Optimal Taxation of Capital Income:

A Mirrleesian Approach to Capital Accumulation *

Ali Shourideh

Wharton School, University of Pennsylvnia

shouride@wharton.upenn.edu

June 28, 2012

Abstract

In this paper, I develop a model of optimal taxation of capital and study its effect on the long run distribution of wealth in presence of idiosyncratic capital income risk. In presence of risk-return trade-offs, I show that it is typically optimal to have progressive saving taxes. Furthermore, in an intergenerational context, we show that bequest taxes should be negative and develop a method for characterization of long run efficient distribution of wealth.

*This is a significantly revised version of the the first chapter of my PhD dissertation at the University of Minnesota. I am indebted to V. V. Chari and Larry Jones for valuable advice. I would also like to thank Laurence Ales, Alessandro Dovis, Andy Glover, Mike Golosov, Roozbeh Hosseini, Patrick Kehoe, Narayana Kocherlakota, Ellen McGrattan, Chris Phelan, Maryam Saeedi, Erick Sager, Florian Scheuer, Maxim Troshkin, and specially Ariel Zetlin-Jones as well as seminar participants in Booth, Carnegie Mellon, Chicago, Columbia GSB, Northwestern, Princeton, Wharton, Yale as well American Economic Association, NBER Macro Public Finance and Midwest Macro Conference for their comments. I acknowledge the support of University of Minnesota Doctoral Dissertation Fellowship.

1 Introduction

How should wealth be taxed? The answer to this question requires taking a stand on the process of wealth accumulation. Much of the optimal economic 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 similar to that observed in the data. It has also been argued that models with entrepreneurs 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 entrepreneurial 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, productivity has two components, a component that is *known* by the entrepreneurs *in advance* 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 insure entrepreneurs against productivity and income risk via redistributive schemes. These redistributive motives together with private information, leads to a trade-off between incentives to invest and insurance as in Mirrlees (1971); hence a Mirrleesian approach to capital accumulation.

In this environment, I first analyze how taxes should be designed to achieve efficiency. 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 incomplete market and constrained efficient allocation. Thus shedding light on the question of what the wealth distribution should look like.

As for the analysis of taxes, there are two main take aways from the model: First, heterogeneity in risk and return leads to progressive taxes on capital income outside the business. The forces behind this result are constant return to scale and moral hazard. Due to constant return to scale, the planner would like to have the most productive unit produce. However, this comes at a cost due to moral hazard. In fact, as the desired investment increases it becomes increasingly costly to provide incentive to invest. This convexity of cost of moral

[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 Quadrini (2000), Cagetti and De Nardi (2006), and Benhabib and Bisin (2009).

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

hazard implies that projects with higher return should have higher investment. This implies that more productive households are bearing higher risk and hence should be discouraged from investing outside the firm through progressive taxes on saving(outside of business).

Second, when ideas do not persist across generations, i.e., productivities are stochastic, bequests should be subsidized. This result is driven by the way savings affect incentive to invest for future generations. In fact, in this model contrary to models with labor income risk, household's consumption increases upon lying. This means that when saving is increased, the value of truth-telling increases by more than value of lying. Hence, saving relaxes incentives to invest in the future.

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. Similar to Benhabib et al. (2011), 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 both an incomplete market version of the model as well the constrained efficient version of the model. The methodology here therefore allows me to compute the efficient distribution of wealth and compare it to data 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. The literature on capital taxation has widely focused on labor income heterogeneity as the main driving factor of inequality[3]. In this paper, I have investigated the other extreme where capital income heterogeneity and risk is the driving factor of income as well as wealth inequality. Given that the results are very different from models with labor income risk, the analysis here is a first step toward a more thorough analysis of capital taxation.

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

This 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 in-

[3] An exception is Piketty and Saez (2012) who develop a model with accidental bequests and study its effect on capital taxation. While their argue that accidental bequests are consist with micro data, they do not investigate the implications of their model about the distribution of wealth in the economy.

come 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 (2006) and Scheuer (2010) are early attempts in studying optimal design of tax system for entrepreneurs. Scheuer (2010) focuses on the decision of entry into entrepreneurship and its implication for differential treatment of entrepreneurs and workers. Albanesi (2006) 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[4]. 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, subsidies are optimal since they relax future incentive constraints.

This paper is related to a body of research that studies power laws in economics, Champernowne (1953), Simon (1955), Gabaix (1999) and Benhabib et al. (2011)[5]. 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.

Finally, from a technical point of view, 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 and we derive conditions under which this first order approach is valid. Since there are two types of private information, this model shares the same structure

[4] 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).

[5] See Gabaix (2009) for an extensive review.

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 intuition via a two period model. Section 3, contains the dynamic extension of the two period model to an overlapping generations model. Section 4, discusses the properties of the long-run distribution of wealth. Section 5 concludes.

2 A Two Period Example

In this section, we focus on a two period economy in order to identify the key economic forces. We start with a two period example to show one of the main results of the paper – progressivity. As we see, the Modified Inverse Euler Equation – an equation governing time series properties of consumption – proves useful in the analysis of the intertemporal wedge. Hence, we derive a version of it for the two period example and further analyze saving wedge for a case with log utility.

Consider a two period economy in which $t = 0, 1$. The economy is populated by a continuum of entrepreneurs. Each entrepreneur is the sole owner of an investment technology or project that is subject to idiosyncratic risk. In particular, entrepreneurs draw a productivity shock, $\theta \in [\underline{\theta}, \bar{\theta}]$, at $t = 0$. I assume that θ is distributed according to the 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 an entrepreneur 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 $G(y|k_1, \theta)(G_y(y|k_1, \theta) = g(y|k_1, \theta))$ where $G(\cdot|\cdot, \cdot)$ is C^1 in all of its argument. Moreover, the mean value of y, given θ, k_1 is given by θk_1, i.e., $\int_0^{\bar{y}} y g(y|k_1, \theta) dy = \theta k_1$. That is the production technology exhibits constant return to scale.mNotice this formulation can stand-in for a more general constant return to scale production function that employs labor and capital with labor being supplied competitively in the labor market[6]. The constant returns to scale assumption makes the ultimate goal of the paper, i.e, a model with fat tail for wealth distribution, easier to establish.[7]

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

[6]Suppose that the production function is given by $y = \varepsilon^{\psi} A \theta k^{\alpha_1} l^{1-\alpha_1}$ where l is labor input and ε is a shock realized once capital is put in place. If managers employ labor at $t = 1$, the profit maximization decision of the firm in $t = 1$ is given by

$$\max_l \varepsilon^{\psi} A \theta^{\alpha} k^{\alpha_1} l^{1-\alpha_1} - wl$$

and therefore, $\alpha_2 \varepsilon^{\psi} A \theta^{\alpha} k^{\alpha_1} l^{-\alpha_1} = w$. Hence, α_2, α_1, ψ, and A can be chosen so that $y = \varepsilon \theta^{\alpha} k^{\alpha}$.

[7]In section 3, I illustrate how an extension of this model to an overlapping generations model can generate a long run distribution of wealth with fat tail.

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

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

The above implies that the distribution of y satisfies

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

We will refer to the p.d.f. of the exponential distribution with parameter η as $h(\varepsilon)$. As we will see, the above assumption together with the log-utility assumption below, implies that optimal consumption in the second period should be linear in income. This would further simplify the analysis in characterizing saving taxes. Note that the above assumption implies that $\frac{g_k(y|k_1,\theta)}{g(y|k_1,\theta)}$ or the likelihood ratio is increasing and linear (and hence it satisfies the conditions in see Jewitt (1988) and Rogerson (1985b).),

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

In addition, entrepreneurs preferences are standard and given by

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

where c_0 and c_1 are consumption of the entrepreneur at each period[8]. Entrepreneurs, therefore, consume in each period and invest at $t = 0$. We assume for simplicity that each agent is endowed with e_0 units of labor income at $t = 0$.

For this economy, an allocation is given by $\{c_0(\theta), c_1(\theta, y), k_1(\theta)\}_{\theta=\underline{\theta}}^{\bar{\theta}}$. An allocation is said to be *feasible* if it satisfies the following:

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

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_0^{\bar{y}} c_1(\theta, y) g(y|k_1(\theta), \theta) dy dF(\theta) \leq \int_{\underline{\theta}}^{\bar{\theta}} \theta k_1(\theta) dF(\theta) \tag{2}$$

2.1 Saving Distortions with Observable Productivity

It is useful to characterize efficient allocations when a planner can observe entrepreneurs' project type, θ, but cannot observe their consumption and investment. Note that when everything is observable, it is optimal to have only the highest productivity entrepreneur $\bar{\theta}$ operate. This is because of constant return to scale and full risk sharing. With consumption and investment unobservable, each project is risky and hence it is optimal to operate all projects.

[8] In the appendix, we also provide how our results would change with $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, $\sigma \neq 1$.

Our goal is to characterize distortions to the saving arising from the moral hazard and how it varies with average returns across projects.

