Optimal Taxation of Entrepreneurial Income:
A Mirrleesian Approach to Capital Accumulation *

Job Market Paper

Ali Shourideh
University of Minnesota
shour004@umn.edu

January 25, 2011

Abstract

In this paper, I study how entrepreneurial income should be taxed. I develop a model in which entrepreneurs are subject to idiosyncratic shocks to their capital income. Shocks to capital income have two components: 1) a component that is known to the entrepreneur at the time of investment, 2) a residual component that is realized after investment. This creates two types of incentive problems: a hidden type problem and a hidden action problem. I show that, absent private markets for insurance of idiosyncratic risk, entrepreneurial and non-entrepreneurial capital income should be taxed differently. Moreover, the government should subsidize non-entrepreneurial capital income when the known component is at its highest and lowest value. Furthermore, for a wide variety of distributions, the optimal tax schedule is progressive with respect to entrepreneurial capital income. Finally, the results regarding taxation of entrepreneurial income depend on the extent to which incentives and insurance are provided by private contracts. In particular, private contracts can approximately implement the efficient allocation if convertible securities are available. The prevalence of these securities in venture capital contracts suggest that the forces identified here are important in practice.

*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, Ben Moll, Chris Phelan, Jose-Victor Rios-Rull, Maryam Saeedi, Erick Sager, Sam Shulhofer-Wohl, Maxim Troshkin, and specially Ariel Zetlin-Jones as well as seminar participants in 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 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 entrepreneurs 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 ask two sets of questions: First, when entrepreneurs cannot insure themselves against idiosyncratic productivity risk, how should the government design the tax schedule? In particular, how should the government tax capital income of entrepreneurs from their businesses and non-entrepreneurial capital income, i.e., financial wealth. Second, if we do not restrict private agents to a particular set of contracts but rather, allow them to sign insurance contracts, can they achieve efficiency? If so, can a set of standard securities implement the optimal allocation? Do we observe these contracts for entrepreneurs?

Regarding the first set of questions, I have two main theoretical results. First, existence of heterogeneity in entrepreneurial ability introduces forces toward subsidies to non-entrepreneurial capital income, i.e., financial wealth. In particular, in an extreme case where there is no residual component, wealth taxes are \textit{negative} for entrepreneurs with lowest and highest ability. When both components are significant, the results are mixed. Using a calibrated version of the model, I find that wealth taxes are negative for entrepreneurs with lowest ability and positive for entrepreneurs with highest ability. Moreover, I show that when the \textit{residual} component of productivity is significant, the tax schedule with respect to business income is \textit{progressive} for a wide variety of

\(^1\)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).
specification for the distribution of shocks.

As for the second set of questions, in this environment, private agents can achieve constrained efficiency when they can sign an unrestricted set of contracts. My main contribution, here, is to show that the optimal allocation can be implemented with a set of standard securities. In fact, one can reinterpret the model as a contract between an entrepreneur and a venture capitalist, with the optimal contract implemented using equity, convertible debt and a credit line/saving account with a variable interest rate.

To derive these results, I first study the properties of the constrained efficient allocations over time and in the cross-section. Using a first order approach, I derive a modified version of the Inverse Euler Equation (see Rogerson (1985a), Golosov et al. (2003)). I use this equation to characterize the optimal distortions to intertemporal margin of saving, i.e., the intertemporal wedge, and hence the marginal tax rate on wealth. In particular, the intertemporal wedge is the highest when current incentive constraints are very tight relative to future constraints and vice versa. Unfortunately, in this environment, our modified version of the Inverse Euler Equation cannot be used to determine the sign of the intertemporal wedge. However, the recursive formulation of the problem can be used to see that when there is no residual productivity shock, the intertemporal wedge is negative for the highest and lowest realizations of productivity.

To provide an intuition for the negative intertemporal wedge result, I describe how different forces are at play when capital income is risky as opposed to a situation where labor income is risky, as is typical in the dynamic Mirrlees literature. When labor income is risky, an extra unit of saving decreases marginal utility of agents in the future and decreases their labor supply, i.e., it tightens incentive constraints in the future. Hence, a planner wants to discourage agents from saving in order to provide incentives for working in the future. When capital income is risky, it is the opposite. An extra unit of saving causes agents to invest more since they have more resources available for consumption and investment. Hence, saving relaxes future incentive constraints. However, saving is not without cost. In fact, it tightens the incentive constraints in the current period. Since the incentive constraints are not binding for the highest and lowest value of productivity, when current productivity is at its highest and lowest value, an extra unit of saving has no effect on current incentives. Hence, the planner wants to encourage saving for the most and least productive agents.

To prove the progressivity result, I characterize the properties of consumption in the cross-section by deriving a simple equation that relates consumption to income. When the utility function is of the CARA form, this equation implies that the inverse of marginal utility is a linear function of the hazard ratio implied by the distribution of shocks to income. For a large class of distributions for the residual component of productivity it can be shown that the hazard ratio is concave in the income realization. Concavity of the hazard ratio implies that the consumption schedule is concave in income and thus the tax schedule with respect to business income is
Although one interpretation of the model is of optimal taxation, I show that it is not necessary for the government to tax entrepreneurs in order to achieve efficiency. In particular, there is a private implementation of the optimal contract using standard securities: equity, convertible debt and a credit line/saving account with a variable interest rate. The role of each security can be associated with the properties of the constrained efficient allocation described above. The presence of convertible debt – a security that is similar to debt but can be converted to equity at a pre-specified price – implies that entrepreneur’s consumption is a concave function of income. Hence, this feature can create the relationship between consumption and income in the constrained efficient allocation. The credit line/saving account with a variable interest introduces an intertemporal wedge in the saving margin of the entrepreneur as in the constrained efficient allocation. The significance of this implementation is that it resembles venture capital contracts. In fact, as noted by Kaplan and Strömberg (2003), Sahlman (1990) and Gompers (1999), a major fraction of securities used in venture capital contracts are in the form of convertible securities, i.e., convertible preferred stock, participating preferred stock, etc. Hence, this implementation sheds light on forces behind the widespread use of convertible securities in venture capital contracts. Moreover, it provides a justification for the forces identified in the model.

In deriving the above results, two implicit assumptions have been made. First, the economy is populated only by entrepreneurs. This feature, however, is not critical regarding the distortions implied by taxes. In particular, it is easy to extend this environment to an environment in which workers and entrepreneurs are distinguishable. In that environment, since a planner can distinguish between workers and entrepreneurs, the efficient allocations can be achieved by a lump sum transfer from entrepreneurs to workers along with taxes/private contracts to achieve efficiency within each group. Second, there is no entry into entrepreneurship. Adding this feature would make the model less tractable thereby making the main forces in the model harder to identify. I leave this extension for future work.

The theoretical results in this paper point to a need for an important empirical question: How successful are credit markets in providing efficient investment incentives for entrepreneurs? As the analysis in this paper shows, the optimal design of non-linear taxes for entrepreneurs depends on the answer to this question. As I have shown, contracts with features similar to venture capital can achieve efficiency. However, since venture capital is a small portion of private equity market, a more rigorous analysis of the credit market contracts is needed to answer this question.

**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

---

3In the two period environment, this result is more general. It holds whenever, \(1/u'(c)\), is a convex function of \(c\). In the special case where \(u(c) = c^{1-\sigma}/(1 - \sigma)\), we must have \(\sigma > 1\).

4See Scheuer (2010) for an analysis of the entry decision in a static economy.
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 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 (2006) and Scheuer (2010) are early attempts in studying optimal design of tax system for entrepreneurs. Albanesi (2006) focuses on specific implementation of optimal contracts and Scheuer (2010) focuses on the decision of entry into entrepreneurship and its implication for differential treatment of entrepreneurs and workers.

An important implication of my paper is the emergence of wealth 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. 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. To my knowledge, this is the first paper that identifies this force.

In deriving optimal progressivity of the tax code with respect to business income, my paper is related to a small number of papers that study optimal progressivity of the tax system (Varian (1980), and Heathcote et al. (2010)). The most related paper is perhaps Varian (1980). In a two period model that shares similar properties to our model, he shows that it is optimal for the government to make the marginal tax rate an increasing function of income. The model in this paper nests his model and extends it to a dynamic environment with productivity risk. Moreover, I show that the progressivity result holds for a large class of distributions.

In this paper, I show that the constrained optimal allocation can be implemented using a set of standard securities that are widely used in venture capital contracts. This result is related to the literature on optimal firm financing and optimal capital structure. DeMarzo and Fishman (2007) and DeMarzo and Sannikov (2006) show that in a dynamic model with non-verifiable income the optimal contract can be implemented using credit lines, equity and debt. Biais et al. (2007) show

\footnote{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).}
that in the same environment the optimal allocation can be implemented using cash reserves, debt and equity and use this implementation to study its implication for dynamics of security prices. Finally, Clementi and Hopenhayn (2006) consider a moral hazard model and show that the optimal allocations can be implemented using short term debt and equity. The implementation in this paper points to the special role of convertible securities, equity buy backs and credit lines in creating the right incentives for the entrepreneur to invest optimally.

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 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 rest of the paper is organized as follows: section 2 describes a two period version of the model in order to identify the key economic forces at play. In section 3, we develop the multi-period model and derive the modified inverse Euler Equation. In section 4, we study the intertemporal wedge. Section 5, generalizes the shape of the tax function

## 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 later extend it to the general environment.

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 \( \theta \) 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, \( \theta \), determines the distribution of returns to individual investment. If an entrepreneur with type \( \theta \) invests \( k_1 \) in his private project, the project will yield an output of \( y \in [0, \bar{y}] \) (\( \bar{y} \in \mathbb{R}^+ \cup \{\infty\} \)) that is distributed according to the c.d.f. function \( G(y|k_1, \theta) \) where \( G(\cdot|\cdot, \cdot) \) is \( C^1 \) in all of its argument. Moreover, the mean value of \( y \), given \( \theta, k_1 \) is given by \( (\theta k_1)^\alpha \), i.e., \( \int_0^{\bar{y}} yg(y|k_1, \theta) dy = (\theta k_1)^\alpha \) with \( \alpha \in (0, 1) \). In other words, a more productive entrepreneur has a higher total output as well as higher marginal product of capital. This formulation of the production function is similar to Lucas (1978) and Evans and Jovanovic (1989). Notice this formulation can stand-in for a more general constant return to scale production function that
employs labor, capital and managerial effort with labor being supplied competitively in the labor market and where managerial effort is inelastically supplied. The decreasing returns to scale assumption implies that in any socially optimal allocation, there should be investment in projects of all productivities. For tractability, I assume that capital fully depreciates over time.

In addition, in order to make the analysis easier and in accordance with the rest of moral hazard literature – see Jewitt (1988) and Rogerson (1985b), we assume that \( g(y|k_1, \theta) \) satisfies the Monotone Likelihood Ratio Property (MLRP):

\[
\frac{\partial}{\partial y} g_k(y|k_1, \theta) \frac{g(y|k_1, \theta)}{g(y|k_1, \theta)} > 0
\]

This assumption is necessary in order for the validity of the first order approach in characterizing incentive compatible allocations. I further assume that \( G(y|k_1, \theta) \) has the following property

\[
G(y|k, \theta) = G(y|\theta k, \theta'), \forall y \in [0, \bar{y}]
\]

or a function \( \tilde{G}(y, \cdot) \) must exists such that \( G(y|k_1, \theta) = \tilde{G}(y, \theta k_1) \). In words, a type \( \theta' \) has the ability to replicate the distribution of output of a type \( \theta \) by investing \( \theta k \). Additionally, the distribution \( G(y|k_1, \theta) \) is an increasing a function of \( k_1 \) and \( \theta \) w.r.t. stochastic first order dominance ordering.

Note that the above formulation of entrepreneurial investment technology is compatible with the literature on entrepreneurial behavior as in Evans and Jovanovic (1989) and Gentry and Hubbard (2004). In particular, they assume that output is given by \( \varepsilon \theta^\alpha k_1^\alpha \) where \( \log \varepsilon \sim N(-\frac{1}{2}\sigma^2_\varepsilon, \sigma^2_\varepsilon) \). This is essentially a special case of the above formulation where \( \bar{y} = \infty \) and \( G(y|k_1, \theta) = \Phi \left( \frac{\log y - \alpha \log \theta k_1 + 1/2\sigma^2_\varepsilon}{\sigma_\varepsilon} \right) \).

In addition, entrepreneurs preferences are standard and given by

\[
u(c_0) + \beta u(c_1)
\]

where \( c_0 \) and \( c_1 \) are consumption of the entrepreneur at each period, where \( u(\cdot) \) is a strictly concave and smooth function satisfying \( u'(0) = \infty \). Entrepreneurs, therefore, consume in each period and invest at \( t = 0 \). We assume for simplicity that each agent is endowed with \( e_0 \) at \( t = 0 \).

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

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

(2)

$$
\int_{\theta}^{\bar{\theta}} \int_{0}^{\bar{y}} c_1(\theta, y) g(y|k_1(\theta), \theta) dy dF(\theta) \leq \int_{\theta}^{\bar{\theta}} \theta^{\alpha} k_1(\theta)^{\alpha-1} dF(\theta) 
$$

(3)

**Efficient Allocations with Full Information.** It is useful to characterize efficient allocations when a planner can observe entrepreneurs’ project type, \( \theta \), as well as their consumption and investment. In such efficient allocations, the planner will equate returns to investment across all types of projects:

$$
\alpha \theta^{\alpha} k_1(\theta)^{\alpha-1} = \alpha \theta^{\alpha} k_1(\theta')^{\alpha-1} = \frac{1}{q}
$$

where \( q \) is the shadow value of consumption at \( t = 1 \) in terms of consumption at \( t = 0 \); formally, \( q \) is the lagrange multiplier on (3) divided by the one on (2). Moreover, if we consider a utilitarian planner that maximizes entrepreneurs’ ex-ante utility before realization of the shock, the efficient allocation must satisfy:

$$
c_0(\theta) = c_0(\theta') = c_0
$$

$$
c_1(\theta, y) = c_1(\theta', y') = c_1
$$

$$
u'(c_0) = \beta q^{-1} u'(c_1) = \beta \alpha \theta^{\alpha} k_1(\theta)^{\alpha-1} u'(c_1)
$$

The first two equations are implied by full risk sharing across types and the third is an Euler Equation for each individual. Hence, with full information, efficiency implies that the rate of return to individual investment should be equated across individuals. It follows that entrepreneurs with higher productivity should invest more than entrepreneurs with lower productivity. Next, I argue that an important assumption for this result is the observability of investment and consumption.

**Private Information.** 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:

$$
u(c_0(\theta)) + \beta \int_{0}^{\bar{y}} u(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy 
\geq \max_{\hat{\theta}, \hat{k}} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_{0}^{\bar{y}} u(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy 
$$

(4)

The RHS of the above inequality is the utility that a type \( \theta \) receives when he reports \( \hat{\theta} \) and invests
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 \( \theta \), have incentive to lie downward about their productivity type even if they invest "the right" amount. By lying downward and investing \( \frac{\partial k_1(\theta)}{\partial \theta} \), 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:

$$
\max_{c_0(\theta), c_1(\theta, y), k_1(\theta)} \int_{\theta}^{\hat{\theta}} \left[ u(c_0(\theta)) + \beta \int_{0}^{\hat{y}} u(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy \right] dF(\theta)
$$

subject to (2), (3), 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, 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 \( \theta \) from truth-telling. Then we must have

$$
U(\theta) = \max_{\theta, k} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_{0}^{\hat{y}} u(c_1(\theta, y)) g(y|\hat{k}, \theta) dy
$$

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

$$
u'(c_0(\theta)) = \beta \int_{0}^{\hat{y}} u(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy$$
$$
u'(c_0(\theta)) \left[ c_0'(\theta) + k_1'(\theta) \right] + \beta \int_{0}^{\hat{y}} u'(c_1(\theta, y)) c_1(y, \theta) g(y|k_1(\theta), \theta) dy = 0
$$

The Envelope condition associated with this problem is given by

$$
U'(\theta) = \frac{\partial}{\partial \theta} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_{0}^{\hat{y}} u(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy \bigg|_{\theta=\theta, k=k_1(\theta)} = \beta \int_{0}^{\hat{y}} u(c_1(\theta, y)) g(y|\hat{k}, \theta) dy
$$

Note that since \( g(y|k_1, \theta) \) is a function of \( \theta k_1 \), 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

\[ U'(\theta) = \frac{1}{\theta} k_1(\theta) u'(c_0(\theta)) \] (7)

We say an allocation is \textit{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. As for the moral hazard problem, there is a series of papers giving providing assumption on fundamentals for validity of the FOA – see Mirrlees (1999), Rogerson (1985b), Jewitt (1988). Regarding the adverse selection problem, there has not been much success in finding general assumptions on primitives that validate the FOA.\footnote{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.}

In Appendix B, 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 B, 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 characterize the properties of the optimal allocations.

Hence, the relaxed problem becomes the following:

\[ \max_{c_0(\theta), c_1(\theta, y), k_1(\theta), U(\theta)} \int_\theta^{\bar{\theta}} U(\theta) dF(\theta) \] (P1)

subject to

\[ \int_\theta^{\bar{\theta}} [c_0(\theta) + k_1(\theta)] dF(\theta) \leq e_0 \] (8)

\[ \int_\theta^{\bar{\theta}} \int_0^{\bar{\theta}} c_1(\theta, y) g(y|k_1(\theta), \theta) dy dF(\theta) \leq \int_\theta^{\bar{\theta}} \theta^\alpha k_1(\theta)\theta^\alpha dF(\theta) \] (9)

\[ U(\theta) = u(c_0(\theta)) + \beta \int_0^{\bar{\theta}} u(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy \]

\[ U'(\theta) = \frac{1}{\theta} k_1(\theta) u'(c_0(\theta)) \] (10)

\[ \beta \int_0^{\bar{\theta}} u'(c_1(\theta, y)) g_k(y|k_1(\theta), \theta) dy = u'(c_0(\theta)) \] (11)
In what follows, we refer to (10) as the *adverse selection* constraint and to (11) as *moral hazard* constraint.

### 2.1 Modified Inverse Euler Equation

In this section, we provide our version of the 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 1 (Modified Inverse Euler Equation).** Suppose that \( c_t, k_1 > 0 \), a.e. Then any solution to (P1) must satisfy

\[
\frac{q}{\beta} \int_0^y \frac{1}{u'(c_1(\theta, y))} g(y|k_1(\theta), \theta) dy = \frac{1}{u'(c_0(\theta))} + \frac{u''(c_0(\theta))}{u'(c_0(\theta))} \left[ \frac{1}{\theta} k_1(\theta) \mu_1(\theta) + \mu_2(\theta) \right]
\]

(12)

where \( q \) is the relative intertemporal price of consumption, \( \mu_1 \) is the costate associated with (10) and \( \mu_2 \) is the lagrange multiplier associated with (11). Both \( \mu_1 \) and \( \mu_2 \) are denominated in \( t = 0 \) consumption.

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 \( MC_t = \frac{1}{\beta u'(c_t)} \) 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 two terms in (12) capture these costs. Here, we give a heuristic derivation of (12).
Consider an infinitesimal perturbation of consumption for type $\theta^8, \{\varepsilon_{c0}, \varepsilon_{c1}(y)\}$ that preserves type $\theta$'s utility along each history path or

$$u(c_0(\theta) + \varepsilon_{c0}) + \beta u(c_1(\theta, y) + \varepsilon_{c1}(y)) = u(c_0(\theta)) + \beta u(c_1(\theta, y)), \forall y \in [0, \bar{y}] \quad (13)$$

There are two types of incentive costs associated with this perturbation. The first is the cost of distorting incentives for truth-telling about $\theta$. By definition, $\mu_1(\theta)$ captures the marginal cost of a unit increase in $U''(\theta)$. The above perturbation increases $U''(\theta)$ by $u''(c_0(\theta)) \frac{1}{\theta} k_1(\theta)$. Hence, the first type of incentive cost in terms of consumption in the first period is given by

$$u''(c_0(\theta)) \frac{1}{\theta} k_1(\theta) \mu_1(\theta) \varepsilon_{c0}$$

The second type of incentive cost is from distortions to the investment decision. Note that the above perturbation leaves the LHS of (11) unchanged. This is due to the fact that the above perturbation shifts utility after any realization of shock $y$ by the same amount. This makes the future marginal benefit from investment unchanged. However, due to the perturbation of consumption at $t = 0$, the incentives for investment at $t = 0$ change and the cost of this change in terms of period 0 consumption is captured by

$$\mu_2(\theta) u''(c_0(\theta)) \varepsilon_{c0}$$

Hence, the total cost of this perturbation is given by

$$q \int_0^\bar{y} \varepsilon_{c1}(y) dy + \varepsilon_{c0} + u''(c_0(\theta)) \frac{1}{\theta} k_1(\theta) \mu_1(\theta) \varepsilon_{c0} + \mu_2(\theta) u''(c_0(\theta)) \varepsilon_{c0}$$

Note that from (13), $\varepsilon_{c1}(y) = - \frac{u'(c_0(\theta))}{u'(c_1(\theta, y))} \varepsilon_{c0}$. Setting the above cost equal to zero leads to the desired MIEE.

