

w o r k i n g  
p a p e r

19 26

**The Optimal Taxation of  
Business Owners**

Thomas Phelan



**FEDERAL RESERVE BANK OF CLEVELAND**

ISSN: 2573-7953

**Working papers** of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

Working papers are available on the Cleveland Fed's website:

**<https://clevelandfed.org/wp>**

## The Optimal Taxation of Business Owners

Thomas Phelan

Business owners in the United States are disproportionately represented among the very wealthy and are exposed to substantial idiosyncratic risk. Further, recent evidence indicates business income primarily reflects returns to the human (rather than financial) capital of the owner. Motivated by these facts, this paper characterizes the optimal taxation of income and wealth in an environment where business income depends jointly on innate ability, luck, and the accumulated past effort exerted by the owner. I show that in (constrained) efficient allocations, more productive entrepreneurs typically bear more risk and that the associated stationary distributions of income, wealth, and firm size exhibit the thick right (Pareto) tails observed in the data. Finally, when owners may save in a risk-free bond and trade shares of their business, I show that the optimal linear taxes in this environment call for positive taxes on firm profits and risk-free savings, and for a tax/subsidy on wealth that may assume either sign.\*

Keywords: Optimal taxation, moral hazard, optimal contracting, human capital.

JEL Codes: D61, D63, E62.

Suggested citation: Phelan, Thomas. 2019. "The Optimal Taxation of Business Owners." Federal Reserve Bank of Cleveland, Working Paper no. 19-26. <https://doi.org/10.26509/frbc-wp-201926>.

---

Thomas Phelan is at the Federal Reserve Bank of Cleveland (tom.phelan@clev.frb.org).

\*The final sentence of the abstract was revised for clarity two days after the paper was initially posted.

# 1 Introduction

It is well documented that the distributions of income and wealth in the United States and other developed nations are skewed to the right, with tails that may be well approximated by a Pareto distribution. For instance, [Atkinson et al. \(2011\)](#) survey empirical studies of top incomes and find that the density  $f$  of incomes in the United States in 2009 satisfied

$$f(z) \sim \frac{C}{z^{1+\alpha}}$$

as  $z \rightarrow \infty$  for some constant  $C$ , with  $\alpha$  approximately equal to 1.53. Further, numerous authors have observed a sharp rise in income inequality in the United States and other developed countries over the past few decades. Using data from the Statistics of Income division of the Internal Revenue Service, [Piketty and Saez \(2003\)](#) estimate that the top 10 percent income share has risen from approximately 34 percent to nearly 50 percent over the period 1980-2015 and that the top 1 percent income share has risen from approximately 8 percent to over 20 percent over the same period. Although the magnitudes of both levels and trends differ depending upon data sources and units of analysis, this significant rise in income inequality has also been observed in survey data<sup>1</sup> and in administrative-level data on labor incomes obtained from the Social Security Administration.<sup>2</sup>

What is perhaps less well known is that much of the recent increase in income inequality is due to the growing importance of business income. [Smith et al. \(2019\)](#) and [Cooper et al. \(2016\)](#) use administrative tax data to document that business income now accounts for a greater share of the top 0.1 percent of income than both non-business capital income and wage income, and that the majority of this growth is due to private, pass-through entities (those not taxed at the firm level), such as partnerships and S-corporations. The importance of business income is reinforced by the observation of [Güvenen and Kaplan \(2017\)](#) that the increase in inequality in the past decade observed in IRS data has not coincided with a similar rise in labor incomes recorded by the Social Security Administration. [Smith et al. \(2019\)](#) also note that top income earners are disproportionately business owners, with households in the top 1 percent of the income distribution 50 times more likely to receive partnership income than households in the bottom half of the income distribution. Further, these authors show business income depends on the owners' active participation rather than passive ownership by documenting that the unexpected death of an owner-manager gives rise to an average fall in profits

---

<sup>1</sup>See, e.g., [Heathcote et al. \(2010\)](#), and references therein.

<sup>2</sup>See [Güvenen et al. \(2014\)](#), [Güvenen and Kaplan \(2017\)](#) and [Song et al. \(2014\)](#).

of 54 percent.

The appropriate policy response to this large degree of inequality and its increase over recent decades depends on the underlying economic mechanisms by which these phenomena arise. Guided largely by the (static) [Mirrlees \(1971\)](#) model, [Diamond and Saez \(2011\)](#) survey the literature on optimal taxation and argue that the top marginal tax rates may be as high as 80 percent, far higher than the current statutory maximum of 39.6 percent. However, the disproportionate role of business income among high earners and its higher level of idiosyncratic risk suggests that appropriate policy responses must consider the role of entrepreneurial activity, a phenomenon absent from the static Mirrlees model. Similarly, although the seminal contributions of [Atkinson and Stiglitz \(1976\)](#), [Chamley \(1986\)](#) and [Judd \(1985\)](#) imply zero capital taxation in the long run, in these models "capital income" is homogeneous and identified with the (safe) returns to savings. They are therefore silent on the question of whether business income ought to be taxed differently from savings, given that the former may depend in part on the past investments of effort undertaken by the owner.

This paper develops theory and numerical techniques to characterize constrained efficient allocations in an environment capable of capturing the above stylized features. Specifically, I assume that both initial ability (whether or not one may run a business) and effort exerted to improve productivity are privately observed only by the agent. In doing so, I draw upon the existing literature on both dynamic contracting and heterogeneous-agent macroeconomics. The first part of the paper extends the principal-agent model of [Sannikov \(2008\)](#) to this richer setting with endogenous productivity. To preserve incentives for high effort, high realizations of productivity growth must be rewarded by either high future consumption or leisure, leading to imperfect risk-sharing and ex-post inequality in all efficient allocations. I then show how this principal-agent model may be used to derive implications for aggregate quantities by decomposing the problem of a benevolent planner facing a continuum of agents with random lifespans into a series of problems of dealing with each generation in isolation, given (shadow) prices for goods and labor that may then be varied until resources balance.

The incorporation of endogenous human capital formation into the agency problem has important implications for both the dynamics of risk-sharing and its distribution across income levels. In any agency model with unobserved effort, high realizations of output must be rewarded with

high consumption or leisure in the future in order to align the incentives of the principal and the agent. In the presence of wealth effects in consumption, it becomes increasingly expensive over time to motivate agents with histories of high output. With fixed productivity, the principal therefore rewards rich agents with leisure rather than consumption, and so, such agents bear little to no risk. With endogenous human capital formation there is a countervailing effect: when agents affect the growth rate, rather than the level, of output, the benefits of motivating an agent rise with their productivity. Although it is still true that the cost of motivating agents rises with their consumption, agents typically obtain high consumption precisely *because* their productivity rose, and so the societal benefit of effort rises along with the costs. I provide sufficient conditions for the former to overwhelm the latter in terms of the underlying preferences. I show that this has first-order effects on the implied stationary distribution, and in particular, the thickness of the right tail of consumption. The tractability of the model then allows for simple comparative statics connecting changes in aggregate technology with efficient distributions of income and firm size. Specifically, in response to any change in resources or technology that increases the marginal productivity of an entrepreneur, a benevolent planner will wish to make incentives for effort more high-powered, which necessitates an increase in inequality ex-post.

In the final section of the paper I connect the abstract characterization above with taxation policy. Prescriptions for taxes depend naturally upon assumptions regarding the degree of risk-sharing present in private markets and the complexity of contracts that agents are assumed capable of signing. To isolate the role for government and its relation to the degree of imperfections in private markets, I consider two *separate* decentralizations that capture opposite extreme assumptions regarding the richness of private financial markets. In the first, agents may sign long-term contracts of arbitrary complexity with a perfectly competitive sector of financial intermediaries. Such contracts are required only to respect the asymmetric information constraints above and so may depend upon the entire history of firm performance. In this environment the decentralization is very simple and calls only for linear taxes on firm profits. The intuition behind this result is an extension of that embedded in [Prescott and Townsend \(1984\)](#): since the government has no obvious advantage over the private sector at overcoming agency frictions, the latter continues to provide utility at the lowest cost possible. However, since the private sector will not achieve the redistributive goals of the government, a tax on firm profits is necessary for precisely this purpose.

This decentralization assumes that the financial intermediaries are capable both of committing to long-term contracts and monitoring the consumption of agents. Since both assumptions may be unreasonably strong, to complement the above I conclude with a restricted-instruments exercise and analyze *linear* taxation in an environment in which agents may only save in a risk-free bond and trade shares of their firms in spot markets. Within this setting the dependence of firm productivity growth on unobserved (and non-contractible) effort implies a novel effect of taxation policy: taxes on capital income affect the incentives for retaining ownership of one's business and hence the degree of risk-sharing present in the private sector. Intuitively, the willingness of investors to purchase shares in a firm will depend upon the incentives of the owner to exert continued effort to improve productivity, and hence upon the outside option, the return on (passive) savings. Within this linear class, the optimal policy calls for taxes on corporate profits, capital income, and wealth. However, interestingly, the taxes on personal capital income and wealth play no redistributive role, and vanish as agency frictions become negligible. Instead, the redistributive role is played by the corporate income tax, as this simply lowers the value of the firm ex-ante, and thereby taxes the elastic quantity of the model: the endowed talent of the firm owner. In contrast, taxes on capital income alter the private returns to effort in one's business and consumption smoothing, respectively.

**Related literature.** A vast literature builds upon the seminal contribution of [Mirrlees \(1971\)](#) to derive optimal taxes in informationally constrained economies. [Rothschild and Scheuer \(2013\)](#) conduct an optimal taxation exercise in an environment in which agents self-select into one of two industries and wages are endogenous. [Scheuer \(2014\)](#) explicitly considers entrepreneurs and firm formation within a static model and allows for both pecuniary externalities across occupations and the possibility of occupational-specific taxation. [Ales et al. \(2017\)](#) adopt a span-of-control technology as in [Rosen \(1982\)](#) and [Lucas \(1978\)](#) and allow firm size to be endogenous. [Scheuer and Werning \(2016\)](#) explore how optimal taxation policy must be altered in the presence of "superstar" effects in the form of assortative matching between individuals and firms. [Ales and Sleet \(2016\)](#) conduct an optimal taxation exercise in a similar environment in which the planner has an explicit concern for the welfare of shareholders, and [Ales et al. \(2015\)](#) consider the role of skill-biased technical change. The above models are all static, and so in contrast with the current paper, no agent bears risk or is subject to a moral hazard problem.

[Stantcheva \(2015\)](#) conducts an optimal taxation exercise in a life-cycle model with risky human

capital, but the nature of the private-information problem differs from that considered in this paper. An agent's wage at any moment is a function of her exogenously evolving stochastic ability and stock of human capital. Ability is unobservable, while the stock of human capital is observable, evolves deterministically and depends on monetary costs expended by the agent. The impediments to risk-sharing are therefore better thought of as hidden type rather than moral hazard, and the incorporation of human capital does not fundamentally alter the incidence of risk-bearing as in this model. Similarly, in papers within the New Dynamic Public Finance literature, such as [Golosov et al. \(2016\)](#) and [Farhi and Werning \(2013\)](#), the private information is in the form of hidden productivity and there is once again no moral hazard problem. Closer to the current paper are the works of [Albanesi \(2006\)](#), [Kapička and Neira \(2013\)](#), and [Best and Kleven \(2012\)](#), who conduct optimal taxation exercises in two-period economies with hidden effort, and [Makris and Pavan \(2019\)](#), who conduct a similar analysis in an environment with learning-by-doing. However, these papers do not derive implications of their framework for long-run distributions of consumption or income. Owing to their two-period nature,<sup>3</sup> these models cannot address how the risk borne by any agent depends upon her history of productivity shocks.

[Jones and Kim \(2014\)](#) characterizes competitive equilibria in an environment in which entrepreneurs are assumed to be unable to save and the growth rate of each agent's productivity depends partly on luck and partly on effort. I follow [Jones and Kim \(2014\)](#) in my modeling of the underlying technology of human capital, but in contrast to these authors, I do not impose a particular market structure and instead require only that allocations respect the underlying informational asymmetry. [Ai et al. \(2016\)](#), [Shourideh \(2013\)](#), and [Phelan \(2019\)](#) all show how a Pareto distribution of consumption may emerge in the presence of asymmetric information with optimal contracting in private insurance markets. The difference here is the nature of the agency problem: instead of allowing entrepreneurs to divert delegated assets to private consumption, in this paper entrepreneurs exert (privately observed) effort to improve (publicly observed) firm productivity.<sup>4</sup> I adopt the specification of random human capital governed by hidden effort because it better captures the aforementioned importance of individual-specific characteristics for business income.

---

<sup>3</sup>[Makris and Pavan \(2019\)](#) calibrate their model to fit the average working lifespan in the US, but the productivity of agents only changes *once*, so the dynamics of consumption and risk-bearing are identical to a two-period model.