Under this specification, an allocation is incentive compatible if

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

Since the specification described above satisfies the conditions for the first order approach described in Jewitt (1988) and Rogerson (1985b), we can replace the above with its first order condition:

$$\frac{1}{c_0\left(\theta\right)}=\beta\int\log\left(c_1\left(\theta,y\right)\right)g_k\left(y|\theta,k_1\left(\theta\right)\right)dy \tag{3}$$

Hence, the planning problem is given by

$$\max_{c_0(\theta),c_1(\theta,y),k_1(\theta)}\int_{\underline{\theta}}^{\bar{\theta}}\left[\log(c_0(\theta))+\beta\int_0^{\bar{y}}\log(c_1(\theta,y))g(y|k_1(\theta),\theta)dy\right]dF(\theta)$$

subject to (3), (1), and (2).

We start characterizing the optimal allocation via the following lemma:

Lemma 1 *Consumption at $t=1$ is a linear function of income, or*

$$c_1\left(\theta,\varepsilon\theta k_1\left(\theta\right)\right)=\phi\left(\theta\right)+\zeta\left(\theta\right)\varepsilon$$

for some $\phi\left(\theta\right),\zeta\left(\theta\right)>0$.

This result is a straightforward implication of the moral hazard model. To see this, consider increasing $\log(c_1\left(\theta,y\right))$ by 1 unit for $y\in[\hat{y},\hat{y}+\delta]$ while decreasing $\log(c_1\left(\theta,y\right))$ for all other y's by $\delta g(\hat{y}|\theta,k_1\left(\theta\right))$ for δ small. This perturbation leaves the expected utility unchanged. Furthermore, it affect the incentives to invest by $\delta g_k\left(\hat{y}|\theta,k_1\left(\theta\right)\right)$. The cost of this perturbation is therefore given by

$$c_1\left(\theta\right)\delta g\left(y|\theta,k_1\left(\theta\right)\right)-\delta g\left(y|\theta,k_1\left(\theta\right)\right)\int c_1\left(\theta,y\right)g\left(y|\theta,k_1\left(\theta\right)\right)dy$$

At the optimal allocation, this cost has to equate the incentive benefits and hence

$$c_1\left(\theta,\hat{y}\right)\delta g\left(\hat{y}|\theta,k_1\left(\theta\right)\right)-\delta g\left(\hat{y}|\theta,k_1\left(\theta\right)\right)\int c_1\left(\theta,y\right)g\left(y|\theta,k_1\left(\theta\right)\right)dy=\mu_2\left(\theta\right)\delta g_k\left(\hat{y}|\theta,k_1\left(\theta\right)\right)$$

and therefore

$$\begin{aligned} c_1(\theta, \hat{y}) &= \int c_1(\theta, y)\, g(y|\theta, k_1(\theta))\, dy + \mu_2(\theta) \frac{g_k(\hat{y}|\theta, k_1(\theta))}{g(\hat{y}|\theta, k_1(\theta))} \\ &= \phi(\theta) + \mu_2(\theta)\, \hat{y} \end{aligned}$$

or

$$c_1(\theta, \varepsilon\theta k_1(\theta)) = \phi(\theta) + \zeta(\theta)\, \varepsilon, \forall \varepsilon \in \mathbb{R}_+$$

for some $\gamma(\theta)$, $\phi(\theta) > 0$.

Evidently this linearity of the consumption schedule is not robust to changes to the distribution of shocks as well the utility function. However, it greatly simplifies the analysis of the distortions to saving and their dependence in θ. In the appendix, we show that some of the results illustrated here extend to general CRRA preferences and gamma distribution.

To characterize distortions to saving, we first define wedges for general securities. Consider, a risky security D with dividends per unit held, d, distributed according to $\Psi(d)$. For this security, we can define the following wedge:

$$\tau_D : u'(c_0) = \beta(1 - \tau_D) E_{(d,y)}[u'(c_1)\, d]$$

Note that this wedge can be thought of as a marginal tax rate on each additional unit of income earned from this security and it affects agent's incentives to invest in that security. For example, consider a bond that pays r units in each state of the world – for all ε, τ_S defined by

$$u'(c_0) = \beta(1 - \tau_S)\, rE\left[u'(c_1)\right]$$

can be thought of as marginal tax rate on gross income from the bond, i.e., principal plus interest. Note that given the above lemma, for a project of type θ, there is a wedge on equity – claim to the revenue of the firm and is equal to $1 - \frac{1}{\theta k_1(\theta)}\zeta(\theta)$. This is because for every unit increase in the revenue from the firm, the entrepreneur's consumption increase by $\gamma(\theta)$. Although, these wedges resemble taxes, there is no reason that market mechanisms cannot achieve efficiency. Later, we consider various market mechanisms and discuss what they imply about taxes.

We further characterize optimal allocations via the following lemma:

Lemma 2 *Average consumption at $t = 1$, $\int_0^\infty c_1(\theta, y)\, g(y|\theta, k)\, dy$, is equated across different types.*

To see this, note that perturbing the allocation via multiplying $c_1(\theta, y)$ by $1 + \delta$ for $\delta > 0$

small, has no effect on incentive since

$$\begin{aligned}\int \log\left(c_1\left(\theta,y\right)\delta\right) g_k\left(y|\theta,k\right) dy &= \int \left[\log\left(c_1\left(\theta,y\right)\right)+\log\delta\right] g_k\left(y|\theta,k\right) dy \\ &= \int \log\left(c_1\left(\theta,y\right)\right) g_k\left(y|\theta,k\right) dy\end{aligned}$$

The marginal cost of such perturbation is $\delta\int_0^\infty c_1\left(\theta,y\right) g\left(y|\theta,k\right) dy$. Since θ is observable, this marginal cost should be equated across types. Given the definition of ϕ and ζ, we must have

$$\phi\left(\theta\right)+\zeta\left(\theta\right)\frac{1}{\eta} = \phi\left(\theta'\right)+\zeta\left(\theta'\right)\frac{1}{\eta}$$

Our main result of this section can be summarized as follows:

Proposition 1 *Suppose that* $q\frac{\theta}{\eta}-1<\bar{x}$ *where* $\bar{x}=\left[\min_{0\leq x\leq 1} x+\beta\left[1-\left(\frac{1}{x}-1\right)e^{1/x-1}\Gamma\left(0,1/x-1\right)\right]\right]^{-1}$, *then* $c_0\left(\theta\right)$, $\frac{k_1(\theta)}{c_0(\theta)}$, $\zeta\left(\theta\right)$, $-\phi\left(\theta\right)$ *as well as* $\tau_S\left(\theta\right)$ *are increasing functions of* θ.

where in the above Proposition, $\Gamma\left(\cdot,\cdot\right)$ is the incomplete gamma function, q is the price of consumption in the second period relative to consumption in the first period, and $\tau_S\left(\theta\right)$ is the wedge associated with a risk free bond with gross rate of return q^{-1}.

Proof. In the Appendix.

The above proposition states that when θ is bounded above, first period consumption, capital-consumption ratio, as well as the riskiness of $c_1(\theta,y)$ increases with type. To see how this implies an increasing wedge $\tau_S\left(\theta\right)$, note that by definition

$$\frac{1}{1-\tau_S\left(\theta\right)} = \beta q^{-1} c_0\left(\theta\right)\int_0^\infty \frac{1}{\phi\left(\theta\right)+\zeta\left(\theta\right)\varepsilon} h\left(\varepsilon\right) d\varepsilon$$

When $-\phi\left(\theta\right)$, $\zeta\left(\theta\right)$, and $c_0\left(\theta\right)$ are increasing in θ, then so is $\tau_S\left(\theta\right)$.

Intuitively, the planner would like to invest as much possible in the highest productivity project. However, due to moral hazard, this comes at a cost. Hence, the planner would like to set the cost of moral hazard equal to the return to the investment, $q\theta/\eta-1$. Since the marginal cost of investment is bounded above, this is possible only up to a certain level. In other words, when the return to investment is very high the cost of moral hazard is not enough to prevent the planner from investing all the funds in the highest productivity project. Furthermore, since the cost of moral hazard is convex in k, optimal investment has to be increasing. For that to happen, the slope of the consumption schedule $\zeta\left(\theta\right)$ has to be increasing with θ. We leave the technical proof of this to the appendix.

More intuitively, note that average consumption is equated across types. In this case, when ζ is increasing in θ, consumption is risker for higher productivity projects, lower utility level. Facing this increased risk, an entrepreneur would like to decrease its investment in the project

and increase its holding of a risk-free bond. Hence in order to achieve the efficient level of investment, the planner should increase distortions on bond holdings.

2.2 Private Information About Productivity

Here we assume that agents are privately informed about their productivities. Moreover, the planner cannot observe consumption and investment by a particular agent at $t = 0$. The planner can only observe income y at $t = 1$. By the Revelation Principle, we can focus on direct mechanisms in which each type reports his productivity. We call an allocation *incentive compatible* if it satisfies the following:

$$\begin{aligned} &\log(c_0(\theta)) + \beta \int_0^\infty \log(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy \\ &\geq \max_{\hat{\theta}, \hat{k}} \log\left(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}\right) + \beta \int_0^\infty \log(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy \end{aligned} \tag{4}$$

The RHS of the above inequality is the utility that a type θ receives when he reports $\hat{\theta}$ and invests $\hat{k}$. Moreover, I call an allocation *incentive feasible*, if it is incentive compatible and feasible.