Our version of Modified Inverse Euler Equation implies when consumption is non-separable from the source of private information, what affects the distortions to intertemporal saving margin is the heterogeneity in second period consumption as well as the tightness of the incentive constraints. In particular, the sign of $\frac{1}{\theta} k_1(\theta) \mu_1(\theta) + \mu_2(\theta)$ which captures the tightness of the incentive constraint, is a key determinant of the distortions to intertemporal saving margin. In section 2.2, we further discuss how the MIEE is useful in characterizing distortions.

Since the perturbation argument given above is independent of specific welfare weights on different individuals, it is straightforward to show that for social welfare functions other than the utilitarian, i.e., when the planner’s objective is $\int G(U(\theta)) dF(\theta)$, the MIEE holds.

---

8Since this is a heuristic derivation, we suppress the technical details. For example, the perturbation has to be over a positive measure of types. However, a continuity assumption on the allocations with respect to $\theta$, makes the above perturbation plausible.
2.2 Wedges

In this section we study the properties of the intertemporal saving wedge implied by the model developed so far. We argue that in this two period model, the intertemporal wedge is positive. We show this by considering the case where the utility function is exponential. Under this assumption, the model becomes more tractable and we can show that intertemporal wedge is positive. For general utility functions, the model is less tractable. However, we can show that when one source of risk is shut down, i.e., either output is not risky or there is no heterogeneity in productivities, again the intertemporal wedge is positive. Although the main result of the paper regarding negative intertemporal wedges cannot be shown in a two period model, the analysis in this section is useful to see the mechanisms in play in the model. Later in section 4.1, we extend the model to more than two periods to show that there are forces toward negative intertemporal wedges when the number of periods increases from two and agents are hit by subsequent productivity, \( \theta \), shocks.

In order to show that the intertemporal wedge is positive, we first show that the incentive costs of utility preserving perturbations, \( \frac{1}{\theta} k_1(\theta) \mu_1(\theta) + \mu_2(\theta) \), are positive. Then using an argument similar to Golosov et al. (2003), we can show that the intertemporal wedge is positive.

Consider the multiplier \( \mu_2(\theta) \) which captures the tightness of the moral hazard constraint. Note that by lemma ??, \( \mu_2(\theta) = \frac{\theta}{u'(c_0)} \text{Cov}_\theta(u(c_1), \frac{1}{u'(c_1)}) \). Since \( u(c_1) \) and \( \frac{1}{u'(c_1)} \) are positively correlated, \( \mu_2(\theta) \) is always positive. As we show in the next section, \( \mu_2(\theta) \) determines the sensitivity of the consumption schedule \( c_1(\theta, y) \) to income realization \( y \). Therefore, this result is equivalent to the consumption schedule \( c_1(\theta, y) \) being increasing in income realization.

Given the sign of \( \mu_2(\theta) \), if we show that the tightness of the adverse selection constraint, \( \mu_1(\theta) \), is positive, then I can show that intertemporal wedge is positive. To do so, I use an argument similar to the argument in Werning (2000) in the context of a static Mirrlees model. In fact, the result that \( \mu_1 \) is positive everywhere is reminiscent of the positive marginal tax result in Mirrleesian contexts. That is, to prove that marginal tax rates are positive in a static Mirrlees economy, one only needs to show that the co-state associated with the incentive constraint is positive. I can do this when the utility function has a CARA form since there are no wealth effects. We can also show it in the case where there is no riskiness in the returns to investment. The positive sign of the co-state, \( \mu_1(\theta) \), intuitively means that the relevant local incentive constraints are the downward incentive constraints.

Hence, we have the following proposition:

**Proposition 2** Suppose that \( u(c) = -\exp(-\psi c) \). Then, \( \mu_1(\theta) \geq 0 \) for all \( \theta \in [\underline{\theta}, \bar{\theta}] \). Moreover, \( \mu_1(\underline{\theta}) = \mu_1(\bar{\theta}) = 0 \) and the above inequality is strict for at least a positive measure of \( \theta \)'s.

Proof can be found in the Appendix.

Using the same proof, I can also show that for general utility functions, when there is no riskiness in returns, i.e., \( G(\cdot|k, \theta) \) puts mass 1 on \( (\theta k)^\alpha \), the co-state \( \mu_1(\theta) > 0 \) is positive – see Appendix for details.
The above discussion on the sign of incentive costs together with the Modified Inverse Euler Equation helps us determine the sign of the intertemporal wedge. That is, since \( \frac{1}{\theta} k_1(\theta) \mu_1(\theta) + \mu_2(\theta) > 0 \), then MIEE together with concavity of the utility function implies that

\[
\frac{q}{\beta} \int_{0}^{\bar{y}} \frac{1}{u'(c_1(\theta, y))} g(y | k_1(\theta), \theta) dy < \frac{1}{u'(c_0(\theta))}
\]

By Jensen’s inequality, we have

\[
\frac{q}{\beta} \int_{0}^{\bar{y}} \frac{1}{u'(c_1(\theta, y))} g(y | k_1(\theta), \theta) dy < \beta \int_{0}^{\bar{y}} \frac{1}{u'(c_1(\theta, y))} g(y | k_1(\theta), \theta) dy < \frac{1}{u'(c_0(\theta))}
\]

or

\[
q^{-1} \beta \int u'(c_1(\theta, y)) g(y | k_1(\theta), \theta) dy > u'(c_0(\theta))
\]

Hence, the intertemporal wedge, defined by

\[
\tau_s(\theta) = 1 - \frac{u'(c_0)}{q^{-1} \beta \int u'(c_1(\theta, y)) g(y | k_1(\theta), \theta) dy}
\]

is positive. One interpretation of positive intertemporal wedge is that in order to provide incentives, the optimal contract encourages consumption early. That is an agent who has access to borrowing and lending at rate \( q^{-1} \), facing the efficient allocation, would like to save. To see the intuition for the above inequality, consider decreasing agent \( \theta \)'s consumption in the first period by \( \varepsilon \) and increasing his consumption by \( q^{-1} \varepsilon \) after any realization in the second period. In addition to the usual direct cost, \( u'(c_1) \varepsilon \), and benefit \( \beta q^{-1} \varepsilon \int u'(c_1) g dy \) of such a perturbation, there are two incentive costs associated with it. The first comes from the moral hazard aspect of the model. Since utility function is concave, such perturbation makes investment relatively unattractive, i.e., it decreases \( \int u(c_1) g dy \). It also increases the current cost of investment to the individual consumer, \( u'(c) \). Hence, an agent of type \( \theta \) will decrease his investment. The second cost associated with this perturbation is that it increases the slope of the schedule \( U(\theta) \), i.e., \( \frac{1}{\theta} u'(c_0) k_1 \). Therefore, the entrepreneurs with higher productivity will find optimal to lie downward and work less. Since the marginal cost of such perturbation should be equal to its marginal benefit, we must have the inequality \((14)\).

Although, this wedge can be interpreted as a tax on saving, it does not directly translate into a marginal tax rate on saving. In fact, the implementation of the efficient allocation requires tax functions that are non-separable between second period income and saving. I discuss this further in section 2.4.

Given the above definition of wedges, it can be shown that a version of Mirrlees (1971)-Saez (2001) tax formulas holds in this economy as well when the returns are deterministic. In fact, I can derive a formula for saving wedge as a function of the skill distribution, intertemporal elasticity of substitution, investment-consumption ratio and distribution of consumption in the second period.
In particular, it can be shown that the following proposition holds:

**Proposition 3** Suppose that $c_t(\theta), k_1(\theta) > 0$, a.e.-$F$. Then any solution to (P1) must satisfy

$$\frac{\tau_s(\theta)}{1 - \tau_s(\theta)} = \frac{1 - F(\theta) k_1(\theta)}{\theta f(\theta) c_0(\theta)} \left( - \frac{u''(c_0(\theta))c_0(\theta)}{u'(c_0(\theta))} \right) \int_{\theta}^{1} \left[ \frac{u'(c_1(\theta))}{u'(c_1(\theta))} - \lambda_0 q^{-1} u'(c_1(\theta)) \right] \frac{dF(\theta)}{1 - F(\theta)}$$

As we can see, these formulas are very similar to Saez’s formulas since they relate marginal income/saving distortions to tail of skill distribution, $\frac{1-F(\theta)}{\theta f(\theta)}$, intertemporal elasticity of substitution, and investment-consumption ratio. Note that in our derivations in the appendix – MIEE and the tax formula, I have not used the fact that the skill distribution is bounded. In particular, the above formulas hold even in the case that $\theta = \infty$. The above formulas are easier to understand for the case where $\theta = \infty$ and $\lim_{\theta \to \infty} \frac{1-F(\theta)}{\theta f(\theta)} > 0$. In this case, saving wedges at the top are non-zero and can be derived explicitly in terms of fundamentals of the model as in Diamond (1998) and Saez (2001). The above analysis implies that the same exercise can be done for our environment.

### 2.3 Shape of the Consumption Schedule

In this section, I provide one of the main results of the paper. That is the possibility of progressive tax schedules. To do so I provide a simple formula for consumption in the second period as function of income realizations in period 2. Using this formula, I can provide conditions under which the consumption schedule is a concave function of income realization. As I argue here and formally show in section 2.4, concavity of the consumption schedule with respect to income implies progressivity of the tax schedule.

I first start by providing a simple formula for consumption in the second period:

**Lemma 1** Consider any solution to (P1) and assume that the allocations are positive almost surely. Then,

$$\frac{1}{u'(c_1(\theta, y))} = \int_{0}^{\hat{y}} \frac{1}{u'(c_1(\theta, \hat{y}))} g(\hat{y}|k_1(\theta), \theta) d\hat{y} + \beta q^{-1} \mu_2(\theta) \frac{g_k(y|k_1(\theta), \theta)}{g(y|k_1(\theta), \theta)}. \quad (15)$$

To intuitively see why this equation holds, consider the following perturbation of the allocation: for $\hat{y} \in [y, y + \varepsilon]$ increase $u(c_1(\theta, \hat{y}))$ by 1 unit and decrease all $u(c_1(\theta, \hat{y}))$ by $\varepsilon g(y|k_1(\theta), \theta)$. Note that this perturbation preserves period 1 utility of type $\theta$. Hence, it does not violate (10). It does, however, change investment incentives for type $\theta$. Note that the uniform decrease in utility for all $\hat{y}$’s does not change the marginal return to investment. As a result, the marginal individual benefit to investment approximately increases by $\beta g_k(y|k_1(\theta), \theta)\varepsilon$. Hence the resource cost of this
Perturbation is given by \( \frac{1}{w(c_1(\theta, y))} g(y|k_1(\theta), \theta) \varepsilon \) while the benefit from lowering consumption and relaxing the incentive constraint is given by

\[
\varepsilon g(y|k_1(\theta), \theta) \int_0^{\hat{y}} \frac{1}{w(c_1(\theta, \hat{y}))} g(\hat{y}|k_1(\theta), \theta) d\hat{y} + q^{-1} \mu_2(\theta) \beta g_k(y|k_1(\theta), \theta) \varepsilon.
\]

Equating the cost and benefit leads to (15). This perturbation is depicted in Figure 1.

Figure 1: Perturbation leading to equation (15) - the black graph is the original schedule, the red line is the perturbed schedule.

Equation (15) implies that the marginal cost of providing utility to income level \( y \), \( \frac{1}{w(c_1(y))} \), is a linear function of the hazard rate. As in Hölmstrom (1979), \( \frac{\partial}{\partial y} \), is the derivative of the likelihood function \( \log g(y|k, \theta) \) where \( k \) can be treated as unobservable from planner’s point of view. Hence, when \( \frac{\partial}{\partial y} \) is the highest, the planner is the most inclined to infer from \( y \) that the agent took the right action. Hence, the rewards to the agent are the highest in those states. Note, also, that the MLRP assumption implies that \( \frac{\partial}{\partial y} \) is an increasing function of \( y \). Since \( u(\cdot) \) is concave, \( \frac{1}{w(c)} \) is an increasing function of \( c \) and therefore from (15), we deduce that the consumption schedule \( c_1(\theta, y) \) is an increasing function of \( y \).

Given the above formula on the consumption schedule, it is rather straightforward to provide sufficient condition under which the consumption schedule is concave. In fact, when \( \frac{\partial}{\partial y} \) is concave in \( y \) and \( \frac{1}{w(c)} \) is a convex function of \( c \), the schedule is concave in \( y \). Notice that when \( u(c) = \frac{c^{1-\sigma}}{1-\sigma} \), the convexity of \( \frac{1}{w(c)} \) requires that \( \sigma > 1 \). That is the intertemporal elasticity of substitution must be bigger than 1. To better understand the requirement on the hazard ratio, \( \frac{\partial}{\partial y} \), I consider the
environment in Evans and Jovanovic (1989)(EJ economy henceforth). Suppose that

\[ y = \varepsilon \theta^\alpha k^\alpha \]

where \( \varepsilon \sim H(\varepsilon) \) with density \( h(\varepsilon) \) and \( \varepsilon \in [0, \infty) \). In this case,

\[ g(y|k, \theta) = \frac{1}{(\theta k)^\alpha} h \left( \frac{y}{(\theta k)^\alpha} \right) \]

and hence

\[ \frac{g_k(y|k, \theta)}{g(y|k, \theta)} = -\alpha k^{-1} \left[ 1 + \frac{y}{(\theta k)^\alpha} h' \left( \frac{y}{(\theta k)^\alpha} \right) \right] \]

Now consider the following examples for the distribution function \( h \):

1. Log-normal distribution: \( h(\varepsilon) = \kappa \varepsilon^{-1} e^{-\frac{(\log \varepsilon - \mu)^2}{2\sigma^2}} \). In this case

\[ \frac{g_k(y|k, \theta)}{g(y|k, \theta)} = \alpha k^{-1} \log y - \alpha \log(\theta k) - \frac{\log y - \alpha \log(\theta k) - \mu}{2\sigma^2} \]

and hence the hazard ratio is concave in \( y \).

2. Gamma distribution: \( h(\varepsilon) = \kappa \varepsilon^{-\zeta - 1} e^{-\varepsilon/\eta} \). In this case,

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

and hence the hazard ratio is linear in \( \varepsilon \).

3. Pareto distribution in the tail: \( h(\varepsilon) = \kappa \varepsilon^{-\zeta - 1} \). In this case, \( \frac{g_k}{g} = \zeta \alpha k^{-1} \) and hence the hazard ratio is constant.

Hence, for the EJ economy, MLRP implies that \( \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \) be decreasing and when \( \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \) is convex, the consumption schedule is concave in income.

The concavity of the consumption schedule in realized income has an important interpretation regarding tax system. In fact, the slope of the consumption function determines the marginal tax rate on income. In particular, when this slope is decreasing, i.e., consumption schedule is concave, the marginal tax rate is increasing and hence the income tax schedule is progressive – I will discuss this in detail in section 2.4. Here, progressivity of the tax system works as an insurance mechanism against income shocks. Due to moral hazard, only partial insurance is feasible and therefore consumption schedule is not fully flat.

The analysis so far points to ways a planner can resolve the two types of informational asymmetries, the moral hazard and the adverse selection problem. Loosely speaking, the intertemporal
wedge induces agents to tell the truth regarding their productivity type. Once the productivity
type is revealed, equation (15) induces the agent to make the right amount of investment.

2.4 Implementation

In this section, I discuss ways for a government to implement the optimal allocations discussed
above. The construction of the tax function below, demonstrates that the tax function is unique
given the market structure imposed. Note that the market structure assumed for the implementa-
tion plays a key role in determining government policy. Here, we assume that the entrepreneurs,
in addition to the individual investment opportunity, have access to a centralized market for risk
free asset in net zero supply. We, then, construct a tax schedule that implements the optimal
allocation. Using the properties of the allocations discussed above, we characterize the properties
of such optimal tax system.

A key assumption in the following implementation is that agents are unable to sign contracts
before realization of their productivity type, \( \theta \). Otherwise, the results in Prescott and Townsend
(1984) imply that private contracts are able to achieve the constrained efficient allocation dis-
cussed above. This assumption gives rise to a need for redistributive policies by the government.
Later, in section 5, we show that if ex-ante contracting is available, the optimal allocation can be
implemented with a set of contracts that are widely used in financial markets and venture capital
contracts. This assumption is in line with the rest of the literature on dynamic public finance.

As mentioned above, we assume that each entrepreneur can invest in his private investment
project and can borrow and save from centralized market. The agent may purchase and sell the
risk free bond at price \( Q \). Hence, the agent’s budget constraint at \( t = 0 \) is given by:

\[
c_0 + k_1 + Qb_0 \leq e_0
\]

The government observes \( b_0 \) and \( y \) at \( t = 1 \) and can tax agents based on observables according
to the tax function \( T(b_0, y) \). Given this tax function, the budget constraint of the agent in the
second period is given by

\[
c_1 \leq y + b_0 - T(b_0, y)
\]

Hence, facing a particular tax function \( T(b_0, y) \), an entrepreneur of type \( \theta \) solves the following
maximization problem

\[
\max_{c_0, c_1(y), k_1, b_0} u(c_0) + \beta \int_0^y u(c_1(y)) g(y|k_1, \theta) dy
\]   (16)

subject to
\[ c_0 + k_1 + Qb_0 \leq e_0 \]
\[ c_1(y) \leq y + b_0 - T(b_0, y) \]

Here, we show that given any incentive compatible allocation \( \{c_0^*(\theta), \{c_1^*(\theta, y)\}, k_1^*(\theta)\} \) together with an intertemporal price of consumption \( q \), there exists a tax system of the above form that implements it. To do so we need to make the following assumption about the allocation:

**Assumption 1** For all \( \theta \neq \theta' \), \( c_0^*(\theta) + k_1^*(\theta) \neq c_0^*(\theta') + k_1^*(\theta') \) and allocations are \( C^1 \) in \( \theta \).

A sufficient condition for the above assumption is that transfers in the first period are increasing in type. In fact, if the allocations are continuous in \( \theta \), the above assumption implies that transfers, \( c_0(\theta) + k_1(\theta) \), are monotone in \( \theta \).

Given this assumption, we can show the following:

**Proposition 4** Consider an incentive compatible allocation \( \{c_0^*(\theta), \{c_1^*(\theta, y)\}, k_1^*(\theta)\}_{\theta \in [\underline{\theta}, \bar{\theta}]} \) together with a risk-free bond price \( q \). If Assumption 1 holds, there is tax function \( T(\cdot, \cdot) \) that implements the allocation. Moreover, the tax function is \( C^1 \).