<sup>4</sup>There are several other modeling differences: the welfare notion and life-cycle structure adopted here also differs from [Shourideh \(2013\)](#), while [Ai et al. \(2016\)](#) pursue an equilibrium analysis.

One virtue of this approach is that unlike Shourideh (2013), Jones and Kim (2014), and Ai et al. (2016), the efficient allocation in this paper does not possess the (counterfactual) property that the Pareto exponents for income and firm size must coincide. To the best of my knowledge, the only other paper with this feature is Ai et al. (2015), where it arises because of a limited commitment problem in which the manager may abscond with a fixed fraction of the firm’s capital stock and productively deploy it outside the firm. Efficient allocations are quite different in these two economies. If the market friction is limited commitment on the part of entrepreneurs, then a government may trivially implement the first-best allocation by levying confiscatory taxes on entrepreneurs’ income above a certain point (independent of firm performance) to effectively eliminate the outside option. In contrast, in this model the *efficient* allocations possess the qualitative features noted above (thick right tails in consumption and income), albeit with tail parameters that need not correspond with their empirical counterparts.

For clarity, the first section of this paper analyzes a principal-agent problem between a risk-averse entrepreneur and a risk-neutral principal in partial equilibrium where the flow output of the entrepreneur depends solely upon her individual productivity. I characterize and compute the policy functions of the principal and show numerically that typically the agent will exert more effort and bear more risk when more productive. In the subsequent overlapping generations economy with a continuum of agents, the productivity of any entrepreneur will depend both on her productivity *and* the endogenously determined shadow price of labor, as the latter determines the societal (resource) cost of assigning workers to entrepreneurs. This allows me to address how exogenous changes in the number of entrepreneurs, or changes in technology that primarily benefit entrepreneurs relative to workers, ought to translate into changes in inequality both *between* workers and entrepreneurs and *among* entrepreneurs. I view this as complementary to the above contributions, as I have a rich dynamic moral hazard component, and, accordingly, simplify the ex-ante heterogeneity for tractability.

The outline of the paper is as follows: Section 2 characterizes the optimal contract between a single agent and principal; Section 3 extends this to an overlapping generations economy with a continuum of agents with heterogeneous ability and shows how to compute stationary distributions of income in a number of example economies; Section 4 provides an implementation of the efficient allocation as well as a complementary restricted-instruments exercise, and Section 5 concludes.

Details of the recursive techniques, numerical implementation, and the welfare notions employed are relegated to the appendix.

## 2 Dynamic principal-agent model

For ease of exposition, I will first proceed in partial equilibrium and characterize the optimal contract between a risk-averse agent operating a risky technology and a risk-neutral principal who may trade at exogenously given prices. In the following section I will show how the problem of a benevolent planner in an overlapping generations economy may be decomposed into a series of principal-agent problems of the above form.

### 2.1 Formal setup

Time is indefinite and continuous. The economy consists of a single risk-averse agent and a risk-neutral principal, both of whom live forever. At any moment in time the agent may consume a flow amount  $c$  of a single good and take an action  $l \in [\underline{l}, 1]$  for some  $\underline{l}$ . The agent has preferences over stochastic sequences of consumption  $c = (c_t)_{t \geq 0}$  and effort  $l = (l_t)_{t \geq 0}$  given by

$$U(c, l) = \rho \int_0^\infty e^{-\rho t} \mathbb{E}[u(c_t, l_t)] dt \quad (1)$$

where the instantaneous flow utility from consumption and effort is

$$u(c, l) = \frac{(c^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} \quad (2)$$

for some constants  $\gamma > 1$  and  $\alpha \in (0, 1)$ . I will refer to  $l = 1$  as "retirement" and assume that the principal may observe when retirement occurs, but is unable to distinguish between all other actions.

At any point in time the agent is associated with a scalar variable  $\theta$  referred to as *productivity*. An agent of productivity  $\theta$  inelastically produces a flow of  $\theta$  units of output per time independently of her actions. The consumption and output produced by the agent are observable, while the productivity of the agent begins at the level  $\theta_0$  and evolves stochastically over time in a manner depending upon their effort. Specifically, there exists a stochastic process  $Z = (Z_t)_{t \geq 0}$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  distributed according to standard Brownian motion, such that if the agent chooses effort levels according to a stochastic process  $(l_t)_{t \geq 0}$  satisfying  $l_t \in \mathcal{L}$

for all  $t \geq 0$ , then productivity follows the law of motion,

$$d\theta_t = \mu(l_t)1_{l_t \neq 0}\theta_t dt + \sigma 1_{l_t \neq 0}\theta_t dZ_t \quad (3)$$

where  $\sigma > 0$  and  $\mu : \mathcal{L} \rightarrow \mathbb{R}$  is given by  $\mu(l) := \bar{\mu} - [\bar{\mu} - \underline{\mu}](1 - l)$  for some  $\bar{\mu} > \underline{\mu} > 0$ . As (3) indicates, the productivity of the agent stops evolving upon retirement. Finally, the principal is risk-neutral and discounts at the rate of time preferences of the agent and so his preferences over output and consumption are

$$U^P(c, l) = \int_0^\infty e^{-\rho t} \mathbb{E}[\theta_t - c_t] dt. \quad (4)$$

The actions taken by both the agent and principal at any date may be an arbitrary function of the output observed up until that time. The following definition formalizes this mathematically.

**Definition 2.1.** An allocation chosen by the principal consists of a pair of  $\mathcal{F}$ -adapted processes<sup>5</sup>  $(c, l^P) = (c_t, l_t^P)_{t \geq 0}$ , while an agent's strategy consists of a single  $\mathcal{F}$ -adapted process  $e = (l_t)_{t \geq 0}$ . For any allocation  $(c_t, l_t^P)_{t \geq 0}$  and agent strategy  $(l_t)_{t \geq 0}$ , the continuation utility of the agent is given by the stochastic process  $W \equiv W^{c, l}$  defined by

$$W_t := \rho \int_t^\infty e^{-\rho[s-t]} \mathbb{E}[u(c_s, l_s) | \mathcal{F}_t] ds \quad (5)$$

for all  $t \geq 0$  almost surely.

Since the effort levels taken by the agent are private information, the principal must restrict his attention to allocations that are incentive compatible. An allocation is incentive compatible if the agent wishes to adhere to the effort recommendations of the principal after every history. To formalize this notion, note that the utility of the agent under a given allocation when adhering to an arbitrary strategy amounts to evaluating the allocation under a *change of measure* associated with his own strategy. That is, the allocation of the principal specifies consumption as a function of every finite history of output, and when choosing a strategy, the agent understands how his actions change the probability of each history and weights them accordingly. For any strategy  $e = (l_t)_{t \geq 0}$  I will write  $P^l$  for this induced measure<sup>6</sup> and  $\mathbb{E}^l$  the associated expectation. The utility of an agent confronted with an allocation  $(c_t, l_t^P)_{t \geq 0}$  when adhering to a strategy  $(l_t)_{t \geq 0}$  is then

$$U(c, l) := \rho \int_0^\infty e^{-\rho t} \mathbb{E}^l [u(c_t, l_t)] dt. \quad (6)$$

---

<sup>5</sup>A sequence  $(x_t)_{t \geq 0}$  is  $\mathcal{F}$ -adapted if and only if there exist  $(\tilde{x}_t)_{t \geq 0}$  such that  $\tilde{x}_t : C([0, t]) \rightarrow \mathbb{R}$  for each  $t \geq 0$  and  $x_t = \tilde{x}_t((Z_s)_{0 \leq s \leq t})$  almost surely, for all  $t \geq 0$ , where for each  $t \geq 0$   $\tilde{x}_t$  is measurable w.r.t. the  $\sigma$ -algebra generated by evaluation maps up until time  $t$ .

<sup>6</sup>The formal construction of these measures is given in Definition A.2 in the appendix.

The formal definition of incentive compatibility is then the following.

**Definition 2.2.** An allocation  $(c_t, l_t^P)_{t \geq 0}$  is incentive compatible if  $U(c, l^P) \geq U(c, l)$  for all strategies  $l$ . The set of all incentive compatible allocations will be denoted  $\mathcal{A}^{IC}$ .

The principal's problem is then the following. It is indexed by the initial productivity  $\theta$  of the agent and the minimal level of promised utility  $W$  necessary for her participation.

**Definition 2.3.** Given outside option  $W$  and productivity  $\theta$  the problem of the principal is

$$\begin{aligned} V(W, \theta) &= \max_{(c, l) \in \mathcal{A}^{IC}} \int_0^\infty e^{-\rho t} \mathbb{E}[\theta_t - c_t] dt. \\ W &= \rho \int_0^\infty e^{-\rho t} \mathbb{E}[u(c_t, l_t)] dt \\ d\theta_t &= \mu(l_t) \theta_t dt + \sigma(l_t) \theta_t dZ_t, \quad \theta_0 = \theta. \end{aligned}$$

Since the increments of Brownian motion are independent, the principal's problem is naturally recursive in the variables  $W$  and  $\theta$ . However, the incentive compatibility constraints in Definition 2.2 implicitly involve a double infinity of constraints: there are infinitely many possible output paths and for each path, the agent has a continuum of actions from which to choose. However, using the techniques of Sannikov (2008) one may show that incentive compatibility is equivalent to requiring that the elasticity of utility to output weakly exceed a minimal level. To make this precise, note that in the deterministic case we have  $\dot{W}_t = \rho(W_t - u(c_t, l_t))$ . Using the martingale techniques of Sannikov (2008), one may show that the evolution of promised utility is the above deterministic term plus a stochastic integral with respect to the noise. The proof is standard and so omitted.

**Proposition 2.1.** *Given any allocation  $(c_t, l_t^P)_{t \geq 0}$  and agent strategy  $(l_t)_{t \geq 0}$ , there exists an ( $\mathcal{F}$ -predictable) process  $S = S^{l^P, C}$  such that the continuation utility  $W_t$  may be written*

$$W_t = W_0 + \rho \int_0^t (W_s - u(c_s, l_s)) ds + \int_0^t S_s dZ_s^l \tag{7}$$

for all  $t \geq 0$  almost surely, where  $Z_t^l$  is distributed as standard Brownian motion.

The content of Proposition 2.1 is that the stochastic increments to continuation utility may be viewed as the increments of the noise process multiplied by the *sensitivity* of utility to output. This representation allows the following continuous-time analogue of the one-shot deviation principle.

**Proposition 2.2.** *An allocation  $(c_t, l_t^P)_{t \geq 0}$  is incentive compatible if and only if the sensitivity process  $S$  given in Proposition 2.1 satisfies*

$$S_t \mu(l_t^P) + \rho \sigma u(c_t, l_t^P) \geq \rho \sigma S_t \mu(l) + u(c_t, l) \tag{8}$$

for all  $l \in [l, 1]$  and  $t \geq 0$ . The optimal choice of  $S_t$  for the planner is then  $S_t = \sigma E(l)(1 - \alpha)(c^{1-\alpha}l^\alpha)^{1-\gamma}$  when  $l > 0$  and  $S_t \equiv 0$  when  $l = 1$ , where

$$E(l) := \frac{\rho\alpha}{(1-\alpha)(\bar{\mu} - \mu)l}.$$

Proposition 2.2 shows that incentive compatibility is equivalent to the elasticity of utility to output being sufficiently large to outweigh the benefits of shirking. It is sometimes referred to as a "skin-in-the-game" constraint. The expression for volatility may be interpreted as the ratio of the benefits of shirking to the effect of a deviation on output, all as a fraction of flow utility. It shows that incentives must be more high-powered (promised utility more responsive to output shocks) when deviations are hard to detect or the benefits of deviation are large. Proposition 2.2 allows us to recast the problem of the principal as an optimal control problem with state  $(\theta, W)$ . Further, the log-linearity of the evolution of productivity, together with the homotheticity of preferences, allows for the following reduction to a single state variable.

