{t+1} = q^{-1}a_t + p \quad (28)$$

$$c_{1t} + a_{t+2} = q^{-1}b_{t+1} + y_{t+1} - T(y_{t+1}, b_{t+1}, a_t) \quad (29)$$

$$a_{t+2} \geq \underline{a} \quad (30)$$

The above equations represent the budget constraints for each household who receives $q^{-1}a_t$ in the form of bequest from her ascendants, has labor income p in the first period and invests in risk free bond as well as risky capital. In the second period, she collects income from these two sources and pays taxes that depend on all three types of capital income: bequests, inside investment, and outside saving. The household is then subject to a limit on how much assets she can leave for her descendants. In the following proposition, I establish that it is possible to choose a tax function so that household's optimal choice implements allocations associated with (27) and $a_t = -qP_c(w_t)$.

Proposition 6 *Consider the allocations created from the policy functions in (27), i.e., $\{c_{0,t}^*, c_{1,t+1}^*, k_{t+1}^*\}$ with the appropriate history dependence. Suppose that $c_0(\theta) + k_1(\theta)$ is monotone in θ . Then a tax function $T(y_{t+1}, b_{t+1}, a_t)$ together with \underline{a} exist so that a households that maximizes utility subject to the budget constraints (28)–(30), would choose $\{c_{0,t}^*, c_{1,t+1}^*, k_{t+1}^*\}$. Furthermore, in the solution to household's problem, $a_t = -qP_c(w_t)$.*

The proof closely follows the proof in Albanesi and Sleet (2006) and will be omitted.²⁰

Note that a similar result holds for allocations generated by the unconstrained planning problem (15). Hence, (19) implies that the sum of human capital as well as financial wealth is a process with random growth – although this implies that this process does not converge. As established by many studies (Kesten (1973) in mathematics as well as Simon (1955) and Gabaix (1999) among others in economics) such processes, when converge, have stationary distributions with fat tails. In what follows, I first establish that the solution to (27) has a stationary distribution. Furthermore, a property similar to (19) holds at the limit for the solution to (27). I will then use a powerful result shown by Mirek (2011) to show that the long-run distribution of wealth has a fat tail.

In order to use Mirek (2011)'s technique, I start by characterizing the slope of the policy function $w'_c(w, \theta, y)$ (the solution to (27)) with respect to w , when $w \rightarrow \sup_{c \geq 0} u(c)$.

²⁰Albanesi and Sleet (2006) construct the tax function using a backward induction argument. The same argument works here and is available upon request.

I do so using contraction mapping theorem. Consider the functional equation (15), its associated value function $P(w)$ and policy function $w'(w, \theta, y)$. I show that if T is the transformation associated with the functional equation (27), and $P_n = T^n P$ with associated policy function $w'_n(w, \theta, y)$, then

$$\lim_{w \rightarrow \infty} \frac{dP_n(w'_n(w, \theta, y))}{dP_n(w)} = \frac{dP(w'(w, \theta, y))}{dP(w)}$$

Note that by Proposition 2, $\frac{dP(w'(w, \theta, y))}{dP(w)}$ is independent of w .²¹ To complete the proof, I assume that the derivative of the policy functions $w'_n(w, \theta, y)$ converges to the derivative of $w'_c(w, \theta, y)$.

Assumption 2 *The derivative of the policy function at iteration n , $\frac{\partial}{\partial w} w'_n(w, \theta, y)$ converges to $\frac{\partial}{\partial w} w'_c(w, \theta, y)$ according to sup-norm.*

Unfortunately, this assumption cannot be shown from first principles. It can be shown, however, that it is equivalent to the convergence of the second derivative of the value function. Hence:

Proposition 7 *Suppose that Assumption 2 holds. Then*

$$\lim_{w \rightarrow \infty} \frac{dP_n(w'_n(w, \theta, y))}{dP_n(w)} = \frac{dP(w'(w, \theta, y))}{dP(w)}$$

Proof can be found in the appendix.

The above result states that inclusion of borrowing constraints does not change the slope of policy functions at infinity. The following theorem establishes the existence of the stationary distribution when interest rates are low, i.e., $q^{-2}\hat{\beta} < 1$:

Theorem 3 *Suppose that Assumption 2 holds. Then for any q such that $q^{-2}\hat{\beta} < 1$, the process w_t generated by the solution to (27) has a stationary distribution. Furthermore, for any stationary measure, μ^* , we must have $\mu^*(\{\underline{w}\}) < 1$.*

²¹When $\sigma = 1$,

$$\frac{dP(w'(w, \theta, y))}{dP(w)} = e^{\frac{1+\beta}{1-\beta}\hat{w}(\theta, y)}.$$

Moreover, when $\sigma \neq 1$,

$$\frac{dP(w'(w, \theta, y))}{dP(w)} = ((1 - \sigma)\hat{w}(\theta, y))^{\frac{1}{1-\sigma}}.$$

Both are independent of w .

Proof can be found in the appendix.

Intuitively, the idea for stationarity is the following: 1. The lower bound \underline{w} is not an absorbing state; for high enough values of y , $w'_c(\underline{w}, \theta, y) > \underline{w}$. 2. For w high enough, the behavior of $w'_c(w, \theta, y)$ is similar to that of $w'(w, \theta, y)$ for the unconstrained problem. Under the assumption that $q^{-2}\hat{\beta} < 1$ and that $\sigma \geq 1$, equations (26)–(25) imply that $w'(w, \theta, y)$ decreases on average. Therefore for low values of w_t , w_{t+2} tend to increase over time while for high values of w_t , w_{t+2} tend to decrease on average.

We also have the following Proposition:

Proposition 8 *Suppose that Assumption 2 holds. Then in any steady state of the above environment, the price of risk free bond, q_c , satisfies $q_c^{-2}\hat{\beta} < 1$. This in turn implies that when $\sigma \geq 1$,*

$$\int_{\Theta \times \mathbb{R}} \lim_{w \rightarrow \infty} \frac{dP_c(w'_c(w, \theta, y))}{dP_c(w)} dG(y|k(\theta), \theta) dF(\theta) < 1.$$

Proof is left for appendix.

The above Proposition states that if a steady state exists, it must be that interest rates are low – lower than the discount rate and hence by Theorem 3, w_t or wealth $-qP_c(w_t)$ is a stationary process with a non-trivial stationary distribution. Furthermore, it shows that $P(w_t)$ satisfies the basic requirements of the main result in Mirek (2011) – Theorem 1.8. Hence, we have the following theorem:

Theorem 4 *Suppose that $\sigma \geq 1$. If Ψ is the long-run distribution of wealth $a_t = -qP_c(w_t)$, it satisfies*

$$\lim_{t \rightarrow \infty} t^{-\nu} \Psi(a > t) = \psi$$

where ψ is a constant and ν satisfies

$$\int_{\Theta \times \mathbb{R}} \left(\frac{dP(w'(w, \theta, y))}{dP(w)} \right)^\nu dG(y|k(\theta), \theta) dF(\theta) = 1 \quad (\star)$$

Proof can be found in the appendix.

The above Theorem is a direct application of the result in Mirek (2011). It is a powerful result that characterizes the tail of the distribution. In particular, there is no need to compute the solution of the constrained problem (27). This is specially simple since the solution of (15) is essentially the same as the solution of the two period problem. In fact, since the margin between c_1 and w' is undistorted, and the formula for the value function is known, w' can be calculated as a function of c_1 . In general we can show that

$$P'(w'(\theta, y)) = \left(\phi(\theta) + \frac{\zeta(\theta)}{\theta k(\theta)} (y - \theta k(\theta)) \right) P'(w)$$

| q | ν^{IM} |
|-------------------------|----------------|
| $0.95^{30} - 0.96^{30}$ | Non-Stationary |
| 1.0309^{30} | 1 |
| 1.0319^{30} | 1.0423 |
| 1.0329^{30} | 1.0872 |

Table 1: The tail behavior of stationary distribution in the incomplete market model

which is a similar property to Lemma 3. In light of the above formula, we can write (★) as

$$\int_{\Theta \times \mathbb{R}} (\phi(\theta) + \zeta(\theta)(\varepsilon - 1))^{\frac{\nu}{\sigma}} dH(\varepsilon) dF(\theta) = 1 \quad (\star\star)$$

The above formula is informative in that it points toward forces determining the tail of the long-run distribution of wealth. As an example, an increase in ζ (less insurance provided) for a positive measure of θ 's causes ν to decrease and hence a heavier tail of stationary distribution. In the online appendix, I show that in an incomplete market version of the model, a similar result holds. Using these formulas, one can compare the tail of the distribution of wealth in the incomplete market model with its efficient counterpart. This is the task I set to do in the next section.