**Proof.** We start by constructing the saving level, \( b_0^*(\theta) \), for each type

\[ b_0^*(\theta) = q^{-1}[e_0 - k_1^*(\theta) - c_0^*(\theta)] \]

Assumption 1 implies that \( b_0^*(\theta) \) is a one-to-one function of \( \theta \). Notice that continuity of the allocations together \( b_0^*(\cdot) \) being one-to-one implies that there exists an interval \([\underline{b}, \bar{b}]\) such that \( b_0^*([\underline{\theta}, \bar{\theta}]) = [\underline{b}, \bar{b}] \) and \( b_0^* \) is a bijection over \([\theta, \bar{\theta}]\). Hence, we can define the following tax function \( T(\cdot, \cdot) \):

\[ T(b, y) = \begin{cases} 
  y + b - c_1((b_0^*)^{-1}(b), y) & b \in [\underline{b}, \bar{b}] \\
  y + b & b \notin [\underline{b}, \bar{b}]
\end{cases} \quad (17) \]

Here, we show that the above tax function implements the desired allocation when the price risk-free bond at \( t = 0 \) are given by \( q \). First, note that if an agent of type \( \theta \), chooses \( c_0^*(\theta), \{c_1^*(\theta, y)\}, k_1^*(\theta), b_0^*(\theta) \), the utility he receives is equal to the utility he receives from the allocation, \( U(\theta) \). Second, it is easy to see that in (16) \( b_0 \in [\underline{b}, \bar{b}] \), otherwise consumption following any income realization is zero. At last, consider a possible solution to (16), \( \{\hat{c}_0, \{\hat{c}_1(y)\}, \hat{k}_1, \hat{b}_0\} \). Since \( b_0^* \) is a bijection, there exists a unique \( \hat{\theta} \in [\underline{\theta}, \bar{\theta}] \) such that \( b_0^*(\hat{\theta}) = \hat{b}_0 \). Then, by definition of \( b^*(\cdot) \),

\[ e_0 - q\hat{b}^*(\hat{\theta}) = c_0^*(\hat{\theta}) + k_1^*(\hat{\theta}) \]

and given the budget constraint at \( t = 0 \), \( \hat{c}_0 + \hat{k}_1 = c_0^*(\hat{\theta}) + k_1^*(\hat{\theta}) \). Moreover, by definition of \( T(\cdot, \cdot) \), \( \hat{b}_0 + y - T(\hat{b}_0, y) = c_1(\hat{\theta}, y) \). Hence, the utility that the agent receives from this allocation is given by

\[ u(c_0^*(\hat{\theta}) + k_1^*(\hat{\theta}) - \hat{k}_1) + \beta \int u(c_1^*(\hat{\theta}, y))g(y|\hat{k}_1, \theta)dy \]
By incentive compatibility (4),

\[ U(\theta) \geq u(c_0^*(\hat{\theta}) + k_1^*(\hat{\theta}) - \hat{k}_1) + \beta \int u(c_1^*(\hat{\theta}, y)) g(y|\hat{k}_1, \theta) dy \]

Therefore, it is optimal for the agent to choose \( \{c_0^*(\theta), \{c_1^*(\theta, y)\}, k_1^*(\theta), b_0^*(\theta)\} \).

A point worth noticing is that given \( q \), the above implementation is unique. In fact, knowing \( q \) and the allocation, one can uniquely pin down saving levels and under Assumption 1, \( T(\cdot, \cdot) \) is uniquely determined by the allocation.

Given the above tax function, properties of the optimal allocation leads to certain properties of the tax function. As we have shown in 2.2, intertemporal wedge is positive. This implies that average value of \( T_b \) weighted by marginal utility is positive. To see this, note that the first order condition from (16) is given by

\[
q^{-1} \beta \int_0^\theta u'(c_1^*(\theta, y))(1 - T_b(b_0^*(\theta), y)) g(y|k_1^*(\theta), \theta) dy = u'(c_0^*(\theta))
\]

Moreover, since \( q^{-1} \beta \int_0^\theta u'(c_1^*(\theta, y)) g(y|k_1^*(\theta), \theta) dy > u'(c_0(\theta)) \) as shown in section 2.2, we must have

\[
\int_0^\theta u'(c_1^*(\theta, y)) T_b(b_0^*(\theta), y) g(y|k_1^*(\theta), \theta) dy > 0
\]

Note that in order to \( T_b \) in for each income realization, we need to know the way \( c_1^*(\theta, y) \) moves as a function of \( \theta \).

As mentioned before, a key result of this paper is that the optimal tax schedule with respect to entrepreneurial income is progressive. To show this, note that in this environment, marginal tax rate on income \( T_y \) is given by

\[
T_y(b_0^*(\theta), y) = 1 - \frac{\partial}{\partial y} c_1^*(\theta, y)
\]

Recall from section 2.3 that,

\[
\frac{1}{u'(c_1(\theta, y))} = a(\theta) + b(\theta) \frac{g_k(y|\theta, k_1(\theta))}{g(y|\theta, k_1(\theta))}
\]

where \( a(\theta), b(\theta) \) are independent of \( y \). Hence, the derivative of the consumption function with respect to \( y \) is given by

\[
\frac{\partial}{\partial y} c_1(\theta, y) = \left( \frac{1}{u'} \right)^{(-1)} \left( a(\theta) + b(\theta) \frac{g_k(y|\theta, k_1(\theta))}{g(y|\theta, k_1(\theta))} \right)
\]

Based on the above formula, when \( \frac{a_k}{g} \) is concave in \( y \) and \( \frac{1}{u'} \) is convex, \( c_1 \) becomes concave in \( y \) and hence \( T_y \) is increasing, i.e., entrepreneurial income taxes are progressive. We have the following
proposition:

**Proposition 5** Suppose that \( \frac{1}{u'(c)} \) is convex and \( g_k \) is concave in \( y \). Then the tax function \( T(\cdot, \cdot) \) defined by (17) is progressive, i.e., \( T_y(b, \cdot) \) is increasing in \( y \).

Note that when \( u(c) = \frac{c^{1-\sigma}}{1-\sigma} \), then convexity of \( \frac{1}{u'(c)} \) is equivalent to \( \sigma \geq 1 \). This result is in contrast with regressivity of the income tax schedule at the top in Mirrlees (1971) when skill distribution is bounded (see Farhi and Werning (2010a) for a dynamic extension). Intuitively, progressivity arises in order to provide the right incentives to invest to the entrepreneur. Hence, the rewards to higher realization, i.e., the slope of consumption schedule, must be higher for low realizations of income. Although, we have shown progressivity in a model of entrepreneurs with capital income risk, this result is not specific to this environment. In particular, it is natural to guess that in a model with risky human capital and private information, same result would hold.

3 Multi-Period Model

In this section, we extend the analysis to a multi-period environment. In this context, we derive the general version of the MIEE and show how results change. We extend the two period model and derive the MIEE. Using the properties of the model, we study the implications of the model on taxation.

Time is discrete and indexed by \( t = 0, 1, \cdots, T \) where \( T \in \mathbb{N} \cup \{ \infty \} \). There is a unit measure of entrepreneurs. The entrepreneurs are endowed with \( e_0 \) units of good at \( t = 0 \). Each entrepreneur has access to a private risky investment technology that evolves as follows: At each date \( t = 0, \cdots, T - 1 \), agent draws a productivity type \( \theta_t \in \Theta = [\underline{\theta}, \bar{\theta}] \) according to a differentiable c.d.f function \( F_t(\theta_t | \theta_{t-1}) \)(with its derivative given by \( f_t(\theta_t | \theta_{t-1}) \)). The initial draw of productivity, \( \theta_0 \), is distributed according to a differentiable c.d.f. function \( F_0(\theta_0) \). At each date \( t \), the agent can privately invest \( k_{t+1} \) in the project. Given the entrepreneur’s investment, \( k_{t+1} \), and productivity type \( \theta_t \), his income, \( y_{t+1} \in Y = [0, \bar{y}] \), realized at \( t + 1 \), is a random variable that is distributed according to a differentiable c.d.f. \( G^{t+1}(y_{t+1}|\theta_t, k_{t+1}) \)(with its derivative given by \( g^{t+1}(y_{t+1}|\theta_t, k_{t+1}) \)).

Similar to the two-period example, the function \( G^{t+1} \) has the following properties:

1. \( G^{t+1}(y_{t+1}|\theta_t, k_{t+1}) \) is strictly decreasing in \( \theta_t \) and \( k_{t+1} \) – hence \( y_{t+1} \) is increasing according to first order stochastic dominance ordering.

2. The mean value of \( y_{t+1} \) is given by

\[
\int_Y y_{t+1} g^{t+1}(y_{t+1}|\theta_t, k_{t+1}) dy_{t+1} = (\theta_t k_{t+1})^\alpha
\]

3. For any \( \theta, \theta', k, \)

\[
G^{t+1}(y|\theta, k) = G^{t+1}(y|\theta, \frac{\theta k}{\theta'}) , \forall y \in [0, \bar{y}]
\]
4. \( G^{t+1}(y|\theta, k) \) satisfies MLRP:

\[
\frac{\partial}{\partial y} \left( \frac{g^{t+1}_k(y|\theta, k)}{g^{t+1}(y|\theta, k)} \right) > 0 \quad \forall y, k, \theta
\]

5. Given \( k_{t+1} \) and \( \theta_t, y_{t+1} \) is independent from \( \theta_{t+1} \).

The 5th assumption above is crucial. It is important that conditional on \( k_{t+1} \) and \( \theta_t, y_{t+1} \) and \( \theta_{t+1} \) are not perfectly correlated. In case of perfect correlation, a deviation over \( k_{t+1} \) at \( t \), must be accompanied by a certain report of \( \theta_{t+1} \) at \( t + 1 \). This further complicates the problem. For analytical tractability, we assume that they are independent. However, the analysis here will go through if we assume that \( y_{t+1} \) and \( \theta_{t+1} \) are partially correlated.

Given this environment, an allocation is given by

\[
\{ c_t(\theta^t, y^t), k_{t+1}(\theta^t, y^t) \}_{t=0}^T
\]

where \( \theta^t = (\theta_0, \ldots, \theta_t) \in \Theta^{t+1} \) and \( y^t = (y_1, \ldots, y_t) \in Y^t \). When \( t = 0 \), \( y^0 \) is the empty history and \( \theta^0 = \theta^{T-1} \) -- there are no draw of productivity at \( T \) and no draw of income at \( 0 \). For ease of notation, we assume that \( \mu_t(\theta^t, y^t; k^{t-1}) \) is the joint distribution of all possible histories at period \( t \) given a sequence of investments \( k^{t-1} = (k_1, \ldots, k_{t-1}) \). Note that by definition,

\[
\mu_t(A_0 \times \cdots \times A_t, B_1 \times \cdots \times B_t; k^{t-1}) = \\
\int_{A_0} \cdots \int_{A_t} f^0(\theta_0) \prod_{\tau=1}^t f^\tau(\theta_\tau|\theta_{\tau-1}) \prod_{\tau=1}^t \left( \int_{B_\tau} g^\tau(y|\theta_{\tau-1}, k_\tau) dy \right) d\theta_t \cdots d\theta_0
\]

An allocation is feasible if it satisfies

\[
\int_{\Theta} \left[ c_0(\theta_0) + k_1(\theta_0) \right] f_0(\theta_0)d\theta_0 \leq e_0 \tag{18}
\]

\[
\int_{\Theta^{t+1} \times Y^t} \left[ c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) \right] d\mu_t(\theta^t, y^t; k^{t-1}(\theta^{t-1}, y^{t-1})) \tag{19}
\]

\[
\leq \int_{\Theta^{t+1} \times Y^t} \left( \theta_t k_{t+1}(\theta^t, y^t) \right)^\alpha d\mu_t(\theta^t, y^t; k^{t-1}(\theta^{t-1}, y^{t-1}))
\]

As before, we assume that the planner observes the income in each period and productivity type as well as consumption and investment is privately known by the agent. Given this assumption about information structure, the nature of the incentive compatibility constraints depends on whether \( y_{t+1} \) is stochastic or deterministic. As it is clear in the two period example, with deterministic returns, agents are not free to deviate to any investment level where as with risky return these deviations are not detectable. Below, we describe an incentive compatible allocation under each set of assumption.
When returns are risky, at each period, an agent can lie about its productivity type and pick a different level of investment. Hence, a deviation strategy by the agent is a reporting strategy \( \sigma = \{ \hat{\sigma}_t(\theta^t, y^t) \} \) as well as an investment strategy \( \hat{k} = \{ \hat{k}_t(\theta^t, y^t) \} \). The utility from a deviation strategy given allocation is given by

\[
U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{k}) = \sum_{t=0}^T \beta^t \int_{\Theta^{t+1} \times Y^t} u(c_t(\hat{\sigma}_t^t(\theta^t, y^t), y^t)) + k_{t+1}(\hat{\sigma}_t^t(\theta^t, y^t), y^t) - \hat{k}(\hat{\sigma}_t^t(\theta^t, y^t), y^t))d\mu_t(\theta^t, y^t; \hat{k}_{t-1}(\theta^{t-1}, y^{t-1}))
\]

Let \( \sigma^* \) be the truth-telling strategy, an allocation is then incentive compatible, if

\[
U(\{c_t, k_{t+1}\}; \sigma^*, k) \geq U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{k}) \quad (20)
\]

When, returns are not risky, agents can freely chose a reporting strategy, \( \sigma = \{ \hat{\sigma}_t(\theta^t, y^t) \} \). They, however, cannot choose any investment level freely. Given \( \hat{\sigma} \), the entrepreneur has to invest \( \frac{\partial}{\partial \theta^t} k_t(\hat{\sigma}_t^t, y^t) \). Since income is observable, any other investment level is in discrepancy with the original report and therefore detectable by the planner. With a little abuse of notation, an allocation is said to be incentive compatible with safe returns if

\[
U(\{c_t, k_{t+1}\}; \sigma^*, k) \geq U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{\sigma}_t^t(\theta^t, y^t)) \quad (21)
\]

Hence, a planner solves the following maximization problem

\[
\max \sum_{t=0}^T \beta^t \int_{\Theta^{t+1} \times Y^t} u(c_t(\theta^t, y^t))d\mu_t(\theta^t, y^t; k^t|\theta^t, y^t) \quad (P)
\]

subject to (20) or (21), (18) and (19).

**First Order Approach to Incentive Compatibility Constraints**

The set of incentive compatibility constraints in the program \( (P) \) is very large. Using the First Order Approach we greatly simplify the set of incentive constraints. Here, we derive necessary first order conditions that any incentive compatible allocation must satisfy.

To do so, let \( U_t(\theta^t, y^t) \) be the utility of agent with history \( (\theta^t, y^t) \),

\[
U_t(\theta^t, y^t) = \sum_{\tau=t}^T \beta^{\tau-t} \int_{\Theta^{\tau+1} \times Y^\tau} u(c_\tau(\theta^\tau, y^\tau))d\mu_\tau(\theta^\tau, y^\tau; k^\tau|\theta^t, y^t)
\]

If we focus on one shot deviations – deviation in one period and telling the truth thereafter –
an allocation is *one-shot incentive compatible* if

\[
\begin{align*}
    u(c_t(\theta^t, y^t)) + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t \\
    \geq \max_{k, \hat{\theta}} u(c_t(\theta^{t-1}, \hat{\theta}, y^t) + k_{t+1}(\theta^{t-1}, \hat{\theta}, y^t) - \hat{k}) \nonumber \\
    + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \hat{\theta}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, \hat{k})f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t
\end{align*}
\]

(22)

When \( T \) is finite, it is easy to see that the above is equivalent to (20). When \( T = \infty \), if the utility from the allocations stays bounded, then (22) is equivalent to (20). This incentive constraint implies that any incentive compatible allocation that is \( C^1 \), must satisfy the following first order condition with respect to investment and Envelope Theorem– similar to (5)-(6):

\[
\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t \\
+ \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \hat{\theta}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t
\]

\[
u'(c_t(\theta^t, y^t)) = \beta \int_{\Theta \times Y} U_{t+1}(\theta^{t+1}, y^{t+1})g_k(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t
\]

(23)

Using property 3 for \( G^{t+1} \), we know that

\[ k_{t+1}g_{t+1}(y_{t+1}|\theta_t, k_{t+1}) = \theta_t g_{t+1}(y_{t+1}|\theta_t, k_{t+1}) \]

Hence, we can rewrite the first equation as

\[
\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \frac{1}{\theta_t} k_{t+1}(\theta^t, y^t)u'(c_t(\theta^t, y^t)) \\
+ \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_t
\]

(24)

Therefore, the relaxed planning problem is the following

\[
\max \int_\Theta U_0(\theta_0)f^0(\theta_0)d\theta_0
\]

(25)

subject to

\[
U_t(\theta^t, y^t) = u(c_t(\theta^t, y^t)) + \beta \int_{\Theta \times Y} U_{t+1}(\theta^{t+1}, y^{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1})f_{t+1}(\theta_{t+1}|\theta_t)d\theta_t dy_{t+1}d\theta_{t+1}
\]

and (18),(19),(23), and (24).
3.1 Modified Inverse Euler Equation

In this section, we derive the general version of the MIEE for the model set up above. We start with a recursive formulation of the model. Since, $\theta_t$ is a first order Markov Process, a result from Fernandes and Phelan (2000) implies that a sufficient statistic for the history is promised utility that is the continuation utility if the agent tells the truth and threat utility, continuation utility when the agent lies – potentially a function when there is a continuum of types. The FOA, implies that a sufficient statistic for threat utility is $\int U_{t+1}f_{\theta_{t-1}}(\theta|\theta_{t-1})d\theta$. Hence, given an allocation, we define the following

$$w_t(\theta^{t-1}, y^t) = \int_\Theta U_t(\theta^t, y^t)f_{\theta_{t-1}}(\theta_{t-1}|\theta_{t-1})d\theta_t$$

$$\Delta_t(\theta^{t-1}, y^t) = \int_\Theta U_t(\theta^t, y^t)f_{\theta_{t-1}}(\theta_{t-1}|\theta_{t-1})d\theta_t$$

where $w_t$ is the promise utility to the agent before realization of productivity shock at $t$ and $\Delta_t$ is the sufficient statistic for keeping track of threat utility. We call $\Delta_t$, the marginal promised utility. Given the above definitions, we can rewrite the local incentive constraints as

$$u'(c_t(\theta^t, y^t)) = \beta \int_Y w_{t+1}(\theta^t, y^{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))dy_{t+1}$$

$$\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \frac{1}{\theta_t}k_{t+1}(\theta^t, y^t)u'(c_t(\theta^t, y^t)) + \beta \int_Y \Delta_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1})g_{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))dy_{t+1}$$

Now, if we let $Q_t$ be the lagrange multiplier on the feasibility constraint – by Theorem 1, Section 8.3 in Luenberger (1969), such multiplier exists. We can interpret these multiplier as price of consumption at period $t$. Conversely, $\frac{Q_{t+1}}{Q_t}$ can be interpreted as a return on a risk free bond at period $t$. Given these prices, we can rewrite the dual of the above planning problem as follows

$$P_t(w, \Delta, \theta_{-1}) = \max_{c,k,w',\Delta',u} \int_\Theta \left[ \frac{Q_{t+1}}{Q_t}(\theta k(\theta))^\alpha - c(\theta) - k(\theta) \right]$$

$$+ \frac{Q_{t+1}}{Q_t} \int_Y P_{t+1}(w'(\theta, y), \Delta'(\theta, y), \theta)g_{t+1}(y|\theta, k(\theta))dy \right] f^t(\theta|\theta_{-1})d\theta$$
subject to

\[ w = \int_{\Theta} U(\theta) f^t(\theta|\theta_{t-1}) d\theta \]
\[ \Delta = \int_{\Theta} U(\theta) f^t_{\theta_{t-1}}(\theta|\theta_{t-1}) d\theta \]
\[ U(\theta) = u(c(\theta)) + \beta \int_{Y} w'(\theta, y) g^{t+1}(y|\theta, k(\theta)) dy \]
\[ \frac{d}{d\theta} U(\theta) = \frac{1}{\theta} k(\theta) u'(c(\theta)) + \beta \int_{Y} \Delta'(\theta, y) g^{t+1}(y|\theta, k(\theta)) dy \quad (26) \]
\[ u'(c(\theta)) = \beta \int_{Y} w'(\theta, y) g^{t+1}(y|\theta, k(\theta)) dy \quad (27) \]

Note that, the first term is the aggregate output for an agent of type \( \theta_t = \theta \) in period \( t + 1 \) and hence it is discounted by \( Q_{t+1}/Q_t \) in order to be in terms of consumption at period \( t \).

The following proposition extends (12) to the environment described above. Technically, it is a result of marginal cost \( P^t_w \) being an Auto Regressive process with autocorrelation \( \frac{\beta Q_t}{Q_{t+1}} \) and how \( P^t_w \) is related to expected reciprocal of marginal utility.