**Lemma 2.3.** *For all  $W$  and  $\theta$  we have  $V(W, \theta) = V(W\theta^{\bar{\gamma}-1}, 1)\theta$  and the policy functions of the planner are functions of  $W\theta^{\bar{\gamma}-1}$ .*

*Proof.* The key observation here is that incentive compatibility is unaffected if we scale consumption in every history by the same scalar. Therefore, for any  $U, \theta$  and  $\mu > 0$ , the values of the two programs

$$\begin{aligned} \max_{(c,l) \in \mathcal{A}^{IC}} \int_0^\infty e^{-\rho t} \mathbb{E}[\theta_t - c_t] dt & & \mu \max_{(\bar{c}, l) \in \mathcal{A}^{IC}} \int_0^\infty e^{-\rho t} \mathbb{E}[\bar{\theta}_t - \bar{c}_t] dt \\ W = \int_0^\infty e^{-\rho t} \mathbb{E}[u(c_t, l_t)] dt & & W\mu^{\gamma-1} = \int_0^\infty e^{-\rho t} \mathbb{E}[u(\bar{c}_t, l_t)] dt \\ d\theta_t = \mu(l_t)\theta_t dt + \sigma(l_t)\theta_t dZ_t & & d\bar{\theta}_t = \mu(l_t)\bar{\theta}_t dt + \sigma(l_t)\bar{\theta}_t dZ_t \\ \theta_0 = \mu\theta & & \bar{\theta}_0 = \theta \end{aligned}$$

coincide, as can be seen by replacing  $\theta_t$  with  $\mu\bar{\theta}_t$  and  $c_t$  with  $\mu\bar{c}_t$ . The program on the left is  $V(W, \mu\theta)$  and that on the right is  $\mu V(W\mu^{\gamma-1}, \theta)$ .  $\square$

In the agency model of Sannikov (2008), output depends only on current actions and the promised utility is sufficient to act as a state variable. This is no longer true here since actions have persistent effects on flow output. However, Lemma 2.3 shows that the principal's choices depend only on promised utility *per unit of output*, and suggests the following simplification.

**Definition 2.4.** Given promised utility  $U$  and productivity  $\theta$ , define normalized promised utility  $u$  by  $u := [(1-\gamma)U]^{\frac{1}{1-\bar{\gamma}}}\theta^{-1}$ , where  $1-\bar{\gamma} = (1-\gamma)(1-\alpha)$ , and the normalized payoff function  $v$  by  $v(u) := V(u^{1-\bar{\gamma}}/(1-\gamma), 1)$ .

Lemma 2.3 shows that the optimal choices of the planner are functions only of normalized promised utility. Similar observations are made in Ai et al. (2016) and He (2008), where the agency problem involves hidden diversion of resources rather than hidden effort. Prior to formulating the problem in a manner suitable to computation, I will first determine from first principles the payoff to the principal when  $u$  is either very low or very high. To this end it is instructive to characterize the value and policy functions associated with the (suboptimal) allocations in which the principal recommends a given effort level for the entirety of an agent’s life but is unconstrained in his consumption choices. I will denote these value functions by  $v_r(u; l)$  for  $l \in [l, 1]$  and refer to them as restricted-action value functions.

**Lemma 2.4.** *For each  $l \in [l, 1]$  the restricted value and consumption functions are given by*

$$v_r(u; l) = \frac{1}{\rho - \mu(l)} + \bar{v}(l)u \qquad c_r(u; l) = \bar{c}(l)u$$

for some  $\bar{v}(l)$  increasing in leisure. Therefore, we have the following limiting behavior near zero

$$\lim_{u \rightarrow 0} v(u) = \frac{1}{\rho - \mu(l)}. \tag{9}$$

Appendix A.3 shows that the policy functions in Lemma 2.4 admit closed-form solutions, although the exact expressions are not crucial for understanding what follows. The important point is that  $\bar{v}(l)$  is increasing in leisure, which captures the intuitive fact that it is more expensive to provide one with the incentive to exert high effort. This helps provide some insight into how optimal effort recommendations vary with normalized promised utility and therefore build intuition for the implied dynamics of risk-bearing. The coefficients of both constant terms and the negative of the linear term in the value function are increasing in the effort recommended to the agent, which captures the fact that higher actions imply a higher growth rate of output but are also more costly. In addition, the limiting value of  $v$  given in Lemma 2.4 shows that the restricted value function associated with the highest action approximates the value of  $v$  near zero, so that the welfare loss from adhering to the restricted-action allocation (rather than the true efficient allocation) falls to zero as productivity rises.

Combining the above with standard results from continuous-time dynamic programming gives the following. The proof amounts to first deriving a PDE for  $V(U, \theta)$  and using the homogeneity in Lemma 2.3 to simplify, and is relegated to the appendix.

**Proposition 2.5.** *The normalized value function of the principal solves an ODE of the form*

$$0 = \max_{\substack{c \geq 0 \\ l \in [\underline{l}, 1]}} H(u, v(u), v'(u), v''(u))$$

*subject to the boundary and smooth-pasting conditions  $v(0) = \rho/(\rho - \mu(\underline{l}))$ ,  $v(\bar{u}) = v_{ret}(\bar{u})$  and  $v'(\bar{u}) = v'_{ret}(\bar{u})$ , where  $v_{ret}(u) = \underline{z} - u$  is the profit per unit of  $\theta$  for retiring the agent.*

The ODE in Proposition 2.5 does not possess a closed-form solution. The numerical methods employed for its computation are outlined in the appendix. Despite the absence of analytical solutions, the sharp characterization of the restricted-action allocations in Lemma 2.4 allows for an intuitive understanding of the joint dynamics of consumption, productivity, and the incidence of risk in the efficient allocation. Since both  $\rho/(\rho - \mu(l))$  and  $\bar{v}(l)$  are decreasing in leisure, Topkis' theorem implies the policy function  $l_r(u)$  associated with the restricted problem  $\bar{v}_r(u) := \max_{l \in [\underline{l}, 1]} v_r(u; l)$  is *decreasing* in  $u$ . In other words, the principal will recommend high effort to individuals with low normalized promised utility and vice versa. Since such individuals have greater incentives to shirk, incentive compatibility requires that they also bear more risk. This implies that agents bear high risk when their productivity is high *relative* to their promised utility. In the restricted-action case, promised utility typically falls over time, while productivity typically grows. It follows that those agents with the highest levels of promised utility are typically those with the highest levels of productivity, and hence low levels of normalized promised utility, and it is the unlucky agents with low promised utility who are retired.

This contrasts with the environments of Phelan and Townsend (1991) and Sannikov (2008), in which productivity is fixed over time and effort affects flow output rather than its growth. In such models agents are typically retired at high levels of utility because it is too costly to motivate them. Although it remains true here that the cost of motivating the agent increases with utility, there is also a simple but novel offsetting effect: agents in this model have high utility precisely *because* they experienced high productivity growth. The cost of motivating the agent may have grown, but since effort affects the growth rate of productivity, so too has the benefit, and Lemma 2.3 shows that the principal cares solely about the *ratio* of these costs to benefits. The desire to smooth consumption implies that promised utility typically remains stable or falls over time, whereas productivity grows exponentially on average. Normalized promised utility therefore typically falls over time and so richer and older agents usually bear more risk. The characterizations of restricted-action allocations in Lemma 2.4 allow us to make this intuition more precise.

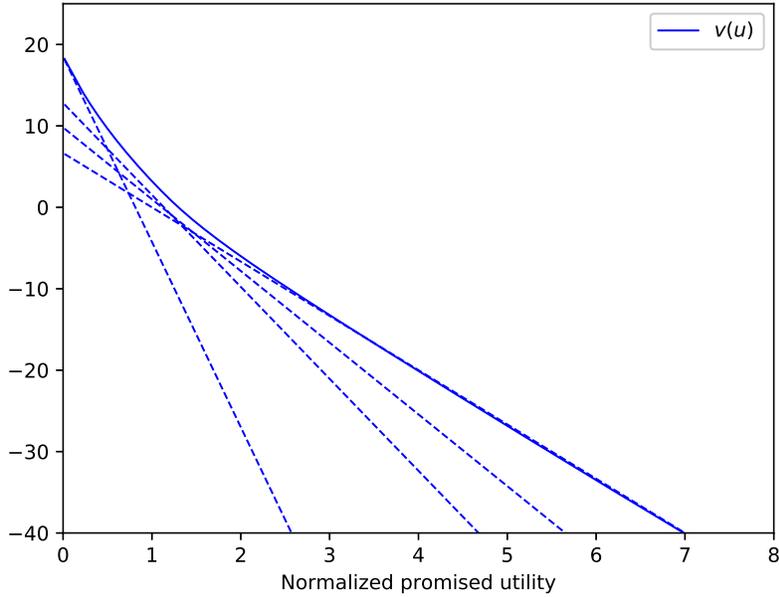


Figure 1: True and restricted value functions

Since none of the above restricted-action value functions are strictly optimal, the true value function must be calculated numerically. Figure 1 plots the true value function alongside the restricted-action value functions associated with a range of intermediate actions, for log utility and the parameters ( $\alpha = 0.3, \rho = .15, r = .15, \sigma = 0.15, \bar{\mu} = 0.12, \underline{\mu} = 0, \underline{l} = 0.2$ ). The steepest dotted line corresponds to the highest level of effort and the flattest to the lowest level. Obviously the true value function must lie above the restricted-action value functions pointwise. Further, as suggested by the above discussion, the true value function is well-approximated by the restricted-action value function associated with the highest action for low values of normalized utility. The above graphs do not give any insight into the pathwise properties of efficient allocations, such as how the incidence of risk evolves in response to productivity shocks. For this we first determine the law of motion of normalized promised utility.

**Lemma 2.6.** *Normalized utility follows the diffusion process  $du_t = \mu_u(u)dt + \sigma_u(u)dB_t$ , where  $\mu_u$*

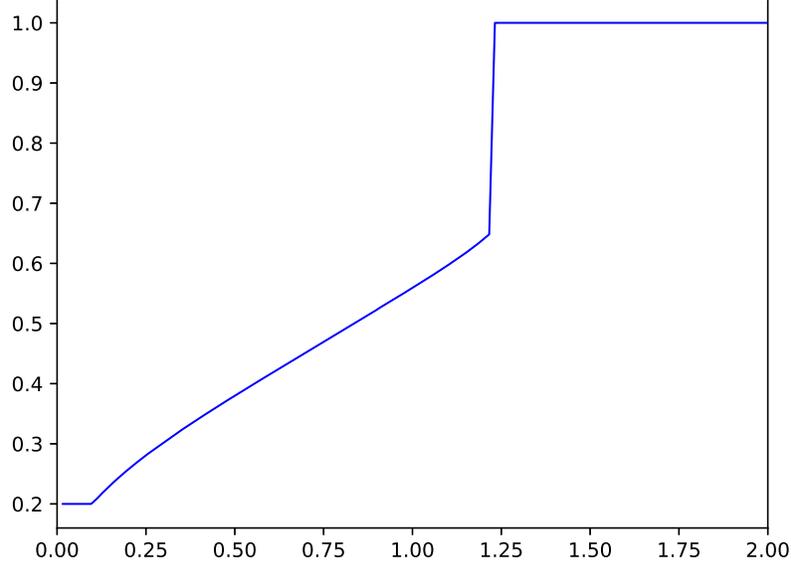


Figure 2: Leisure function

and  $\sigma_u$  are given by

$$\mu_u(u; \bar{c}, l) = \rho \left( \frac{1 - (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma}}{1 - \bar{\gamma}} \right) + \frac{\sigma^2 E(l(u))^2}{2} (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma} \bar{\gamma} - \mu(l(u))$$

$$\sigma_u(u; \bar{c}, l_t) = \sigma \left( E(l(u)) (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma} - 1 \right)$$

In the restricted-actions case, both  $\bar{c}$  and  $l$  are constant and so the law of consumption simply coincides with the law of  $u_t \theta_t$ ,

$$\mu_c = \rho \left( \frac{1 - (\bar{c}(l)^{1-\alpha} l^\alpha)^{1-\gamma}}{1 - \bar{\gamma}} \right) + \frac{\sigma^2 E(l)^2}{2} (\bar{c}(l)^{1-\alpha} l^\alpha)^{1-\gamma} \bar{\gamma}$$

$$\sigma_c = \sigma E(l) (\bar{c}(l)^{1-\alpha} l^\alpha)^{1-\gamma}.$$

High realizations of productivity will imply a fall in normalized promised utility in the restricted action allocation if and only if

$$\rho < \rho/E(l) + \frac{\sigma^2}{2} (2\bar{\gamma} - 1)(\bar{\gamma} - 1). \quad (10)$$

The requirement (10) in Lemma 2.6 appears quite modest. For instance, for a sufficiently large degree of risk aversion, it is satisfied for *all* values of leisure, regardless of the parameters that determine the function  $E$ .

Before proceeding to the macroeconomic setting with a continuum of agents, it is instructive to compare the above optimal contract with those derived in similar environments with hidden actions, and with the law of motion of income that obtains in economies with exogenously incomplete markets. This will also serve as a prelude to why thick right tails of consumption and firm size emerge in this environment and highlight the novel forces present in this model. [Shourideh \(2013\)](#) and [Phelan \(2019\)](#) both conduct optimal taxation exercises in economies with idiosyncratic capital risk and hidden consumption. In these models, productivity is fixed over time and risk-sharing is impeded by the ability of entrepreneurs to divert delegated capital to private consumption. To preserve incentives for investment, the risk borne by entrepreneurs must scale with the benefits of diverting capital to private consumption. Homotheticity implies that the delegated capital, consumption and labor, assigned to each entrepreneur are all proportional to one another. As a result, individuals with low marginal utility of consumption also control high amounts of capital, with the net effect on incentives to deviate ambiguous. It turns out that these exactly counteract one another, and that the elasticity of consumption with respect to output is common across agents, and in particular, does not vanish for agents in the right tail.