4.1 Numerical Example

Here I provide a numerical example, in order to illustrate the forces in determining the efficient distribution of wealth. For simplicity, I assume that $\alpha = 1$ – no labor input and I calculate the policy functions for various values of interest rate q . I assume that each generation is around three years implying a discount factor of $\beta = (0.95)^{30}$. Furthermore, I assume that $\Theta = [0, \bar{\theta}]$ where $\bar{\theta} = 6.075$ which is associated with an annual rate of return of approximately %5.5²². In our exercise, I vary the risk-free interest rate q^{-1} and study its effect on the wealth distribution. Note that changing q is equivalent to changing the lower bound on promised utility. Since the formulas are more insightful with variable q we will use this approach. Figure 4 illustrates the tail of the distribution of wealth for different values of q for the constrained efficient allocations. Table 1, contains related information for the incomplete market model.

Note that as it can be seen, with incomplete market, the tail of the wealth distribution is non-existent for q close to β – it can be shown that with $q < \beta$, the process is non-stationary. Moreover, for q high enough, the solution to the constrained efficient

²²The rest of the parameters are as follows, θ is Pareto distributed with parameter $a = 2$ – implied variance of .5. $\phi = 1$ – an implied variance of 1.

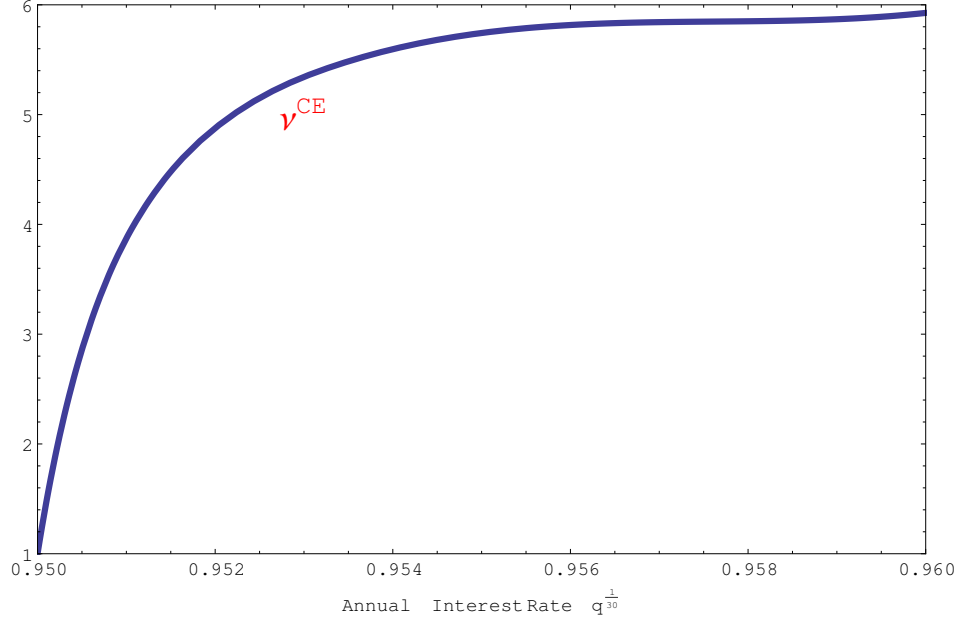


Figure 4: Tail of the stationary distribution of wealth in the constrained efficient allocation

allocation does not exist. This is due to the fact that the marginal cost of investment arising from moral hazard is bounded from above and with q high enough it is optimal to invest only in the most productive project.

As it can be seen, in the constrained efficient allocation, the tail of the distribution is more responsive to changes in the interest rate compared to the incomplete market model. This is perhaps because there is more variability in consumption with incomplete market. The following figures show average consumption as well as the slope of the consumption schedule with respect to ε . As for constrained efficient allocation, average consumption is roughly constant in θ while the slope is increasing in θ . On the other hand, consumption is much more volatile in the incomplete market model. Both average value of consumption as well as its slope with respect to ε . Note that in this model there is also a selection effect: as q varies, the cutoff for projects used is given by θ_l where $q\kappa\theta_l = 1$. With incomplete markets this effect is more pronounced. When q is high enough, there are many low productivity projects selected. The high variability of consumption implies a lower ν and hence a fatter tail for the distribution of wealth with incomplete market.