**Theorem 1** Any solution to (P2) must satisfy the following Modified Inverse Euler Equation:

\[ \frac{1}{u'(c_t)} + \frac{1}{u'(c_t)} \left[ \frac{1}{\theta_t} k_{t+1} \mu_{1t} + \mu_{2t} \right] = \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} + \frac{u''(c_{t+1})}{u'(c_{t+1})} \left[ \frac{1}{\theta_{t+1}} k_{t+1} \mu_{1t+1} + \mu_{2t+1} \right] \right\} \quad (28) \]

where \( \mu_{1t} \) is the costate associated with (26) and \( \mu_{2t} \) is the lagrange multiplier associated with (27).

Proof can be found in the appendix.

Notice that \( \mu_{1t} \) and \( \mu_{2t} \) represent the tightness of the incentive constraints at period \( t \). Similar to lemma ??, we can show that

\[ \mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} \text{Cov} \left( P^t_w, w_{t+1} | (\theta^t, y^t) \right) \quad (29) \]

Hence, when \( P^t_w \) is decreasing with respect to \( w_{t+1} \) – an example of this is the case where \( P^t_w \) is concave and \( \theta_t \) is i.i.d., \( \mu_{2t} \) is always positive\(^9\).

What the above equation implies is that the sign of distortions on saving is affected not only by the heterogeneity of consumption, as shown by Golosov et al. (2003), but also it depends on relative tightness of incentive constraints across periods. In particular, if this tightness increases or decreases in expectation, it might change the sign of the distortions. Saving distortions are the highest when this difference is the highest. Here we perform a heuristic analysis of the above equation. In particular, suppose that \( \mu_{1t} \) is always positive and that project returns are

\(^9\)In the appendix, we provide a formula for \( \mu_{1t} \) similar to lemma ??.
deterministic, i.e., local downward incentive constraints are binding. Note that we always have, 
\( \mu_{1t}(\theta^{t-1}, \bar{\theta}, y^t) = 0 \). In this case, (28) implies that

\[
\frac{1}{u'(c_t)} < \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} \right\}
\]

Since \( \theta_t = \bar{\theta} \), incentive constraints are relatively tighter in the future. In this case, reciprocal of marginal utility should increase. This creates a force toward decreasing the intertemporal wedge. Later, we show that the intertemporal wedge is in fact negative at the top. On the other hand, when current incentive constraints are tighter relative to future incentive constraints, we must have

\[
\frac{1}{u'(c_t)} > \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} \right\}
\]

This creates a force toward increasing the intertemporal wedge and therefore the intertemporal wedge is positive. To our knowledge, this feature is new to this model\(^\text{10}\). What it implies is that contrary to previous results as in Golosov et al. (2003), there is a possibility of saving subsidies in a model with capital income risk. The closest result to the above is perhaps Albanesi (2006). She shows that in an environment with moral hazard, there is a possibility of negative taxes. However, in that environment, since the source of private information is separable from consumption, Inverse Euler Equation is satisfied and intertemporal wedge is always positive. The negative tax result, however, is specific to the particular implementation rather than being a property of the optimal allocation.

4 Optimal Taxes

In this section, we show the main result of the paper regarding negative intertemporal wedges. Moreover, we show that the progressivity result extends to the dynamic model. Finally, we provide a tax schedule that implements the optimal allocations.

4.1 Intertemporal Wedge

In this section we focus on the intertemporal wedge implied by the efficient allocation discussed above. In particular, we derive conditions under which its sign is negative. First, in a model with deterministic returns and i.i.d. shocks, we show that the intertemporal wedge is negative at the top and bottom and positive in the middle of the distribution of returns. Moreover, we show that when ex-ante heterogeneity is shut down, i.e., \( \theta \) is not risky and the only source of risk is in returns to investment, the intertemporal wedge is positive. When both types of risk are present, these two forces act against each other and when ex-ante heterogeneity is sufficiently high, the intertemporal wedge is negative. To our knowledge, this feature is new to this model\(^\text{10}\).

\(^\text{10}\) The same approach can help us characterize saving distortions in a model in which period utility function is non-separable in consumption and leisure. I suspect, a similar result holds in that environment.
wedge is negative at the top. Throughout, this section, we assume that utility function has the CARA form for which there are no wealth effect.

**Assumption 2** The period utility function has the CARA form \( u(c) = -\exp(-\psi c) \).

To prove our main result, negative intertemporal wedge at the top, we start with the model with safe returns and i.i.d. shocks.

### 4.1.1 Safe Returns – A Negative Wedge Result

Here we discuss the case where \( \theta_t \) is i.i.d. over time and the returns to investment is deterministic. That is the return to investment is \( (\theta_t k_{t+1})^\alpha \) at \( t + 1 \). In this case, since the income from the project is not risky once \( \theta \) is known, \( \mu_{2t} = 0 \) and therefore

\[
\frac{1}{u'(c_t)} - \psi \frac{1}{\theta_t} k_{t+1} \mu_{1t} = \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} - \psi \frac{1}{\theta_t} k_{t+2} \mu_{1t+1} \right\}
\]

Moreover, in this case we can show that

\[
\frac{1}{\theta_t} \mu_{1t} = \left[ \frac{Q_{t+1}}{Q_t} \alpha \theta_t^\alpha k_{t+1}^{\alpha-1} - 1 \right] \frac{1}{u'(c_t)}
\]

That is \( \mu_{1t} \) measure the distortions to productive efficiency – how different is the marginal return of an individual project from the economy-wide rate of return. In particular, using the same argument as in 2, we can show that \( \mu_{1t} \geq 0 \). This implies that the inside return from the project, \( \alpha \theta_t^\alpha k_{t+1}^{\alpha-1} \) is higher than the outside return \( \frac{Q_{t+1}}{Q_{t+1}} \). That is the entrepreneurs in this model look “borrowing constrained”, i.e., the investment in the project is less than what it would have been without frictions. This result is related to a strand of literature in corporate finance that deals with Modigliani-Miller theorem and its determinants (see Tirole (2006).) Hence, the sign of the intertemporal wedge depends on how distortions to productive efficiency evolve over time. In particular, when distortions to productive efficiency are higher relative to future, \( \frac{1}{u'(c_t)} > \frac{Q_{t+1}}{\beta Q_t} E_t \frac{1}{u'(c_{t+1})} \) and hence intertemporal wedge is positive. When, the distortion to productive efficiency is lower relative to future, namely at the top(bottom) of the distribution of \( \theta \) where it is zero, we must have

\[
\frac{1}{u'(c_t)} < \frac{Q_{t+1}}{\beta Q_t} E_t \frac{1}{u'(c_{t+1})}.
\]

Unfortunately, this inequality cannot be used to determine the sign of the intertemporal wedge. Therefore, we use a direct argument using the recursive formulation of the problem to show that the intertemporal wedge is negative at the top(bottom). In particular, we show the negativity in two steps:

1. The margin between \( c_{t-1}(w_t, \bar{\theta}) \) and \( w_t(w_{t-1}, \bar{\theta}) \) is undistorted, i.e., \( u'(c_{t-1}(w, \bar{\theta})) P^t_w(w_t(w_{t-1}, \bar{\theta})) = -\frac{\beta Q_{t-1}}{Q_t} \)
2. The marginal utility of increasing cost \( P \) by one unit, \(-\frac{1}{P_{tw}}\), is more than the marginal utility from increasing consumption at each state, \( E_t u'(c_t) \).

The first step is a natural implication of the no-distortions-at-the-top result. That is, since no other type wants to pretend to be the highest type, the margin between \( c_t - 1 \) and \( w_t \) is undistorted.

The marginal cost of increasing utility in the future by one unit, \(-\frac{Q_{t+1}}{\beta Q_t} P_{tw}^{t+1} \), is equal to the marginal benefit of decreasing utility in the current period by one unit, \( \frac{1}{u'(c_{w, \bar{\theta}})} \). Step 2 implies that a unit of saving relaxes incentive constraints in the future.

To show step 2, note that

\[
P_t(w) = A_t \psi \log(-w) + B_t \quad \text{where} \quad A_t = \frac{1}{Q_t} \sum_{s=t}^{T} Q_s
\]

and hence

\[
P_t(w) = A_t w.
\]

Moreover, it is easy to see that the margin between \( c_t \) and \( w_{t+1} + 1 \) is distorted downward for all \( \theta \) or

\[
-\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} P_{tw}^{t+1}(w_{t+1}(w_t, \theta)) \leq \frac{1}{u'(c_t(w_t, \theta))} \tag{31}
\]

Consider a perturbation that increases \( u(c_t) \) by one unit and decreases \( w_{t+1} \) by \( \frac{1}{\beta} \), this perturbation relaxes the incentive constraint (26) by decreasing marginal utility. The cost of such perturbation is \( \frac{1}{u'(c_t)} \) and its benefit is \( -\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} P_{tw}^{t+1} \) plus the benefit from relaxing the incentive constraints. Therefore, we must have the above inequality and equality holds at the top and the bottom since there are no distortions. Using the fact that \( u(c) = -e^{-\psi c} \), we know that \( P_{tw} = \frac{A_t}{\psi w} \) and \( \frac{1}{u'(c_t)} = -\frac{1}{\psi u(c_t)} \). Hence, the above inequality implies that

\[
-\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{w_{t+1}(w_t, \theta)} \leq -\frac{1}{u'(c_t(w_t, \theta))} \tag{32}
\]

or

\[
\frac{Q_{t+1}}{Q_t} A_{t+1} u'(c_t(w_t, \theta)) \geq \beta w_{t+1}(w_t, \theta) \tag{33}
\]

Integrating the above inequality and using promise keeping constraint implies that

\[
\left[ \frac{Q_{t+1}}{Q_t} A_{t+1} + 1 \right] \int u'(c_t(w_t, \theta)) f^{t}(\theta) d\theta > w
\]

or

\[
-\frac{A_t}{\psi} \int u'(c_t(w_t, \theta)) f^{t}(\theta) d\theta > w
\]

and therefore

\[
\int u'(c_t(w_t, \theta)) f^{t}(\theta) d\theta < \frac{-1}{P_{tw}^{t}(w_t)} \tag{34}
\]

By step 1,

\[
u'(c_{t-1}(w, \bar{\theta})) = -\frac{\beta Q_{t-1}}{Q_t} \frac{1}{P_{w}^{t}(w_{t-1}, \theta)}
\]

and by step 2

\[
u'(c_{t-1}(w, \bar{\theta})) = -\frac{\beta Q_{t-1}}{Q_t} \frac{1}{P_{w}^{t}(w_{t-1}, \theta)} > \int u'(c_t(w_t, \theta)) f^{t}(\theta) d\theta
\]
We summarize the above discussion in the following theorem:

**Theorem 2** Suppose that assumption 2 holds. Then any solution to program \((P2)\) satisfies the following

\[
\frac{\beta Q_{t-1}}{Q_t} \int u'(c_t(w_t(w_{t-1}, \bar{\theta}), \theta)) f^t(\theta) d\theta < u'(c_{t-1}(w_{t-1}, \bar{\theta}))
\]

Proof is given in the appendix.

There are two key ingredients in the above argument for step 2. The first ingredient is inequality (33). This inequality implies that between current utility, \(u(c)\), and promised utility, \(\beta w'\), the planner allocates more to \(u(c)\) relative to their weight in the objective \(\frac{Q_{t+1}}{Q_t} A_{t+1}\) function. Note that, for a general utility function, inequality (32) holds whenever \(P^t\) is concave in \(w\). The fact that (33) is implied by (32) is a direct consequence of Assumption 2. The second ingredient is the fact that with CARA utility \(u'(c)\) is proportional to \(u(c)\) and hence using the promise keeping constraint we can show that (34) holds.

As noted before, this result is in contrast with the seminal result of Golosov et al. (2003) where intertemporal wedge is always positive. In what follows, we illustrate how the two models are different and what leads to negative wedges in this model. To do so, it is useful to switch to a model with finite number of types \(\theta_1 < \cdots < \theta_N\). For illustrative reasons, we also make two other assumptions: 1. only local downward constraints are binding, 2. future promised utility is increasing in \(\theta\). Note that with local downward constraints binding, we must have

\[
u(c_i) + \beta w'_i = u(c_{i-1} + k_{i-1}(1 - \frac{\theta_{i-1}}{\theta_i})) + \beta w'_{i-1}
\]

Since \(w_i\) is increasing \(i\), the above equality implies that

\[
c_{i-1} + k_{i-1}(1 - \frac{\theta_{i-1}}{\theta_i}) < c_i
\]

This inequality implies that the current utility of an agent increases when he lies \(11\). Now consider an \(\varepsilon\) increase in \(c_i\) for all \(i\)'s. Then the LHS of the above inequality goes up by \(u'(c_i)\varepsilon\) while the RHS is increased by \(u'(c_{i-1} + k_{i-1}(1 - \theta_{i-1}/\theta_i))\varepsilon\). Above inequality and concavity of \(u\) imply that the incentive constraints are relaxed by such perturbation. The added cost of this perturbation is \(\varepsilon\) while overall utility is increased by \(\varepsilon E_{t-1} u'(c_t)\). Hence, if we set \(\varepsilon = \frac{1}{E_{t-1} u'(c_t)}\), cost increases by \(\frac{1}{E_{t-1} u'(c_t)}\) and overall utility increases by 1. Optimality of the allocations then implies that \(-P^t_w < \frac{1}{E_{t-1} u'(c_t)}\). That is the implied increase in cost from a unit increase in promised utility, \(-P^t_w\), must be less than the added cost from a uniform increase in consumption \(\frac{1}{E_{t-1} u'(c_t)}\).

In other words, saving relaxes incentive constraints.

\(11\)In the model where \(\theta \in [\bar{\theta}, \bar{\theta}]\), this inequality becomes \(c'(<bar{\theta}) < \frac{1}{\bar{\theta}} k(\bar{\theta})\). That is current utility from lying \(u(c(\bar{\theta}) + k(\bar{\theta})(1 - \bar{\theta}/\theta))\) is decreasing in \(\bar{\theta}\) when \(\bar{\theta} = \theta\).
In contrast, consider the model in Golosov et al. (2003) with the same assumptions: discrete types, local incentive constraints and increasing promised utility. In this model dis-utility of effort is separable from consumption. Hence, the local downward incentive constraints become the following

\[ u(c_i) - v(l_i) + \beta w'_i = u(c_{i-1}) - v(\frac{\theta_{i-1}}{\theta_i}l_i) + \beta w'_{i-1} \]

Note that in this model, since consumption is separable from the source of private information, the margin between \( c \) and \( w' \) is undistorted. Hence, the fact that \( w_i \) is increasing in \( i \) implies that \( c_i \) is also increase in \( i \). Now consider an \( \varepsilon \) increase in \( c_i \) for all \( i \)'s, as before. The RHS of the above constraint increases by \( u'(c_i)\varepsilon \) while its LHS is increased by \( u'(c_{i-1})\varepsilon \). Concavity of \( u \) together with \( c_i > c_{i-1} \) then implies that this perturbation tightens the set of incentive constraint. That is saving tightens the incentive constraints and therefore intertemporal wedges are positive.

The above analysis also suggests that when \( u'(\theta) \) is increasing in \( (P2) \), intertemporal wedges are negative at the top and the bottom. In fact, in the Appendix, we show that this result is true when only downward incentive constraints are binding or \( \mu_1(\theta) \geq 0 \). That is, we have the following proposition:

**Proposition 6** Suppose that \( \theta_t \) is i.i.d. Moreover, suppose that in the solution to \( (P2) \), \( w'(\theta) \) is increasing in \( \theta \) and the co-state \( \mu_1(\theta) \) is always positive. Then, the intertemporal wedge is negative at the top, i.e.,

\[
\frac{\beta Q_{t-1}}{Q_t} \int u'(c_t(w_t(w_{t-1}, \bar{\theta}, \theta))) f_t(\theta|\theta_{t-1}) d\theta < u'(c_{t-1}(w_{t-1}, \bar{\theta}))
\]

Proof can be found in the appendix.

So far, we have assumed that the process for productivity is i.i.d. The case where \( \theta \) is persistent is worth discussing. In this case, we can do the same perturbation as above. We again assume that when an agent lies his current utility increases, i.e., \( \frac{1}{\theta} k(\theta) > c'(\theta) \), then a uniform increase in consumption in all states relaxes the incentive constraints. However, this perturbation increase the overall utility, \( w \), and it changes the overall marginal promised utility, \( \Delta \). What this implies is that

\[
-P_w \int u'(c(\theta)) f_t(\theta|\theta_-) d\theta - P_{\Delta} \int u'(c(\theta)) f_{t-1}(\theta|\theta_-) d\theta < 1
\]

Hence, the sign of the intertemporal wedge depends on the sign of \( P_{\Delta} \) and \( \int u'(c(\theta)) f_{t-1}(\theta|\theta_-) d\theta \). However, it is implied by the local approach that \( P_{\Delta}(w_t, \Delta_t, \bar{\theta}) = 0 \). That is, the threat keeping constraint at \( t \) is slack for the entrepreneur with the highest shock at period \( t + 1 \). This result is an implication of the first order approach. An implication of the first order approach is that no other agent wants to pretend to be the highest type. Hence, the threat keeping constraint is not

---

One can show that \( \mu_1(\theta) \geq 0 \) whenever the value function is concave. In the appendix, we provide conditions under which the value function is concave and show how concavity of the value function leads to a positive sign for \( \mu_1(\theta) \).
binding at the top, i.e., \( P_{\Delta}^t(w, \Delta, \bar{\theta}) = 0 \). This implies that we again have

\[-P_w^t \int u'(c(\theta)) f'(\theta|\theta_-) d\theta < 1\]

and hence the intertemporal wedge is negative. Therefore, we have the following proposition:

**Proposition 7** Suppose that in the solution to \( (P2) \), \( \frac{d}{d\theta} w'((\theta) > \Delta'(\theta) \) and the co-state \( \mu_1(\theta) \) is always positive. Then, the intertemporal wedge is negative at the top, i.e.,

\[\frac{\beta Q_{t-1}}{Q_t} E_{t-1} u'(c_t(\theta^t)) < u'(c_{t-1}(\theta^{t-2}, \bar{\theta}))\]

### 4.1.2 The Role of Residual Component

So far, we have shown that when there is no residual component to productivity, intertemporal wedge is negative at the top and bottom. Here, we discuss how inclusion of residual shocks affect the sign of the intertemporal wedge. To do so, we start from a special case where productivity is only residual and there is no heterogeneity in \( \theta \), what we call a *pure moral hazard economy*. In this example and under Assumption 2, we can show that the movements in tightness of the incentive constraint only depends on the movements in \( Q_t \) over time and hence, in steady state MIEE becomes the same as the Inverse Euler Equation. Therefore, saving wedges are positive.

Consider a version of the model in section 3, where \( \theta_t = \theta \) is fixed and known and \( g^t = g \) is time independent. Then the recursive problem becomes

\[
P^t(w) = \max \frac{Q_{t+1}}{Q_t} \theta^\alpha k^\alpha - c - k + \frac{Q_{t+1}}{Q_t} \int P_{t+1}^t(w'(y)) g(y|k) dy
\]

subject to

\[
\begin{align*}
    u(c) + \beta \int w'(y) g(y|k) dy &= w \\
    \beta \int w'(y) g_k(y|k) dy &= u'(c_0)
\end{align*}
\]

In this problem, as before, the value function satisfies

\[P^t(w) = B_t + A_t \log(-w)\]

where \( A_t = \frac{1}{\bar{\psi}Q_t} \sum_{s=1}^{T} Q_s \). Moreover, since there are no wealth effects, the policy functions satisfy the following

\[
\begin{align*}
    c_t(\theta, w) &= -\frac{1}{\bar{\psi}} \log(-w) + c^*_t \\
    w'_t(w, y) &= (-w) \cdot w^*_{t+1}(y)
\end{align*}
\]
where \( c^*_t \) and \( w^*_t \) are independent of \( w \) but dependent on time. Recall the Modified Inverse Euler Equation. In this case, since there is no heterogeneity in \( \theta \), \( \mu_{1t} = 0 \). Moreover, we know that

\[
\mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} \text{Cov}(P_{w}^{t+1}, w_{t+1}|y^t)
\]

Given the above properties of the policy functions, it is easy to see that \( \text{Cov}(P_{w}^{t+1}, w_{t+1}|y^t) \) is independent of individual history, \( w_t \). This is due to the fact that \( P_{w}^{t+1} \) is proportional to \( \frac{1}{w_t} \) and \( w_{t+1} \) is proportional to \( w_t \). Hence \( \text{Cov}(P_{w}^{t+1}, w_{t+1}) \) is independent of \( w_t \). That is

\[
\mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} \text{Cov}(A_{t+1}^{w_t+1}, w_{t+1}^*)
\]

Therefore, the Modified Inverse Euler Equation becomes

\[
\frac{1}{u'(c_t)} \left[ 1 + \psi \frac{Q_{t+1}}{Q_t} \text{Cov}(A_{t+1}^{w_t+1}, w_{t+1}^*) \right] = \frac{Q_{t+1}}{Q_t} \beta E_t \frac{1}{u'(c_{t+1})}
\]

Note that the terms in the brackets are time dependent but independent of the individual history. In particular they depend on the aggregate state of the economy represented by \( Q_t \). However, if we assume that \( T = \infty \) and the economy is on aggregate in Steady State, i.e., \( Q_t/Q_{t+1} \) is constant over time, the term in the bracket becomes constant, since the Bellman equation described above becomes time independent. Hence, in steady state, the usual Inverse Euler Equation emerges and we have

\[
\frac{1}{u'(c_t)} = \frac{Q_{t+1}}{\beta Q_t} E_t \frac{1}{u'(c_{t+1})}
\]

Therefore, the intertemporal wedge must be positive in steady state. Notice that during transition to steady state, the sign of the wedge might change since the tightness of the incentive constraint depends on the aggregate state of the economy.