Since the technology adopted above is identical to that in [Jones and Kim \(2014\)](#), it is also worth contrasting the lessons drawn from this optimal contracting environment with their exogenously incomplete-markets model. In their model entrepreneurs are unable to save and so make a single choice of effort at any moment. Homotheticity ensures that all entrepreneurs select the same level of effort regardless of their productivity. In contrast, the above shows that agents with high productivity exert higher levels of effort in the efficient allocation. This is driven by the fact that relative to a low-productivity individual with the same promised utility, the actions of a high-productivity type have greater marginal benefit. In the perpetual youth environment that follows in [Section 3](#), the stationary distributions of both productivity and consumption exhibit thick right Pareto tails. The mechanism in the model that generates this is the same as that noted by [Jones and Kim \(2014\)](#), exponential growth for an exponential period of time leads to a Pareto distribution. This does not depend upon the specific nature of the cross-sectional distribution of shocks and is even true when growth rates are deterministic: if  $d \ln U(t) = \mu_U dt$  for some  $\mu_U > 0$  and the agent dies at rate  $\rho_D$ , then the stationary distribution satisfies

$$P(U \geq x) = P(T \geq \mu_U^{-1}(x/U_0)) = (x/U_0)^{-\rho_D/\mu_U}.$$

Numerous authors<sup>7</sup> in both the mathematics and economics literatures have demonstrated that a fat tail emerges whenever growth rates (rather than levels) are random. The crucial property necessary for the emergence of a fat tail is that the law of motion of the state variable be subject to proportional shocks. As the above numerical example demonstrates, the true law of motion of promised utility is everywhere bounded between two linear functions, which hints at the emergence of a thick right tail in consumption and promised utility. Although I cannot show this formally, this observation is borne out in all numerical simulations.

Before proceeding to the general equilibrium context, recall that the flow output of the agent in the above was simply assumed to be  $\theta$ . If flow output is scaled by a constant  $c$  then the principal behaves as if confronted with an agent of productivity  $c\theta$ . This has the following simple (but important) consequence.

**Lemma 2.7** (Homogeneity in productivity). *For any scale parameter  $c > 0$  denote the normalized value function associated with this level of productivity by  $v(u; c)$ . Then for all  $c, u \geq 0$  we have  $v(u; c) = cv(u/c; 1)$ .*

Lemma 2.7 will prove useful in the general equilibrium setting when the productivity of entrepreneurs depends upon an endogenously determined shadow price of labor. In what follows I will write  $v(\cdot) \equiv v(\cdot; 1)$  for the normalized payoff function associated with unitary productivity.

### 3 Efficient stationary allocations

The preceding section analyzed the optimal contract between a single risk-averse agent and a risk-neutral principal. This section builds on this analysis by showing how the problem of a benevolent planner in an overlapping-generations economy may be decomposed into a series of one-on-one principal-agent problems identical in form to those considered above. This allows me to derive the effects of private information on the long-run distribution of consumption. I first describe the environment and specify the welfare notion. The preferences and technology will be identical to those considered in the previous section. General equilibrium concerns affect both the level of promised utility that may be given to a generation and the productivity of entrepreneurs, because the latter depends on the number of agents working for their business.

---

<sup>7</sup>See, for instance, [Benhabib et al. \(2011\)](#) and the references therein.

### 3.1 Environment

Time is again continuous and indefinite. At any instant there is a continuum of agents alive in the economy with subjective discount factor  $\rho_S$  who die at rate  $\rho_D$ . A flow of  $\rho_D$  agents are born every unit of time so that the total population is fixed at unity. The utility function of all agents continues to be of the form given in (1) and (2) with the discount rate now  $\rho := \rho_S + \rho_D$ . Every agent alive at the initial date is indexed by a single variable  $v \in V \subseteq \mathbb{R}$  identified with promised utility. To each  $v$ -agent there is an associated process  $Z^v = (Z_t^v)_{t \geq 0}$  distributed according to standard Brownian motion and referred to as the *noise* process for agent  $v$ . These processes are independent of one another, and so by a law of large numbers for a continuum of agents,<sup>8</sup> the ex-post distribution across agents will coincide with the ex-ante distribution faced by a single agent. The noise processes of agents of subsequent generations will be indexed by agents' dates of birth rather than their promised utility and so agents are only distinguished by date of birth and possibly type.

Agents may either run their own firm or work for someone else. All agents have a common, fixed level of productivity  $\theta_0 = \bar{\theta}$  at birth. The productivity  $(\theta_t)_{t \geq 0}$  evolves according to the same law of motion given in the partial equilibrium setting

$$d\theta_t^v = \mu(l_t)1_{l_t \neq 0}\theta_t^v dt + \sigma 1_{l_t \neq 0}\theta_t^v dZ_t^v$$

where  $(l_t)_{t \geq 0}$  denotes the leisure enjoyed by the agent at any time  $t$ . In the partial equilibrium analysis the flow output of an agent simply coincided with her productivity  $\theta_t$ . In contrast, in this section the output of an entrepreneur is a function of both her productivity and the total effective labor assigned to her. All agents inelastically supply a flow  $L$  of labor per unit of time, irrespective of the effort employed to improve their productivity. If an entrepreneur of productivity  $\theta$  is assigned  $L$  units of effective worker labor, then the flow output produced is given by

$$F(\theta, L) = Z\theta^{1-\beta}L^\beta \tag{11}$$

for some fixed  $Z > 0$ . An allocation must specify the consumption, recommended effort exerted and effective labor assigned to every member of the initial generation throughout her (random) life as a function of the initial promised utility, type, and history of output, together with analogous quantities for all subsequent generations as functions of their date of birth and initial type. Implicit in this definition is the fact that effective worker labor assigned to an agent is non-zero if and only

---

<sup>8</sup>Subject to the usual measurability caveats.

if she is engaged in the risky activity. Because of the presence of the initial generation, allocations will be indexed by distributions  $\Phi$  over triples  $(v, \theta, i)$  of promised utility, productivity, and type. The formal definition is then the following.

**Definition 3.1.** Given a distribution  $\Phi$  over promised utility, productivity and types, an allocation consists of consumption, leisure, and labor assignments

$$\left\{ \left( c_t^{v,\theta,i}, l_{E,t}^{v,\theta,i}, l_{W,t}^{v,\theta,i}, L_t^{v,\theta,i} \right)_{t \geq 0} \mid (v, \theta, i) \in \text{supp}(\Phi) \right\} \\ \left\{ \left( c_t^{T,i}, l_{E,t}^{T,i}, l_{W,t}^{T,i}, L_t^{T,i} \right)_{t \geq T \geq 0} \mid i = E, W \right\}.$$

for initial and subsequent generations.

In the above  $1 - l_{W,t}$  refers to effort supplied as a worker and  $1 - l_{E,t}$  refers to effort supplied as an entrepreneur, with the understanding that only one of these may be non-zero. To ease notation I will write  $l_t$  where no ambiguity will arise. For agents not yet alive at the initial date, the superscript refers to birth date and the subscript to calendar time. I will write  $A$  for an arbitrary allocation and denote the set of all allocations by  $\mathcal{A}$ . Also, for any  $T \geq 0$  I will write  $\mathcal{A}_T$  for the set of all stochastic processes associated with agents born at time  $T$ . Note that aggregate consumption  $C_t$ , labor  $L_t$ , and output  $Y_t$  at any date comprise contributions from the initial generations and from subsequent generations. Formal definitions are contained in the appendix. The (physical) resource constraints are then the following.

**Definition 3.2.** An allocation satisfies the goods resource constraints if  $C_t \leq Y_t$  for all  $t \geq 0$  and satisfies the labor resource constraints if  $L_t^l \leq L$  for all  $t \geq 0$ . An allocation will be termed *resource feasible* if it satisfies both of the resource constraints. The set of all resource feasible allocations will be denoted  $\mathcal{A}^{RF}$ .

Finally, I will suppose that the weight placed on an agent born at time  $T$  is  $\alpha(T) = e^{-\rho s T}$ . This may be interpreted as a generalized utilitarian objective over generations, since it implies that the planner values the utility of an agent at any given date the same regardless of the agent's date of birth. It also ensures that social preferences assume a simple form suitable for recursive analysis. Lemma A.5 shows that it is equivalent to the objective function

$$U^P(A) = \int_0^\infty \left( e^{-\rho t} \underline{U}_t + \int_0^t e^{-\rho[t-T]} e^{-\rho s T} U_t^T dT \right) dt.$$

### 3.2 Recursive analysis

The previous section specified the physical resource constraints. This section will define the incentive compatibility and promise-keeping constraints and relate the planner's problem to the principal-agent problems of Section 2.

**Definition 3.3.** Given an initial distribution  $\Phi$  over promised utility and types, an allocation  $A$  satisfies promise-keeping if  $U(c^{v,\theta}, l^{v,\theta}) = v$  for all  $(v, \theta)$  in the support of  $\Phi$ .

Notice that promise-keeping is only relevant for the first generation. Incentive-compatibility requirements are of two separate types: an agent must be induced to reveal her type at birth (if not in the first generation) and an entrepreneur must be induced to follow the planner's effort recommendations. I will refer to the first type as *type-revelation* constraints and the second type as *incentive-compatibility* constraints. The first captures *adverse selection* and the second *moral hazard*.

**Definition 3.4** (Incentive compatibility). An allocation  $(c_t^E, c_t^W, l_t^E, l_t^W)_{t \geq 0}$  for a particular generation is incentive compatible if for all strategies  $l$ ,

$$U_E(c^E, l^E) \geq \max \{U_E(c^E, l), U_W(c^W, 0)\}.$$

Since there are two dimensions of private information, incentive compatibility must account for the possibility of double deviations, in which an agent simultaneously misreports her type and subsequently deviates from the recommended actions. However, the simple specification of hidden types in the preferences ensures that the analysis remains tractable, because regardless of one's type, preferences over consumption and leisure for agents assigned to the risky activity are common knowledge. This also ensures that there is no loss in supposing that all entrepreneurs and all workers are offered the same contracts; so an allocation need only specify the utility to a worker and the utility of an entrepreneur if engaged in the risky activity. The presence of hidden types therefore simply imposes the sole additional requirement on the planner that the initial promised utility of entrepreneurs be sufficiently high to induce truthful revelation. I will write  $\mathcal{A}^{RF} \equiv \mathcal{A}^{RF}(\Phi)$  for the set of resource feasible allocations and  $\mathcal{A}^{IC} \equiv \mathcal{A}^{IC}(\Phi)$  for the set of incentive compatible allocations, viewed as functions of the initial joint distribution of types, productivity and promised utility. Finally, define

$$\mathcal{A}^{IF} := \mathcal{A}^{RF} \cap \mathcal{A}^{IC}$$

for the set of all *incentive feasible* allocations. I can now state the planner's problem then show how it may be broken down into more tractable one-on-one principal-agent problems. The arguments

employed here generalize those employed by [Farhi and Werning \(2007\)](#) in an endowment setting with no technological interdependence among agents.

**Definition 3.5.** Given an initial distribution  $\Phi$ , the planner's problem is

$$V(\Phi) = \max_{A \in \mathcal{A}^{IF}(\Phi)} U^P(A).$$

The planner's problem appears intractable for an arbitrary distribution of utility, even in the case of full information and no heterogeneity in types. I will therefore focus on *stationary* solutions, in which the implied distributions of productivity and promised utility are constant over time. To solve these I will first consider the simpler problem of a planner who may trade both goods and labor intertemporally at the rate of time preference, which I will term the *relaxed* planner's problem.

**Definition 3.6.** Given an initial distribution  $\Phi$ , the relaxed problem of the planner is

$$\begin{aligned} V^R(\Phi) = \max_{A \in \mathcal{A}^{IC}} & \int_0^\infty \left( e^{-\rho t} \underline{U}_t + \int_0^t e^{-\rho[t-T]} e^{-\rho T} U_t^T dT \right) dt \\ & \int_0^\infty e^{-\rho t} (C_t - Y_t) dt \leq 0 \\ & \int_0^\infty e^{-\rho t} (L_t^l - L_t^W) dt \leq 0. \end{aligned}$$

Stationary solutions to the relaxed planner's problem will solve the planner's problem.