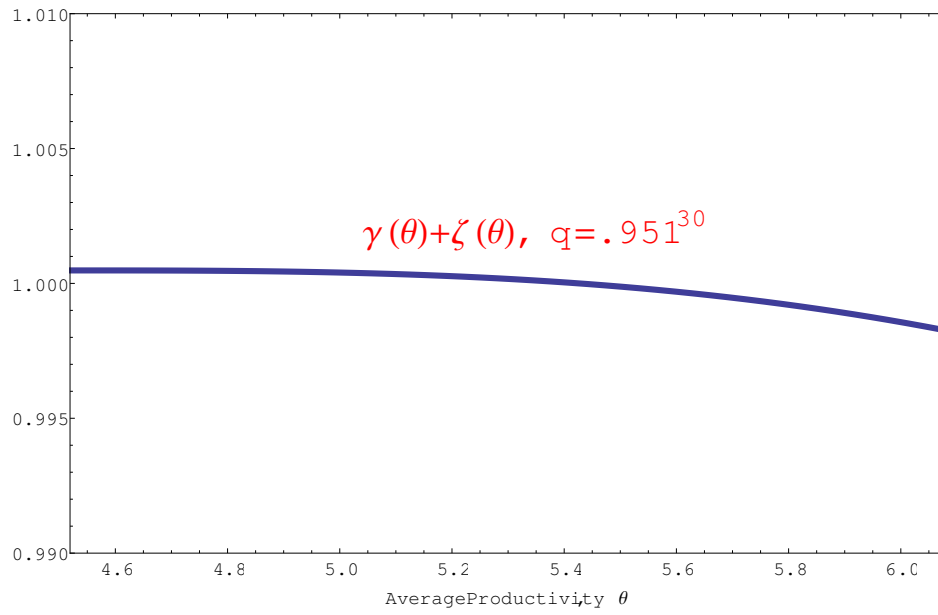


Figure 5: Average consumption with constrained efficiency

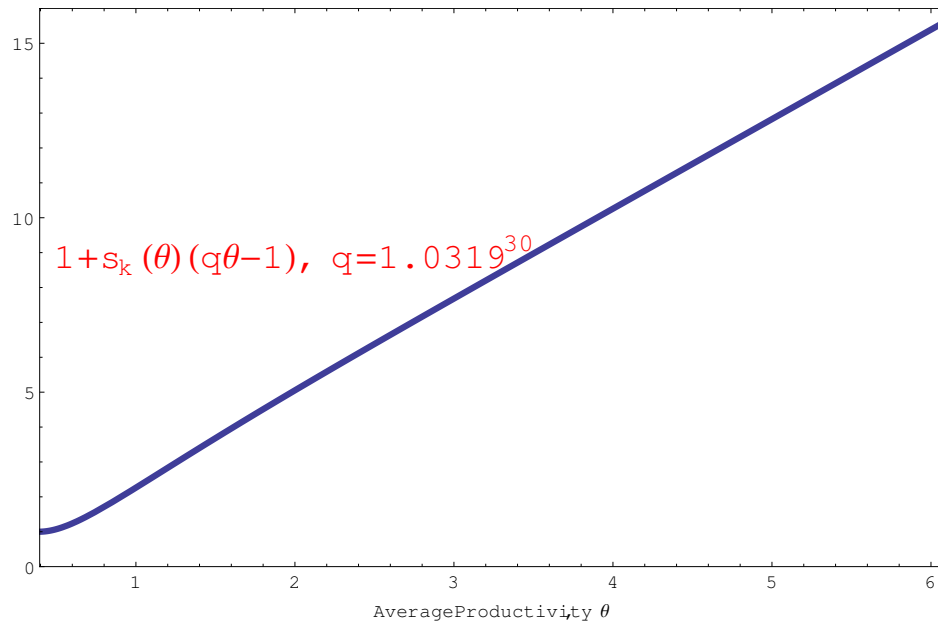


Figure 6: Average consumption with incomplete market

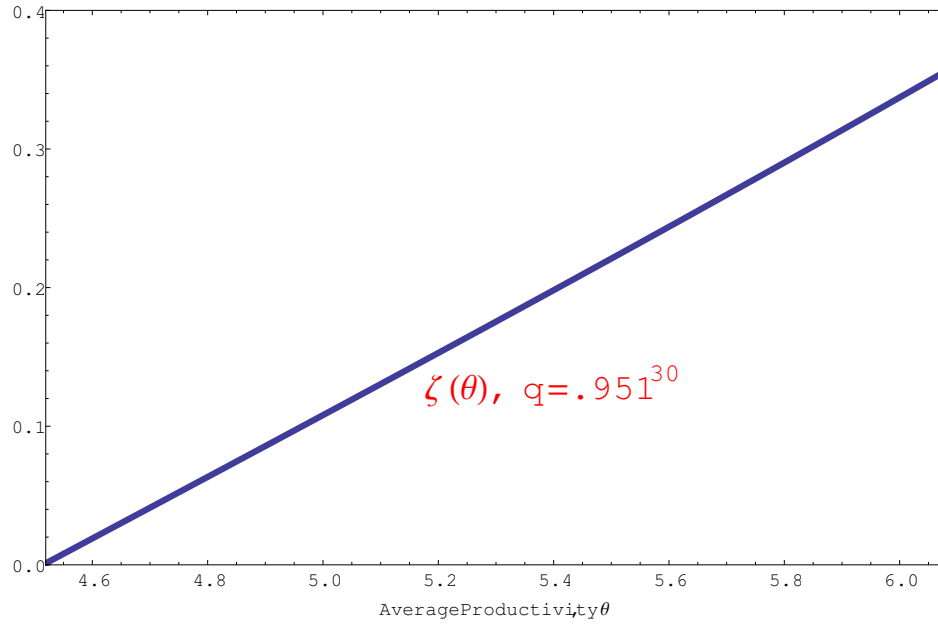


Figure 7: Slope of consumption with respect to ϵ , constrained efficient

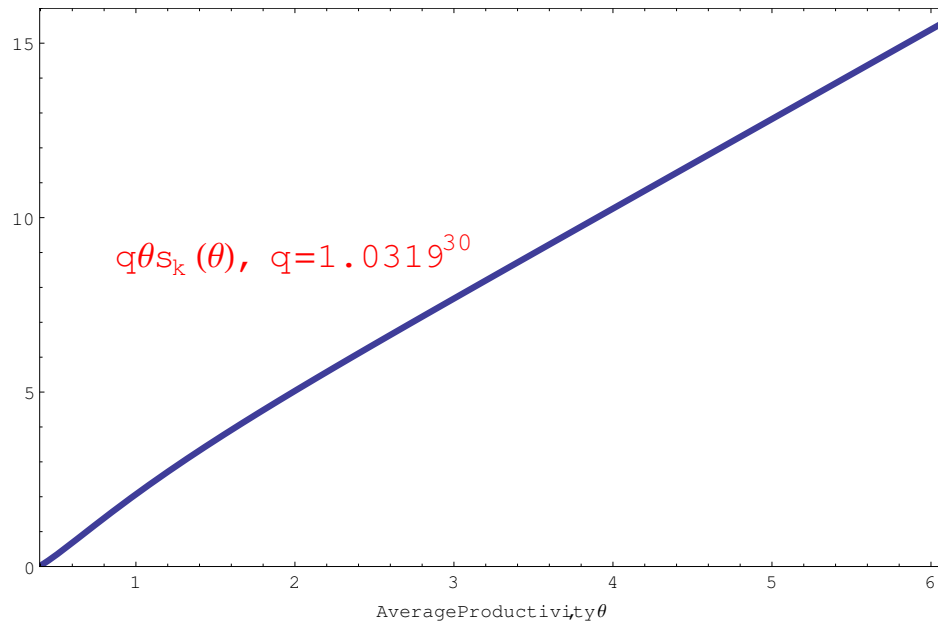


Figure 8: Slope of consumption with respect to ϵ , incomplete market

5 Conclusion

In this paper, I have studied optimal taxation of capital income in presence of capital income risk. I have shown that allowing households to invest in businesses, thereby being subject to idiosyncratic investment risk, changes the standard results on taxation of wealth and capital income. I have also shown how the model can be used to study efficient distribution of wealth.

This paper can be thought of as a first step toward designing capital income taxes for the wealthy. The US economy is perhaps comprised of households with significant capital income risk and households with labor income risk. When the tax authority has the ability to distinguish between the two types of households, designing optimal taxes are easy: it is optimal to use the results of this paper for households with major capital income risk and use the implications of models with labor income risk on the rest. However, it is perhaps unrealistic to assume such an ability for the tax authority. Absent this ability, the tax authority can use wealth as a proxy for the likelihood of being each type of households: wealthy households are perhaps more subject to capital income risk and should be taxed according to insights developed here. It would be an important extension of this paper, to study optimal taxes in such an environment.

Finally, although I have interpreted the agents in the model as households subject to capital income risk, the model can be used for a variety of issues. In particular, it can be interpreted as a model with risky human capital and private information. Hence, its implications can be used to draw policy implication for labor income for top labor earners as well.

References

- AIYAGARI, S. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 659–684. 2
- ALBANESI, S. (2011): "Optimal Taxation of Entrepreneurial Capital with Private Information," *Working Paper*. 4, 5
- ALBANESI, S. AND C. SLEET (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73, 1–30. 31, 32
- ANGELETOS, G. (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30. 29
- ATKESON, A. AND R. LUCAS (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59, 427–53. 30
- (1995): "Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance," *Journal of Economic Theory*, 66, 64–88. 31
- BENHABIB, J. AND A. BISIN (2009): "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *NBER Working Paper*.
- BENHABIB, J., A. BISIN, AND S. ZHU (2011): "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *Econometrica*, 79, 123–157. 2, 6
- CAGETTI, M. AND M. DE NARDI (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114, 835–870. 2
- (2009): "Estate Taxation, Entrepreneurship, and Wealth," *The American Economic Review*, 99, 85–111. 5
- CHAMLEY, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica: Journal of the Econometric Society*, 54, 607–622. 5
- CHAMPERNOWNE, D. (1953): "A Model of Income Distribution," *The Economic Journal*, 63, 318–351. 6
- CONESA, J., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea After All!" *The American Economic Review*, 99, 25–48. 5
- DIAMOND, P. AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–190. 9

- EVANS, D. AND B. JOVANOVIĆ (1989): "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *The Journal of Political Economy*, 97. 2
- FARHI, E. AND I. WERNING (2007): "Inequality and Social Discounting," *Journal of Political Economy*, 115, 365–402. 31
- (2008): "The Political Economy of Nonlinear Capital Taxation," Mimeo. 5
- (2010a): "Insurance and Taxation over the Life Cycle," Working Paper, MIT. 5
- (2010b): "Progressive Estate Taxation," *Quarterly Journal of Economics*, 125, 635–673. 5
- (2012): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 120, 398–445. 29
- FONG, K. (2009): "Evaluating Skilled Experts: Optimal Scoring Rules for Surgeons," Working Paper. 6
- GABAIX, X. (1999): "Zipf's Law for Cities: an Explanation," *The Quarterly Journal of Economics*, 114, 739–767. 6, 32
- (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294. 6
- GARRETT, D. AND A. PAVAN (2010): "Dynamic Managerial Compensation: On the Optimality of Seniority-based Schemes," Mimeo. 6
- GENTRY, W. AND G. HUBBARD (2004): "Entrepreneurship and Household Saving," *Advances in Economic Analysis and Policy*, 4. 2, 4
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587. 5, 27
- GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2010): "Optimal Dynamic Taxes," . 5
- HÖLMSTROM, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10, 74–91. 12
- JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 1177–1190. 7, 11, 12

- JUDD, K. (1985): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28, 59–83. 5
- KESTEN, H. (1973): "Random difference equations and renewal theory for products of random matrices," *Acta Mathematica*, 131, 207–248. 32
- KITAO, S. (2008): "Entrepreneurship, Taxation and Capital Investment," *Review of Economic Dynamics*, 11, 44–69. 5
- KOCHERLAKOTA, N. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73, 1587–1621. 5
- (2010): *The New Dynamic Public Finance*, Princeton Univ Press. 9
- LAFFONT, J. AND J. TIROLE (1986): "Using Cost Observation to Regulate Firms," *The Journal of Political Economy*, 94, 614–641. 6
- LUCAS, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 508–523. 14
- MEYN, S. P. AND R. L. TWEEDIE (2009): *Markov Chains and Stochastic Stability*, Cambridge University Press. 57
- MIREK, M. (2011): "Heavy Tail Phenomenon and Convergence to Stable Laws for Iterated Lipschitz Maps," *Probability Theory and Related Fields*, 151, 705–734. 6, 32, 34, 60
- MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208. 2
- MOLL, B. (2012): "Inequality and Financial Development: A Power-Law Kuznets Curve," Mimeo, Princeton University. 6
- PANOUSI, V. (2009): "Capital Taxation with Entrepreneurial Risk," Working Paper. 5
- PAVAN, A., I. SEGAL, AND J. TOIKKA (2009): "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure," Mimeo. 6
- PIKETTY, T. AND E. SAEZ (2012): "A Theory of Optimal Capital Taxation," NBER Working Paper. 4
- QUADRINI, V. (2000): "Entrepreneurship, Saving, and Social Mobility," *Review of Economic Dynamics*, 3, 1–40. 2, 4

- ROGERSON, W. P. (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357–1367. 7, 11
- SCHEUER, F. (2010): "Entrepreneurial Taxation, Occupational Choice, and Credit Market Frictions," Mimeo. 5
- SIMON, H. (1955): "On a Class of Skew Distribution Functions," *Biometrika*, 42, 425–440. 6, 32
- SLEET, C. AND Ş. YELTEKIN (2006): "Credibility and Endogenous Societal Discounting," *Review of Economic Dynamics*, 9, 410–437. 31
- THOMAS, J. AND T. WORRALL (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-agent Problem," *Journal of Economic Theory*, 51, 367–390. 30
- WERNING, I. (2010): "Nonlinear Capital Taxation," Working Paper, MIT. 5

Appendix

A Proofs

Proof of Lemma 1.