A comparison with the model with productivity shocks provides better intuition regarding the differences that cause the change in the sign of the wedge. The best way to describe the negative wedge result in the model with productivity shocks is to consider a small decrease in current consumption by \( \varepsilon \) accompanied by a uniform increase in consumption in the next period by \( q^{-1} \varepsilon \). Such perturbation has two effects: utility effect and incentive effect. The utility effects are standard: there is a utility benefit \( \varepsilon q^{-1} \beta E_t u'(c_{t+1}) \) and a utility cost \( \varepsilon u'(c_t) \). As for incentive effects, when productivity is currently at the top (bottom), the incentive constraint is not binding currently. Hence, such a perturbation does not have any effect on current incentive constraints. However, due to the reasons discussed above it relaxes incentive constraints in the next period. Hence, the utility benefit of this perturbation, \( \varepsilon q^{-1} \beta E_t u'(c_{t+1}) \) must be less than its utility cost \( \varepsilon u'(c_t) \) and the intertemporal wedge has to be negative.

We can apply the same perturbation in the pure moral hazard economy. The difference is that the incentive effects of such perturbation are more complicated. In fact in the pure moral
hazard model, since the incentive constraint is always binding, such perturbation has an effect on current incentives. The perturbation tightens up the incentive constraint since it increases current marginal utility and decreases the slope of promised utility profile in the next period. This perturbation also affects incentives in the next period and relaxes the incentive constraints since it decreases marginal utility. The above analysis shows that in steady state the current costs from tightening the incentive constraints are higher than the future benefit from relaxing the incentive constraints. Hence, the intertemporal wedges are positive.

So far, we have analyzed two extreme cases: when the residual risk is shot down and when entrepreneurs do not know anything in advance about their future productivity. We have shown that the two extreme cases have different implications on the intertemporal wedge. The novel result of this paper is in fact that intertemporal wedge is negative at the top and bottom when productivity has no residual component. As we have seen, when the known component of productivity is shut down, in Steady State, Inverse Euler Equation emerges and intertemporal wedge is positive. Hence, the result on the general model is indeterminate. As the perturbation argument shows, the sign of the wedge depends on the relationship between current incentive costs of decreasing consumption versus benefits from relaxing the incentive constraints in the future. In fact, in the general model as in the pure moral hazard model, would relax both types of incentive constraints in the future and would tighten the current moral hazard constraint. In section 6, we use a reasonably calibrated version of the model and show that intertemporal wedge is negative at the bottom and positive at the top.

4.2 Progressive Taxes on Entrepreneurial Income

In this section, we study whether the progressivity of the tax schedule with respect to entrepreneurial income generalizes to the dynamic model. We do so, by characterizing the shape of consumption in each period as a function of income. As, the two period example in section 2.3 shows, movements in current consumption as a function of current income, depends on the EIS and the hazard rate \( g_k \). In this section, we study the economy with exponential utility function. We show that in the general model described above, when \( \theta \) is i.i.d., the inverse of marginal utility is a linear function of the hazard rate. Since, the inverse of marginal utility is a convex function, the shape of the consumption schedule, i.e., the shape of the tax function, is solely determined by the shape of the hazard rate. In particular, when the hazard ratio is concave, the consumption is concave in income realization and tax schedule is progressive. Moreover, when \( \theta \) is persistent, the consumption schedule is concave for the highest and lowest value of productivity.

Consider the economy in section 3. The first order conditions imply that at each date

\[
- P_{w_{t+1}} = a_t + b_t \frac{g_k(y_{t+1}|k_{t+1}, \theta_t)}{g(y_{t+1}|k_{t+1}, \theta_t)} \tag{37}
\]

where \( a_t \) and \( b_t \) are independent of \( y_{t+1} \). In fact, \( b_t = \beta \mu_{2t} \) and \( a_t \) is a function of the tightness of the
first incentive constraint. These multipliers, depend solely on the past history of shock as well as
the current realization of $\theta_t$. When $\theta_t$ is i.i.d., the value function satisfies $P^t(w) = A_t \log(-w) + B_t$
and consumption policy function satisfies
\[ c_t(w, \theta) = -\frac{1}{\psi} \log(-w) + \hat{c}_t(\theta) \]
This implies that $u'(c_t(w, \theta)) = (-w)u'(\hat{c}_t(\theta))$ and that $P_w^t = \frac{A_w}{w}$. Therefore, (37) becomes the
following
\[ \frac{A_{t+1}u'(\hat{c}_{t+1}(\theta))}{u'(c_{t+1}(w, \theta))} = a_t + b_t \frac{g_k(y_{t+1}|k_{t+1}, \theta_t)}{g(y_{t+1}|k_{t+1}, \theta_t)} \]
Hence, $\hat{a}_t$ and $\hat{b}_t$ exist that are history dependent and independent of $y_{t+1}$ such that
\[ \frac{1}{u'(c_{t+1}(w, \theta))} = \hat{a}_t + \hat{b}_t \frac{g_k(y_{t+1}|k_{t+1}, \theta_t)}{g(y_{t+1}|k_{t+1}, \theta_t)} \] (38)
Therefore, since $c_{t+1}$ is concave in $y_{t+1}$ when $\frac{g_k}{g}$ is concave in $y$. In particular, for the examples
given in section 2.3, the same analysis holds and consumption schedule is concave in $y$.
When $\theta_t$ is persistent, we can show that the value function $P^t(w, \Delta, \theta_-) = A_t \log(-w) + B_t(\Delta, \theta_-)$. Hence, in this case, the shape of $B_t$ affects $P_w$ and hence the above analysis does not apply. However, when $\theta_- = \bar{\theta}$, we have $P^t\Delta = 0$. Therefore, in that case, the above analysis applies and (38) determines the shape of the consumption schedule.
An important assumption above is that given history of actions, $y_t$ and $\theta_t$ are independent.
This implies that the marginal cost of increase utility by a unit, $P_w^t$, is related to the inverse of
marginal utility in a way described above. However, when $y_t$ and $\theta_t$ are perfectly correlated, this
is not necessarily true. Considering a correlation between $y_t$ and $\theta_t$ would further complicate the
model and we do not pursue this idea here.
Given the above analysis, an investigation of entrepreneurial income processes is required in
order to determine the progressivity of the tax schedule. Moreover, a key assumption that leads to
this result on marginal tax rates is that, markets are unable to provide any insurance. However,
as shown by Kaplan and Strömb erg (2003)’s analysis of Venture Capital contracts and Bitler et al.
(2005)’s analysis of SCF data shows, certain features of observed private equity contracts are
consistent with the optimal contracting theory. This evidence suggests that markets are able to
provide some insurance. Hence, a natural question is what is the role of government in providing
insurance. Although an important question, this question is beyond the scope of this paper. In
Shourideh (2010), we partially try to address this question by considering an environment where
there is a role for government and study its implication for optimal taxation.

35
4.3 A Tax Implementation

In this section, we analyze the implementation of efficient allocations discussed above. In particular, we show that the tax functions used in the two period example can be extended to the multi-period model. To do so, we impose that agents have only access to a risk free asset \( b_t \) at each date traded at price \( Q_t \). We also assume that the planner can observe \( b_t \) as well as income at each period \( t \). Thus, our aim is to find a tax schedule \( \{ T_t(y^t, b^t) \}_{t=0}^T \), where \( b^t \) is the history of asset holdings for each agent and \( y^t \) is the history of income. The value \( T_t \) is the tax paid by the agent at period \( t \). Later, we discuss how the properties of the optimal allocations discussed above translate into properties of the tax function.

Facing such schedule, each agent solves the following maximization problem

\[
\max_{c_t(\theta^t, y^t), k_t+1(\theta^t, y^t)} \beta \sum_{t=0}^{T} \int_{\Theta^{t+1} \times Y^t} u(c_t(\theta^t, y^t))d\mu_t(\theta^t, y^t; k^t(\theta^{t-1}, y^{t-1}))
\]

subject to

\[
c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) + \frac{Q_{t+1}}{Q_t} b_{t+1}(\theta^t, y^t) \leq b_t(\theta^{t-1}, y^{t-1}) + y_t - T_t(y^t, b^{t+1}(\theta^t, y^t))
\]

Now consider the optimal allocation \( \{ c_t^*(\theta^t, y^t), k_{t+1}^*(\theta^t, y^t) \}_{t=0}^T \) which is the solution to the planning problem \((P)\). A difficulty, in extending the implementation to a multi-period model is an indeterminacy in the level of risk free asset. Note that in the two period model, the asset level at period 0 is pinned down. For each agent, the level of non-entrepreneurial wealth is given by \( q_0 - c_0(\theta) - k_1(\theta) \). When the number of periods is more than 2 and in an intermediate period, the allocation only pins down the number \( Q_t b_{t+1} + T_t(y^t, b^{t+1}) \). In order to show that the tax system is progressive in income, we construct \( b_{t+1} \) such that its slope as a function of \( y_t \) is the same as the slope of \( -\frac{Q_{t+1}}{Q_t} k_{t+1} \). This assumption helps us later in proving that the tax schedule is progressive.

Formally, the construction of \( b_t^* \) and \( T_t \) given the optimal allocation is as follows:

1. At \( t = 0 \), non-entrepreneurial wealth is defined as

\[
b_1^*(\theta_0) = \frac{Q_0}{Q_1} (e_0 - c^*_0(\theta_0) - k^*_1(\theta_0))
\]

2. At intermediate period \( 1 \leq t \leq T - 1 \), given a history \( (\theta^{t-1}, y^t) \) and the optimal allocation:

\[
b_{t+1}^*(\theta^t, y^t) = -\frac{Q_t}{Q_{t+1}} k_{t+1}^*(\theta^t, y^t) - \xi_t(\theta_t; \theta^{t-1}, y^{t-1}),
\]

The function \( \xi_t \) is arbitrary \( C^1 \) function that makes \( b_{t+1}^*(\theta^t, y^t) \) monotone. Obviously we must
have

\[
\int_{\Theta^{t+1} \times Y^t} \left[ \xi_t(\theta^t, y^{t-1}) + \frac{Q_t}{Q_{t+1}} k^*_{t+1}(\theta^t, y^t) \right] d\mu_t(\theta^t, y^t; k^*(\theta^{t-1}, y^{t-1})) = 0.
\]

The above construction of risk-free asset holdings, simply pins down the tax function. However, in order for the tax function to be well-defined, the following assumption on allocations is needed:

**Assumption 3** For all histories \((\theta^{t-1}, y^t), b^*_{t+1}(\theta^{t-1}, \theta_t, y^t) \neq b^*_{t+1}(\theta^{t-1}, \hat{\theta}_t, y^t), \forall \hat{\theta}_t \neq \theta_t. Moreover, \(b^*_{t+1}(\theta^{t-1}, \theta_t, y^t)\) is continuous in \(\theta_t\) for all \(t\).

The first part of assumption 3 is similar to an assumption in Kocherlakota (2005) regarding income – Assumption 1, page 1601. However introduction of \(\xi_t\) makes it easier to ensure that the above assumption is satisfied. The second part, ensures that \(b^*_{t+1}(\theta^{t-1}, \cdot, y^t)\) belongs to an interval \(I_t(\theta^{t-1}, y^t)\).

Note that by intermediate value theorem, for any value \(\hat{b} \in I_t\), there is a unique \(\hat{\theta}_t\) such that \(b^*_{t}(\theta^{t-1}, \hat{\theta}_t, y^t) = \hat{b}\). Given the above assumption, we can define the following well-defined tax function

\[
T_t(y^t, b^{t+1}) = \begin{cases} 
    b^*_{t}(\theta^{t-1}, y^{t-1}) + y_t - c^*_t(\theta^t, y^t) \\
    -k^*_{t+1}(\theta^t, y^t) - \frac{Q_t}{Q_{t+1}} b^*_{t+1}(\theta^t, y^t) \\
    2(y_t + b_t)
\end{cases}
\]

if \(\exists \theta^t \in \Theta^{t+1} \ni b^{t+1} = b^{t+1}(\theta^t, y^t)\)

otherwise

where \(b^{t+1} = (b_1, \ldots, b_{t+1})\). Note that taxes paid in the current period depend on the current level of non-entrepreneurial asset holding \(b_{t+1}\).

In what follows, we show that the above tax function implements the optimal allocation. That is given above \(T_t\), the optimal allocation is a solution to (39).

Given assumption 3, it is easy to see how the above tax system implements the allocation. First, notice that the above definition of \(T_t(\cdot, \cdot)\) ensures that \(b_{t+1} \in I_t\), otherwise the agent loses all his income. Hence, we only restrict attention to asset choices \(b_{t+1} \in I_t\). In particular, suppose an agent picks a sequence of non-entrepreneurial asset levels \(\{\hat{b}_{t+1}(\theta^t, y^t)\} \neq \{b^*_{t+1}(\theta^t, y^t)\}\) where \(\hat{b}_{t+1}(\theta^t, y^t) \in I_t(\theta^{t-1}, y^t)\). We first show that such sequence of asset holding is equivalent to a reporting strategy \(\hat{\sigma}_t(\theta^t, y^t)\). Then, incentive compatibility of the optimal allocation implies that such strategy is weakly dominated by truth-telling or \(\{b^*_{t}(\theta^t, y^t)\}\). Starting from period 0 and the fact that \(\hat{b}_1(\theta_0) \in I_0\), implies that \(\hat{b}_1(\theta_0) = b^*_1(\hat{\sigma}_0(\theta_0))\) for some function \(\sigma_0\). Given \(\sigma_0\) and \(\hat{b}_2(\theta^1, y^1)\), there must exist a \(\hat{\sigma}_1(\theta^1, y^1)\) such that \(\hat{b}_2(\theta^1, y^1) = b^*_2(\hat{\sigma}_0(\theta_0), \hat{\sigma}_1(\theta^1, y^1), y^1)\). Similarly, we can construct a reporting strategy \(\hat{\sigma}_t(\theta^t, y^t)\) for all possible histories. Hence the choice of asset positions \(\{\hat{b}_{t+1}(\theta^t, y^t)\}\) is equivalent to the choice of a reporting strategy \(\hat{\sigma}^t\).

Note that given \(\hat{b}_{t+1}(\theta^t, y^t)\) the budget constraint for the agent is given by

\[
c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) + \frac{Q_{t+1}}{Q_t} b^*_{t+1}(\theta^t, y^t) \leq b^*_{t}(\hat{\sigma}^t(\theta^{t-1}, y^{t-1}), y^{t-1}) + y_t - T_t(y^t, b^{t+1}(\hat{\sigma}^t(\theta^t, y^t), y^t))
\]
Hence by construction of the tax function in (41), the total amount available for consumption and investment is given by

\[
b^*_t(\hat{\sigma}^t(\theta^t, y^t), y^t) + y_t - T_t(y^t, b^*_{t+1}(\hat{\sigma}^t(\theta^t, y^t), y^t)) - \frac{Q_{t+1}}{Q_t} b^*_t(\hat{\sigma}^t(\theta^t, y^t), y^t) = c^*_t(\hat{\sigma}^t(\theta^t, y^t), y^t) + k^*_{t+1}(\hat{\sigma}^t(\theta^t, y^t), y^t)
\]

That is given \(\hat{b}^t\) and \(\hat{\sigma}^t\), the agent solves the following problem

\[
\max_{t=0}^T \beta^t \int_{\Omega^{t+1} \times Y^t} u(c_t(\theta^t, y^t))d\mu_t(\theta^t, y^t; k^t(\theta^{t-1}, y^{t-1}))
\]

subject to

\[
c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) \leq c^*_t(\hat{\sigma}^t(\theta^t, y^t), y^t) + k^*_{t+1}(\hat{\sigma}^t(\theta^t, y^t), y^t)
\]

Therefore, the highest utility achievable to the agent is

\[
\max_k U(\{c^*_t, k^*_{t+1}\}; \hat{\sigma}, \hat{k})
\]

Hence, by incentive compatibility the utility received by the agent is the highest if the agent chooses \(\{b^*_{t+1}(\theta^t, y^t)\}\), i.e., tells the truth, and follows the optimal investment. This implies that the tax function implements the optimal allocation. We summarize this discussion in the following theorem:

**Theorem 3** Consider the optimal allocation \(\{c^*_t(\theta^t, y^t), k^*_{t+1}(\theta^t, y^t)\}\) and suppose that assumption 3 holds. Then the tax function as defined by (41) implements the optimal allocation.

This implementation is similar to Kocherlakota (2005)’s implementation in an economy with labor income risk. In both implementations, the tax function at a period is a function of income and risk free asset holdings in that period. However, as Werning (2010) has shown, for the economy with labor income risk there is another implementation that separates the tax paid on asset holding and on labor income. However, in our economy this separation is not possible since allocations are not separable in \(\theta\) and \(y\). Hence this non-separability is necessary for implementation of the optimal allocation.

To see how the properties of the allocations discussed above translate into properties of the tax function. To do so, we assume that the optimal allocations are \(C^1\) w.r.t. histories. That is \(c^*_t(\theta^t, y^t)\) and \(k^*_{t+1}(\theta^t, y^t)\) are continuously differentiable with respect to each \(\theta_s\) and \(y_s\) for all \(0 \leq s \leq t\). By definition, \(b^*_t\) and \(T_t\), are also continuously differentiable by definition since they are constructed from \(c^*_t\) and \(k^*_t\) using \(C^1\) transformations. Hence, any solution of (39) must satisfy
the following first order condition:

\[ u'(c_t(\theta^t, y^t)) = \beta \frac{Q_t}{Q_{t+1}} \int_{\Theta \times Y} u'(c_{t+1}(\theta^{t+1}, y^{t+1}))(1-T_{b_{t+1}}(y^{t+1}, b^{t+2}))(\theta_t, k_{t+1})f^{t+1}(\theta_{t+1}|\theta_t)d\theta_t d\theta_{t+1} \]

Therefore, the intertemporal wedge is given by

\[ \tau^*_t(\theta^t, y^t) = \frac{E_t[u'(c_{t+1})T_{b_{t+1}}(y^{t+1}, b^{t+2})]}{E_t[u'(c_{t+1})]} \]

and is a weighted average of the derivative of the function \( T(\cdot, \cdot) \) weighted by marginal utility. Hence, our result on negative marginal tax rates imply that on average \( T_b \) must be negative whenever \( b_{t+1} = b^{t+1}_*(\theta^{t-1}, \tilde{\theta}, y^t) \).