**Lemma 3.1.** *Suppose that for some distribution  $\Phi$  of initial promised utility, productivity and activities the allocation  $A$  solves the relaxed planner's problem with intertemporal price  $\rho$  and that the sequence of implied distributions of promised utility, productivity and activities  $(\Phi^t)_{t \geq 0}$  is constant. Then  $A$  also solves the planner's problem with initial distribution  $\Phi$ .*

*Proof.* If  $A$  solves the relaxed planner's problem, then by definition it satisfies the promise-keeping and incentive compatibility constraints of the original problem. Further, since consumption and output at any date are solely a function of the distribution, we know that both  $C_t - Y_t$  and  $L_t^l - L_t^W$  are constant over time; so the discounted resource constraint implies the resource constraint.  $\square$

The relaxed planner's problem still takes the entire joint distribution of productivity and promised utility as its argument. However, there are now only two constraints, and so the only interdependence between the decisions made by the planner regarding agents in different generations or engaged in different activities may be captured by Lagrange multipliers on resource constraints. Formally,

by the general theorem of Lagrange, there exist multipliers  $\lambda := (\lambda_R, \lambda_L)$  such that the optimal allocation solves

$$V_\lambda(\Phi) = \max_{A \in \mathcal{A}^{IC}} \int_0^\infty e^{-\rho t} \left( U_t + \lambda_R \left[ Y_t - C_t + \lambda_L \left( L_t^W - L_t^l \right) \right] \right) dt \quad (12)$$

where I have written

$$U_t = e^{-\rho D t} \underline{U}_t + \int_0^t e^{-\rho D [t-T]} U_t^T dT$$

for the total flow utility experienced at time  $t$ . For a given  $\lambda$ , the consumption and labor assignments for the initial generation do not affect the consumption and labor assignments of future generations. Therefore, solving (12) simply amounts to maximizing the components of the integral relevant to each generation *in isolation*. In this way the problem of the planner decomposes into a series of one-on-one problems similar in form to the principal-agent problems analyzed in the above partial equilibrium setting. I will refer to the problem of a planner who takes  $\lambda$  as given and wishes to maximize the weighted average of utility of all agents born at a given instant as a *generational planner problem*. These sub-problems are similar to (but distinct from) the component planning problems introduced in the taste shocks settings of [Farhi and Werning \(2007\)](#) and [Atkeson and Lucas \(1992\)](#). Notice that the objects of choice in the following are the utilities, output, consumption, labor supply, and labor demand of agents born at a particular date. Recall that an allocation for a particular generation is an element of  $\mathcal{A}_T$  and denoted  $\left\{ (c_t^{T,\psi}, l_t^{T,\psi}, L_t^{T,\psi})_{t \geq T} \mid \psi \in \Psi \right\}$ .

**Definition 3.7.** Given  $\lambda = (\lambda_R, \lambda_L)$ , the generational planner is

$$V_\lambda^G = \max_{\mathcal{A}_T^{IF}} \int_0^\infty e^{-\rho t} \left( U_t^T + \lambda_R \left[ Y_t^T - C_t^T + \lambda_L \left( L_t^{W,T} - L_t^{E,T} \right) \right] \right) dt.$$

Now define  $\Pi$  to be the minimum cost of providing a weighted level of utility  $W$  to a generation of newborns when the multiplier on the labor resource constraint is  $\lambda_L$ ,

$$\begin{aligned} \Pi(W, \lambda_L) &= \min_{\mathcal{A}_T^{IF}} \int_0^\infty e^{-\rho t} \left[ C_t - Y_t + \lambda_L \left( L_t^l - L_t^W \right) \right] dt \\ W &= \bar{U}_E [1 - G] + \bar{U}_W G \\ \bar{U}_E &\geq \bar{U}_W \\ (\bar{U}_E, \bar{U}_W) &= (U_E(c^E, l^E), U_W(c^W, l^W)). \end{aligned}$$

where  $\mathcal{A}_T^{IF}$  denotes the set of incentive feasible allocations for a particular generation. Notice that the only qualitative difference between the above problem and the principal-agent problem analyzed in Section 2 is the presence of an additional constraint requiring that the utility of an entrepreneur be sufficiently high to ensure truthful revelation. Then we have the following.

**Proposition 3.2.** *The solution to the generational planner's problem with multiplier  $\lambda = (\lambda_R, \lambda_L)$  is  $V_\lambda^G = \max_W W + \lambda_R \Pi(W, \lambda_L)$ .*

**Definition 3.8.** For a given pair  $\lambda := (\lambda_R, \lambda_L)$  of multipliers, denote the associated stationary distribution over  $\Omega := Z \times \Theta$  by  $\Phi_\lambda$ . The stationary form of the goods and labor resource constraints then reduce to the following pair of equations

$$\begin{aligned} 0 &= \int_{\Omega} \mathbb{E} \left[ c^{v,\theta,\psi} - F \left( \theta^{v,\theta,\psi}, L^{v,\theta,\psi} \right) \right] \Phi_\lambda(d\omega) \\ G &= \int_{\Omega} \mathbb{E} \left[ L^{v,\theta} \right] \Phi_\lambda(d\omega). \end{aligned} \tag{13}$$

The following summarizes the relationship between the planner's problem and the component planners' problems and is the culmination of the formal analysis.

**Lemma 3.3.** *If the multiplier  $\lambda$  satisfies (13), then the solution to the relaxed planner's problem when the initial distribution of promised utility, productivity and types is  $\Phi_\lambda$  amounts to adhering to the solutions to the generational planner's problem.*

Proposition 3.3 concludes the discussion of the planner's problem and the manner in which it is related to simpler component planners' problems. I will now combine these decomposition theorems with the optimal contract characterized earlier in partial equilibrium to infer properties of the long run distributions of consumption and productivity. This amounts to solving generational planners' problems and then varying the shadow prices of goods and labor until the resource constraints holds in the associated stationary distribution. For simplicity, I assume only workers may exert effort. The fraction of potential entrepreneurs will be denoted  $1 - G$ , where  $G \in (0, 1)$ . Types are permanent and unobserved but productivity evolves stochastically.

### 3.3 No workers

First suppose all agents have the ability to be entrepreneurs and worker labor is useless in production. The multiplier on the labor resource constraint is then zero and the problem facing the generational planner is

$$V_\lambda^G = \max_U U + \lambda_R \Pi(U, \theta_0). \tag{14}$$

In this case only the initial level of promised utility must be varied until the resource constraint is satisfied. The problem of a generational planner confronted with a generation of newborns is then

$$\max_{\substack{A \in \mathcal{A}^{IC} \\ U = U_E(c,l)}} U + \lambda_R \int_0^\infty e^{-\rho t} \mathbb{E} [Z\theta_t - C_t] dt.$$

Conditional on the choice of  $U$  this problem is identical to the principal-agent problem analyzed in the partial equilibrium setting. For each initial  $U$  we find the associated stationary distributions of consumption and production and then vary  $U$  until the resource constraint is satisfied. This would seem to require that we solve for a two-dimensional distribution for each initial level of promised utility. However, just as the homotheticity of preferences and log-linearity of the law of motion allowed for a simplification of the principal's problem, so too do the linear policy functions imply that when calculating aggregate quantities we need only restrict our attention to a one-dimensional distribution.

**Definition 3.9.** Given a stationary distribution  $\Phi$  over productivity and normalized utility, the summary measure for any  $B \subseteq [0, \infty)$  is defined by

$$m(B) = \int_B \int_0^\infty \theta \Phi(\theta, u) d\theta du.$$

The homogeneity of the policy functions ensures that aggregate quantities may be expressed in terms of this *summary* measure. For instance, since  $C(\theta, u) = c(u)\theta$ , aggregate consumption is

$$\bar{C}(U) = \int_0^\infty \int_0^\infty C(\theta, u) \Phi(\theta, u) d\theta du = \int_0^\infty \int_0^\infty c(u) \theta \Phi(\theta, u) d\theta du = \int_0^\infty c(u) m(u) du$$

while aggregate output is simply  $\int_0^\infty m(u) du$ . The following, taken from Ai ?, shows that this summary measure solves a version of the Kolmogorov forward equation for the *single variable*  $u$ .

**Lemma 3.4.** *If  $(u_t, \theta_t)_{t \geq 0}$  evolves according to  $(du_t, d\theta_t) = (\mu_u(u)dt + \sigma_u(u)dZ_t, \mu_\theta(u)\theta_t dt + \sigma_\theta(u)\theta_t dZ_t)$  for some  $\mu, \sigma_u, \mu$  and  $\sigma_\theta$ , then  $m(u) := \int_0^\infty \theta \Phi(\theta, u) d\theta$  solves the ODE*

$$0 = -(\rho_D - \mu_\theta(u))m(u) - [(\mu_u(u) + \sigma_\theta(u)\sigma_u(u))m(u)]' + \frac{1}{2}[\sigma_u^2(u)m(u)]''. \quad (15)$$

A proof is given in Section A.5. The boundary conditions for the ODE (15) are not obvious *a priori* but need not be determined to find the stationary distribution. One can interpret (15) as the stationary distribution of a diffusion process and solve for the stationary distribution of a Markov chain "close" in distribution to the original process following the method of Kushner and Dupuis (1992).<sup>9</sup> For any initial normalized  $\bar{u}$ , denote the implied density by  $m_{\bar{u}}$  and write the average productivity and consumption of agents

$$\begin{aligned} M(\bar{u}) &:= \int_0^\infty m_{\bar{u}}(u) du \\ C(\bar{u}) &:= \int_0^\infty c(u) m_{\bar{u}}(u) du. \end{aligned} \quad (16)$$

---

<sup>9</sup>Further details of this procedure are given in the appendix.

Solving for the efficient allocations in this economy then amounts to solving a single non-linear equation:

$$M(u) = C(u). \quad (17)$$

### 3.4 Workers and entrepreneurs

I will now relax the assumption of no technological dependence amongst agents and suppose only a fixed fraction  $1 - G < 1$  of agents may be entrepreneurs. The output of any entrepreneur now depends on the total amount of worker labor that has been assigned to her. In contrast with the case with only entrepreneurs, a generational planner facing a population of newborns must internalize the effect that the effort levels recommended to entrepreneurs have on the shadow price of labor. Suppose that the flow output per unit of time of an entrepreneur of productivity  $\theta$  who has been assigned  $L$  units of labor is now given by (11) for some  $\beta \in (0, 1)$ . Extending the analysis to a more general constant-elasticity-of-substitution production poses no difficulty. Notice that at any moment the optimal assignment of workers to an entrepreneur depends solely upon the shadow price of labor and the entrepreneur's productivity. It is the purely static problem

$$Y(\theta, \lambda_L) := \max_{L \geq 0} Z\theta^{1-\beta}L^\beta - \lambda_L L$$

with the solution requiring only elementary algebra.

**Lemma 3.5.** *Given the shadow price  $\lambda_L$ , the labor assigned to an entrepreneur of productivity  $\theta$  is*

$$L(\theta) = [Z\beta/\lambda_L]^{1/(1-\beta)} \theta. \quad (18)$$

*Flow output is  $Z^{1/(1-\beta)}(\beta/\lambda_L)^{\beta/(1-\beta)}\theta$ , and flow output net of labor resource costs  $\lambda_L L$  is given by  $Y(\theta, \lambda_L) = \bar{Z}\lambda_L^{-\beta/(1-\beta)}\theta$  where  $\bar{Z} = Z^{1/(1-\beta)}\beta^{\beta/(1-\beta)}(1-\beta)$ .*

Workers have no preferences over the entrepreneurs to whom they are assigned so the above assignment may be made independently of other choices. The key point of Lemma 3.5 is that changes in the shadow price for labor simply translates into changes in the productivity of every entrepreneur, with payoffs remaining proportional to  $\theta$ . It is convenient to define

$$x := \bar{Z}^{-1}\lambda_L^{\beta/(1-\beta)}u$$

for the promised utility of an agent per unit of productivity, given the shadow value of labor. In this way the problem may be rewritten

$$\Pi(\lambda) = \lambda_R \lambda_L G + \max_{x \geq 0} \bar{Z}^{1-\bar{\gamma}} \lambda_L^{-\frac{\beta(1-\bar{\gamma})}{1-\beta}} \frac{x^{1-\bar{\gamma}}}{1-\bar{\gamma}} + \bar{Z} \lambda_L^{-\frac{\beta}{1-\beta}} \lambda_R ([1-G]v(x) - Gx) \quad (19)$$

so that for any pair  $\lambda = (\lambda_R, \lambda_L)$ , the problem of the planner reduces to choosing an initial level of normalized promised utility  $x$ . As with the case without workers, for any such choice there will be an associated stationary density of normalized utility  $m_x(du)$  that may be computed using only the policy functions associated with  $v$  (corresponding to *unitary* productivity). The optimal choice of  $x$  depends on the multipliers  $\lambda_L$  and  $\lambda_R$ . For each choice of  $x$  we determine excess labor demand and excess consumption as functions of the multipliers and find conditions under which these both vanish. The labor resource constraint will be satisfied if and only if

$$G = [1 - G]M(x)(Z\beta/\lambda_L)^{\frac{1}{1-\beta}} \quad (20)$$

where  $M(x)$  is given by (16). To see this, note that

$$\begin{aligned} \text{RHS of (20)} &= (\text{average } \theta \text{ per entrepreneur in stationary distribution}) \\ &\quad \times (\text{mass of entrepreneurs}) \\ &\quad \times (\text{labor demand per unit of productivity (18)}) \end{aligned}$$

$$\text{LHS of (20)} = \text{labor supply by workers.}$$

Further simplification then gives the following.