The assumption about private information features two type of incentive problems: a hidden type problem and a hidden action problem. The hidden type problem implies that, when facing the full information efficient allocation, agents with higher productivity – θ, have incentive to lie downward about their productivity type even if they invest "the right" amount. By lying downward and investing $\frac{\hat{\theta} k_1(\hat{\theta})}{\theta}$, higher productivity agents can enjoy higher consumption in the first period. Moreover, the hidden action problem implies that even if the agents tell the truth, the full insurance in the second period leads to under-investment in the first period.

Given above definitions, a utilitarian planner that maximizes entrepreneurs' ex-ante utility solves the following problem:

$$\max_{c_0(\theta), c_1(\theta, y), k_1(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} \left[\log(c_0(\theta)) + \beta \int_0^\infty \log(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy\right] dF(\theta)$$

subject to (1), (2), and (4).

First Order Approach. As can be seen, the set of incentive compatibility constraints is large and this complicates the characterization of optimal allocations. Here, as before, I appeal to the first order approach to simplify the set of incentive compatibility constraints and discuss the validity of this approach in this environment. In particular, let $U(\theta)$ be the utility of type

θ from truth-telling. Then we must have

$$U(\theta) = \max_{\hat{\theta},\hat{k}} \log\left(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}\right) + \beta \int_0^\infty \log(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy$$

If we assume that the allocations are C^1 in θ and y, then incentive compatibility yields the following first order conditions and Envelope condition:

$$\frac{1}{c_0(\theta)} = \beta \int_0^\infty u(c_1(\theta, y)) g_k(y|k_1(\theta), \theta) dy \tag{5}$$

$$\frac{1}{c_0(\theta)} \left[c_0'(\theta) + k_1'(\theta)\right] + \beta \int_0^\infty \frac{1}{c_1(\theta, y)} c_{1\theta}(\theta, y) g(y|k_1(\theta), \theta) dy = 0 \tag{6}$$

The Envelope condition associated with this problem is given by

$$\begin{aligned} U'(\theta) &= \frac{\partial}{\partial \theta} \log\left(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}\right) + \beta \int_0^\infty \log(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy \Big|_{\hat{\theta}=\theta, \hat{k}=k_1(\theta)} \\ &= \beta \int_0^\infty \log(c_1(\hat{\theta}, y)) g_\theta(y|\hat{k}, \theta) dy \end{aligned}$$

Note that since $g(y|k_1, \theta)$ is a function of θk_1, I can write $g_\theta(y|k_1, \theta) = \frac{k_1}{\theta} g_k(y|k_1, \theta)$. Hence, the above envelope condition combined with the first order condition simplifies to

$$U'(\theta) = \frac{1}{\theta} \frac{k_1(\theta)}{c_0(\theta)} \tag{7}$$

We say an allocation is *locally incentive compatible* if it satisfies (5) and (7).

The above conditions are necessary for incentive compatibility. However, it is not clear that they are sufficient for incentive compatibility. Our aim, here, is to provide sufficient conditions under which the local incentive compatibility implies incentive compatibility, i.e., the First Order Approach(FOA) is valid. As mentioned before, there are two frictions in this model: an adverse selection problem and a moral hazard problem. Regarding the adverse selection problem, there has not been much success in finding general assumptions on primitives that validate the FOA[9]. In Appendix **??**, in line with Pavan et al. (2009), we provide monotonicity conditions on *endogenous* allocations that can be easily checked and are sufficient to ensure that FOA is valid.

Given the above discussion and conditions provided in Appendix **??**, in what follows, we relax the set of incentive compatible constraints and only impose local incentive compatibility. This further simplifies the analysis of the planning problem and enables us to further

[9]There are special cases for which assumptions on fundamentals exist. For example Myerson (1981) and Guesnerie and Laffont (1984) show that when principal and agent are both risk neutral, a monotone likelihood ratio assumption on the distribution of types validates the FOA.

characterize the properties of the optimal allocations.

Hence, the relaxed problem becomes the following:

$$\max_{c_0(\theta),c_1(\theta,y),k_1(\theta),U(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} U(\theta)dF(\theta) \tag{P1}$$

subject to

$$\begin{aligned}
\int_{\underline{\theta}}^{\bar{\theta}} \left[c_0(\theta) + k_1(\theta)\right] dF(\theta) &\leq e_0 &(8)\\
\int_{\underline{\theta}}^{\bar{\theta}} \int_0^{\infty} c_1(\theta,y) g(y|k_1(\theta),\theta)dydF(\theta) &\leq \int_{\underline{\theta}}^{\bar{\theta}} \eta^{-1}\theta k_1(\theta)dF(\theta) &(9)\\
U(\theta) &= \log(c_0(\theta)) + \beta \int_0^{\infty} \log(c_1(\theta,y)) g(y|k_1(\theta),\theta)dy &\\
U'(\theta) &= \frac{1}{\theta}\frac{k_1(\theta)}{c_0(\theta)} &(10)\\
\beta \int_0^{\infty} \log(c_1(\theta,y)) g_k(y|k_1(\theta),\theta)dy &= \frac{1}{c_0(\theta)} &(11)
\end{aligned}$$

In what follows, we refer to (10) as the *adverse selection* constraint and to (11) as *moral hazard* constraint.

Unfortunately, the analytical characterization of saving wedge proves impossible with private information about θ. However, here we provide partial characterization by providing a modified inverse Euler Equation that will prove useful in characterizing taxes and wedges. We call this the Modified Inverse Euler Equation. We have the following proposition:

Proposition 2 ***(Modified Inverse Euler Equation).*** *Suppose that $c_t, k_1 > 0$, a.e. Then any solution to (P1) must satisfy*

$$k_1(\theta)\left(q\frac{\theta}{\eta} - 1\right) + \frac{q}{\beta}\int_0^{\bar{y}} \frac{1}{u'(c_1(\theta,y))} g(y|k_1(\theta),\theta)dy = \frac{1}{u'(c_0(\theta))} \tag{12}$$

where q is the relative intertemporal price of consumption, and $u(c) = \log c$.

The proof can be found in the appendix.

This equation extends the results in Rogerson (1985a) and Golosov et al. (2003) to the described environment. A key condition in deriving the IEE in Golosov et al. (2003) is the fact that marginal utility is observable by the planner. In general optimality of allocations implies that a perturbation of the allocations that keeps utility of all types unchanged must keep the cost unchanged. In particular, any such perturbation at any given period t should imply that

$$MC_t + MC_{t+1} = 0$$

where MC_t is the marginal cost of such perturbation. When marginal utility is observable, i.e., consumption is separable from the source of private information, a perturbation in consumption that keeps utility unchanged along every history does not change incentives – $\beta^t u(c_t(h^t)) + \beta^{t+1} u(c_{t+1}(h^t, h_{t+1}))$ is unchanged for all h_{t+1}. Since, the source of private information is separate from consumption and the utility from consumption has not changed, the perturbed allocation must be incentive compatible. This implies that the marginal cost of the perturbation at period t is given by $MC_t = \frac{1}{\beta^t u'(c_t)}$ while at period $t+1$, it is given by $MC_{t+1} = -q_{t+1} E\left[\frac{1}{\beta^{t+1} u'(c_{t+1})} | h^t\right]$ – with q_{t+1} being the relative shadow value of aggregate consumption. In our environment, however, consumption is non-separable from the source of private information. Hence, a perturbation in consumption alone will induce some agents to lie and breaks the incentive compatibility requirement. Therefore, there are incentive cost associated with such perturbations. The last terms in (12) capture these costs. Here, we give a heuristic derivation of (12).

Consider a proportional perturbation in $c_0(\theta), k_1(\theta)$ by $1+\delta$, and $\phi(\theta), \zeta(\theta)$ by $(1+\delta)^{-\frac{1}{\beta}}$. Note that this perturbation, does not change the incentive compatibility constraints nor the utility of type θ. To see this, consider the utility of type θ:

$$\begin{aligned} \log\left(c_0(\theta)(1+\delta)\right) + \beta \int_0^\infty \log\left(\left(\phi(\theta) + \zeta(\theta)\varepsilon\right)(1+\delta)^{-1/\beta}\right) h(\varepsilon)\, d\varepsilon &= \log c_0(\theta) + \beta \int_0^\infty \log\left(\phi(\theta) + \zeta(\theta)\varepsilon\right) h(\varepsilon)\, d\varepsilon \\ &= U(\theta) \end{aligned}$$

Furthermore,

$$\begin{aligned} \int_0^\infty u(\tilde{c}_1)\, g_k dy &= \frac{1}{\tilde{k}_1(\theta)} \int_0^\infty \log\left(\left(1+\delta\right)\left(\phi(\theta) + \zeta(\theta)\varepsilon\right)\right)(\eta\varepsilon - 1)\, h(\varepsilon)\, d\varepsilon \\ &= \frac{1}{1+\delta}\frac{1}{k_1(\theta)} \int_0^\infty \log\left(\left(\phi(\theta) + \zeta(\theta)\varepsilon\right)\right)(\eta\varepsilon - 1)\, h(\varepsilon)\, d\varepsilon \\ &= \frac{1}{1+\delta} \int_0^\infty u(c_1)\, g_k dy = \frac{1}{\tilde{c}_0(\theta)} \end{aligned}$$

and $\frac{k_1(\theta)}{c_0(\theta)}$ is unchanged as a result of the perturbation. This implies that the cost of such perturbation should be equal to its benefits. In terms of $t=0$ consumption, the benefit is given by

$$\left(q\frac{\theta}{\eta} - 1\right) k_1(\theta)\,\delta + q\left[1 - (1+\delta)^{-1/\beta}\right] \int_0^\infty c_1(\theta, y)\, g\left(y | k_1(\theta), \theta\right) dy$$

while the cost of the perturbation is $\delta c_0(\theta)$. Equating the cost and the benefit would result in equation (12).