Next, we turn to the shape of the tax function with respect to income realization. Given the way we have constructed non-entrepreneurial asset holdings, we know that

\[ \frac{\partial}{\partial y_t} b^{t+1}_*(\theta^t, y^t) = -\frac{Q_t}{Q_{t+1}} \left[ \frac{\partial}{\partial y_t} k^{t+1}_*(\theta^t, y^t) \right] \]

Hence, using (41), we must have

\[ \frac{\partial}{\partial y_t} T(b^t, y^t) = 1 - \frac{\partial}{\partial y_t} c^{t}_*(\theta^t, y^t) - \frac{\partial}{\partial y_t} k^{t+1}_*(\theta^t, y^t) - \frac{Q_{t+1}}{Q_t} \frac{\partial}{\partial y_t} b^{t+1}_*(\theta^t, y^t) \]

That is, the shape of the tax function with respect to income realization is the same as the shape of the consumption schedule as a function of income realization. Hence, whenever consumption is concave in income, the marginal tax rate or \( \frac{\partial}{\partial y_t} T(b^t, y^t) \) is increasing and therefore the tax schedule is progressive.

## 5 Implementation with Private Contracts

In this section, we study how private contracts can implement the optimal allocation. We do so by showing that there is an implementation of the optimal allocation that uses the types of contracts used in typical venture capital contracts as documented by Kaplan and Strömberg (2003). We do this in the context of the pure moral hazard model described in section 4.1.2. Moreover, we assume that assumption 2 holds and the hazard ratio, \( g_k \), is concave in \( y \).

We first describe the types of securities used in our implementation – throughout we assume that the environment is comprised of an outside lender and the entrepreneur:

1. Equity: The equity holders collect dividends paid in each period. At each period, the entrepreneur and the lender own parts of the company and the ownership is evolving over
2. Short Term Convertible Debt: this security is risk free debt together with $N$ options. Upon exercising option $i$, $1 \leq i \leq N$, the holder can buy a certain number of shares –fraction $\hat{s}^i$ of total equity, at a pre-specified price, $e^i$. Both sequences $\{\hat{s}^i\}_{i=1}^N$ and $\{e^i\}_{i=1}^N$ are increasing in $i$.

3. Credit Line/Saving Account: A bank account that the entrepreneur can borrow and save with a variable interest rate. The interest rate only depends on the ownership structure of the firm, i.e., the fraction of the equity owned by the entrepreneur.

These securities are very standard and as documented by many authors\textsuperscript{13}, are widely used in venture capital contracts. In what follows, we show that the above securities can approximately implement the optimal allocation – as $N$ tends to $\infty$, the implemented allocation converges to the optimal allocation. Given the above securities, the timing is as follows:

1. At the beginning of the each period and before realization of income, the entrepreneur buys all the shares from the outside lender.
2. Income is realized.
3. The outside lender decides whether to convert the convertible debt.
4. Investment is made by the entrepreneur.
5. Dividends are paid out.
6. The entrepreneur can decide to save or draw funds from the credit line and new convertible debt is issued by the outside lender.

Given the above timing, it is useful to introduce a bit of notation:

- Amount of convertible debt issued by the outside lender: $D$, with price $p$,
- Total equity value of the firm before realization of income in each period: $V_t$,
- Share of the entrepreneur in the company: $s_t$,
- Entrepreneur’s debt level: $B_t$; negative values are associated with saving.
- Interest rate on credit line/saving account: $R(s_t)$,
- Conversion decision at option $i$: $j_t(i) \in \{0, 1\}$. Note that the outside lender can only exercise one of the options and therefore $\sum_{i=1}^N j_t(i) \in \{0, 1\}$.

\textsuperscript{13}See Kaplan and Strömberg (2003), Sahlman (1990), and Gompers (1999), among others.
The securities defined above and the sequence of actions lead to the following budget constraint for the entrepreneur:

\[ \sum_{i=1}^{N} e^{i} j_{i}(i) + y_{t} + B_{t+1} - B_{t} + pD = R(s_{t-1})B_{t} + d_{t} + k_{t+1} + D \left( 1 - \sum_{i=1}^{N} j_{i}(i) \right) + (1 - s_{t-1})V_{t} \]

Moreover, \( c_{t} = s_{t}d_{t} \). Note, also, that due to the buy-back of the stocks in the beginning of the period \( s_{t} = 1 - \sum_{i=1}^{N} \hat{s}^{i}j_{i}(i) \). The LHS of the budget constraint is revenue available to the firm from various sources: revenue from sale stock in case of conversion, income, credit drawn from the credit line, and money raised through issuance of convertible, respectively. The RHS of the budget constraint is the expenses paid: interest payment on the credit balance, dividends, investment, payments to convertible debt holders in case of no conversion, and the cost of share buy-back.

We need to further describe the conversion decision by the outside lender. Since this debt matures every period, the conversion decision can be easily described by a one time optimality condition. The holder of the debt will convert at option \( i \) if and only if the value of converting:

\[ (1 - \hat{s}^{i}) (d_{t} + qV_{t+1}) - e^{i} \]

exceeds the face value of the debt \( D \). Moreover, the value of the stock to outside lender is given by

\[ V_{t} = E_{t-1} \sum_{\tau=0}^{\infty} q^{-\tau} d_{t+\tau} \]

In order to show that the above implementation works, we first show that there exists a fixed interest rate \( \hat{R} \) and a transfer function \( T(y) \) from the entrepreneur, such that any stationary efficient solution to \( (P3) \) can be implemented where entrepreneur can freely borrow and save at rate \( \hat{R} \) and \( T(y) \) is taken away from the entrepreneur in each period.

**Lemma 2** Consider a solution to \( (P3) \) where \( Q_{t+1}/Q_{t} = \hat{q} \) given by allocations \( \{c_{t}(y^{t}), k^{*}, w_{0}\} \). Then, there exists a function \( T(y) \), interest rate \( \hat{R} \) and debt level \( B_{0} \) such that the allocation is the solution to the following optimization problem

\[ \max_{k_{t+1}, c_{t}, B_{t+1}} \sum_{t=0}^{\infty} \beta^{t} \int_{Y^{t}} u(c_{t}(y^{t}))d\mu_{t}(y^{t}; k^{t}) \quad (P4) \]

subject to

\[ c_{t} + k_{t+1} + (1 + \hat{R})B_{t} = B_{t+1} + y_{t} - T(y_{t}) \]

Proof can be found in the appendix.

The idea behind this lemma is simple. It can be shown that in the pure moral hazard model, due to no wealth effect, the intertemporal wedge is constant in all states. This implies that facing an interest rate \( 1 + \hat{R} = (1 - \tau_{s})q^{-1} \), the agents Euler equation will be satisfied. Moreover, given the policy functions in (35)-(36), income at period \( t \) affects \( c_{t}(y^{t}) \) in an additively separate way. Moreover, stationarity implies that the tax function is independent of time. Note that the concavity of the hazard ratio, implies that \( T(\cdot) \) is a convex function of income.
Lemma 2 guides us toward our main implementation result. That is, we show that the above securities can replicate the above transfer function and interest rate. In order to prove our main theorem, we need to make one further assumption and that is:

**Assumption 4** The optimal allocation satisfies \( \frac{\partial}{\partial y_t} c_t(y_t) \leq 1 \), \( \forall y_t \in [0, \bar{y}] \).

The above assumption implies that the in each period, the total payment to the outside lender \( y_t - c_t - k_{t+1} \) is increasing in \( y_t \). When this assumption is violated, there will be a region such that the slope is bigger than 1. In that case the payment to outside lenders decreases following an increasing in \( y_t \). Although it is possible to modify the above implementation in order to implement the optimal allocation, for simplicity we make the above assumption.

The following theorem, contains our main implementation result:

**Theorem 4** Consider a sequence \( y^1 < \cdots < y^N \) and a solution to \((P3)\) where \( Q_{t+1}/Q_t = \hat{q} \) given by allocations \( \{c_t(y^t), k^*, w_{0}\} \). Then there exists \( \{e^1\}, \{s^t\}, D, p, R(s) \) and \( B_0 \) such that the above security structure exactly implements the allocation for all histories \( y^t \in \{y^1, \cdots, y^N\}^t \).

Proof can be found in the appendix.

The idea behind this implementation can be seen from lemma 2. First, we note that by concavity of hazard ratio \( T(y) \) is convex and by assumption 4 its slope is always positive. Moreover, the payoff schedule resulting from the convertible debt is a piece-wise linear function with increasing slope. Therefore, the role of the convertible debt is to approximate the function \( T(y) \). However, conversion implies that the outside lenders will be equity holders is in the future. This reduces the incentive for the entrepreneur to invest optimally in the firm. The role of share buy-back is dispose of this problem. Since, the buy-back is done before new investment is made, Finally, since the ownership of the entrepreneur is changing over time the interest rate needs to be changing. In fact, the Euler equation from entrepreneur’s decision problem is given by

\[
s_t^{-1}u'(c_t) = \beta R(s_t) E_t s_{t+1}^{-1}u'(c_{t+1})
\]

Given the policy functions in \((35)-(36)\), \( R(s_t) = \frac{1}{\beta s_t E_t [s_{t+1}^{-1}(-w_{t+1})]} \).

The above theorem implies that our security structure can approximately implement the optimal allocation. So, we have the following corollary:

**Corollary 1** As \( N \to \infty \), the allocation implemented by the above security structure converges to the optimal allocation.

Note that the above implementation is not unique. In fact, we can combine all of the above securities into one security. We can, also, implement the optimal allocation using only debt/saving with a variable interest rate, as used in Quadrini (2000). The value of this implementation is that it
points to the role of convertible debt and share buy back\textsuperscript{14}. Moreover, it shows that the allocation can be approximately implemented using securities widely used in venture capital contracts.

The above analysis shows that the implication of the model on taxation is mixed. In fact, we have shown that it is possible for market arrangements that are commonly used in financial contracts to achieve constrained efficiency. Given such arrangements, there is no reason for the government to use taxes to achieve constrained efficiency. In that case, government only crowds out private markets. In Shourideh (2010), we try to resolve this issue by allowing for unobservable trades among entrepreneurs. In this case, the price of the risk free bond affects entrepreneur’s incentive for investment and hence private contracts cannot implement constrained efficient allocations. We analyze the optimal policy in a two period model and show that optimal tax policy is linear tax function on income.

6 Numerical Simulations

In this section, to fully characterize the properties of optimal allocations, I use a calibrated version of the model in order to calculate optimal intertemporal wedges as well as taxes on entrepreneurial income. To do so, I consider an EJ economy in which \(\theta_t\) is i.i.d and \( \log \theta_t \sim N(\log A - \frac{1}{2} \sigma_\theta^2, \sigma_\theta^2) \). Moreover, I assume that \( \varepsilon \sim \Gamma(\sigma_\varepsilon^{-2}, \sigma_\varepsilon^2) \). I keep the assumption that the utility function is exponential, \(u(c) = -e^{\psi c}\). Next, I describe how each parameter is calibrated.

To calibrate the economy, I need to calibrate the parameters \((\alpha, \beta, \psi, A, \sigma_\theta, \sigma_\varepsilon)\). I assume that \(\beta = .96\) so that each period is associated with a year. As we have mentioned before, \(\alpha\) should be thought of as a span of control parameter. Hence, to calibrate \(\alpha\), we consider an entrepreneur with production function \(e(k^{\eta l^{1-\eta}})^{\nu}\) who adjusts the labor input, \(l\), once shock \(e\) is realized\textsuperscript{15}. This maximization decision implies that the production function can be written as \(\tilde{e}k^{\frac{\eta \nu}{1-(1-\eta)\nu}}\). Hence, the implied share of inputs (other than managerial talent) as a fraction of income is given by \(\frac{\eta \nu}{1-(1-\eta)\nu}\). Further, notice that since we have assumed that capital fully depreciates, we need to adjust \(\alpha\) in order to take that into account. Hence, in this model \(\alpha\) is given by

\[
\alpha = \frac{\text{Payments to factors other than managerial talent} + \text{K-Depreciation}}{\text{Output} + \text{K-Depreciation}}
\]

As we have discussed above,

\[
\text{Payments to factors other than managerial talent} = \frac{\eta \nu}{1-(1-\eta)\nu} \text{Output}
\]

Given the value \(\alpha\), I pick \(A\) so that output for the average firm is normalized to 1.

\textsuperscript{14}Green (1984) in a two period model shows that a convertible debt with one conversion option does better than non-convertible debt in providing investment incentives to the shareholders. His results have the same flavor as ours. He, however, does not provide optimal security design based on underlying frictions.

\textsuperscript{15}See footnote 6.
Calibrating the precise value for cross-sectional variance of productivity is problematic, since there are no precise estimate for this process. Moskowitz and Vissing-Jørgensen (2002), study the private-equity returns using the Survey of Consumer Finances but are unable to provide precise estimates for variances of returns at individual level. For their benchmark calculations, they use 0.3 for cross-sectional standard deviation of private-equity firms. Angeletos (2007) uses 0.2 in a model where only residual risk in productivity is present. I, assume that cross-sectional standard deviation of productivity \( \varepsilon \theta^\alpha \) can take value of \{0.2, 0.3, 0.4, 0.5\}. Moreover, inspired by the estimates of Evans and Jovanovic (1989), I assume that \( \sigma_{\varepsilon} = \alpha \sigma_{\theta} \). This assumption pins down the value of \( \sigma_{\varepsilon} \) and \( \sigma_{\theta} \). For the risk aversion parameter, I use \( \psi = 10^{16} \).

Using these parameter values, I compute the model. In doing so, I use a truncated distribution for \( \theta \). As noted above, I use a first order approach to simplify the set of incentive constraints. I assume that the model is in steady state, i.e., \( \frac{Q_{t+1}}{Q_t} = q \) is constant. Since I have assumed an exponential utility function, the policy functions satisfy the following:

\[
\begin{align*}
    c(w, \theta) &= -\frac{1}{\psi} \log(-w) + \hat{c}(\theta) \\
    w'(w, \theta, y) &= (-w)\hat{w}(\theta, y)
\end{align*}
\]

This implies that the difference in consumption across periods is given by

\[
\int_{\Theta} \int_{Y} \frac{1}{\psi} \log(-\hat{w}(\theta, y))g(y|\theta, k(\theta))f(\theta)dyd\theta
\]

I find \( q \) so that the above integral is zero. This implies that the total expenditures in the economy do not change over time. Whether total income \( \int \theta^\alpha k^\alpha \) is greater or less than total expenditure \( \int c + k' \), depends on the initial value of promised utility. Because, I am considering exponential utility, the distortions, i.e., intertemporal wedge and the slope of consumption schedule, do not depend on the initial value of the promised utility.

**Intertemporal Wedge.** Figure 2, below shows how intertemporal wedge depends on productivity.

In fact, the intertemporal wedge is negative and quite large for low values of \( \theta \) and positive for high values. To understand these results, we should note that intertemporal wedge is closely related to the variance of growth rate of consumption. In this model, there are two forces that create variability for consumption. First effect comes from that fact that \( \theta \) is private information. This implies that consumption should depend on \( \theta \) in order to give incentive for more productive types to invest. The second effect is a result of moral hazard in addition to heterogeneity in \( \theta \). Moral hazard implies that the planner, should given incentive to entrepreneurs to invest optimally by creating spread in their consumption in future as well as increasing their consumption in the

\[16\]I have considered various values for \( \psi \), in the range of 1-20. Surprisingly, the results do not change that much for these parameter values.
Figure 2: Intertemporal wedge, full model - x-axis is $\theta$ in units of standard deviation from the mean.
current period. Moreover, due to decreasing return to scale, the planner wants higher productivity
types to invest more in their business. This suggests that the moral hazard problem is tighter for
higher ability types. Hence, it is optimal to have current consumption positively correlated with
$\theta$. In order to investigate the contribution of each of these effects, we consider two extreme cases:
A case where we shut down the residual risk, $\varepsilon$, and a case where $\varepsilon$ is present and $\theta$ is public
information. For the case where there is no residual risk, the intertemporal wedge is shown in
Figure 3.

![Figure 3: Intertemporal wedge, absent residual component - x-axis is $\theta$ in units of standard
deviation from the mean.](image)

We can see that relative to Figure 2, the distortions are very small. This implies that the
variability in consumption is small in a model with no residual risk. Note that, these number are
still large in comparison to Farhi and Werning (2010a) or Golosov et al. (2010), since productivity
shocks are i.i.d. Persistence decreases variability of consumption growth rate at an individual
level. Next, we consider the case where $\theta$ is public information. Figure 4 depicts the intertemporal
wedge in this environment.

We can see that similar to Figure 2, the intertemporal wedge is negative and large for low
values of $\theta$ and positive and large for high values of $\theta$. That is the variability of consumption as
a function of $\theta$ has a big effect on variability of consumption and therefore, intertemporal wedges
are very large at extremes. In fact, since the distortions from heterogeneity in ability are small,
Figure 4: Intertemporal wedge, ability is public info - x-axis is $\theta$ in units of standard deviation from the mean.
this case is almost identical to the case with private information. Figure 5 plots the intertemporal wedge for the case that $\theta$ is fully persistent (this case is equivalent to the pure moral hazard model discussed in section 4.1.2). We can see that as shown before, intertemporal wedge is positive and small relative to the case where $\theta$ is i.i.d. This analysis suggests that one should analyze the model where $\theta$ is stochastically evolving over time but persistent – still in progress.

![Figure 5: Intertemporal wedge, fully persistent ability - x-axis is $\theta$ in units of standard deviation from the mean.](image)

7 Conclusion

In this paper, I have studied optimal taxation of entrepreneurial income. 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. Although the model can be interpreted as one of optimal taxation, I have shown that standard securities commonly used in venture capital contracts can implement efficient allocations.

Although, I have interpreted the agents in the model as entrepreneurs subject to capital income risk, the model can be used 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. Moreover, I have assumed away fixed costs associated with
investment. In presence of fixed costs of investment, the model can be interpreted as a model of innovation and it can be used for optimal patent policy evaluations. Hence, the techniques developed in this paper are usefully in analyzing a wide variety of questions.
References


Appendix

A Proofs

Proof of Lemma 3.

We define $\tilde{k}(\theta, \hat{\theta})$ as in the text. That is the optimal value that an agent of type $\theta$ who pretends to be $\hat{\theta}$ invests optimally given the transfer $c_0(\hat{\theta}) + k_1(\hat{\theta})$ in the first period and the schedule $c_1(\hat{\theta}, y)$ in the second period. By assumption 1, this value is unique and is given by

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) = \beta \int_0^\theta u(c_1(\hat{\theta}, y))g_k(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy$$

We first prove the claim for the case with $\hat{\theta} > \theta$. As in the example in the text, we start by showing $\tilde{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta})$. Suppose not. That is, $\tilde{k}(\theta, \hat{\theta}) > k_1(\hat{\theta})$. this implies that

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) > u'(c_0(\hat{\theta}))$$

Therefore,

$$\int_0^\theta u(c_1(\hat{\theta}, y))g_k(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy > \int_0^\theta u(c_1(\hat{\theta}, y))g_k(y|k_1(\hat{\theta}), \hat{\theta})dy$$

(44)

By Assumption 1, the function $\Psi(k, \theta; \hat{\theta}) = \int_0^\theta u(c_1(\hat{\theta}, y))g_k(y|k, \theta)dy$ is increasing in $\theta$ and decreasing in $k$. Since $\hat{\theta} > \theta$ and $\tilde{k}(\theta, \hat{\theta}) > k_1(\hat{\theta})$, $\Psi(k_1(\hat{\theta}), \hat{\theta}) \geq \Psi(\tilde{k}(\theta, \hat{\theta}), \theta)$ which is a contradiction to (44). Therefore, we must have $\tilde{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta})$. Therefore, from assumption 2 we must have

$$u'(c_0(\hat{\theta}) + k_1(\theta) - \tilde{k}(\theta, \hat{\theta})) \left[ c'_0(\hat{\theta}) + k'_1(\theta) \right] \leq u'(c_0(\hat{\theta})) \left[ c'_0(\hat{\theta}) + k'_1(\hat{\theta}) \right]$$

Moreover, $\tilde{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta})$ and $\hat{\theta} > \theta$ implies that $G(y|k_1(\hat{\theta}), \hat{\theta}) \succ_{FOSD} G(y|\tilde{k}(\theta, \hat{\theta}), \theta)$. Hence, from assumption 3,

$$\int_0^\theta u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy \leq \int_0^\theta u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|k_1(\hat{\theta}), \hat{\theta})dy$$

The above inequalities together with the local incentive constraint (6),

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c'_0(\hat{\theta}) + k'_1(\theta) \right] \leq \int_0^\theta u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy \leq 0$$

(45)
Note that if we define \( U(\theta, \hat{\theta}) \) as follows

\[
U(\theta, \hat{\theta}) = \max_k u(c_0(\hat{\theta}) + k(\hat{\theta}) - \hat{k}(\hat{\theta})) + \int_0^\hat{\theta} u'(c_1(\hat{\theta}, y)) g(y|k, \theta)dy
\]

then \( \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \) is given by the LHS of the above inequality (45). Hence, for all \( \hat{\theta} > \theta \), \( \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \leq 0 \) and therefore

\[
U(\theta, \hat{\theta}) \leq U(\theta, \theta) = U(\theta)
\]

When \( \hat{\theta} < \theta \), a similar argument as above shows that \( \bar{k}(\theta, \hat{\theta}) > k_1(\hat{\theta}) \). Therefore, \( G(y|k_1(\hat{\theta}), \hat{\theta}) \leq_FOSD G(y|\bar{k}(\theta, \hat{\theta}), \theta) \) and hence

\[
\int_0^\theta u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|\bar{k}(\theta, \hat{\theta}), \theta)dy \geq \int_0^\theta u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|k_1(\hat{\theta}), \hat{\theta})dy
\]

Moreover,

\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \bar{k}(\theta, \hat{\theta})) [c'_0(\hat{\theta}) + k'_1(\hat{\theta})] \geq u'(c_0(\hat{\theta})) [c_0(\hat{\theta}) + k'_1(\hat{\theta})]
\]

Hence

\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \bar{k}(\theta, \hat{\theta})) [c'_0(\hat{\theta}) + k'_1(\hat{\theta})] + \int_0^\theta u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|\bar{k}(\theta, \hat{\theta}), \theta)dy \geq 0
\]

That is, for \( \hat{\theta} < \theta \), \( \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \geq 0 \). Hence, \( U(\theta, \hat{\theta}) \leq U(\theta, \theta) = U(\theta) \) for all \( \hat{\theta} < \theta \).