**Proposition 3.6.** *The stationary  $x$  is a solution to the equation*

$$\left(\frac{1-G}{1-\beta}\right)M(x) = (1-G)C(x) + Gx \quad (21)$$

with the associated level of  $u$  given by  $u = Z(1-\beta)x([1/G - 1]M(x))^{-\beta}$ .

*Proof.* From Lemma 3.5 the output of each entrepreneur's firm is  $Z\theta^{1-\beta}L(\theta)^\beta = Z^{\frac{1}{1-\beta}}(\beta/\lambda_L)^{\frac{\beta}{1-\beta}}\theta$ .

Integrating over all entrepreneurs and using (21), output in the stationary distribution is

$$Y(x) = Z^{\frac{1}{1-\beta}}(\beta/\lambda_L)^{\frac{\beta}{1-\beta}}(1-G)M(x) = \bar{Z}\lambda_L^{-\frac{\beta}{1-\beta}}\left(\frac{1-G}{1-\beta}\right)M(x). \quad (22)$$

Aggregate consumption is simply the sum of consumption from both entrepreneurs and workers, or

$$\bar{Z}\lambda_L^{-\frac{\beta}{1-\beta}}[1-G]C(x) + \bar{Z}\lambda_L^{-\frac{\beta}{1-\beta}}Gx = \bar{Z}\lambda_L^{-\frac{\beta}{1-\beta}}([1-G]C(x) + Gx). \quad (23)$$

Dividing (22) and (23) by  $\bar{Z}\lambda_L^{-\frac{\beta}{1-\beta}}$  then gives the left- and right-hand sides of (21), respectively.

Rearranging (20) gives  $\lambda_L^{-\frac{\beta}{1-\beta}} = ([1/G - 1]M(x))^{-\beta}(Z\beta)^{-\frac{\beta}{1-\beta}}$ , and so the implied level of normalized promised utility is equal to  $u = \bar{Z}x\lambda_L^{-\frac{\beta}{1-\beta}} = \bar{Z}x([1/G - 1]M(x))^{-\beta}(Z\beta)^{-\frac{\beta}{1-\beta}}$  which reduces to the claimed expression.  $\square$

Note that to determine the constrained efficient stationary distribution, the value function need only be calculated *once* even though there is a continuum of agents and two resource constraints. Proposition 3.6 now allows for simple derivations of comparative statics for the initial level of  $x$ .

**Corollary 3.7.** *The stationary level of  $x$  is decreasing in  $G$ , increasing in  $\beta$ , and independent of  $Z$ .*

*Proof.* Define  $H(x, G) := M(x)(1 - \beta)^{-1} - C(x) - xG(1 - G)^{-1}$  and note that the stationary  $x$  is a root of  $H(\cdot, G)$ . The first two claims follow from Topkis' theorem and the fact that  $H$  is decreasing in  $G$  and increasing in  $\beta$ , and the last claim follows from the fact that  $Z$  appears nowhere in the equation defining  $x$ .  $\square$

The third claim in Corollary 3.7 may be viewed as a neutrality result. It shows that (Hicks-neutral) changes in total factor productivity (constant  $Z$  in the above) have no effect on the stationary value of  $x$ . They therefore have no effect on inequality in the associated stationary distributions of promised utility and consumption, as these quantities are simply scaled for all agents after every history by the same proportion.

### 3.5 Restricted-action allocations

I showed earlier that the restricted value and policy functions associated with the highest action are good approximations to the true value and policy functions for low levels of normalized promised utility. Since agents with low normalized utility have high productivity and so typically high consumption, one expects the right tail of the consumption distribution to look similar to that associated with the restricted-action allocation for the highest effort level. It is therefore instructive to consider the stationary distribution associated with these restricted policy and value functions as they may be calculated in closed form. Not only does this allow a clearer characterization of the implied distributions of wealth and firm size, it will serve as a prelude to the decentralization results of Section 4.2. The only change to the problem (19) facing the generational planner is that  $v$  is replaced with  $v_r(x) := \max_{l \in [l, 1]} v_r(x; l)$ , and the stationary  $x$  is now the solution to the analogue of (21),

$$\left( \frac{1 - G}{1 - \beta} \right) M_r(x) = (1 - G)C_r(x) + Gx, \quad (24)$$

where  $M_r$  and  $C_r$  denote aggregate output and consumption in the optimal restricted-action allocation, given the initial level of normalized utility  $x$ . Now recall that Lemma 2.4 shows that the restricted-action normalized value and policy functions have the explicit solutions  $v_r(u; l) =$

$(\rho - \mu(l))^{-1} - \bar{v}(l)u$  and  $c_r(u; l) = \bar{c}(l)u$ . Since  $C(U, \theta) = c(u)u\theta$  and  $u = [(1 - \gamma)U]^{\frac{1}{1-\gamma}}\theta^{-1}$ , we have  $C(U_t, \theta_t) = \bar{c}(l)u_t\theta = \bar{c}(l)[(1 - \gamma)U_t]^{\frac{1}{1-\gamma}}$ . Lemma ?? shows that consumption then follows the law of motion  $dc_t = \mu_c(l)c_t dt + \sigma_c(l)c_t dZ_t$  for some constants  $\mu_c(l)$  and  $\sigma_c(l)$ . This leads to the following simple characterization. The proof is given in Section A.5.

**Proposition 3.8.** *The optimal restricted-action allocation is that corresponding to leisure level  $l_r(x)$ , where  $x$  solves*

$$\frac{(1 - \beta)^{-1}}{\rho_D - \mu(l_r(x))} = \frac{\bar{c}(l_r(x))}{\rho_D - \mu_c(l_r(x))} + \frac{Gx}{1 - G}. \quad (25)$$

For each  $l \in [\underline{l}, 1]$  the stationary distribution of consumption associated with the restricted-action allocation for the leisure level  $l$  is double-Pareto with the tail parameters given by

$$\alpha_{\pm}(l) = -\frac{\bar{\gamma}}{2} \pm \frac{\bar{\gamma}}{2} \sqrt{1 + \frac{4\rho_D}{\rho(1-x)}(1 - 1/\bar{\gamma})(2 - 1/\bar{\gamma})}$$

where  $x = x(l)$  solve  $0 = E(l)^2(\bar{\gamma} - 1)(\bar{\gamma} - 1/2)\sigma^2 x^2 + \rho x - \rho$ , at least for  $\gamma > 1$ , while for  $\gamma = 1$  the expression for tail parameters tends to

$$\alpha_{\pm}(l) = -\frac{1}{2} \pm \frac{1}{2} \sqrt{1 + \frac{8\rho_D}{E(l)^2\sigma^2}}.$$

The stationary distribution of a firm associated with the restricted-action allocation for the leisure level  $l$  is also double-Pareto, with the tail parameters now given by

$$\beta_{\pm}(l) = \mu(l)/\sigma^2 - 1/2 \pm \sqrt{(\mu(l)/\sigma^2 - 1/2)^2 + 2\delta/\sigma^2}. \quad (26)$$

Proposition 3.8 shows that the characterization of the stationary distribution given in Proposition 3.6 is even simpler in the case of restricted-action allocations. Furthermore, when combined with the discussion following Proposition 2.5, the characterization gives us insight into the tail of the (unrestricted) efficient stationary distribution. Intuitively, individuals in the right tail in the stationary distribution are those who have experienced rapid productivity growth, and hence typically have low normalized utility. Since  $v(0) \approx (\rho - \mu(l))^{-1}$ , the welfare loss to the principal from adhering to the restricted-action allocation for the lowest leisure (and highest effort) level when confronted with such agents becomes arbitrarily small as  $u \rightarrow 0$ . We therefore expect that the tail of the stationary distribution looks like the tail associated with the lowest leisure level, and this is precisely what is observed in every simulation.

## 4 Implementation

The above analysis has focused on the forces that shape the long run efficient levels of inequality in an economy with repeated moral hazard, with no discussion of a market structure. All quantities were implicitly specified by a social planner administering the direct mechanism. In this final section I will discuss the implications of the analysis for taxes, given assumptions on the market structure.

### 4.1 Unrestricted contracts

I first consider how efficient allocations may be implemented by a government when agents may write contracts of arbitrary complexity with private financial intermediaries. In this case, despite the rich history dependence of the optimal contract, a constant linear tax on the consumption of entrepreneurs, together with a constant linear subsidy on the consumption of workers, will suffice to implement the efficient allocation. Suppose that a fraction  $1-G$  of agents are potential entrepreneurs and that a fraction  $G$  may only be workers, with the entrepreneurs obtaining utility  $\bar{\psi}U$  from the entrepreneurial activity when given utility  $U$  from consumption and effort. I will also suppose that intermediaries may borrow and lend at the rate of time preference. I will first define the problem of an intermediary faced with an agent to whom he has promised a certain amount of utility, given prices for labor and intertemporal trade. This problem has a similar form to that of the principal considered in the above partial equilibrium setting. For ease of notation, in what follows I will write

$$U_E((1-\tau)c, l) := \rho \int_0^\infty e^{-\rho t} \mathbb{E}[u((1-\tau_t)c_t, l_t)] dt$$

for the utility of an entrepreneur alive at the first date from the sequences  $(c_t, l_t)_{t \geq 0}$  when facing a sequence of taxes  $(\tau_t)_{t \geq 0}$ , together with analogous quantities for all future generations and workers.

**Definition 4.1** (Intermediaries' problem). Given sequences of wages  $w = (w_t)_{t \geq 0}$ , interest rates  $r = (r_t)_{t \geq 0}$  and consumption taxes  $(\tau_{C,t}^E, \tau_{C,t}^W)_{t \geq 0}$  on entrepreneurs and workers, respectively, the problem of a financial intermediary facing an agent with outside option  $U$  and productivity  $\theta$  is

$$\Pi_E(U, \theta; w, r, \tau^E) = \max_{(c, l) \in \mathcal{A}^{IC}} \int_0^\infty e^{-\int_0^t (r_s + \rho_D) ds} \mathbb{E} \left[ \bar{Z}(\beta, e) w_t^{-\frac{\beta}{1-\beta}} \theta_t - c_t \right] dt.$$

$$U = U_E((1-\tau^E)c, l)$$

$$d\theta_t = \mu(l_t)\theta_t dt + \sigma(l_t)\theta_t dZ_t$$

$$\theta_0 = \theta.$$

The problem of a financial intermediary facing an agent operating the risk-free technology with outside option  $U$  and productivity  $\theta$  is defined to be

$$\Pi_W(U, \theta; w, r, \tau^W) = \max_{(c, l) \in \mathcal{A}^{IC}} \int_0^\infty e^{-\int_0^t (r_s + \rho_D) ds} \mathbb{E} [w_t l_t \theta - c_t] dt.$$

$$U = U_W((1 - \tau^W)c, l).$$

**Definition 4.2.** Given sequences  $(\tau^E, \tau^W) := (\tau_t^E, \tau_t^W)_{t \geq 0}$  of linear taxes on the consumption of entrepreneurs and workers, a competitive equilibrium consists of sequences of wages  $(w_t)_{t \geq 0}$  and interest rates  $(r_t)_{t \geq 0}$ , together with an allocation

$$A := \left\{ (c_t^{v, \theta}, l_{E,t}^{v, \theta}, l_{W,t}^{v, \theta}, L_t^{v, \theta})_{t \geq 0} \mid (v, \theta) \in \Omega \right\}, \left\{ (c_t^{T, \psi}, l_{E,t}^{T, \psi}, l_{W,t}^{T, \psi}, L_t^{T, \psi})_{t \geq T \geq 0} \right\}_{\psi \in \Psi}$$

and contracts  $\left\{ (c_{E,t}^{v, \theta}, l_{E,t}^{v, \theta}, c_{W,t}^{v, \theta}, l_{W,t}^{v, \theta})_{t \geq 0} \mid (v, \theta) \in \Omega \right\}$  and  $(c_{E,t}^T, l_{E,t}^T, c_{W,t}^T, l_{W,t}^T, L_t^T)_{t \geq T \geq 0}$  such that:

- Given the levels of promised utility

$$v = U_E \left( (1 - \tau^E) c_E^{v, \theta}, l_E^{v, \theta} \right) \qquad v = U_W \left( (1 - \tau^W) c_W^{v, \theta}, l_W^{v, \theta} \right)$$

$$U_E^T = U_E \left( (1 - \tau^E) c_E^{T, \theta}, l_E^{T, \theta} \right) \qquad U_W^T = U_W \left( (1 - \tau^W) c_W^{T, \theta}, l_W^{T, \theta} \right)$$

the contracts chosen by the intermediaries minimize the costs of attaining  $v, U_E^T$  and  $U_W^T$  given prices  $w$  and  $r$ . This requires that for all  $(v, \theta) \in \Omega$ , the contracts  $(c_E^{v, \theta}, l_E^{v, \theta}, L^{v, \theta})$  and  $(c_W^{v, \theta}, l_W^{v, \theta})$  solve  $\Pi_E(v, \theta; w, r)$  and  $\Pi_W(v, \theta; w)$ , respectively, and for all  $T \geq 0$ , the contracts  $(c^{T, E}, l^{T, E}, L^T)$  and  $(c^{T, W}, l^{T, W})$  solve  $\Pi_E(U_E^T, \theta_0; w, r)$  and  $\Pi_W(U_W^T, \theta_0; w, r)$ , respectively.