Consider the First Order Conditions from the program (7). The FOC associated with c_1 implies:

$$\beta \frac{1}{c_1} - \lambda_1 + \hat{\zeta} \beta \frac{1}{c_1} \left(-1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \right) = 0 \quad (31)$$

where $\hat{\zeta} k_1$ is the multiplier associated with (6). Hence,

$$\begin{aligned} c_1 &= \frac{\beta}{\lambda_1} \left(1 + \hat{\zeta} \left(-1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \right) \right) \\ &= \frac{\beta}{\lambda_1} \left(1 + \hat{\zeta} \frac{1}{\eta} (\varepsilon - 1) \right) \end{aligned}$$

where the last equality is derived using the properties of the Gamma distribution. This proves part 1.

Integrating (31) over $\varepsilon \in \mathbb{R}_+$, and using the fact that

$$\int_0^\infty \left(-1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \right) dH(\varepsilon) = 0,$$

we have

$$\int_0^\infty c_1(\theta, \varepsilon) dH(\varepsilon) = \frac{\beta}{\lambda_1}$$

which proves part 2.

Q.E.D.

Proof of Theorem 1.

In light of Lemma 1 the constraint (6) can be written as

$$\rho \frac{k_1}{c_0} = \beta \int_0^\infty \log(\phi(\theta) + \zeta(\theta)\varepsilon) (\eta\varepsilon - \eta) dH(\varepsilon)$$

Note that, one can write the objective as

$$\int \left[\log c_0 + \beta \int \log(\phi + \zeta\varepsilon) dH(\varepsilon) \right] dF(\theta)$$

Given this modification of the constraint, the first order conditions associated with problem (7) are given by

$$\frac{1}{c_0} - \lambda_0 + \rho \hat{\zeta} \frac{k_1}{c_0^2} = 0 \quad (32)$$

$$\beta \int_0^\infty \frac{1}{\phi + \zeta \varepsilon} dH(\varepsilon) - \lambda_1 + \hat{\zeta} \beta \int \frac{\eta \varepsilon - \eta}{\phi + \zeta \varepsilon} dH(\varepsilon) = 0 \quad (33)$$

$$\beta \int_0^\infty \frac{\varepsilon}{\phi + \zeta \varepsilon} dH(\varepsilon) - \lambda_1 + \hat{\zeta} \beta \int \frac{\varepsilon(\eta \varepsilon - \eta)}{\phi + \zeta \varepsilon} dH(\varepsilon) = 0 \quad (34)$$

$$\lambda_1 \theta - \lambda_0 - \rho \hat{\zeta} \frac{1}{c_0} = 0 \quad (35)$$

If we multiply the second equation with ϕ and the third one with ζ and add them together, we have

$$\beta = \lambda_1 (\phi + \zeta)$$

If we replace the above in the second FOC, we have

$$\beta \int \frac{1 + \hat{\zeta} \eta (\varepsilon - 1)}{\phi + \zeta + \zeta (\varepsilon - 1)} dH = \lambda_1 \rightarrow \beta \int \frac{1 + \hat{\zeta} \eta (\varepsilon - 1)}{\lambda_1^{-1} \beta + \zeta (\varepsilon - 1)} dH = \lambda_1$$

This implies that

$$\int_0^\infty \frac{1 + \hat{\zeta} \eta (\varepsilon - 1)}{1 + \beta^{-1} \lambda_1 \zeta (\varepsilon - 1)} dH = 1$$

The left hand side of the above equation is a strictly decreasing function of $\hat{\zeta}$ and the equation is satisfied for $\hat{\zeta} = \frac{\lambda_1 \zeta}{\beta \eta}$. Furthermore, the last equation implies that

$$\lambda_1 \theta - \lambda_0 = \rho \hat{\zeta} \frac{1}{c_0}.$$

Multiplying the first FOC by $\frac{\hat{\zeta}}{c_0}$, we have

$$\frac{1}{\hat{\zeta}} - \lambda_0 \frac{c_0}{\hat{\zeta}} + \rho \frac{k_1}{c_0} = 0 \quad (36)$$

Furthermore

$$\begin{aligned}
\rho \frac{k_1}{c_0} &= \beta \int_0^\infty \log(\phi + \zeta \varepsilon) (\eta \varepsilon - \eta) dH(\varepsilon) \\
&= \beta \int_0^\infty \log\left(\beta \lambda_1^{-1} + \beta \lambda_1^{-1} \hat{\zeta} \eta (\varepsilon - 1)\right) (\eta \varepsilon - \eta) dH(\varepsilon) \\
&= \beta \int_0^\infty \log(1 + \hat{\zeta} \eta (\varepsilon - 1)) (\eta \varepsilon - \eta) dH(\varepsilon) = F_1(\hat{\zeta})
\end{aligned}$$

where $F_1(\zeta)$ is an increasing function of ζ . Hence, (36) becomes

$$\frac{1}{\hat{\zeta}(\theta)} + F_1(\hat{\zeta}(\theta)) = \frac{\rho \lambda_0}{\lambda_1 \theta - \lambda_0} \quad (37)$$

The LHS of the above is given by

$$\frac{1}{\hat{\zeta}(\theta)} + \beta \eta e^{\frac{1}{\hat{\zeta}(\theta)} - 1} \left(\frac{1}{\hat{\zeta}(\theta)} - \eta \right)^\eta \Gamma\left(-\eta, \frac{1}{\hat{\zeta}(\theta)} - \eta\right)$$

where $\Gamma(a, z) = \int_z^\infty x^{a-1} e^{-x} dx$ is the incomplete Gamma function. This confirms that $\hat{\zeta}(\theta)$ exists when $\frac{\lambda_1}{\lambda_0} \theta - 1$ is less than the minimum of $\rho \left[\frac{1}{x} + \beta \eta e^{\frac{1}{x} - 1} \left(\frac{1}{x} - \eta \right)^\eta \Gamma\left(-\eta, \frac{1}{x} - \eta\right) \right]^{-1}$. It can be shown that the above function is U-shaped. If we assume that for the values of θ that (37) has two solutions, the lower solution is chosen (since it provides more insurance at the same investment incentive) an increase in θ increases $\hat{\zeta}(\theta)$. This shows that $\hat{\zeta}(\theta)$ is increasing in θ . This implies that $\frac{k_1}{c_0}$ is also an increasing function of θ . We can rewrite (36) as

$$\lambda_0 c_0(\theta) = 1 + \rho \frac{k_1(\theta)}{c_0(\theta)} \hat{\zeta}(\theta)$$

which implies that $c_0(\theta)$ is an increasing function and hence so is $k_1(\theta)$.

Obviously, for values of $\theta < 1/q$, it is not optimal to invest at all. This concludes the proof.

Q.E.D.

Proof of Lemma 2.

One can simply rewrite the definition of τ_K as

$$\tau_K(\theta) = 1 - \frac{1}{\theta} \frac{\zeta(\theta)}{c_0(\theta)} \frac{c_0(\theta)}{k_1(\theta)}$$

Recall the FOCs from program (7):

$$k_1 : \lambda_1 \theta - \lambda_0 - \hat{\zeta} \frac{\rho}{c_0} = 0$$

and note that

$$\zeta = \frac{\beta}{\lambda_1 \eta} \hat{\zeta}$$

Hence,

$$\frac{\zeta}{c_0} = \frac{1}{\eta} \beta \frac{\lambda_1 \theta - \lambda_0}{\rho \lambda_1}$$

So,

$$\tau_K(\theta) = 1 - \frac{1}{\eta \rho} \frac{\beta q \theta - 1}{q} \frac{c_0(\theta)}{k_1(\theta)}$$

This concludes the proof.

Q.E.D.

Proof of Proposition 1.