Equation (12) is useful in partial characterization of the saving wedge. Note that when capital consumption ratio $\frac{k_1(\theta)}{c_0(\theta)}$ is increasing in θ, one can show that the saving wedge is increasing. Mechanically this is because both $\frac{k_1}{c_0}$ and $E\left[u'(c_1)\right] E\frac{1}{u'(c_1)}$ are both increasing

functions of $\frac{\zeta}{\gamma}$. Hence, we have the following proposition:

Proposition 3 *Suppose that in the solution to (P1), investment-consumption ratio, $\frac{k_1(\theta)}{c_0(\theta)}$, is increasing in θ. Then $\tau_S^{PI}(\theta)$ is increasing in θ.*

where τ_S^{PI} is the saving wedge implied by the solution to (P1). Intuitively, when the planner would like to achieve higher saving rate in the form of risky capital by higher productivity types, it needs to tax other forms of saving heavier, i.e., increasing $\tau_S(D)$.

Note also that as it is common in problems with hidden information, the local incentive constraint (10) does not bind at the extreme values of $\theta = \underline{\theta}, \bar{\theta}$. This implies that the distortion to saving is equal to those characterized in section 2.1. So we have the following corollary:

Corollary 1 *Consider the solution to (P1) and its implied saving wedge $\tau_S^{PI}(\theta)$. Then $\tau_S^{PI}(\theta) = \tau_S^{FI}(\theta)$ for $\theta = \underline{\theta}, \bar{\theta}$. Furthermore, $\tau_S^{PI}(\bar{\theta}) > \tau_S^{PI}(\underline{\theta})$.*

Numerical Example. Here, to shed more light on mechanisms at play, we provide a numerical example[10]. Figure 1, displays average consumption at $t = 1$ and how it moves with θ. Note that due to private information, average consumption in the second period cannot be equated across types. This is due to the fact that in the full information economy, higher productivity agents are bearing higher risk and hence have an incentive to pretend to be of lower type.

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 lower than the one with full information. This is because in the model with private information, the planner cannot achieve as high level of saving rate (in the risky capital) as in the model with full information. Hence, the motives for taxing bond holdings decreases and so does the saving wedge.

Finally, Figure 3 shows the slope of $y - c_1(\theta, y)$ which can be interpreted as the marginal tax rate on income from the business. This figure implies that although the slope of c_1 with respect to ε is increasing in ε, since average income, $\frac{\theta}{\eta}k(\theta)$, is increasing in θ and increases faster than $\zeta(\theta)$. Furthermore, comparison with $\tau_S(\theta)$ suggests that perhaps a progressive tax on total income can almost implement the optimal allocation.

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

[10]We 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 25%. The other parameter values are $q = \beta = 0.95$, 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.We will discuss this in more detail in section 3.

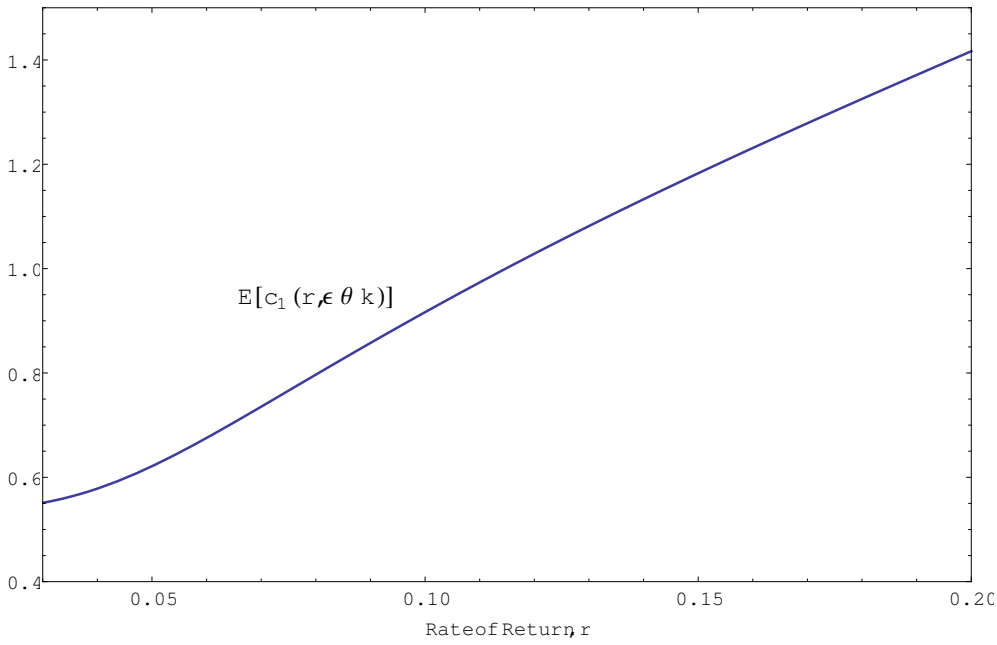


Figure 1: Average Consumption as a Function of rate of return

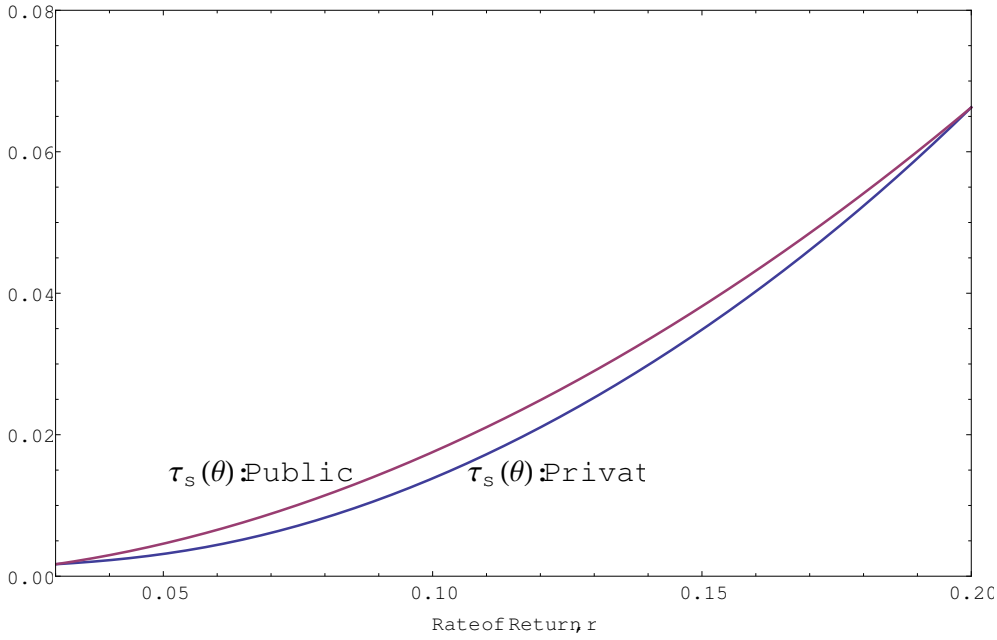


Figure 2: Saving Wedge with Public and Private Productivity

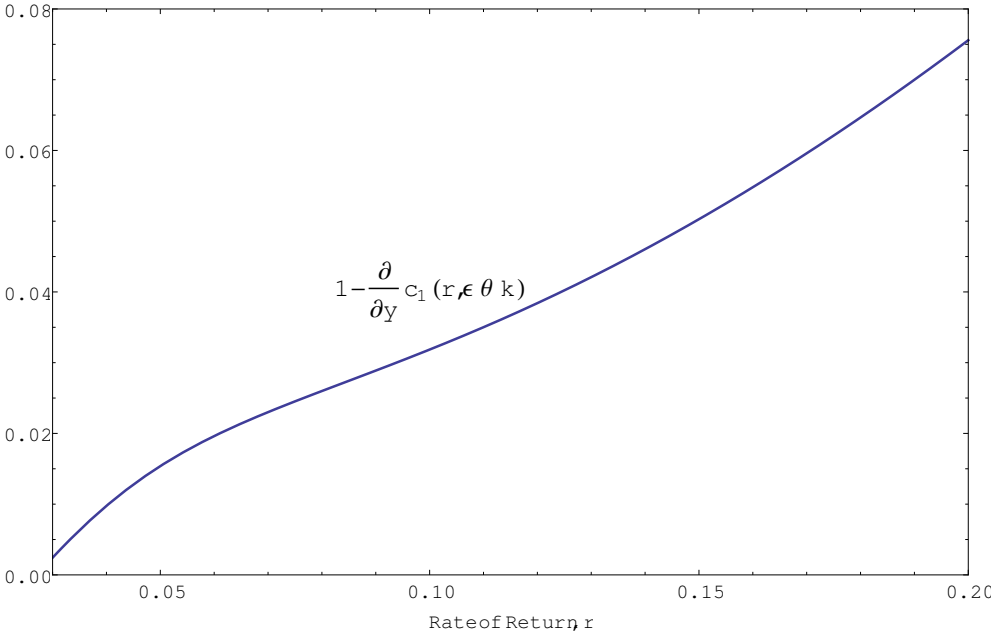


Figure 3: $1 - \frac{\partial}{\partial y} c_1(\theta, y)$

taxes should be consistent with dynamic wealth and its distribution in the society. In this section, I develop a fully dynamic model of optimal capital taxes based on the the two period economy described above. Later in section I discuss the implication of the model on optimal taxes and optimal distribution of wealth.