- Financial intermediaries make zero discounted expected profits from all contracts written at or after the initial date:  $\Pi_E(U_E^T, \theta_0; w, r, \tau^E) = \Pi_W(U_W^T, \theta_0; w, r, \tau^W) = 0$  for all  $T \geq 0$ . Further, there is no contract that is not offered but would make positive profits if offered.
- All agents not in the first generation choose the contract that maximizes their utility: for all  $T \geq 0$  this requires  $U_E \left( (1 - \tau^E) c_E^{T, \theta}, l_E^{T, \theta} \right) \geq U_W \left( (1 - \tau^W) c_W^{T, \theta}, l_W^{T, \theta} \right)$ .
- labor markets and goods markets clear every instant.

There is no need to impose the requirement that the government budget balance because of Walras' law. In keeping with the focus of this paper on long run inequality, I will restrict attention to *stationary* equilibria in which wages, interest rates and the distributions of consumption and productivity are constant.

The competitive equilibrium without taxes need not coincide with the utilitarian efficient allocation characterized above. Competition amongst firms to provide insurance to workers will ensure that the level of promised utility is set at the point at which firms make zero expected discounted profits, and nothing forces this level to coincide with the constrained efficient level of ex-ante utility assigned to entrepreneurs. However, the form of the preferences ensures that it is easy to describe the effects of a constant linear consumption tax on utility. Specifically, if consumption is taxed at rate  $\tau$  and  $(c, l) := (c_t, l_t)_{t \geq 0}$  denotes the consumption and recommended effort levels specified in the contract between the firm and the entrepreneur, then the utility of the agent is given by

$$U((1 - \tau)c, l) = \rho \int_0^\infty e^{-\rho t} \mathbb{E}[u((1 - \tau)c_t, l_t)] dt = (1 - \tau)^{1-\gamma} U(c, l). \quad (27)$$

Homotheticity therefore implies that the imposition of a linear tax simply scales the utility experienced by the entrepreneur state-by-state, leaving incentive compatibility unchanged. From the perspective of the intermediary, with the introduction of a linear tax on entrepreneurs' income  $\tau$  the profit from contracting with an agent with normalized promised  $u$  changes from  $v(u)$  to  $v(u[1 - \tau]^{-1})$ . The level of normalized utility  $u(\tau)$  that obtains in the competitive equilibrium is then

$$u(\tau) = (1 - \tau)x$$

where  $x$  solves  $v(x) = 0$ . This implies the following.

**Proposition 4.1.** *Denote by  $U^E$  and  $U^W$  the levels of utility associated with a newborn entrepreneur and worker, respectively, in the efficient allocation characterized above. Then the efficient allocation may be implemented as a competitive equilibrium with interest rate  $r = \rho$ , constant linear taxes  $\tau^E$  and  $\tau^W$ , where  $\tau^E$  satisfies  $u(\tau^E) = [(1 - \gamma)U^E]^{\frac{1}{1-\gamma}}$ , and  $\tau^W$  satisfies*

$$\Pi_W(U_W^T, \theta_0; \lambda_L, \tau^W) = 0,$$

where  $\lambda_L$  is the multiplier on the labor resource constraint found in the planning problem.

In light of the decentralization results of [Prescott and Townsend \(1984\)](#), [Atkeson and Lucas \(1992\)](#), and [Golosov and Tsyvinski \(2007\)](#), the fact that the desire of the taxation authority is reduced when agents are unrestricted in the contracts they may write with intermediaries is not surprising. However, it is worth noting that [Proposition 4.1](#) shows that even when one introduces ex-ante heterogeneity, the redistributive role of the government does not *necessarily* imply progressivity of taxation.

## 4.2 Stock market equilibrium

Proposition 4.1 serves as a natural benchmark as it assumes the government is no more capable of providing insurance than the private sector. However, the assumption of a competitive sector of financial intermediaries able to commit to (history-dependent) contracts for the entirety of an agent's life may be unreasonably strong. Not only are these contracts highly history dependent and executed over long periods of time, they implicitly assume that the intermediary is capable of perfectly observing the asset positions of the agent at all times. For this reason I now complement the above by fixing a simpler market structure and characterizing the optimal taxation policy of the government within a particular parametric class. Specifically, I will assume that agents may only save in a risk-free bond and sell shares of their firm in competitive *spot* markets, and that the government may only levy *linear* taxes on all forms of income and wealth.

The state variable for the agent now consists of firm productivity  $\theta_t$  and wealth  $a_t$ . The latter is the sum of risk-free savings  $b_t$  and ownership of the asset  $p_t x_t$ , where  $x_t$  is the fraction of the firm owned by the agent and  $p_t$  the total price of the firm, so

$$a_t := b_t + p_t x_t. \quad (28)$$

Wealth now depends upon the endogenously determined price of the firm. Since the stock market is assumed to be competitive, I will assume that this price always equals the expected discounted value of firm profits, given the outside investors' expectations regarding the effort of the owner. The equilibrium notion outlined below will then impose the consistency requirement that these beliefs are correct, so that the outside investors break even in expectation. When the agent adheres to the strategy  $\hat{l} = (\hat{l}_t)_{t \geq 0}$  this present-discounted value is equal to

$$p_t = \int_t^\infty e^{-r(s-t)} (1 - \tau_d) \mathbb{E}^{\hat{l}}[\theta_s] ds.$$

If  $\hat{l}_t$  is constant over time, the Gordon growth formula implies the explicit expression

$$p_t = \frac{(1 - \tau_d)\theta_t}{r - \mu(\hat{l})}$$

for all  $t \geq 0$  almost surely. When effort expectations are constant and the government levies *constant* taxes  $(\tau_s, \tau_d, \tau_{cg}, \tau_a)$  on savings, profits, capital gains and wealth, the law of motion of wealth becomes

$$da_t = \left[ (1 - \tau_s)r - \tau_a - \bar{c}_t + \left( \tau_s r + (1 - \tau_{cg})\mu(l) - \mu(\hat{l}) \right) \bar{l}_t \right] a_t dt + \sigma(1 - \tau_{cg}) \bar{l}_t a_t dZ_t \quad (29)$$

where  $\bar{c}_t$  and  $\bar{l}_t$  denote the fraction of the agent's wealth consumed and invested in her business, respectively.<sup>10</sup> Note the distinction in (29) between  $\hat{l}$  and  $l$ : the former refers to the level of effort expected by investors, while the latter refers to the true effort exerted by the agent. The two must coincide in equilibrium but only the latter is affected by the agent directly. Note also that the tax on firm profits  $\tau_d$  appears nowhere in the law of motion (29) since the sole effect of this tax is to reduce the value of the firm (and hence the lifetime income of the agent).

**Definition 4.3.** Given an interest rate  $r$ , linear taxes  $\tau = (\tau_s, \tau_d, \tau_{cg}, \tau_a)$  and expectations of effort exerted  $(\hat{l}_t)_{t \geq 0}$ , the problem of an agent with firm size  $\theta$  and assets  $a$  is

$$\begin{aligned} V(a, \theta) &= \max_{c, l, \iota} \int_0^\infty \rho e^{-\rho t} \mathbb{E}[u(c_t, l_t)] dt \\ da_t &= [(1 - \tau_s)r - \tau_a - \bar{c}_t]a_t dt + \bar{l}_t a_t dR_t(l_t) \\ d\theta_t &= \mu(l_t)\theta_t dt + \sigma\theta_t dZ_t \\ (a_0, \theta_0) &= (a, \theta) \end{aligned}$$

where  $dR_t(l_t) = (\tau_s r + (1 - \tau_{cg})\mu(l) - \mu(\hat{l}))dt + \sigma(1 - \tau_{cg})dZ_t$ .

The notion of a competitive equilibrium is essentially standard: agents optimize, markets clear, and expectations are consistent with individual incentives.

**Definition 4.4** (Stock market equilibrium with linear taxes). Given an interest rate  $r$  and linear taxes  $\tau$ , a competitive stock market equilibrium consists of effort expectations  $\hat{l}$ , value functions  $V$  and policy functions  $(c, l, \iota) = (c_t, l_t, \iota)_{t \geq 0}$  for consumption, leisure and investment, such that the following hold:

- The functions  $(c, l, \iota)$  solve the consumer problem in Definition 4.3 given expectations  $\hat{l}$ .
- The outside investors break even in expectation, or  $l_t = \hat{l}$  for all  $t \geq 0$  almost surely.

A stationary competitive stock market equilibrium is one in which the cross-sectional distributions of wealth and firm size are constant over time.

The first observation here is that if  $\hat{l}$  does not depend upon  $\theta$  then the latter variable drops out of the individual agent problem and it becomes a standard portfolio problem of Merton-Samuelson type. The proof is simple algebra and so relegated to the appendix.

---

<sup>10</sup>To aid the reader, a discrete-time formulation of the law of motion is contained in Section B.1.

**Lemma 4.2.** *With linear taxes and constant expectations of effort, the policy functions for consumption and investment in the consumer problem are of the form  $c(a) = \bar{c}a$  and  $i(a) = \bar{i}a$  for some constants  $\bar{c}$  and  $\bar{i}$ , and the choice of leisure  $l$  is constant over time. The pair  $(\bar{c}, l)$  solves the pair*

$$\begin{aligned}\frac{\bar{c}}{\rho} &= \frac{(1 - \tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r}{\sigma^2(1 - \tau_{cg})\bar{\gamma}E(l)} \\ \frac{\rho - \bar{c}}{1 - \bar{\gamma}} &= (1 - \tau_s)r - \tau_a - \bar{c} + \frac{1}{2} \frac{[(1 - \tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]^2}{\sigma^2(1 - \tau_{cg})^2\bar{\gamma}}\end{aligned}\tag{30}$$

and  $\bar{i}$  is given by

$$\bar{i} = \frac{(1 - \tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r}{\sigma^2(1 - \tau_{cg})^2\bar{\gamma}}.$$

If we impose the requirement  $\bar{l} = l$ , the ensuing system (30) completely determines the equilibrium  $\bar{c}$  and  $l$ , given the taxes  $\tau_{cg}, \tau_s$  and  $\tau_a$ , with  $r = \rho$  for the same reasons as those given in the discussion of the characterization of the planner's problem in Section 3. Note that since the allocations with linear taxes imply constant effort by the agent, given the planner's objective function they are by definition weakly dominated by the restricted-action allocations characterized in Lemma 2.4. Further, since the policy functions for consumption and investment are linear in wealth, they share the property of the restricted-action allocations that they are characterized by constant mean and volatility of growth rates. It follows that if we can find linear taxes such that the mean and volatility of consumption growth in the competitive equilibria coincide with their counterparts in the restricted-action allocations, then we will have found the optimal linear taxes.

**Proposition 4.3.** *The optimal allocation possible with linear taxes may be implemented with a tax  $\tau_d$  on dividends, common taxes on capital income  $\tau_c = \tau_s = \tau_{cg}$ , and a tax  $\tau_a$  on wealth, where  $(w, \tau_c, \tau_a)$  are solutions to the system of equations*

$$\begin{aligned}0 &= \sigma^2 E(l)^2 (\bar{\gamma} - 1/2)(\bar{\gamma} - 1)w^2 + \rho\bar{\gamma}^2(w - \bar{\gamma}^2) \\ \tau_a &= \frac{\sigma^2}{\bar{\gamma}^2} \left( E(l)w - \frac{\rho\bar{\gamma}^2}{\bar{\gamma}(\rho - \mu(l)) + \sigma^2 E(l)w} \right) E(l)w. \\ \tau_c &= \frac{\sigma^2 E(l)w}{\bar{\gamma}(\rho - \mu(l)) + \sigma^2 E(l)w}\end{aligned}$$

and  $l$  is the optimal solution  $l_r(u)$  associated with the restricted problem  $v_r(u) := \max_{l \in [l, 1]} v(u; l)$ . The tax on dividends  $\tau_d$  is simply set at the level at which agents are indifferent between being an entrepreneur or a worker.