Equation (32) can be written as

$$\gamma - \lambda_0 c_0 + (\lambda_1 \theta - \lambda_0) \frac{k_1}{c_0} c_0 = 0 \Rightarrow c_0 = \frac{\gamma(\theta)}{\lambda_0 - (\lambda_1 \theta - \lambda_0) \frac{k_1(\theta)}{c_0(\theta)}}$$

Also, as before, we must have $\beta \lambda_1^{-1} \hat{\zeta}(\theta) \eta$. Hence, we can rewrite the saving wedge as

$$\begin{aligned} \frac{1}{1 - \tau_s(\theta)} &= \beta \frac{\lambda_0}{\lambda_1} \frac{\gamma(\theta)}{\lambda_0 - (\lambda_1 \theta - \lambda_0) \frac{k_1(\theta)}{c_0(\theta)}} \int_0^\infty \frac{1}{\beta \lambda_1^{-1} \gamma(\theta) + \beta \lambda_1^{-1} \hat{\zeta}(\theta) \eta (\varepsilon - 1)} dH(\varepsilon) \\ &= \lambda_0 \frac{1}{\lambda_0 - (\lambda_1 \theta - \lambda_0) \frac{k_1(\theta)}{c_0(\theta)}} \int_0^\infty \frac{1}{1 + \frac{\hat{\zeta}(\theta)}{\gamma(\theta)} \eta (\varepsilon - 1)} dH(\varepsilon) \end{aligned}$$

Note that from the moral hazard constraint, $\frac{k_1}{c_0}$ is an increasing function of $\frac{\hat{\zeta}}{\gamma}$. Hence when $\frac{k_1}{c_0}$ is increasing, so is $\frac{\hat{\zeta}(\theta)}{\gamma(\theta)}$. Therefore, the above should be increasing in θ .

Q.E.D.

Proof of Proposition 2.

Consider the case in which $\sigma = 1$. One can guess that the value function $P_{t+2}(w)$ takes the following form

$$-B_{t+2} e^{\frac{1+\beta}{1-\beta} w} + H_{t+2}$$

Now consider the policy functions $c_{0,t}(\theta, w)$, $k_{t+1}(\theta, w)$, $c_{1,t}(\theta, y, w)$, $w'_t(\theta, y, w)$, $U_t(\theta, w)$

and define the following:

$$\begin{aligned}
\hat{c}_{0,t}(\theta, w) &= c_{0,t}(\theta, w) e^{-\frac{1+\beta}{1-\beta}w} \\
\hat{c}_{1,t}(\theta, y, w) &= c_{1,t}(\theta, y, w) e^{-\frac{1+\beta}{1-\beta}w} \\
\hat{k}_{t+1}(\theta, w) &= k_{t+1}(\theta, w) e^{-\frac{1+\beta}{1-\beta}w} \\
\hat{w}_t(\theta, y, w) &= w'_t(\theta, y, w) - w \\
\hat{U}_t(\theta, w) &= U_t(\theta, w) - w
\end{aligned}$$

Then $\hat{c}_{0,t}(\theta, w)$, $\hat{k}_{t+1}(\theta, w)$, $\hat{c}_{1,t}(\theta, y, w)$, $\hat{w}_t(\theta, y, w)$, $\hat{U}_t(\theta, w)$ must solve the following problem

$$\max \int_{\Theta} \left[q_t \kappa_{t+1} \theta \hat{k}_1(\theta) - \hat{c}_0(\theta) - \hat{k}_1(\theta) - q_t \int_0^{\infty} \left[\hat{c}_1(\theta, y) + q_{t+1} B_{t+2} e^{\frac{1+\beta}{1-\beta} \hat{w}(\theta, y)} \right] dG(y|\theta, k_1(\theta)) \right] dF(\theta)$$

subject to

$$\begin{aligned}
\int_{\Theta} \hat{U}(\theta) dF(\theta) &= 0 \\
\log \hat{c}_0(\theta) + \int_0^{\infty} [\beta \log(c_1(\theta, y)) + \hat{\beta} \hat{w}(\theta, y)] dG(y|\theta, k_1(\theta)) &= \hat{U}(\theta) \\
\hat{U}'(\theta) &= \frac{1}{\theta} \frac{\hat{k}_1(\theta)}{\hat{c}_0(\theta)} \\
\int [\beta \log \hat{c}_1(\theta, y) + \hat{\beta} \hat{w}'(\theta, y)] g_k(y|\theta, \hat{k}_1(\theta)) dy &= \frac{1}{\hat{c}_0(\theta)}
\end{aligned}$$

This implies that these functions are independent of w . Furthermore, if the value associated with the above program is $-B_t$, the following should hold

$$\begin{aligned}
P_t(w) &= -B_t e^{\frac{1+\beta}{1-\beta}w} + p_t + q_t q_{t+1} H_{t+2} \\
&= -B_t e^{\frac{1+\beta}{1-\beta}w} + H_t
\end{aligned}$$

The case with $\sigma \neq 1$ can be similarly shown. A guess for the value function $P_{t+2}(w)$ can be

$$-B_{t+2}((1-\sigma)w)^{\frac{1}{1-\sigma}} + H_{t+2}$$

Then, we define the following

$$\begin{aligned}
\hat{c}_{0,t}(\theta, w) &= c_{0,t}(\theta, w) ((1 - \sigma) w)^{-\frac{1}{1-\sigma}} \\
\hat{c}_{1,t}(\theta, y, w) &= c_{1,t}(\theta, y, w) ((1 - \sigma) w)^{-\frac{1}{1-\sigma}} \\
\hat{k}_{t+1}(\theta, w) &= k_{t+1}(\theta, w) ((1 - \sigma) w)^{-\frac{1}{1-\sigma}} \\
\hat{w}_t(\theta, y, w) &= w'_t(\theta, y, w) ((1 - \sigma) w)^{-1} \\
\hat{U}_t(\theta, w) &= U_t(\theta, w) ((1 - \sigma) w)^{-1}
\end{aligned}$$

Note that since the constraint set is homogeneous in w , under the above change of variables it becomes:

$$\begin{aligned}
\int_{\Theta} \hat{U}_{0,t}(\theta) dF(\theta) &= \frac{1}{1 - \sigma} \\
\frac{\hat{c}_{0,t}(\theta)^{1-\sigma}}{1 - \sigma} + \int_0^\infty \left[\beta \frac{\hat{c}_{1,t}(\theta, y)^{1-\sigma}}{1 - \sigma} + \hat{\beta} \hat{w}_{0,t}(\theta, y) \right] dG(y|\theta, k_1(\theta)) &= \hat{U}_{0,t}(\theta) \\
\hat{U}'_{0,t}(\theta) &= \frac{1}{\theta} \hat{k}_{1,t+1}(\theta) \hat{c}_{0,t}(\theta)^{-\sigma} \\
\int \left[\beta \frac{(\hat{c}_{1,t}(\theta, y))^{1-\sigma}}{1 - \sigma} + \hat{\beta} \hat{w}_{0,t}(\theta, y) \right] g_k(y|\theta, \hat{k}_{1,t}(\theta)) dy &= \hat{c}_{0,t}(\theta)^{-\sigma}
\end{aligned}$$

Furthermore, the objective in (15) becomes

$$\int_{\Theta} \left[q_t \kappa_{t+1} \theta \hat{k}_{1,t}(\theta) - \hat{c}_{0,t}(\theta) - \hat{k}_{1,t}(\theta) - q_t \int_0^\infty \left[\hat{c}_{1,t}(\theta, y) + q_{t+1} B_{t+2} ((1 - \sigma) \hat{w}_{t+2}(\theta, y))^{\frac{1}{1-\sigma}} \right] dG(y|\theta, \hat{k}_{1,t}(\theta)) \right] + H_t.$$

As in the log-case the solution to optimization problem in new variables is independent of w and the above function verifies the initial guess for the value function.

Q.E.D.

Proof of Proposition 4.