The basic model is an overlapping generations extension of the two period model. Time is discrete, $t = 0, 1, \cdots$. 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 $t + 1$-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 prodcutivity is drawn from a distribution $F(\theta)(\theta \in \Theta = [\underline{\theta}, \bar{\theta}])$ and ε_{it+1} is distributed according to an exponential distribution with parameter η. Throughout the paper, we refer to the distribution of ε as $H(\varepsilon)$ and call the distribution of y_{t+1} induced by k_{t+1} and l_{t+1} as $\hat{G}(y|\theta, k, l)$. 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.[11]

When old at period $t+1$, households collect revenue from the production technology, consume and leave bequests. Hence a household born at t, consumes $c_{0,t}$ at t and $c_{1,t+1}$ 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:

$$V_t = E_t \log c_{0,t} + \beta \log c_{1,t} + \beta^2 V_{t+2}$$

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

Given this environment, an allocation is given by

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

where

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

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

$$\begin{aligned} &\int_{\Theta^{\lceil (t+1)/2\rceil}\times\mathbb{R}_+^{\lceil t/2\rceil}} \left[c_{0,t}\left(\theta^t, y^{t-1}\right) + k_{t+1}\left(\theta^t, y^{t-1}\right)\right] d\mu_{0,t}\left(\theta^t, y^{t-1}\right) \\ &\quad + \int_{\Theta^{\lceil t/2\rceil}\times\mathbb{R}_+^{\lceil (t-1)/2\rceil}} c_{t-1,t}\left(\theta^{t-1}, y^t\right) d\mu_{1,t-1}\left(\theta^{t-1}, y^t\right) \\ &\quad \leq \int_{\Theta^{\lceil t/2\rceil}\times\mathbb{R}_+^{\lceil (t-1)/2\rceil}} y_t d\mu_{1,t-1}\left(\theta^{t-1}, y^t\right) \end{aligned} \tag{13}$$

$$\int_{\Theta^{\lceil t/2\rceil}\times\mathbb{R}_+^{\lceil (t-1)/2\rceil}} l_t\left(\theta^{t-1}, y^{t-1}\right) d\mu_{1,t-1}\left(\theta^{t-1}, y^t\right) = 1 \tag{14}$$

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. I assume that the agent privately observes θ_t,ε_{t+1} as well as investment and

[11] Note 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. This interpretation is preferred and throughout the paper, I use this interpretation of the model.

consumption. The planner can observe the output from the project as well as labor input and total transfers $\left\{\tau_{0,t}\left(\theta^t, y^{t-1}\right), \tau_{1,t}\left(\theta^t, y^{t+1}\right)\right\}$ – used for consumption and investment, to the households. By revelation principle we can focus on direct mechanisms. So given a transfer scheme $\boldsymbol{\tau}_t\left(\theta^{t-2}, y^{t-1}\right) = \left\{\tau_{\tau,\tau}\left(\theta^\tau, y^{\tau-1}\right), \tau_{\tau,\tau+1}\left(\theta^\tau, y^{\tau+1}\right), \theta^\tau \succcurlyeq \theta^{t-2}, y^{\tau-1} \succcurlyeq y^{t-1}\right\}_{\tau \geq t}$, each agent solves the following:

$$\begin{aligned} V_t\left(\boldsymbol{\tau}_t\left(\theta^{t-2}, y^{t-1}\right)\right) &= \max_{c_{0,t}, c_{1,t}, k_{t+1}, l_{t+1}, \hat{\theta}} \int_{\Theta \times \mathbb{R}_+} \left[\log c_{0,t} + \beta \log c_{1,t}\right. \\ &\qquad \left. + \beta^2 V_{t+2}\left(\boldsymbol{\tau}_{t+2}\left(\theta^{t-1}, \hat{\theta}, y^{t+1}\right)\right)\right] dF\left(\theta_t\right) d\hat{G}\left(y_{t+1} | \theta_t, k_{t+1}, l_{t+1}\right) \end{aligned}$$

subject to

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

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.

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

$$V_0 + \beta V_1$$

that is the planner cares about the welfare of the first two generations. This welfare criterion, although time-inconsistent, implies that within each date, the planner puts equal weight on all the agents alive.

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

Labor Demand. Before starting to characterize incentive efficient allocations, we show how one can solve for 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 14 to 13. Hence,

$$(1-\alpha)\,\hat{y}_t l_t^{-\alpha} = p_t \tag{15}$$

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

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

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

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

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

$$\begin{aligned} \frac{1}{c_{0,t}} &= \beta \int_{\mathbb{R}_+} \left[\log c_{1,t}\left(\theta_t, y_{t+1}\right) + \beta w_{t+1}\left(\theta_t, y_{t+1}\right)\right] g_k\left(y_t|\theta_t, k_{t+1}\right) dy_t \\ \frac{\partial}{\partial \theta_t} U_t\left(\theta^t, y^{t-1}\right) &= \frac{1}{\theta_t} \frac{k_{t+1}}{c_{0,t}} \end{aligned}$$

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

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

subject to

$$\begin{aligned} \int U(\theta)\, dF(\theta) &= w \\ \log c_0(\theta) + \beta \int \left[\log\left(c_1(\theta,y)\right) + \beta w'(\theta,y)\right] dG\left(y|\theta,k(\theta)\right) &= U(\theta) \\ U'(\theta) &= \frac{1}{\theta}\frac{k(\theta)}{c_0(\theta)} \qquad (16) \\ \beta \int_0^\infty \left[\log\left(c_1(\theta,y)\right) + \beta w'(\theta,y)\right] g_k\left(y|k(\theta),\theta\right) dy &= \frac{1}{c_0(\theta)} \qquad (17) \end{aligned}$$

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 (15). Furthermore, q_t, κ_t as well as w_0 and w_1 are determined so that (13) and (14) 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 (P) and discuss the main properties of the distortions. The first point to be noted in characterizing the solution to (P) is that the problem is homogenous in w. That is, we only needs to study properties of the optimal allocation for when $w = 0$. We have the following theorem:

Theorem 1 *Given $\{q_t, \kappa_t\}$, the value function and policy functions associated with ((P)) satisfy the following*

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

where A_t, c_{0t}, c_{1t}, k_t, $\hat{w}_t$ and U_t solve the following sequence of problems

$$\begin{aligned} B_t &= \max \int_\Theta \left[\frac{1}{\eta} \alpha q_t \kappa_{t+1} \theta k(\theta) - c_0(\theta) - k(\theta) \right. \\ &\qquad \left. + q_t \int_0^\infty \left[-c_1(\theta, y) + q_{t+1} B_{t+2} e^{w'(\theta,y)(1-\beta)} \right] dG(y|\theta, k(\theta)) \right] dF(\theta) \end{aligned} \tag{18}$$

subject to

$$\begin{aligned} \int U(\theta)\, dF(\theta) &= 0 \\ \log c_0(\theta) + \beta \int \left[\log(c_1(\theta, y)) + \beta w'(\theta, y) \right] dG(y|\theta, k(\theta)) &= U(\theta) \\ U'(\theta) &= \frac{1}{\theta} \frac{k(\theta)}{c_0(\theta)} \\ \beta \int_0^\infty \left[\log(c_1(\theta, y)) + \beta w'(\theta, y) \right] g_k(y|k(\theta), \theta)\, dy &= \frac{1}{c_0(\theta)} \end{aligned}$$

The above theorem, implies that stationary allocation solve problem (18) in its stationary form, i.e, when A_t, q_t are constant.

It is worth noting that (18) is very similar to (P1). The only additional instrument available to the planner to provide incentive is $w'(\theta, y)$. However, one can show that since the margin between $w'(\theta, y)$ and $c_1(\theta, y)$ is undistorted, the above problem can be written in a similar fashion to (P1) using a different discount factor $\hat{\beta}$ and intertemporal price $\hat{q}_t$. In particular, lemma **??** in the appendix shows that in the objective the last integral is replaced by

$$\frac{1}{1-\beta} \int -c_1(\theta, y)\, dG$$

and continuation utility for the old agent becomes

$$\frac{\beta}{1-\beta} \int \log c_1(\theta, y)\, dG.$$

We provide this formulation in the appendix. Given this formulation, one can show that similar to the two period example, $c_1(\theta, y)$ is linear in y.