QED.

**Proof of Proposition 1.** Recall the planning problem (P1). Suppose the lagrange multiplier on feasibility at period \( t \) is \( \lambda_t \), the multiplier on the third constraint is given by \( \gamma(\theta) \), the multiplier on (11) is given by \( \mu_2(\theta) \lambda_0 f(\theta) \), and the co-state associated with (10) is given by \( \mu_1(\theta) f(\theta) \lambda_0 \).

Then the Lagrangian for this problem is given by

\[
\mathcal{L} = \int_\theta^\theta \left\{ U(\theta) + \lambda_0 [c_0(\theta) - c_1(\theta) - k_1(\theta)] + \lambda_1 \left[ \theta^\alpha k_1(\theta)^\alpha - \int c_1(\theta) g(y|k_1(\theta), \theta) \right]
\]

\[+ \gamma(\theta) \left[ u(c_0(\theta)) + \beta \int_0^\theta u(c_1(\theta, y)) g(y|k_1(\theta), \theta)dy - U(\theta) \right] + \lambda_0 \mu_1(\theta) \left[ U'(\theta) - \frac{1}{\theta} k_1(\theta) u'(c_0(\theta)) \right]
\]

\[+ \lambda_0 \mu_2(\theta) \left[ \beta \int_0^\theta u(c_1(\theta, y)) g_k(y|k_1(\theta), \theta)dy - u'(c_0(\theta)) \right] \} f(\theta)d\theta
\]
An application of the integration by parts formula implies that

\[
\mathcal{L} = \int_\theta^\theta \left\{ U(\theta) + \lambda_0 [c_0 - c_0(\theta) - k_1(\theta)] + \lambda_1 \left[ \theta^\alpha k_1(\theta)^\alpha - \int c_1(\theta)g(y|k_1(\theta), \theta) \right] + \gamma(\theta) \left[ u(c_0(\theta)) + \beta \int_0^y u(c_1(\theta, y))g(y|k_1(\theta), \theta)dy - U(\theta) \right] - \lambda_0 \mu_1(\theta) \left[ \frac{1}{\theta} k_1(\theta)u'(c_0(\theta)) \right] \right. \\
+ \lambda_0 \mu_2(\theta) \left[ \beta \int_0^y u(c_1(\theta, y))g_k(y|k_1(\theta), \theta)dy - u'(c_0(\theta)) \right] \left\} f(\theta)d\theta \\
+ \lambda_0 U(\theta)f(\theta)\mu_1(\theta) d\theta - \lambda_0 \int_\theta^\theta U(\theta) \left[ \mu_1(\theta)f(\theta) + \mu_1(\theta)f'(\theta) \right] d\theta
\]

By a theorem from Luenberger (1969) (section §9.3, Theorem 1), in order for an allocation \( \{c_0(\theta), c_1(\theta, y), U(\theta), k_1(\theta)\}_{(\theta, y) \times \in [\theta, \bar{\theta}] \times [0, \bar{y}]} \) to attain a local optimum in program (P1), the multipliers \( \mu_1, \mu_2 \) and \( \gamma \) must exist such that the Fréchet derivative of the above Lagrangian is zero. More technically, we assume that our underlying space is the space of bounded continuous functions \( \{c_0(\theta), c_1(\theta, y), U(\theta), k_1(\theta)\}_{(\theta, y) \times \in [\theta, \bar{\theta}] \times [0, \bar{y}]} \) and the Fréchet derivative is taken with respect to a member of this Hilbert space. Given the assumption that all allocation are interior, the theorem implies that the following conditions should hold a.e.-\( F \):

\[
-\lambda_0 + \gamma(\theta)u'(c_0(\theta)) - \lambda_0 \mu_1(\theta) \frac{1}{\theta} k_1(\theta)u''(c_0(\theta)) - \lambda_0 \mu_2(\theta)u''(c_0(\theta)) = 0 \quad (46) \\
-g(y|k_1(\theta), \theta)\lambda_1 + \beta u'(c_1(\theta, y)) \left[ g(y|k_1(\theta), \theta) + \beta \lambda_0 \mu_2(\theta)g_k(y|k_1(\theta), \theta) \right] = 0 \quad (47) \\
1 - \gamma(\theta) - \lambda_0 [\mu_1(\theta)f(\theta) + \mu_1(\theta)f'(\theta)] = 0 \quad (48) \\
\lambda_1 \alpha k_1(\theta)^{\alpha-1} - \lambda_0 - \lambda_1 \int c_1(\theta, y)g_k(y|k_1(\theta), \theta)dy + \gamma(\theta)\beta \int_0^y u(c_1(\theta, y))g(y|k_1(\theta), \theta)dy \\
- \frac{1}{\theta} u'(c_0(\theta))\lambda_0 \mu_1(\theta) + \lambda_0 \mu_2(\theta)\beta \int_0^y u(c_1(\theta, y))g_k(y|k_1(\theta), \theta)dy = 0 \quad (49)
\]

Moreover, continuity of allocations in \( \theta \) implies that

\[
\mu_1(\theta) = \mu_1(\bar{\theta}) = 0
\]

Note that by definition \( q = \frac{\lambda_1}{\lambda_0} \). From above, we have

\[
\gamma(\theta) = \frac{\lambda_0}{u'(c_0(\theta))} + \lambda_0 \frac{u''(c_0(\theta))}{u'(c_0(\theta))} \left[ \frac{1}{\theta} k_1(\theta)\mu_1(\theta) + \mu_2(\theta) \right] \quad (50)
\]

\[
\gamma(\theta)g(y|k_1(\theta), \theta) = \frac{\lambda_1}{\beta u'(c_1(\theta, y))}g(y|k_1(\theta), \theta) - \lambda_0 \mu_2(\theta)g_k(y|k_1(\theta), \theta)
\]

Integrating the last equation over \( y \) and using \( \int_0^y g_k(y|k_1(\theta), \theta)dy = 0 \), we get

\[
\gamma(\theta) = \lambda_1 \int_0^y \frac{1}{\beta u'(c_1(\theta, y))}g(y|k_1(\theta), \theta)dy \quad (51)
\]
Combining (50) and (51), gives us the desired result.

QED.

**Proof of Lemma ??**.

If we multiply (47) by \( u(c_1(\theta, y)) \) and divide it by \( u'(c_1(\theta, y)) \), we have – we suppress \( \theta \):

\[
\lambda_1 u(c_1) \times \frac{1}{u'(c_1)} g = \beta \gamma u(c_1) g + \beta \lambda_0 \mu_2 u(c_1) g_k
\]

Integrating over \( y \), we get

\[
\lambda_1 \int u(c_1) \times \frac{1}{u'(c_1)} g dy = \beta \gamma \int u(c_1) g dy + \beta \lambda_0 \mu_2 \int u(c_1) g_k dy
\]

Using (51) and (11), we get

\[
\lambda_1 \int u(c_1) \times \frac{1}{u'(c_1)} g dy = \lambda_1 \int \frac{1}{u'(c_1)} g dy \int u(c_1) g dy + \lambda_0 \mu_2 u'(c_0)
\]

and therefore

\[
\mu_2 \lambda_0 u'(c_0) = \lambda_1 \left[ \int u(c_1) \times \frac{1}{u'(c_1)} g dy - \int \frac{1}{u'(c_1)} g dy \int u(c_1) g dy \right]
\]

\[
\Rightarrow \mu_2 = \frac{q}{u'(c_0)} \text{Cov}_\theta(u(c_1), \frac{1}{u'(c_1)})
\]

Replacing the above in (49) gives the formula for \( \mu_1 \).

QED.

**Proof of Proposition 2.** From above, we know that an optimal allocation must satisfy (46)-(49). Now suppose to the contrary that \( \mu_1(\theta) < 0 \) for some \( \theta \). Since \( \mu_1(\bar{\theta}) = \mu_1(\bar{\theta}) = 0 \) and \( \mu_1(\theta) \) is continuously differentiable, there must exist \( \theta_1 < \theta_2 \) such that \( \mu_1(\theta_1) = \mu_1(\theta_2) = 0 \) and \( (\mu_1 f)'(\theta_1) \leq 0 \leq (\mu_1 f)'(\theta_2) \). Hence the equations (46)-(48) evaluated at \( \theta_1 \) and \( \theta_2 \) become

\[
\begin{align*}
-\lambda_0 + \gamma u'(c_0) - u''(c_0) \lambda_0 \mu_2 &= 0 \\
-g \lambda_1 + \beta \gamma u'(c_1) g + \beta \lambda_0 \mu_2 u'(c_1) g_k &= 0 \\
1 - \gamma - \lambda_0 (\mu_1 f)' &= 0
\end{align*}
\]

The last equation implies that \( \gamma(\theta_1) \geq 1 \geq \gamma(\theta_2) \). Since \( u(c) = -\exp(-\psi c) \), the above can be rewritten as

\[
\begin{align*}
-\lambda_0 - \gamma \psi u(c_0) + \psi u'(c_0) \lambda_0 \mu_2 &= 0 \\
-g \lambda_1 - \beta \gamma \psi u(c_1) g - \beta \mu_2 \lambda_0 \psi u(c_1) g_k &= 0
\end{align*}
\]

(52)
Replacing the above equation in (52)

\[
\lambda_0\mu_2 = -\frac{\lambda_1}{\psi u'(c_0)} - \gamma \frac{\beta \int u(c_1)gdy}{u'(c_0)}
\]

Replacing the above equation in (52)

\[-\lambda_0 - \gamma \psi u(c_0) - \lambda_1 - \psi\gamma \beta \int u(c_1)gdy = 0\]

or

\[u(c_0) + \beta \int u(c_1)gdy = -\frac{\lambda_0 + \lambda_1}{\psi\gamma}\]

Therefore, \(U(\theta) = -\frac{\lambda_0 + \lambda_1}{\psi\gamma(\theta)}\) when \(\theta = \theta_1, \theta_2\). Hence, we have

\[
\gamma(\theta_1) \geq \gamma(\theta_2) \Rightarrow \frac{\lambda_0 + \lambda_1}{\psi\gamma(\theta_1)} \leq \frac{\lambda_0 + \lambda_1}{\psi\gamma(\theta_2)}
\]

\[
\Rightarrow U(\theta_1) \geq U(\theta_2)
\]

(53)

Note however that by (10), \(U(\theta_2) - U(\theta_1) = \int_{\theta_1}^{\theta_2} \frac{1}{\psi} u'(c_0(\theta)) k_1(\theta) d\hat{\theta} > 0\) which is in contradiction with (53). This proves that \(\mu_1(\theta) \geq 0\). The above analysis also shows also that \(\mu_1(\theta)\) needs to be non-zero for at least a positive measure of \(\theta\)'s. Otherwise, for almost all \(\theta\)'s, \(\mu_1(\theta) = 0\) and continuous differentiability of \(\mu_1(\theta)\) implies that \(\mu'_1(\theta) = 0\). In this case, the above analysis implies that \(U(\theta)\) is constant which violates the adverse selection incentive constraint (10).

QED.

**Proof of Proposition 1.**

The FOC from program (P2) are given by

\[-f^t(\theta|\theta_{-1}) + \gamma(\theta) f^t(\theta|\theta_{-1}) u'(c(\theta)) + \mu(\theta) f^t(\theta|\theta_{-1}) \frac{1}{\theta} k(\theta) u''(c(\theta)) + \mu_2(\theta) u''(c(\theta)) = 0\]

\[f^t(\theta|\theta_{-1}) \frac{Q_{t+1}}{Q_t} P_{w_{t+1}}^{t+1}(\theta) g^{t+1}(y|k(\theta), \theta) + \beta \gamma(\theta) f^t(\theta|\theta_{-1}) g^{t+1}(y|k(\theta), \theta) - \beta \mu_2(\theta) f^t(\theta|\theta_{-1}) g_k^{t+1}(y|k(\theta), \theta) = 0\]

\[\lambda f^t(\theta|\theta_{-1}) + \lambda' f_{\theta_{-1}}^t(\theta|\theta_{-1}) - \gamma(\theta) f^t(\theta|\theta_{-1}) + \mu(\theta) f^t(\theta|\theta_{-1}) + \mu(\theta) f_{\theta_{-1}}^t(\theta|\theta_{-1}) = 0\]

where \(\mu(\theta)\) is the costate associated with (26), \(\mu_2(\theta)\) is the Lagrange Multiplier on (27) and \(\gamma(\theta) f^t(\theta|\theta_{-1})\) is the Lagrange Multiplier on the third constraint. Integrating the second equation with respect to \(y\) implies that

\[
\frac{Q_{t+1}}{Q_t} \int_0^\beta P_{w_{t+1}}^{t+1}(\theta) g^{t+1}(y|k(\theta), \theta) dy + \beta \gamma(\theta) = 0
\]
Moreover, the first FOC can be written as

\[ \gamma(\theta) = \frac{1}{w'(c(\theta))} - \mu(\theta) \frac{1}{\theta} k(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} - \mu_2(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} \]

Hence,

\[ \frac{1}{w'(c_t)} - \mu_t(\theta) k_{t+1} \frac{u''(c_t)}{w'(c_t)} - \mu_{2t} \frac{u''(c_t)}{w'(c_t)} + \frac{Q_{t+1}}{\beta Q_t} \int_0^\theta P_{w}^{t+1}(w_{t+1}, \Delta_t, \theta_t) g^{t+1}(y|k_{t+1}, \theta_t) dy = 0 \quad (54) \]

Moreover, from the envelope theorem, we know that \( P_w^t = -\lambda \). Integrating the first and the last FOC w.r.t \( \theta \) implies that

\[
-\int_\theta^0 \frac{1}{w'(c(\theta))} f^t(\theta|\theta-1) d\theta + \int_\theta^0 \gamma(\theta) f^t(\theta|\theta-1) d\theta + \int_\theta^0 \mu(\theta) f^t(\theta|\theta-1) \frac{1}{\theta} k(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} d\theta
\]

\[ + \int_\theta^0 \mu_2(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} d\theta = 0 \]

\[ \lambda - \int_\theta^\theta \gamma(\theta) f^t(\theta|\theta-1) d\theta = 0 \]

Hence, \( \frac{Q_{t+1}}{\beta Q_t} E_t P_{w}^{t+1} = P_w^t \).

From (29), \( \mu_{2t} = f^t(\theta|\theta-1) \frac{Q_{t+1}}{Q_t} \frac{1}{w'(c_t)} Cov(P_{w}^{t+1}, w_{t+1}|(\theta^t, y^t)) \). Therefore,

\[
P_w^t = E \left[ \frac{1}{w'(c_t)} + \mu_t k_{t+1} \frac{u''(c_t)}{\theta_t u'(c_t)} + \frac{Q_{t+1}}{Q_t} \frac{u''(c_t)}{w'(c_t)} Cov(P_{w}^{t+1}, w_{t+1}|(\theta^t, y^t)) \right]
\]

Applying the above formula to period \( t + 1 \) and using (54) gives the MIEE.

QED.

**Proof of Theorem 2.**

Given the steps provided in the text, we only need to show that the value function has the form \( P(w) = \frac{A_t}{\psi} \log(-w) + B_t \) where \( A_t = \frac{1}{Q_t} \sum_{s=t}^T Q_s \) and that inequality (31) holds.

To show that the value function has the claimed form, we use induction. Notice that at \( t = T \), since there are no shocks, simply, we have

\[ P_T^t(w) = \frac{1}{\psi} \log(-w). \]

Now suppose that the claim holds for \( t + 1 \), then we show that it holds for \( t \). Given this assumption, planning problem (P2) becomes

\[
P_t^t(w) = \max_{\theta} \int_\theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha - c(\theta) - k(\theta) + \frac{Q_{t+1}}{\psi} \frac{A_t + 1}{\psi} \log(-w'(\theta)) \right] f^t(\theta) d\theta
\]

59
subject to
\[
\begin{align*}
    w &= \int U(\theta) f^t(\theta) d\theta \\
    U(\theta) &= u(c(\theta)) + \beta w'(\theta) \\
    \frac{\partial}{\partial \theta} U(\theta) &= \frac{1}{\theta} k(\theta) u'(c(\theta))
\end{align*}
\]

We guess that the policy functions have the form \( c(w, \theta) = \hat{c}(\theta) - \frac{1}{\psi} \log(-w) \), \( k(w, \theta) = k(\theta) \), \( w'(w, \theta) = (-w) \cdot \hat{w}(\theta) \), and \( U(\theta) = (-w) \cdot \hat{U}(\theta) \). Under these assumption, \( u(c(w, \theta)) = (-w) \cdot u(\hat{c}(\theta)) \) and \( u'(c(w, \theta)) = (-w) u'(\hat{c}(\theta)) \). Hence,

\[
P^t(w) = \max \int_{\Theta} \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha + \frac{1}{\psi} \log(-w) - \hat{c}(\theta) - k(\theta) + \frac{Q_{t+1}}{Q_t} A_{t+1} \frac{1}{\psi} \log((-w)(-\hat{w}(\theta))) \right] f^t(\theta) d\theta
\]

subject to
\[
\begin{align*}
    -1 &= \int_{\Theta} \hat{U}(\theta) f^t(\theta) d\theta \\
    \hat{U}(\theta) &= u(\hat{c}(\theta)) + \beta \hat{w}(\theta) \\
    \frac{\partial}{\partial \theta} \hat{U}(\theta) &= \frac{1}{\theta} k(\theta) u'(\hat{c}(\theta))
\end{align*}
\]

and the objective becomes

\[
\int_{\Theta} \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^{\alpha - \hat{c}(\theta) - k(\theta)} + \frac{Q_{t+1}}{Q_t} A_{t+1} \frac{1}{\psi} \log(-\hat{w}(\theta)) \right] f^t(\theta) d\theta + \frac{1}{\psi} \log(-w) + \frac{Q_{t+1}}{Q_t} A_{t+1} \frac{1}{\psi} \log(-w)
\]

This means that the above maximization problem is independent of \( w \) and therefore \( P^t(w) = \frac{A_t}{\psi} \log(-w) + B_t \) where \( A_t = 1 + \frac{Q_{t+1}}{Q_t} A_{t+1} \).