Consider the maximization problem in (15). If we write the allocation in terms of ε

as opposed to y , the following are its associated FOCs:

$$\begin{aligned}
\alpha q_t \kappa_{t+1} \theta_t - 1 - \zeta \frac{u'(c_0)}{k} &= 0 \\
-1 + \gamma u'(c_0) - \zeta u''(c_0) &= 0 \\
-q_t + \gamma \beta u'(c_1) + \zeta \beta u'(c_1) \left(-1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \right) &= 0 \\
q_t q_{t+1} P'_{t+2}(w') + \gamma \hat{\beta} + \zeta \hat{\beta} \left(-1 - \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \right) &= 0 \\
\gamma &= \lambda
\end{aligned}$$

together with $P'_t(w) = -\lambda$ as Envelope condition. After some manipulations, the following hold

$$q_{t+1} P'_{t+2}(w') = -\frac{\hat{\beta}}{\beta} \frac{1}{u'(c_1)}$$

as well as

$$\begin{aligned}
\gamma &= \frac{1}{u'(c_0)} + \zeta \frac{u''(c_0)}{u'(c_0)} \\
&= \frac{1}{u'(c_0)} + (\alpha q_t \kappa_{t+1} \theta_t k - k) \frac{u''(c_0)}{u'(c_0)^2} \\
\gamma &= \int \frac{q_t}{\beta u'(c_1)} dH \\
P'_t(w) &= - \int \gamma dF
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\hat{\beta}}{q_{t+1} \beta} \frac{1}{u'(c_{1,t+1})} &= -P'_{t+2}(w_{t+2}) = E_{t+1} \left[\frac{1}{u'(c_{0,t+2})} + (\alpha q_{t+2} \kappa_{t+3} \theta_{t+2} k_{t+3} - k_{t+3}) \frac{u''(c_{0,t+2})}{u'(c_{0,t+2})^2} \right] \\
\frac{q_t}{\beta} E_t \left[\frac{1}{u'(c_{1,t+1})} \right] &= \frac{1}{u'(c_{0,t})} + (\alpha q_t \kappa_{t+1} \theta_t k_{t+1} - k_{t+1}) \frac{u''(c_{0,t})}{u'(c_{0,t})^2}
\end{aligned}$$

which are the desired equations.

Q.E.D.

Proof of Theorem 2.

Consider the component planning problem (15). The FOC w.r.t to c_0 , U , c_1 , and w' ,

respectively, are given by

$$-1 + \gamma u'(c_0) - \hat{\zeta} u''(c_0) = 0 \quad (38)$$

$$-\lambda + \gamma = 0$$

$$-q + \beta \gamma u'(c_1) + \beta \hat{\zeta} u'(c_1) \frac{g_k}{g} = 0 \quad (39)$$

$$q^2 P'_{t+2}(w') + \hat{\beta} \gamma + \hat{\beta} \hat{\zeta} \frac{g_k}{g} = 0 \quad (40)$$

where $\hat{\zeta}(\theta) f(\theta)$ is the multiplier on the incentive constraint and $\gamma(\theta) f(\theta)$ is the multiplier on (17). By the Envelope Theorem, $P'_t(w) = -\lambda$. The last two FOCs combined, imply:

$$\beta u'(c_1) = -\frac{\hat{\beta}}{q P'_{t+2}(w')}$$

To prove the above claim, we first show the following lemma:

Lemma 4 *The multiplier $\hat{\zeta}$ is positive.*

Proof. If we multiply (40) by $w'g$ and (40) $\frac{u(c_1)}{u'(c_1)}g$, we have

$$\begin{aligned} -q_t \frac{u(c_1)}{u'(c_1)} g + \beta \gamma u(c_1) g + \beta \hat{\zeta} u(c_1) g_k &= 0 \\ q_t q_{t+1} P'_{t+2}(w') w' g + \hat{\beta} \gamma w' g + \hat{\beta} \hat{\zeta} w' g_k &= 0 \end{aligned}$$

By adding and integrating over $y > 0$,

$$\begin{aligned} -q_t \int \frac{u(c_1)}{u'(c_1)} dG + \gamma \int \beta u(c_1) dG + \hat{\zeta} \int \beta u(c_1) g_k dy &= 0 \\ q_t q_{t+1} \int P'_{t+2}(w') w' dG + \gamma \int \hat{\beta} w' dG + \hat{\zeta} \int \beta w' g_k dy &= 0 \end{aligned}$$

Note that by integrating (39) and (40), we must have

$$q_t q_{t+1} \int P'_{t+2}(w') dG + \hat{\beta} \gamma = 0, -q_t \int \frac{1}{u'(c_1)} dG + \beta \gamma = 0$$

Replacing in the above, it implies that

$$\begin{aligned} -q_t \int \frac{u(c_1)}{u'(c_1)} dG + q_t \int \frac{1}{u'(c_1)} dG \int u(c_1) dG + \hat{\zeta} \int \beta u(c_1) g_k dy &= 0 \\ q_t q_{t+1} \int P'_{t+2}(w') w' dG - q_t q_{t+1} \int P'_{t+2}(w') dG \int w' dG + \hat{\zeta} \int \beta w' g_k dy &= 0 \end{aligned}$$

or

$$\begin{aligned} -q_t \text{Cov}\left(\frac{1}{u'(c_1)}, u(c_1)\right) + \hat{\zeta} \int \beta u(c_1) g_k dy &= 0 \\ q_t q_{t+1} \text{Cov}(P'_{t+2}(w'), w') + \hat{\zeta} \int \beta w' g_k dy &= 0 \end{aligned}$$

Adding the two constraint and using the incentive constraint, we have

$$\hat{\zeta} u'(c_0) = -q_t q_{t+1} \text{Cov}(P'_{t+2}(w'), w') + q_t \text{Cov}\left(\frac{1}{u'(c_1)}, u(c_1)\right)$$

Since $u(c_1)$ and $P_{t+2}(w)$ are concave functions – this is a direct implication of theorem 2, $\frac{1}{u'(c_1)}$ is increasing in $u(c_1)$ and $P'_{t+2}(w')$ is decreasing in w' . Hence, the RHS of the above equations is positive and $\zeta > 0$. ■

Given the above Lemma, we show that

$$-\frac{1}{P'_t(w)} = \frac{1}{\lambda} > \int_{\Theta} u'(c_0) dF \quad (41)$$

This would imply that:

$$\frac{\beta}{\hat{\beta}} q_{t+1} u'(c_{1,t+1}) > -\frac{1}{P'_{t+2}(w_{t+2})} = \frac{1}{\lambda} > E_{t+1} u'(c_{0,t+2})$$

and the claim follows. To see (41), note that from (38):

$$u'(c_0) = \frac{1}{\lambda} + \frac{\zeta}{\lambda} u''(c_0)$$

Since $\zeta > 0$ and that $u'' < 0$, we must have

$$u'(c_0) < \frac{1}{\lambda}$$

This proves, (41).

Q.E.D.

Proof of Proposition 5.

First recall equation (23):

$$\int_{\Theta \times \mathbb{R}} \rho(\hat{w}(\theta, y)) dG dF = 1$$

where

$$\rho(w) = \begin{cases} e^{\frac{1+\beta}{1-\beta}w} & \sigma = 1 \\ ((1-\sigma)w)^{\frac{1}{1-\sigma}} & \sigma \neq 1 \end{cases}$$

Consider the case with $\sigma = 1$: Recall equation (24)

$$\int_{\Theta \times \mathbb{R}} P'_{t+2}(w'_t(w, \theta, y)) dGdF = \frac{\hat{\beta}}{q_t q_{t+1}} P'_t(w)$$

In steady state, q_t , κ_t and p_t are all constant. Furthermore,

$$\begin{aligned} P'_t(w) &= \frac{1+\beta}{1-\hat{\beta}} B e^{\frac{1+\beta}{1-\hat{\beta}}w} \\ w'_t(w, \theta, y) &= \hat{w}(\theta, y) + w \end{aligned}$$

Hence, the above equation implies that

$$\int_{\Theta \times \mathbb{R}} e^{\frac{1+\beta}{1-\hat{\beta}}\hat{w}(\theta, y)} dGdF = \frac{\hat{\beta}}{q^2}$$

Hence, steady state implies that $\hat{\beta} = q^2$.

When $\sigma > 1$, (24) becomes

$$\int_{\Theta \times \mathbb{R}} ((1-\sigma)\hat{w}(\theta, y))^{\frac{\sigma}{1-\sigma}} dGdF = \frac{\hat{\beta}}{q^2}.$$

Note that since $\sigma > 1$, the function x^σ is convex in x . Hence, by Jensen's inequality

$$\frac{\hat{\beta}}{q^2} = \int_{\Theta \times \mathbb{R}} ((1-\sigma)\hat{w}(\theta, y))^{\frac{\sigma}{1-\sigma}} dGdF < \left(\int_{\Theta \times \mathbb{R}} ((1-\sigma)\hat{w}(\theta, y))^{\frac{1}{1-\sigma}} dGdF \right)^\sigma$$

So in a steady state, we must have $\hat{\beta} < q^2$.

When $\sigma < 1$, the function x^σ is concave in x . Hence, a similar argument as above shows that $\hat{\beta} > q^2$ in steady state.

Q.E.D.

Proof of Corollary 3.