What the above characterization suggests is that in the stationary version of the problem, distortions or wedges are independent of history w. When the problem is non-stationary, wedges depend on time but not on specific history. Hence, they inherit the correlation structure of the shocks, i.e., they are i.i.d.

Given that the dynamic problem is very similar to the problem discussed in section 2, similar properties are satisfied and we do not further characterize the saving wedge for the young, i.e., this is equivalent to τ_S in section twoperiod. However, an object of interest is

distortions to saving for the old. One can interpret these distortions as marginal tax rate on bequest. The following modified inverse Euler equation sheds more light on the forces in determining the sign of the bequest tax:

Theorem 2 *The solution to ((P)). Then optimal consumption must satisfy the following equation:*

$$\begin{aligned} \frac{q_{t+1}}{\beta} E_{t+1} \frac{1}{u'(c_{0,t+2})} &= \frac{1}{u'(c_{1,t})} + \frac{q_{t+1}}{\beta} E_t \left(\alpha q_{t+2} \kappa_{t+3} \frac{\theta_{t+2}}{\eta} - 1 \right) \\ \left(\alpha q_t \kappa_{t+1} \frac{\theta_t}{\eta} - 1 \right) + \frac{q_t}{\beta} E_t \frac{1}{u'(c_{1,t})} &= \frac{1}{u'(c_{0,t})} \end{aligned}$$

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, a 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. In Shourideh (2012), I show that when $c_{0,t}$ is increasing in θ_t and that incentive constraints are downward binding, i.e., the multiplier associated with (16) is positive, then the bequest tax is negative.

3.1 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 are given by

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

Furthermore, from labor demand, we must have

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

and hence aggregate labor demand is given by

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

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

$$\int e^{w(1-\beta)} d\Psi_{0,t+2}(w) = \int e^{\hat{w}(\theta,y)(1-\beta)} dG(y|\theta, k(\theta))\, dF(\theta) \int e^{w(1-\beta)} d\Psi_{0,t}(w)$$

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

$$\int e^{\hat{w}(\theta,y)(1-\beta)} dG(y|\theta, k(\theta))\, dF(\theta) = 1$$

One can show that in order for $\hat{w}(\theta, y)$ to be the solution to (18), we must have $\int e^{\hat{w}(\theta,y)(1-\beta)} dG(y|\theta, k(\theta))$ $\frac{\beta^2}{q^2}$[12]. Therefore, for stationarity, we must have $\beta = q$. This result is similar to a result in Farhi and Werning (2012). They show that in a standard Mirrleesian model, stationarity and log-utility implies that rate of interest has to be equal to the discount factor. Both results follow

[12]This is coming from the fact that $P'(w)$ is a martingale, i.e., $P'(w_t) = \frac{q^2}{\beta^2} E_t P'(w_{t+2})$.

from inverse Euler equation. Here one should note that

$$\frac{q^2}{\beta^2} E_t \frac{1}{u'(c_{0,t+2})} = \frac{1}{u'(c_{0,t})}$$

and hence, stationarity of consumption for the young implies that $q = \beta$.

Also note that similar to most model of dynamics with private information, the stationary distribution is trivial, almost all the agents will be at the lowest possible utility level, $-\infty$. To see this, note that $w_{t+2} = w_t + \hat{w}(\theta_t, \varepsilon_{t+1})$. Moreover, $Ee^{\hat{w}(\theta_t, \varepsilon_{t+1})} = 1$and therefore, by Jensen's inequality, $E\hat{w}(\theta_t, \varepsilon_{t+1}) < 0$. This means that in the long run, w_t converges to $-\infty$. So the so called immiseration result holds. As we will see in the next section, this property holds in a model with incomplete market. In the following section, we introduce a method to resolve this issue similar to Albanesi and Sleet (2006). Furthermore, we provide formulas for characterization of the tail of the wealth distribution.

4 Long Run Distribution of Wealth

In this section, I discuss how the model developed above can generate a long-run distribution of wealth with a Pareto tail and what an optimal Pareto tail looks like. I first illustrate, how an incomplete market version of the above model can generate a Pareto tail for the distribution of wealth. We then will illustrate how this tail is affected with optimal taxes.

4.1 An Incomplete Market Model

Our incomplete market version of the model is very similar to Angeletos (2007). Suppose that when young, households have two options for investment: invest in the risky project with production function $(A_{t+1}k_{t+1})^{\alpha} l_{t+1}^{1-\alpha}$, or to borrow and lend using a risk free bond. When old, households hire labor to produce output form a competitive labor market with wage ω_t. When old, the households leave bequest for their descendants. Given this market structure, the budget constraints for the young and old households are given by

$$\begin{aligned} c_{0,t} + k_{t+1} + b_{t+1} &= R_t a_t + \omega_t \\ c_{1,t} + a_{t+2} &= (\varepsilon_{t+1}\theta_t k_{t+1})^{\alpha} l_{t+1}^{1-\alpha} - \omega_{t+1} l_{t+1} + R_{t+1} b_{t+1} \end{aligned}$$

An equilibrium is defined as the solution to the following problem

$$V_t(a) = \max \int \left[\log c_0 + \beta \log c_1 + \beta^2 V_{t+2}(a')\right] dGdF \qquad \text{(P')}$$

subject to the budget constraints above as well as

$$a_t \geq -h_t = -\sum_{j=0}^{\infty} \frac{\omega_{t+2j}}{R_t \cdots R_{t+2j}}.$$

Note that the above borrowing constraint is a natural debt limit and h_t is the present value of labor income by future generations and it can be interpreted as human capital. Further R_t and ω_t are determined so that

$$\begin{aligned} \int b_{t+1}(a, \theta_t)\, dF(\theta_t)\, d\psi_{0,t}(a) + \int a d\psi_{1,t-1}(a) &= 0 \\ \int l_{t+1}(a, \theta_t, \varepsilon_{t+1})\, dH(\varepsilon_{t+1})\, dF(\theta_t)\, d\psi_{1,t}(a) &= 1 \end{aligned}$$

where $\psi_{0,t}$ and $\psi_{1,t-1}$ are the distributions of asset for the young and the old at period t. As before, profit maximization implies that

$$(1-\alpha)(\varepsilon_{t+1}\theta_t k_{t+1})^{\alpha} l_{t+1}^{-\alpha} = \omega_{t+1}$$

and hence

$$\begin{aligned} \pi_t = (\varepsilon_{t+1}\theta_t k_{t+1})^{\alpha} l_{t+1}^{1-\alpha} - \omega_{t+1} l_{t+1} &= \alpha \left(\frac{1-\alpha}{\omega_{t+1}}\right)^{\frac{1-\alpha}{\alpha}} \varepsilon_{t+1}\theta_t k_{t+1} \\ &= \hat{\kappa}_t \varepsilon_{t+1}\theta_t k_{t+1} \end{aligned}$$

Once labor demand is determined, the above problem is a classic portfolio problem studied by Samuelson (1969) . The utility function is homothetic while the budget set is linear in allocations. This means that the policy functions are linear in an appropriate state variable. Because of the existence of labor income, assets are not the state variable. However, we can show that if we define

$$\begin{aligned} \hat{a}_t &= a_t + h_t \\ \hat{b}_{t+1} &= b_t + \frac{h_{t+2}}{R_{t+1}} \end{aligned}$$

Then the budget constraints become

$$\begin{aligned} c_{0,t} + k_{t+1} + \hat{b}_{t+1} &= \hat{a}_t R_t \\ c_{1,t} + \hat{a}_{t+2} &= \pi_t + R_{t+1}\hat{b}_{t+1} \end{aligned}$$

where $\hat{a}_t$ and $\hat{b}_t$ are physical asset together with present value of future generations labor income – what can be interpreted as human capital. Given this definition, we have the following

theorem[13]:

Theorem 3 *The policy functions in (P'), satisfy the following*

$$\begin{aligned}
k_t(\theta, a) &= s_{k,t}(\theta, R_{t+1})\,\beta R_t(a + h_t) \\
b_{t+1}(\theta, a) &= s_{b,t}(\theta, R_{t+1})\,\beta R_t(a + h_t) - \frac{h_{t+2}}{R_{t+1}} \\
c_{0,t}(\theta, a) &= (1-\beta) R_t(a + h_t) \\
c_{1,t}(\varepsilon, \theta, a) &= (1-\beta)\beta\left(\hat{\kappa}_{t+1}\varepsilon\theta s_{k,t}(\theta, R_{t+1}) + R_{t+1} s_{b,t}(\theta, R_{t+1})\right)(a + h_t) \\
a_{t+2}(\varepsilon, \theta, a) &= \beta^2\left(\hat{\kappa}_{t+1}\varepsilon\theta s_{k,t}(\theta, R_{t+1}) + R_{t+1} s_{b,t}(\theta, R_{t+1})\right)(a + h_t) - h_t
\end{aligned}$$

where $s_{k,t}(\theta, R_{t+1}) + s_{b,t}(\theta, R_{t+1}) = 1$ *and*

$$\int_0^\infty \frac{\hat{\kappa}_{t+1}\varepsilon\theta - R_{t+1}}{s_{k,t}(\theta, R_{t+1})\left(\hat{\kappa}_{t+1}\theta\varepsilon - R_{t+1}\right) + R_{t+1}} dH(\varepsilon) = 0 \tag{19}$$

The above result is familiar from Samuelson (1969) as well more recently Angeletos (2007). With log utility, the total saving rate is β. Furthermore, the break-down between bond and equity is given by the portfolio choice equation (19). In the appendix, we show that $s_{k,t}(\theta, R)$ is increasing in θ.