In order to show (31), we show that \( \mu_1(\theta) \geq 0 \). In fact, the proof that \( \mu_1(\theta) \geq 0 \) is identical to the proof of Proposition 2. That is, if \( \mu_1(\theta) < 0 \), there must exists \( \theta_1 < \theta_2 \) such that \( \mu_1(\theta_1) = \mu_1(\theta_2) = 0 \) and \( \frac{d}{d\theta} (\mu_1(\theta_1) f^t(\theta_1)) \leq 0 \leq \frac{d}{d\theta} (\mu_1(\theta_2) f^t(\theta_2)) \). Note that under this assumption the FOCs evaluated at \( \theta_1 \) and \( \theta_2 \) are given by

\[
\begin{align*}
    -f^t + \gamma u'(c) &= 0 \\
    f^t \frac{Q_{t+1}}{Q_t} A_{t+1} \frac{1}{\psi} - \gamma \beta &= 0 \\
    \lambda - \gamma - \frac{1}{f^t d\theta} (\mu_1 f^t) &= 0
\end{align*}
\]

Hence \( \gamma(\theta_1) \geq \gamma(\theta_2) \) and therefore \( c(\theta_1) \geq c(\theta_2) \) and \( w(\theta_1) \geq w(\theta_2) \) which contradicts the incentive
constraint. Hence, $\mu_1(\theta) \geq 0$. Note that from the FOCs we have

$$-1 + \gamma(\theta)u'(c(\theta)) - \mu_1(\theta)\frac{1}{\theta}k(\theta)u''(c(\theta)) = 0$$

$$\frac{Q_{t+1}}{\beta Q_t}P_{w}^{t+1}(w'(\theta)) + \gamma(\theta) = 0$$

and therefore

$$-\frac{Q_{t+1}}{\beta Q_t}P_{w}^{t+1}(w'(\theta)) = \frac{1}{u'(c(\theta))} + \mu_1(\theta)\frac{1}{\theta}k(\theta)\frac{u''(c(\theta))}{u'(c(\theta))}$$

Since $\mu_1 \geq 0$ and $u''(c) < 0$, we have the inequality (31).

QED.

**Proof of Proposition 6.**

The recursive problem with safe returns is given by

$$P^t(w) = \max \int_{\Theta} \left[ \frac{Q_{t+1}}{Q_t}(\theta k(\theta))^{\alpha} - c(\theta) - k(\theta) + \frac{Q_{t+1}}{Q_t}P_{w}^{t+1}(w'(\theta)) \right] f^t(\theta)d\theta$$

subject to

$$w = \int_{\Theta} U(\theta)f^t(\theta)d\theta$$

$$U(\theta) = u(c(\theta)) + \beta w'(\theta)$$

$$\frac{\partial}{\partial \theta}U(\theta) = \frac{1}{\theta}k(\theta)u'(c(\theta))$$

The Envelope condition (55) can be written as

$$u'(c(\theta))c'(\theta) + \beta \frac{d}{d\theta}w'(\theta) = \frac{1}{\theta}k(\theta)u'(c(\theta))$$

By assumption, $\frac{d}{d\theta}w'(\theta) > 0$. Hence, we must have $\frac{1}{\theta}k(\theta) > c'(\theta)$. The term $\frac{1}{\theta}k(\theta) - c'(\theta)$ can be thought of as incremental increase in consumption in the current period when the agent lies locally. Hence, when $w'(\theta)$ is increasing, lying downward increases consumption.

The FOC of the above planning problem are given by

$$-f^t(\theta) + \gamma(\theta)f^t(\theta)u'(c(\theta)) - \mu_1(\theta)f^t(\theta)\frac{1}{\theta}k(\theta)u''(c(\theta)) = 0$$

$$\lambda f^t(\theta) - \gamma(\theta)f^t(\theta) - \frac{d}{d\theta} \left( \mu_1(\theta)f^t(\theta) \right) = 0$$

Combining these equations implies that

$$-f^t(\theta|\theta_{-1}) + \left[ \lambda f^t(\theta) - \frac{d}{d\theta} \left( \mu_1(\theta)f^t(\theta) \right) \right] u'(c(\theta))$$

$$-\mu_1(\theta)f^t(\theta)\frac{1}{\theta}k(\theta)u''(c(\theta)) = 0$$
Moreover, by assumption $\mu_1(\theta) > 0$. Hence, we must have
\[
\mu_1(\theta) \frac{1}{\theta} k(\theta) u''(c(\theta)) < \mu_1(\theta) c'(\theta) u''(c(\theta))
\]

Therefore,
\[
-f^t(\theta) + \left[ \lambda f^t(\theta) - \frac{d}{d\theta} (\mu_1(\theta) f^t(\theta)) \right] u'(c(\theta)) - \mu_1(\theta) f^t(\theta) c'(\theta) u''(c(\theta)) < 0
\]

or
\[
-f^t(\theta) + \lambda f^t(\theta) u'(c(\theta)) - \frac{d}{d\theta} \mu_1(\theta) f^t(\theta) u'(c(\theta)) < 0.
\]

Integrating the above inequality over $\theta$,
\[
-1 + \lambda \int u'(c(\theta)) f^t(\theta)d\theta - \mu_1(\theta) f^t(\theta) u'(c(\theta)) |_{\bar{\theta}}^{\theta} < 0
\]

and since $\mu_1(\bar{\theta}) = \mu_1(\theta) = 0$, hence
\[
\int u'(c(\theta)) f^t(\theta)d\theta < \frac{1}{\lambda}
\]

Note that by Envelope theorem, $\lambda = -P^t_w(w)$, hence
\[
\int u'(c(\theta)) f^t(\theta)d\theta < -\frac{1}{P^t_w(w)}
\]

The rest is identical to the prove given in the text.

QED.

**Proof of Lemma 2.**

Note that since the problem is stationary, i.e., $Q_{t+1}/Q_t = \hat{q}$, the policy functions are time independent and therefore
\[
c_t(w_t) = -\frac{1}{\psi} \log(-w_t) + c^*
\]
\[
c_{t+1}(w_t, y_{t+1}) = -\frac{1}{\psi} \log(-w_t) - \frac{1}{\psi} \log(-w^*(y_{t+1})) + c^*
\]

Hence, the intertemporal wedge is given by
\[
1 - \tau_s = \frac{qu'(c_t)}{\beta E_t u'(c_{t+1})} = \frac{q(-w_t)u'(c^*)}{\beta(-w_t)u'(c^*) \int_0^q (-w^*(y_{t+1})) g(y_{t+1}|k^*)dy_{t+1}}
\]
\[
= \frac{\beta \int_0^\hat{q} (-w^*(y_{t+1})) g(y_{t+1}|k^*)dy_{t+1}}{q}
\]
Let $1 + \hat{R} = \frac{1 - q}{q} = \frac{1}{\beta \int_0^\beta (-w^*(y_{t+1})) g(y_{t+1}|k^*) dy_{t+1}}$. Note by definition that $\beta \int_0^{\beta} (-w^*(y_{t+1})) g(y_{t+1}|k^*) dy_{t+1} = 1 + u(c^*) < 1$. Hence, $\hat{R} > 0$. Next, for any convex and smooth function $T(y)$, consider the following recursive formulation of (P4):

$$V(a) = \max u(c) + \beta \int_0^\beta V(y - T(y) - (1 + \hat{R})B') g(y|k') dy$$

subject to

$$c + k' - B' = a$$

Using a guess and verify method, we show that $V(a) = e^{-\frac{a}{1 + \hat{R}} \psi a + \varphi}$ for a constant number $\varphi$ that depends on the choice of $T(y)$. Moreover, the policy functions implied by the above maximization problem is

$$c(a) = \frac{\hat{R}}{1 + \hat{R}} a + \zeta_1$$

$$B'(a) = \frac{\hat{R}}{1 + \hat{R}} a + \zeta_2$$

for constants $\zeta_1$ and $\zeta_2$. Furthermore, $k'$ solves the following equation:

$$\psi \beta \hat{R} \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi [y_0 - T(y_0)]} g(y|k') dy = - \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi [y_0 - T(y_0)]} g_k(y|k') dy$$

and hence independent of $a$. Since $a = y - T(y) - (1 + \hat{R}) B$, $c_t = \frac{\hat{R}}{1 + \hat{R}} \int y_t - T(y_t) - (1 + \hat{R}) B_t + \zeta_1$. Note that from (56), $c_t = -\frac{1}{\psi} \log(-w_{t-1}) - \frac{1}{\psi} \log(-w^*(y_t)) + c^*$. Hence, if the tax function $T(\cdot)$ is to implement the optimal allocation, we must have

$$T(y) = y + \frac{1 + \hat{R}}{\hat{R}} \frac{1}{\psi} \log(-w^*(y)) + \kappa$$

for some constant $\kappa$. Therefore, to complete the proof we need to find $B_0$ and $\kappa$ given the value of income realization at $t = 0$ and $w_0$. Note that for the implementation to work, we must have

$$V(y_0 - T(y_0) - (1 + \hat{R}) B_0) = e^{-\frac{\hat{R}}{1 + \hat{R}} \psi [y_0 - T(y_0) - (1 + \hat{R}) B_0] + \varphi} = w_0$$

Further analysis of the recursive problem implies that we must have

$$e^{-\psi c} = \beta \hat{R} \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi [y_0 - T(y_0) - (1 + \hat{R})(c + k' - a)] + \varphi} g(y|k') dy$$

Taking log from both sides

$$-\psi c = \log(\beta \hat{R}) + \psi \hat{R} c - \hat{R} \psi a - \hat{R} \psi k' + \varphi + \log \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi [y_0 - T(y_0)]} g(y|k') dy$$
Hence,
\[-\psi_c = \frac{1}{1+\hat{R}} \log(\beta \hat{R}) \int e^{-\frac{\hat{R}}{1+\hat{R}}\psi[y_0-T(y_0)]+\varphi} g(y|k')dy - \psi \frac{\hat{R}}{1+\hat{R}} a\]

Replacing in the value function, we get
\[e^{\varphi} = (1 + \frac{1}{\hat{R}}) \left( \beta \hat{R} \right) \frac{1}{1+\hat{R}} \int e^{-\frac{\hat{R}}{1+\hat{R}}\psi[y_0-T(y_0)]+\varphi} g(y|k')dy\]

Therefore,
\[e^{\frac{\hat{R}}{1+\hat{R}} \varphi} = (1 + \frac{1}{\hat{R}}) \left( \beta \hat{R} \right) \frac{1}{1+\hat{R}} \int (-w^*(y)) \frac{1}{1+\hat{R}} e^{\frac{\hat{R}}{1+\hat{R}} \psi} g(y|k')dy\]

which gives \( \varphi \) as a function of \( \kappa \). Hence, for any value of \( \kappa \), \( B_0 \) must be chosen so that
\[(-w^*(y)) e^{\frac{\hat{R}}{1+\hat{R}} \psi} - \hat{R} B_0 + \varphi = w_0\]

This completes the proof.
QED.

Proof of Theorem 4
TO BE COMPLETED.

B Sufficient Conditions for FOA

In this section, we provide sufficient conditions for validity of the FOA. To gain insight, we start from a two shock example and extend the derived sufficient conditions to the general case.

A Two Shock Example.

To address the adverse selection problem, we consider a simple example in which the FOA is evidently valid regarding the moral hazard problem. Suppose that the output from the project can only take two value \( \{0, \bar{y}\} \) where \( \Pr(y = \bar{y}|\theta,k) = \theta^\alpha k^\alpha \). In this case, the local incentive compatibility constraints become

\[u'(c_0(\theta)) = \beta \alpha \theta^\alpha k_1(\theta)^{\alpha-1} [u(c_1(\theta, \bar{y})) - u(c_1(\theta, 0))] \]  
(57)

\[u'(c_0(\theta)) [c_0'(\theta) + k_1'(\theta)] + \beta \theta^\alpha k_1(\theta)^\alpha u'(c_1(\theta, \bar{y})) c_1(\theta, \bar{y}) + \beta (1 - \theta^\alpha k_1(\theta)^\alpha) u'(c_1(\theta, 0)) c_1(\theta, 0) = 0 \]  
(58)

In this case, if an agent of type \( \theta \) pretends to be \( \hat{\theta} \), his optimal investment is given by \( \tilde{k}(\theta, \hat{\theta}) \) where

\[u'(c_0(\theta) + k_1(\theta) - \tilde{k}(\theta, \hat{\theta})) = \beta \alpha \theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha-1} [u(c_1(\theta, \bar{y})) - u(c_1(\theta, 0))] \]  
(59)

We claim that in this case, if \( c_0(\theta) + k_1(\theta) \) and \( u(c_1(\theta, \bar{y})) - u(c_1(\theta, 0)) \) are both increasing functions of \( \theta \), then the FOA is valid. To show the validity of FOA under this assumption, we
must show that
\[
\frac{u(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}(\theta, \hat{\theta})) + \beta \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha u(c_1(\hat{\theta}, \bar{y})) + \beta (1 - \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha) u(c_1(\hat{\theta}, 0))}{U(\theta)} \leq U(\theta)
\]

We illustrate this for the case where \( \hat{\theta} > \theta \) by showing that the LHS is decreasing in \( \theta \). The case with \( \hat{\theta} < \theta \) can be shown in the same way. To do this, we first show that \( \hat{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta}) \). Suppose not. Then,
\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}(\theta, \hat{\theta})) \geq u'(c_0(\hat{\theta}))
\]

Note that from (59) and (57), we must have
\[
\theta^\alpha \hat{k}(\theta, \hat{\theta})^{\alpha - 1} \geq \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha - 1}
\]
or
\[
\frac{k_1(\hat{\theta})}{k(\theta, \hat{\theta})} \geq \left( \frac{\hat{\theta}}{\theta} \right)^{\frac{\alpha}{\alpha - 1}} > 1
\]
which is a contradiction. Hence \( \hat{k}(\theta, \hat{\theta}) < k_1(\hat{\theta}) \). Note that given this, \( \theta^\alpha \hat{k}(\theta, \hat{\theta})^{\alpha - 1} < \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha - 1} \) and hence, \( \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha < \hat{\theta}^\alpha k_1(\hat{\theta})^\alpha \). This results is intuitive. In fact, if an agent with a lower productivity pretends to have a higher productivity, since his marginal return is lower, he will invest less than the agent with high productivity and will enjoy more consumption today. Therefore, since \( c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \geq 0 \) by assumption,
\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] \leq u'(c_0(\hat{\theta})) \left[ c_0'(\theta) + k_1'(\theta) \right]
\]

Moreover, since \( u(c_1(\theta, \bar{y})) \) is increasing, \( u'(c_1(\theta, \bar{y})) c_{1\theta}(\theta, \bar{y}) > u'(c_1(\theta, 0)) c_{1\theta}(\theta, 0) \).

Hence,
\[
\beta \theta^\alpha k_1(\hat{\theta}, \hat{\theta})^\alpha u'(c_1(\hat{\theta}, \bar{y})) c_{1\theta}(\hat{\theta}, \bar{y}) + \beta (1 - \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha) u'(c_1(\hat{\theta}, 0)) c_{1\theta}(\hat{\theta}, 0)
\]
\[
\leq \beta \hat{\theta}^\alpha k_1(\hat{\theta}, \hat{\theta})^\alpha u'(c_1(\hat{\theta}, \bar{y})) c_{1\theta}(\hat{\theta}, \bar{y}) + \beta (1 - \hat{\theta}^\alpha k_1(\hat{\theta})^\alpha) u'(c_1(\hat{\theta}, 0)) c_{1\theta}(\hat{\theta}, 0)
\]
The above inequalities together with (58) implies that
\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] + \beta \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha u'(c_1(\hat{\theta}, \bar{y})) c_{1\theta}(\hat{\theta}, \bar{y})
\]
\[
+ \beta (1 - \theta^\alpha \hat{k}(\theta, \hat{\theta})^\alpha) u'(c_1(\hat{\theta}, 0)) c_{1\theta}(\hat{\theta}, 0) \leq 0
\]

The above expression coincides with \( \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \). Hence, for all \( \hat{\theta} > \theta \), \( \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \leq 0 \) and therefore, \( U(\theta, \hat{\theta}) \leq U(\theta) \). A similar argument works for \( \hat{\theta} < \theta \). So the required assumptions on endogenous variables that can be checked are
1. Total transfers in the first period must be increasing with type, i.e., $c'_0(\hat{\theta}) + k'_1(\hat{\theta}) \geq 0$.

2. $u(c_1(\theta, \hat{y})) - u(c_1(\theta, 0))$ must be increasing in $\theta$.

QED.

The above example is useful since it identifies the main forces leading to validity of the FOA regarding adverse selection. In fact, there are two steps in proving the validity of the FOA. First, we need to show that the optimal choice of $\tilde{k}(\theta, \hat{\theta})$ when agent of type $\theta$ pretends to be $\hat{\theta}$ is monotone decreasing in $\hat{\theta}$. This imposes certain restrictions on the schedule $c_1(\theta, y)$. Second, we should show that given the monotonicity of $\tilde{k}(\theta, \hat{\theta})$, $\int^\theta_0 u'(c_1(\hat{\theta}, y))g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy$ is monotone decreasing in $\hat{\theta}$. In this case, a sufficient assumption is for $u'(c_1(\theta, y))c_1(\theta, y)$ to be increasing in $\theta$. We can summarize this discussion in the following lemma:

**Lemma 3** Suppose that an allocation $\{c_0(\theta), c_1(\theta, y), k_1(\theta)\}_{(\theta, y) \in [\hat{\theta}, \theta] \times [0, \bar{y}]}$ satisfies the following:

1. The function $\frac{\partial}{\partial k} \int^\theta_0 u'(c_1(\hat{\theta}, y))g(y|k, \theta)dy$ is increasing in $\theta$ and decreasing in $k$, for all $\hat{\theta}$, $\theta$, and $k$,

2. Transfers in first period, $c_0(\theta) + k_1(\theta)$, is increasing in $\theta$,

3. The function $u'(c_1(\theta, y))c_1(\theta, y)$ is increasing in $y$ for all $\theta$, $y$,

4. The allocation is locally incentive compatible,

Then, the allocation is incentive compatible.

Proof is given in the appendix.

Note that condition 1 implies that given a report $\hat{\theta}$ (possibly $\theta$), there is a unique level of investment that maximizes utility. Hence, this assumption resolves the moral hazard issue as well. In the two shock example given above, it is satisfied since for any $c_1(\hat{\theta}, \hat{y}) > c_1(\hat{\theta}, 0)$, since the function $\frac{\partial}{\partial k} \{\theta^2 k^a u(c_1(\hat{\theta}, \hat{y})) + (1 - \theta^2 k^a)u(c_1(\hat{\theta}, 0))\}$ is increasing in $\theta$ and decreasing in $k$ due to decreasing returns to scale. See Jewitt (1988) for an extensive discussion of assumptions on fundamentals that lead to assumption 1. Unfortunately, condition 1 is a complicated condition that cannot be easily checked. Below, we provide further restriction on the distribution function $g(y|k, \theta)$ that makes checking condition 1 easier.

**Lemma 4** Suppose that for all $\hat{\theta}$, $c_1(\hat{\theta}, y)$ is increasing in $y$ and $g(\cdot|k, \theta)$ satisfies the following:

1. The function $\frac{g_{y\theta}(y|k, \theta)}{g(y|k, \theta)}$ is increasing in $y$,

2. The function $\frac{g_{y\theta}(y|k, \theta)}{g(y|k, \theta)}$ is decreasing in $y$.

Then, condition 1 in lemma 3 is satisfied.
**Proof.** First, note that
\[
\int_0^y g(y|k, \theta) dy = 1
\]
Therefore
\[
\int_0^y g_k(y|k, \theta) dy = \int_0^y g_{k\theta}(y|k, \theta) dy = \int_0^y g_{kk}(y|k, \theta) dy = 0
\]
Now define \( \Psi(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_k(y|k, \theta) dy \). Then,
\[
\Psi_k(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_{kk}(y|k, \theta) dy = \text{Cov}(u(c_1(\theta, y)), \frac{g_{kk}(y|\theta, k)}{g(y|\theta, k)})
\]
\[
\Psi_\theta(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_{k\theta}(y|k, \theta) dy = \text{Cov}(u(c_1(\theta, y)), \frac{g_{k\theta}(y|\theta, k)}{g(y|\theta, k)})
\]
Therefore, the above assumptions imply that \( \Psi_k < 0 \) and \( \Psi_\theta > 0 \).