When $\sigma = 1$,

$$w_{t+2} = w_t + \hat{w}(\theta_t, y_{t+1})$$

Hence,

$$w_t = w_{t \bmod 2} + \sum_{s=0}^{\lfloor t/2 \rfloor} \hat{w}(\theta_{t-2s}, y_{t+1-2s})$$

Now, note that in steady state

$$\int_{\Theta \times \mathbb{R}} e^{\frac{1+\beta}{1-\beta} \hat{w}(\theta, y)} dGdF = 1$$

Hence, by Jensen's inequality:

$$\int \hat{w}(\theta, y) dGdF < 0$$

Hence, applying law of large numbers to w_t implies that

$$\frac{w_t}{\lfloor t/2 \rfloor} \rightarrow_{a.s.} \int \hat{w}(\theta, y) dGdF < 0 \Rightarrow w_t \rightarrow_{a.s.} 0$$

When $\sigma \neq 1$,

$$\begin{aligned} w_{t+2} &= (1 - \sigma) w_t \hat{w}(\theta_t, y_{t+1}) \\ \Rightarrow \log((1 - \sigma) w_t) &= \log((1 - \sigma) w_{t \bmod 2}) + \sum_{s=0}^{\lfloor t/2 \rfloor} \log((1 - \sigma) \hat{w}(\theta_{t-2s}, y_{t+1-2s})) \end{aligned}$$

Note that in steady state:

$$\int_{\Theta \times \mathbb{R}} ((1 - \sigma) \hat{w}(\theta, y))^{\frac{1}{1-\sigma}} dGdF = 1$$

Taking log from both sides and using Jensen's inequality implies that

$$\frac{1}{1 - \sigma} \int_{\Theta \times \mathbb{R}} \log((1 - \sigma) \hat{w}(\theta, y)) dGdF < \log \left(\int_{\Theta \times \mathbb{R}} ((1 - \sigma) \hat{w}(\theta, y))^{\frac{1}{1-\sigma}} dGdF \right) = 0$$

Hence, when $\sigma > 1$

$$\int_{\Theta \times \mathbb{R}} \log((1 - \sigma) \hat{w}(\theta, y)) dGdF > 0$$

and when $\sigma < 1$,

$$\int_{\Theta \times \mathbb{R}} \log((1 - \sigma) \hat{w}(\theta, y)) dGdF < 0$$

A simple application of law of large number implies that

$$\begin{aligned}\log((1-\sigma)w_t) &\rightarrow_{a.s.} 0, \sigma < 1 \\ \log((1-\sigma)w_t) &\rightarrow_{a.s.} \infty, \sigma > 1\end{aligned}$$

Hence

$$\begin{aligned}w_t &\rightarrow_{a.s.} 0, \sigma < 1 \\ w_t &\rightarrow_{a.s.} -\infty, \sigma > 1\end{aligned}$$

Finally that either $\log((1-\sigma)w_t)$ or w_t are random walks implies that its unconditional variance converges to infinity.

Q.E.D.

Proof of Proposition 7.

I start by showing the first claim. Consider the n -th program associated with $P_n(w)$. I show that there exists \hat{w}_n such that

$$\begin{aligned}w'_n(w, \theta, y) &= w'(w, \theta, y), \forall w \geq \hat{w}_n, \\ P_n(w) &= P(w), \forall w \geq \hat{w}_n.\end{aligned}$$

I establish this for the case with $\sigma = 1$ and the other cases can be shown similarly. I prove this claim via induction. When $n = 1$, note that the unconstrained policy function is given by $w'(w, \theta, y) = \hat{w}(\theta, y) + w$. Since $\frac{\varepsilon h'}{h}$ is bounded below, $\hat{w}(\theta, y)$ is bounded below. Hence, \hat{w}_1 exists so that for $w \geq \underline{w} - \inf_{\theta, y} \hat{w}_1(\theta, y) (= \hat{w}_1)$

$$w'(w, \theta, y) \geq \underline{w}$$

Hence,

$$w'_1(w, \theta, y) = w'(w, \theta, y), \forall (\theta, y) \in \Theta \times \mathbb{R}^+$$

and other policy functions also coincide with the unconstrained policy functions. Furthermore, by definition of P , we must have that $P_1(w) = P(w)$, for all $w \geq \hat{w}_1$. Now consider the planning problem associated with n -th iteration. Note that when $w \geq \underline{w} - n \inf_{\theta, y} \hat{w}(\theta, y) = \hat{w}_n$, then $w'(w, \theta, y) = w + \hat{w}(\theta, y) \geq \hat{w}_{n-1}$. Hence, for $w \geq \hat{w}_n$, the policy functions associated with the unconstrained program solve the con-

strained planning problem. Hence,

$$w'_n(w, \theta, y) = w'(w, \theta, y), \forall (\theta, y) \in \Theta \times \mathbb{R}^+$$

Furthermore, $P_n(w) = P(w)$, for all $w \geq \hat{w}_n$. Hence,

$$\frac{dP_n(w'_n(w, \theta, y))}{dP_n(w)} = \frac{dP(w'(w, \theta, y))}{dP(w)} = e^{\frac{1+\beta}{1-\beta}\hat{w}(\theta, y)}, \forall (\theta, y) \in \Theta \times \mathbb{R}^+.$$

The rest of the first claim follows from the fact that $\frac{\partial}{\partial w}\hat{w}'_n(w, \theta, y)$ converges to $\frac{\partial}{\partial w}w'_c(w, \theta, y)$ in sup-norm. This can be seen from the following lemma:

Lemma 5 Consider a sequence of functions $f_n : \mathbb{R} \rightarrow \mathbb{R}$ with the property that $\lim_{x \rightarrow \infty} f_n(x) = a$ for all $n \in \mathbb{N}$ and $a \in \mathbb{R}$. Suppose that there exists a function $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |f_n(x) - f(x)| = 0$. Then $\lim_{x \rightarrow \infty} f(x) = a$.

Proof. Note that $|f(x) - a| \leq |f(x) - f_n(x)| + |f_n(x) - a|, \forall x$. By definition of sup-norm convergence, for any ε , there exists \hat{n} such that $n > \hat{n}, |f_n(x) - f(x)| < \varepsilon/2, \forall x \in \mathbb{R}$. Now for any $n > \hat{n}$, there must exist \tilde{x}_n such that for all $x > \tilde{x}_n, |f_n(x) - a| < \varepsilon/2$. Now let, $\tilde{x} = \inf_{n \geq \hat{n}} \tilde{x}_n$. For any $x > \tilde{x}$, there exists $m > \hat{n}$ such that $x > \tilde{x}_m$, and hence $|f_m(x) - a| < \varepsilon/2$. Furthermore, $|f_m(x) - f(x)| < \varepsilon/2$. Hence

$$|f(x) - a| \leq |f(x) - f_n(x)| + |f_n(x) - a| < \varepsilon, \forall x > \tilde{x}_m$$

Therefore, $\lim_{x \rightarrow \infty} f(x) = a$. ■

Applying the above lemma to $\frac{\partial}{\partial w}w'_n(w, \theta, y)$ and $\frac{\partial}{\partial w}w'_c(w, \theta, y)$ proves the result.

Q.E.D.

Proof of Proposition 3.

I focus on the case with $\sigma = 1$. The rest of the cases, can be shown in a similar way.

Given the policy function $w'_c(w, \theta, y)$, let

$$(T\mu)(A) = \int_{[\underline{w}, \infty)} \mathbf{1}[w'_c(w, \theta, y) \in A] d\mu(w)$$

A stationary measure, μ^* , is given by

$$T\mu^* = \mu^*$$

I show that the sequence $T^n \delta_{w_0}$ is tight, that is, for all $w_0 \in \mathbb{R}$ and $\varepsilon > 0$, there exists a

compact set $A_\varepsilon \subset \mathbb{R}$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n T^i \delta_{w_0} (A_\varepsilon) \geq 1 - \varepsilon.$$

Then by theorem 12.1.2 in [Meyn and Tweedie \(2009\)](#), this would imply that T has a stationary measure. Take any point w_0 and let w_t be the stochastic process constructed from repeated application of $w'_c(w, \theta, y)$ to w_0 . By law of large number, $E[w_n] = E[T^n \delta_{w_0}]$. Note that from Proposition 7,

$$\forall \hat{\varepsilon} > 0, \exists w^*; \forall w \geq w^*, |w'_c(w, \theta, y) - \hat{w}(\theta, y) - w| \leq \hat{\varepsilon}$$

From Proposition 7 and under $\hat{\beta} < q^2$,

$$\int \hat{w}(\theta, y) dGdF < 0$$

Hence,

$$w_{t+2} \leq w_t + \hat{w}(\theta_t, y_{t+1}) + \varepsilon, \forall w_t \geq w^*$$

So

$$E_t[w_{t+2}] \leq w_t + \int \hat{w}(\theta, y) dGdF + \varepsilon, \forall w_t \geq w^*$$

If ε is small enough,

$$E_t[w_{t+2}] < w_t - \varepsilon/2, \forall w_t > w^*$$

Now, define

$$\tilde{w}_t = \max\{w_t, w^*\}$$

Note that $E[\tilde{w}_t]$, the unconditional mean of \tilde{w}_t , by definition is higher than $E[T^n \delta_{w_0}]$. Note that since w_t decreases in expectation above w^* , for t large enough, it must be that $E[\tilde{w}_t]$ is bounded above and hence so is $E[T^n \delta_{w_0}]$. Suppose that this upper bound is given by \tilde{w} . Without loss of generality, we can assume that $\tilde{w} > 0$. Then, let $A = [\underline{w}, \tilde{w} + a]$ for some $a > 0$. Then, we have the following inequalities;

$$\tilde{w} > E[T^n \delta_{w_0}] \geq (1 - T^n \delta_{w_0}(A))(\tilde{w} + a) + T^n \delta_{w_0}(A) \underline{w}$$

Hence,

$$T^n \delta_{w_0}(A) \geq \frac{a}{\tilde{w} + a - \underline{w}} > 0$$

This proves that $T^n \delta_{w_0}$ is a tight sequence of measure and hence so is $\frac{1}{n} \sum_{i=1}^n T^i \delta_{w_0}$.