As it can be seen in the above, total value of financial assets and human capital is a random process with i.i.d. growth. As shown in Champernowne (1953), Gabaix (1999), among others, such process deliver stationary distribution with Pareto tail. However, pure random growth usually does not deliver stationary distribution. In particular, for any random variable ζ_t and stochastic process X_t with $X_{t+1} = \zeta_{t+1} X_t$, $\log X_{t+1}$ is a random walk and hence its variance converges to ∞.Furthermore, $\log X_t$ converges to 0 or ∞ almost surely depending on whether $E \log \zeta$ is negative or positive.

In the above model, steady state implies that

$$\beta^2 \int \left(\hat{\kappa}\varepsilon\theta s_k(\theta, R) + R\left(1 - s_k(\theta, R)\right)\right) dF(\theta)\, dH(\varepsilon) = 1$$

and hence from Jensen's inequality

$$\int \log\left[\beta^2\left(\hat{\kappa}\varepsilon\theta s_k(\theta, R) + R\left(1 - s_k(\theta, R)\right)\right)\right] dF(\theta)\, dH(\varepsilon) < 0$$

This implies that $a_t + h_t$ converges almost surely to zero. That is since households can borrow against their descendants labor income, over time, they accumulate debt so that their financial wealth, i.e., bequest, is negative and equal to the negative of their human capital.

[13]The analysis here closely follows that of Angeletos (2007).

That households accumulate debt over time suggests that a simple borrowing constraint would fix this problem. In fact, it can be shown that if we impose $a_t \geq -\bar{a}$, where $\bar{a} > -h_t$, then a stationary distribution exists. Furthermore, a result by Mirek (2011) allows us to characterize what the tail of the distribution looks like. We present a slightly simplified version of the theorem here:

Theorem 4 *Consider a stochastic process $X_{t+1} = \psi(X_t, \lambda_t)$ with $X_0 = x$ where $\lambda_t \in \Lambda$ are i.i.d. according to some measure μ on the metric space Λ and $X_t \in \mathbb{R}_+$. Suppose that ψ satisfies the following conditions:*

1. *ψ is continuously differentiable with ψ_x bounded above and that for all λ, x, $\lim_{t\to 0} t\psi\left(t^{-1}x, \lambda\right)$ exists and is finite. Furthermore, suppose that $M(\lambda)$ exists such that*

$$\lim_{t\to 0} t\psi\left(t^{-1}x, \lambda\right) = M(\lambda)\, x,$$

2. *For all $\lambda \in \Lambda$, there is a random variable N_λ such that*

$$|\psi(x, \lambda) - M_\lambda x| \leq |N_\lambda|, \forall x \in \mathbb{R}_+$$

3. *$\log|M_\lambda|$ does not have support of the form $r\mathbb{Z}$,*

4. *There exists some $\alpha > 0$, such that*

$$E\left(|M_\lambda|^\alpha\right) = 1$$

5. *M satisfies $E\left(|M|^\alpha \log|M|\right) < \infty$,*

6. *For the random variable N defined above, $E\left(|N|^\alpha\right) < \infty$.*

Theorem 5 *Then, there exists a unique random variable X with measure ν exists such that*

$$\lim_{t\to\infty} t^\alpha \Pr(\{|X| > t\}) = \xi$$

where ξ is non-zero if one of the following two conditions are satisfied:

$$\lim_{s\to s_\infty} \frac{E\left(|N|^s\right)}{E\left(|M|^s\right)} = 0 \textit{(for } s_\infty < \infty\textit{) or } \lim_{s\to\infty} \left(\frac{E\left(|N|^s\right)}{E\left(|M|^s\right)}\right)^{\frac{1}{s}} < \infty$$

and X_t converges to X in probability.

Condition 1 are worth discussing. Applying the L'hôpital's rule implies that

$$\lim_{t\to 0} t\psi\left(t^{-1}x, \lambda\right) = x\psi_x(\infty, \lambda)$$

Hence $M_\lambda = \psi_x(\infty, \lambda)$. Hence, the determinant of the tail of the stationary distribution is the slope of $\psi(x, \lambda)$ with respect to x.

Applying this theorem to the incomplete market model with borrowing constraint, it implies that as long as $\frac{\partial}{\partial a}a'(\varepsilon, \theta, a)$ is bounded above and $B(\varepsilon, \theta)$ exists such that

$$\left| a'(\varepsilon, \theta, a) - \frac{\partial}{\partial a}a'(\varepsilon, \theta, \infty)\, a \right| \leq B(\varepsilon, \theta)$$

and if $\frac{\partial}{\partial a}a'(\varepsilon, \theta, \infty)$ and $B(\varepsilon, \theta)$ satisfy all the assumptions above, then a unique stationary distribution exists with pareto tail parameter given by ν where α satisfies

$$\int_{\Theta \times \mathbb{R}_+} \left(\frac{\partial}{\partial a}a'(\varepsilon, \theta, \infty) \right)^{\nu} dF(\theta)\, dH(\varepsilon) = 1$$

Note that $\frac{\partial}{\partial a}a'(\varepsilon, \theta, \infty)$ is the generational saving rate when wealth converges to infinity. If one is to assume that the ex-ante probability of binding borrowing constraint vanishes as wealth converges to infinity, $\frac{\partial}{\partial a}a'(\varepsilon, \theta, \infty)$, is associated with an unconstrained agent's saving rate. Now if we suppose that the value function $V(\hat{a})$ is proportional to $\log \hat{a}$ for large values of $\hat{a}$, the analysis is identical to the above. Unfortunately, we have to make these assumptions in order to proceed with our characterization. Most numerical simulations suggest that they are reasonable assumptions[14].

Assumption 2 *The value function associated with incomplete market model with borrowing constraint has the following properties:*

1. $\lim_{\hat{a}\to\infty} \hat{a}V'(\hat{a}) = \frac{1}{1-\beta}$.
2. $\lim_{\hat{a}\to\infty} \Pr(\{(\varepsilon, \theta) : \hat{a}'(\varepsilon, \theta, \hat{a}) = -\bar{a} + h\}) = 0$.

Under the above assumption, we can show the following:

Theorem 6 *Suppose that 2 holds and that R_t and ω_t are constant over time. Then we have*

$$\lim_{\hat{a}\to\infty} \frac{\partial}{\partial \hat{a}}\hat{a}'(\varepsilon, \theta, \hat{a}) = \beta^2 \left(s_k(\theta, R)\left(\hat{\kappa}\theta\varepsilon - R\right) + R \right)$$

where $s_k(\theta, R)$ satisfies

$$\int_0^\infty \frac{\kappa\varepsilon\theta - R}{(\kappa\varepsilon\theta - R)\, s_k + R} dH(\varepsilon) = 0$$

[14] One of way of dealing with this problem is a la Benhabib et al. (2011). They assume that each generation gets a utility of the form $\log a_{t+2}$ while this is not the utility that their descendants receive. Their model is slightly different in that it includes labor income to get around the the convergence problem. But the basic insight remains the same.

Now, if we assume that $\int_0^\infty \log\left(\beta^2\left(\left(\kappa\varepsilon\theta - R\right)s_k + R\right)\right) dFdG < 0$, the the above theorem suggest that the stationary distribution of wealth is given by α such that

$$\int_0^\infty \left[\beta^2\left(\left(\kappa\varepsilon\theta - R\right)s_k + R\right)\right]^\alpha dF\left(\theta\right) dH\left(\varepsilon\right) = 1 \qquad (\bigstar)$$

The formula ($\bigstar$) is useful in characterizing the tail of the income distribution. Later we use this to compare the implied tail of the distribution by incomplete market to that of constrained efficient allocation.

4.2 Efficient Allocations and Stationary Distribution

In this section, we use 4's powerful result to characterize the tail of the wealth distribution implied by optimal taxes. We use a similar technique as above to characterize the efficient distribution of wealth at the top.

Note that given the above characterization of the policy functions in theorem 1, the cost of delivering promised utility, $P\left(w_t\right)$ is a stochastic process with i.i.d. random growth. That is[15]

$$P\left(w_{t+2}\right) = P\left(w_t\right) e^{\hat{w}\left(\theta_t, y_{t+1}\right)}$$

Note that $P\left(w_t\right)$ is the present value of income less consumption and investment expenditure. While this can be thought of as wealth, in an environment with taxes, wealth is the present value of after tax income less expenditure. Hence, in order to define wealth in the constrained efficient allocation, we have to define taxes.