That the stationary distribution does not put full mass on \underline{w} follows from the fact that the likelihood ratio converges to ∞ as ε converges to ∞ . This concludes the proof.

Q.E.D

Proof of Proposition 8.

Note that in the constrained problem, we must have

$$P'_c(w_t) = q^2 \hat{\beta}^{-1} E_t P'_c(w_{t+2}) + E_t \xi_{t+2}$$

where ξ_{t+2} is the Lagrange multiplier on $w_{t+2} \geq \underline{w}$ and hence positive. In any steady state and under the stationary measure μ^* , we must have

$$\begin{aligned} \int P'_c(w) d\mu^*(w) &= q^2 \hat{\beta}^{-1} \int \int_{\Theta \times \mathbb{R}_+} P'_c(w'_c(w, \theta, y)) dG(y|\theta, k) dF(\theta) d\mu^*(w) \\ &\quad + \int \int_{\Theta \times \mathbb{R}_+} \xi(w, \theta, y) dG(y|\theta, k) dF(\theta) d\mu(w^*) \end{aligned}$$

By definition of stationary measure, we can write the above as

$$(1 - q^2 \hat{\beta}^{-1}) \int P'_c(w) d\mu^*(w) = \int \int_{\Theta \times \mathbb{R}_+} \xi(w, \theta, y) dG(y|\theta, k) dF(\theta) d\mu(w^*)$$

Since P'_c is negative and ξ is positive, we must have $q^2 \hat{\beta}^{-1} > 1$ or $\hat{\beta} q^{-2} < 1$.

The rest of the claim follows from Proposition 7. From Proposition 7, we have that

$$\lim_{w \rightarrow \infty} \frac{dP_c(w'_c(w, \theta, y))}{dP_c(w)} = \frac{dP(w'(w, \theta, y))}{dP(w)}$$

When $\sigma = 1$,

$$\frac{dP(w'(w, \theta, y))}{dP(w)} = \frac{dB e^{\frac{1+\beta}{1-\beta}(w+\hat{w}(\theta, y))}}{dB e^{\frac{1+\beta}{1-\beta}w}} = e^{\frac{1+\beta}{1-\beta}\hat{w}(\theta, y)}$$

where $\hat{w}(\theta, y)$ comes from Proposition 2. Note further that, equation (24) implies

$$\int_{\Theta \times \mathbb{R}_+} P'(w'(w, \theta, y)) dG dF = \hat{\beta} q^{-2} P'(w),$$

or

$$\begin{aligned} \int_{\Theta \times \mathbb{R}_+} e^{\frac{1+\beta}{1-\beta}\hat{w}(\theta, y)} dG dF &= \hat{\beta} q^{-2} < 1 \\ \int_{\Theta \times \mathbb{R}_+} \lim_{w \rightarrow \infty} \frac{dP_c(w'_c(w, \theta, y))}{dP_c(w)} dG dF &= \hat{\beta} q^{-2} < 1 \end{aligned}$$

When $\sigma \neq 1$,

$$\frac{dP(w'(w, \theta, y))}{dP(w)} = \frac{dB\left((1-\sigma)^2 w \hat{w}(\theta, y)\right)^{\frac{1}{1-\sigma}}}{dB\left((1-\sigma) w\right)^{\frac{1}{1-\sigma}}} = ((1-\sigma) \hat{w}(\theta, y))^{\frac{1}{1-\sigma}}$$

As before and from equation (24),

$$\int_{\Theta \times \mathbb{R}_+} P'(w'(w, \theta, y)) dGdF = \hat{\beta} q^{-2} P'(w),$$

or

$$\int_{\Theta \times \mathbb{R}_+} ((1-\sigma) \hat{w}(\theta, y))^{\frac{\sigma}{1-\sigma}} dGdF = \hat{\beta} q^{-2} < 1$$

When $\sigma > 1$, an application of Jensen's inequality implies

$$\int_{\Theta \times \mathbb{R}_+} ((1-\sigma) \hat{w}(\theta, y))^{\frac{\sigma}{1-\sigma}} dGdF > \left(\int_{\Theta \times \mathbb{R}_+} ((1-\sigma) \hat{w}(\theta, y))^{\frac{1}{1-\sigma}} dGdF \right)^\sigma$$

Therefore, we must have

$$\begin{aligned} 1 &> \int_{\Theta \times \mathbb{R}_+} ((1-\sigma) \hat{w}(\theta, y))^{\frac{1}{1-\sigma}} dGdF \\ 1 &> \int_{\Theta \times \mathbb{R}_+} \lim_{w \rightarrow \infty} \frac{dP_c(w'_c(w, \theta, y))}{dP_c(w)} dGdF \end{aligned}$$

which concludes the proof.

Q.E.D.

Proof of Theorem 4:

Note that the process for wealth is defined by

$$\begin{aligned} a_{t+2} &= -qP_c\left(w'_c\left((P_c)^{-1}\left(-q^{-1}a_t\right), \theta_t, y_{t+1}\right)\right) \\ &= \psi(a_t, \theta_t, y_{t+1}) \end{aligned}$$

Note that if $a = -qP_c(w)$, then

$$\psi_a(a, \theta, y) = \frac{dP_c(w'_c(w, \theta, y))}{dP_c(w)}$$

Hence, we can define $M(\theta, y) = \lim_{a \rightarrow \infty} \psi_a(a, \theta, y) = \lim_{w \rightarrow \infty} \frac{dP(w'(w, \theta, y))}{dP(w)}$. Furthermore,

given that w is bounded below by \underline{w} . There must exist, $N(\theta, y)$ such that

$$M(\theta, y) a - N(\theta, y) < \psi(a, \theta, y) < M(\theta, y) a + N(\theta, y)$$

The conditions required in [Mirek \(2011\)](#) for existence of a Pareto stationary measure are – see theorem 1.8:

- Condition 1: There must exist $\nu > 0$ such that $E[M(\theta, y)^\nu] = 1$: To see this note that, note that from Proposition 3 $E[M(\theta, y)] < 1$. Furthermore by Jensen's inequality the function $E[M(\theta, y)^\nu]$ is an increasing function of ν . Moreover, since the likelihood ratio $\frac{g_k}{g}$ converges to ∞ as y tends to ∞ and that $P'(\hat{w}(\theta, y)) = q^{-1}\hat{\beta}\left(\phi(\theta) + \zeta(\theta)\frac{g_k}{g}\right)$, there is a positive measure of (θ, y) 's for which $M(\theta, y) > 1$. Hence, $E[M(\theta, y)^\nu]$ tends to ∞ as ν tends to ∞ . So there must exist ν such that $E[M(\theta, y)^\nu] = 1$.
- Condition 2: $\log M(\theta, y)$ should not have support of the form $r\mathbb{Z}$: This is simply shown from the fact that (θ, y) have continuous distributions and that $w'(w, \theta, y)$ is continuous in (θ, y) .
- Condition 3: $M(\theta, y)$ satisfies $E[M^\nu \log M] < \infty$: This is simply seen from the fact that any polynomial or exponential function under gamma distribution has bounded expectation. Note that $P'(\hat{w}(\theta, y))$ is a linear function of y . Since P' is either an exponential or power function, the expected value of $P'(\hat{w}(\theta, y))^\alpha$ must exist for any α and so $E[M(\theta, y)^\nu \log M(\theta, y)]$ must also exist.
- Condition 4: random variable $N(\theta, y)$ must satisfy $E[N^\nu] < \infty$: This is again followed by the fact that P_c can be bounded above and below by an exponential function and hence $E[N^\nu]$ must exist.

This concludes the proof.

Q.E.D.