Taxes. The tax system that we use to implement the optimal allocation is consisted of a tax function of the following form:

$$T\left(\frac{y}{a+h}, \frac{b+h}{h}\right) a$$

where a is financial wealth, b is the saving in risk free bond across periods, and h is human capital. Given this, each generation's budget constrains are given by

$$\begin{aligned} c_{0,t} + k_{t+1} + b_{t+1} &= Ra_t + \omega \\ c_{1,t} + a_{t+2} &= Rb_{t+1} + y_t - T\left(\frac{y}{a+h}, \frac{b+h}{a+h}\right)(a+h) \end{aligned}$$

The following theorem establishes that we can implement a stationary constrained efficient allocation, using the above tax system:

Theorem 7 *Consider a stationary constrained efficient allocation associated. Then one can find T to implements the optimal allocation with $R = \beta$.*

[15] We have assumed stationarity and hence the value function is time-independent.

The idea behind this implementation is simple, the dependence of T on b makes sure that productivity types are revealed while the dependence on y provides incentive for efficient investment. Furthermore, the planning problem is homogenous in $a+h$ and hence a_t+h is a process with random growth rate. In the appendix, we show that when $q=\beta$, $a_t+h=-\beta P(w_t)$. Hence the growth rate of a_t+h is given by $e^{\hat{w}(\theta,y)(1-\beta)}$.

Evidently, the model described above does not generate a stationary distribution of wealth. However, it is helpful in guiding us toward possible modifications of the model to achieve stationarity. In fact, the modification is very similar to the one in 4.1. We assume that there exists a lower bound $\underline{w}>-\infty$ on the set of promised utilities. This is a similar technique as in Atkeson and Lucas (1995). What this constraint implies is that the lower bound on promised utility is not an absorbing state any more. Note that when $\underline{w}=-\infty$, the lower bound is an absorbing state and therefore, a downward drift in promised utility drives everyone's utility to the lower bound. With a finite $\underline{w}$, however, due to spreading of promise utility, $\underline{w}$ is not an absorbing state anymore and hence there exists a non-trivial stationary distribution.

In presence of the lower bound, the component planning problem together with the lower bound constraint is given by:

$$\begin{aligned} P_c(w) &= \max \int_\Theta \left[\frac{1}{\eta}\alpha q\kappa\theta k(\theta) - c_0(\theta) - k(\theta) \right. \\ &\qquad \left. + q\int_0^\infty \left[-c_1(\theta,y) + qP_c\left(w'(\theta,y)\right)\right] dG\left(y|\theta,k(\theta)\right)\right] dF(\theta) \end{aligned}$$

subject to

$$\begin{aligned} \int U(\theta)\, dF(\theta) &= w \\ \log c_0(\theta) + \beta\int \left[\log\left(c_1(\theta,y)\right) + \beta w'(\theta,y)\right] dG\left(y|\theta,k(\theta)\right) &= U(\theta) \\ U'(\theta) &= \frac{1}{\theta}\frac{k(\theta)}{c_0(\theta)} \\ \beta\int_0^\infty \left[\log\left(c_1(\theta,y)\right) + \beta w'(\theta,y)\right] g_k\left(y|k(\theta),\theta\right) dy &= \frac{1}{c_0(\theta)} \\ w'(\theta,y) &\geq \underline{w}. \end{aligned}$$

Notice that the problem is not homogeneous in w anymore. However, the following assumption helps us in applying theorem 4.

Assumption 3 *The value function $P_c(w)$ with the constraint $w\geq\underline{w}$, has the following property:*

$$\begin{aligned} \lim_{w\to\infty} P_c'(w)(1-\beta)B^{-1}e^{-w(1-\beta)} &= 1 \\ \lim_{w\to\infty} \Pr\left\{(\theta,y): w'(\theta,y,w)=\underline{w}\right\} &= 0 \end{aligned}$$

Under the above condition, the following theorem can be shown:

Theorem 8 *Suppose 3 holds. Then we must have*

$$\lim_{w\to\infty}\frac{\partial}{\partial w}e^{(1-\beta)w'(\theta,y,w)}=e^{(1-\beta)\hat{w}(\theta,y)}$$

where $\hat{w}(\theta,y)$ is the solution to the constrained efficient problem (18).

Proof can be found in the appendix.

For the case with lower bound constraint, we can show that the tax implementation takes a form of a tax function $T(y,b+h,a+h)$ paid by the old. Furthermore in our implementation as before, $a_t+h=-qP(w_t)$. Hence, a straightforward application of 4 implies that the tail of the distribution is given by $\hat{\nu}$ where ν satisfies

$$\int e^{\hat{\nu}\hat{w}(\theta,y)(1-\beta)}dG(y|k(\theta),\theta)\,dF(\theta)=1 \qquad (\bigstar\bigstar)$$

Formula ($\bigstar\bigstar$) is the formula that determines the tail of the optimal distribution of wealth.

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 personal income. I have also shown how the model can be used to study efficient distribution of wealth.

Although, I have interpreted the agents in the model as entrepreneurs 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.

References

Aiyagari, S. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 659–684. 2

Albanesi, S. (2006): "Optimal taxation of entrepreneurial capital with private information," *NBER working paper*. 4

Albanesi, S. and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73, 1–30. 24

Angeletos, G. (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30. 24, 26

Atkeson, A. and R. Lucas (1995): "Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance," *Journal of Economic Theory*, 66, 64–88.

Benhabib, J. and A. Bisin (2009): "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *NBER Working Paper*. 2

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. 3, 4, 28

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. 3

Chamley, C. (1986): "Optimal taxation of capital income in general equilibrium with infinite lives," *Econometrica*, 54, 607–622. 4

Champernowne, D. (1953): "A Model of Income Distribution," *The Economic Journal*, 63, 318–351. 4, 26

Conesa, J., S. Kitao, and D. Krueger (2009): "Taxing Capital? Not a Bad Idea After All!" *The American Economic Review*, 99, 25–48. 4

Evans, D. and B. Jovanovic (1989): "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *The Journal of Political Economy*, 97. 2

Farhi, E. and I. Werning (2008): "The Political Economy of Nonlinear Capital Taxation," Mimeo. 4

——— (2010a): "Insurance and Taxation over the Life Cycle," Working Paper, MIT. 3

——— (2010b): "Progressive Estate Taxation," *Quarterly Journal of Economics*, 125, 635–673. 4

——— (2012): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 120, 398–445. 23

Fong, K. (2009): "Evaluating Skilled Experts: Optimal Scoring Rules for Surgeons," Working Paper. 5

Gabaix, X. (1999): "Zipf's Law for Cities: an Explanation," *The Quarterly Journal of Economics*, 114, 739–767. 4, 26

——— (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294. 4

Garrett, D. and A. Pavan (2010): "Dynamic Managerial Compensation: On the Optimality of Seniority-based Schemes," Mimeo. 5

Gentry, W. and G. Hubbard (2004): "Entrepreneurship and Household Saving," *Advances in Economic Analysis and Policy*, 4. 2

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587. 3, 12

Golosov, M., M. Troshkin, and A. Tsyvinski (2010): "Optimal Dynamic Taxes," . 3

Guesnerie, R. and J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369. 11

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 1177–1190. 6, 7

Judd, K. (1985): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28, 59–83. 4

Kitao, S. (2008): "Entrepreneurship, Taxation and Capital Investment," *Review of Economic Dynamics*, 11, 44–69. 3

Kocherlakota, N. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73, 1587–1621. 4

Laffont, J. and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *The Journal of Political Economy*, 94, 614–641. 5

Mirek, M. (2011): "Heavy Tail Phenomenon and Convergence to Stable Laws for Iterated Lipschitz Maps," *Probability Theory and Related Fields*, 151, 705–734. 4, 27

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208. 2

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73. 11

Panousi, V. (2009): "Capital Taxation with Entrepreneurial Risk," Working Paper. 3

Pavan, A., I. Segal, and J. Toikka (2009): "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure," Mimeo. 4, 11

Piketty, T. and E. Saez (2012): "A Theory of Optimal Capital Taxation," NBER Working Paper. 3

Quadrini, V. (2000): "Entrepreneurship, Saving, and Social Mobility," *Review of Economic Dynamics*, 3, 1–40. 2

Rogerson, W. P. (1985a): "Repeated Moral Hazard," *Econometrica*, 53, 69–76. 12

——— (1985b): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357–1367. 6, 7

Samuelson, P. (1969): "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *The Review of Economics and Statistics*, 51, 239–246.

Scheuer, F. (2010): "Entrepreneurial Taxation, Occupational Choice, and Credit Market Frictions," Mimeo. 4

Shourideh, A. (2012): "Essays on Dynamic Macroeconomic Policy," University of Minnesota Dissertation. 22

Simon, H. (1955): "On a Class of Skew Distribution Functions," *Biometrika*, 42, 425–440. 4

Werning, I. (2010): "Nonlinear Capital Taxation," Working Paper, MIT. 4