The system in Proposition 4.3 may be simplified when utility is logarithmic. In this case the

first equation in Proposition 4.3 implies  $w = 1$  and so taxes become

$$\tau_c = \frac{\sigma^2 E(l)}{\rho - \mu(l) + \sigma^2 E(l)} \quad \tau_a = \sigma^2 E(l) \left( E(l) - \frac{\rho}{\rho - \mu(l) + \sigma^2 E(l)} \right).$$

To determine the total taxes paid, note that the law of motion of wealth in Definition 4.3 becomes

$$da_t = [(1 - \tau_c)\rho(1 - \bar{l}) - \tau_a - \bar{c} + \bar{l}(\rho - \mu(l)) + (1 - \tau_c)\mu(l)]a_t dt + \sigma(1 - \tau_c)\bar{l}a_t dZ_t. \quad (31)$$

In this case the investment function reduces to  $\bar{l} = \tau_c(\rho - \mu(l))[\sigma^2(1 - \tau_c)^2\bar{\gamma}]^{-1}$ . Finally, using

$$\frac{\tau_c}{1 - \tau_c} = \frac{\sigma^2 E(l)}{\rho - \mu(l) + \sigma^2 E(l) - \sigma^2 E(l)} = \frac{\sigma^2 E(l)}{\rho - \mu(l)}$$

we see that the expected taxes paid on interest, wealth, and capital gains are

$$\begin{aligned} \rho\tau_c(1 - \bar{l}) + \tau_a + \tau_c\mu(l)\bar{l} &= \rho\tau_c + \tau_a + \tau_c(\mu(l) - \rho)\bar{l} = \rho\tau_c + \tau_a - \frac{\tau_c^2}{(1 - \tau_c)^2} \frac{(\rho - \mu(l))^2}{\sigma^2} \\ &= \sigma^2 E(l)^2 - \sigma^2 E(l)^2 = 0. \end{aligned}$$

It follows that the dividends tax is the only tax that raises revenue for the government, while the taxes on capital income and wealth are only necessary because of the presence of the agency problem. Intuitively, the capital income tax is chosen so that *effort* is set at the constrained efficient level, while the wealth tax is present to ensure that the drift in *consumption* is set at the constrained efficient level.

## 5 Conclusion

This paper develops theory and numerical techniques to determine the constrained efficient levels of inequality in an economy with moral hazard and stochastic growth in human capital. There were two principal findings from the analysis. First, the presence of endogenous human capital accumulation has important implications for the efficient bearing of risk and hence the degree of inequality in the implied stationary distribution. In dynamic agency models with fixed productivity and a (standard) wealth effect in preferences, it becomes too expensive to motivate agents with high realizations of output and consumption, and so they are eventually retired. In contrast, in the model in this paper agents become richer *precisely* because they experienced rapid productivity growth, and so the benefits of further effort rise along with the costs. In Proposition 2.6 I provide sufficient conditions for this second force to overwhelm the first, and in Proposition 3.8, I characterize the upper tail of the stationary distribution.

Second, I provided two decentralizations (one complete and one partial with simple instruments) of the efficient allocations that differ in the degree of risk-sharing present in financial markets. In the first, agents may write long term contracts with financial intermediaries restricted only by informational frictions. In this case Proposition 4.1 shows that the government may achieve its redistributive goals simply with a linear tax on firm profits. In the second, I suppose that agents may only trade shares or save in a bond, and characterize the optimal *linear* taxes of the government. In this case Proposition 4.3 uncovers a novel role for taxes in this environment: a tax on (personal) capital income alters the incentives of agents to retain ownership of their firm, and hence to exert continued *effort* to improve productivity. The optimal linear taxation policy therefore calls for three *distinct* kinds of capital income taxes — corporate profits, personal, and wealth — serving three *distinct* purposes. The tax on corporate profits serves redistributive purposes, the tax on personal savings ensures that effort is set at the efficient level, while the wealth tax ensures that the drift in consumption is at the efficient level.

## References

- Ai, Hengjie, Dana Kiku, and Rui Li (2015). “A mechanism design model of firm dynamics: The case of limited commitment.” URL <http://www.eief.it/files/2015/08/ai.pdf>.
- Ai, Hengjie, Dana Kiku, and Rui Li (2016). “Quantifying the impact of moral hazard: Evidence from a structural estimation.” *SSRN Electronic Journal*. doi:10.2139/ssrn.2801116.
- Albanesi, Stefania (2006). “Optimal taxation of entrepreneurial capital with private information.” Working Paper 12419, National Bureau of Economic Research. doi:10.3386/w12419.
- Ales, Laurence, Antonio Andrés Bellofatto, and Jessie Jiaxu Wang (2017). “Taxing atlas: Executive compensation, firm size, and their impact on optimal top income tax rates.” *Review of Economic Dynamics*, 26, pp. 62–90. doi:10.1016/j.red.2017.02.007.
- Ales, Laurence, Musab Kurnaz, and Christopher Sleet (2015). “Technical change, wage inequality, and taxes.” *American Economic Review*, 105(10), pp. 3061–3101. doi:10.1257/aer.20140466.
- Ales, Laurence and Christopher Sleet (2016). “Taxing top CEO incomes.” *American Economic Review*, 106(11), pp. 3331–3366. doi:10.1257/aer.20151093.

- Atkeson, Andrew and Robert E. Lucas (1992). “On efficient distribution with private information.” *The Review of Economic Studies*, 59(3), pp. 427–453. doi:[10.2307/2297858](https://doi.org/10.2307/2297858).
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez (2011). “Top incomes in the long run of history.” *Journal of Economic Literature*, 49(1), pp. 3–71. doi:[10.1257/jel.49.1.3](https://doi.org/10.1257/jel.49.1.3).
- Atkinson, Anthony Barnes and Joseph E. Stiglitz (1976). “The design of tax structure: Direct versus indirect taxation.” *Journal of Public Economics*, 6(1-2), pp. 55–75. doi:[10.1016/0047-2727\(76\)90041-4](https://doi.org/10.1016/0047-2727(76)90041-4).
- Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011). “The distribution of wealth and fiscal policy in economies with finitely lived agents.” *Econometrica*, 79(1), pp. 123–157. doi:[10.3982/ecta8416](https://doi.org/10.3982/ecta8416).
- Best, Michael Carlos and Henrik Jacobsen Kleven (2012). “Optimal income taxation with career effects of work effort.” doi:[10.2139/ssrn.2051214](https://doi.org/10.2139/ssrn.2051214).
- Chamley, Christophe (1986). “Optimal taxation of capital income in general equilibrium with infinite lives.” *Econometrica*, 54(3), pp. 607–622. doi:[10.2307/1911310](https://doi.org/10.2307/1911310).
- Cooper, Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick (2016). “Business in the United States: Who owns it, and how much tax do they pay?” *Tax Policy and the Economy*, 30(1), pp. 91–128. doi:[10.1086/685594](https://doi.org/10.1086/685594).
- Cvitanić, Jakša, Xuhu Wan, and Jianfeng Zhang (2008). “Optimal compensation with hidden action and lump-sum payment in a continuous-time model.” *Applied Mathematics and Optimization*, 59(1), pp. 99–146. doi:[10.1007/s00245-008-9050-0](https://doi.org/10.1007/s00245-008-9050-0).
- Diamond, Peter and Emmanuel Saez (2011). “The case for a progressive tax: From basic research to policy recommendation.” *Journal of Economic Perspectives*, 25(4), pp. 165–190. doi:[10.1257/jep.25.4.165](https://doi.org/10.1257/jep.25.4.165).
- Farhi, Emmanuel and Iván Werning (2007). “Inequality and social discounting.” *Journal of Political Economy*, 115(3), pp. 365–402. doi:[10.1086/518741](https://doi.org/10.1086/518741).
- Farhi, Emmanuel and Iván Werning (2013). “Insurance and taxation over the life cycle.” *The Review of Economic Studies*, 80(2), pp. 596–635. doi:[10.1093/restud/rds048](https://doi.org/10.1093/restud/rds048).

- Fleming, Wendell H. and Halil Mete Soner (2006). *Controlled Markov Processes and Viscosity Solutions*, volume 25. Springer-Verlag. doi:[10.1007/0-387-31071-1](https://doi.org/10.1007/0-387-31071-1).
- Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016). “Redistribution and social insurance.” *American Economic Review*, 106(2), pp. 359–386. doi:[10.1257/aer.20111550](https://doi.org/10.1257/aer.20111550).
- Golosov, Mikhail and Aleh Tsyvinski (2007). “Optimal taxation with endogenous insurance markets.” *The Quarterly Journal of Economics*, 122(2), pp. 487–534. doi:[10.1162/qjec.122.2.487](https://doi.org/10.1162/qjec.122.2.487).
- Guvenen, Fatih and Greg Kaplan (2017). “Top income inequality in the 21st century: Some cautionary notes.” Working Paper 23321, National Bureau of Economic Research. doi:[10.3386/w23321](https://doi.org/10.3386/w23321).
- Guvenen, Fatih, Greg Kaplan, and Jae Song (2014). “The glass ceiling and the paper floor: Gender differences among top earners, 1981-2012.” Working Paper 20560, National Bureau of Economic Research. doi:[10.3386/w20560](https://doi.org/10.3386/w20560).
- He, Zhiguo (2008). “Optimal executive compensation when firm size follows geometric brownian motion.” *Review of Financial Studies*, 22(2), pp. 859–892. doi:[10.1093/rfs/hhn037](https://doi.org/10.1093/rfs/hhn037).
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante (2010). “Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006.” *Review of Economic Dynamics*, 13(1), pp. 15–51. doi:[10.1016/j.red.2009.10.010](https://doi.org/10.1016/j.red.2009.10.010).
- Jones, Charles and Jihee Kim (2014). “A Schumpeterian model of top income inequality.” Working Paper 20637, National Bureau of Economic Research. doi:[10.3386/w20637](https://doi.org/10.3386/w20637).
- Judd, Kenneth L. (1985). “Redistributive taxation in a simple perfect foresight model.” *Journal of Public Economics*, 28(1), pp. 59–83. doi:[10.1016/0047-2727\(85\)90020-9](https://doi.org/10.1016/0047-2727(85)90020-9).
- Kapička, Marek and Julian Neira (2013). “Optimal taxation in a life-cycle economy with endogenous human capital formation.” Working paper, University of California, Santa Barbara. URL <http://www.fas.nus.edu.sg/ecs/events/seminar/seminar-papers/05Nov13.pdf>.
- Kushner, Harold J. and Paul G. Dupuis (1992). *Numerical Methods for Stochastic Control Problems in Continuous Time*. Number 24 in Stochastic Modelling and Applied Probability. Springer US. doi:[10.1007/978-1-4684-0441-8](https://doi.org/10.1007/978-1-4684-0441-8).
- Lucas, Robert E. (1978). “On the size distribution of business firms.” *The Bell Journal of Economics*, 9(2), pp. 508–523. doi:[10.2307/3003596](https://doi.org/10.2307/3003596).

- Makris, Miltiadis and Alessandro Pavan (2019). “Taxation under learning-by-doing.” Working Paper. URL <http://faculty.wcas.northwestern.edu/~apa522/LBD.pdf>.
- Mirrlees, James A. (1971). “An exploration in the theory of optimum income taxation.” *The Review of Economic Studies*, 38(2), pp. 175–208. doi:[10.2307/2296779](https://doi.org/10.2307/2296779).
- Phelan, Christopher and Robert M. Townsend (1991). “Computing multi-period, information-constrained optima.” *The Review of Economic Studies*, 58(5), pp. 853–881. doi:[10.2307/2297941](https://doi.org/10.2307/2297941).
- Phelan, Thomas (2019). “On the optimality of differential asset taxation.” Working Paper 19-17, Federal Reserve Bank of Cleveland. doi:[10.26509/frbc-wp-201917](https://doi.org/10.26509/frbc-wp-201917).
- Piketty, Thomas and Emmanuel Saez (2003). “Income inequality in the United States, 1913-1998.” *The Quarterly Journal of Economics*, 118(1), pp. 1–41. doi:[10.1162/00335530360535135](https://doi.org/10.1162/00335530360535135).
- Prescott, Edward C. and Robert M. Townsend (1984). “Pareto optima and competitive equilibria with adverse selection and moral hazard.” *Econometrica*, 52(1), pp. 21–45. doi:[10.2307/1911459](https://doi.org/10.2307/1911459).
- Rosen, Sherwin (1982). “Authority, control, and the distribution of earnings.” *The Bell Journal of Economics*, 13(2), pp. 311–323. doi:[10.2307/3003456](https://doi.org/10.2307/3003456).
- Rothschild, Casey and Florian Scheuer (2013). “Redistributive taxation in the Roy model.” *The Quarterly Journal of Economics*, 128(2), pp. 623–668. doi:[10.1093/qje/qjs076](https://doi.org/10.1093/qje/qjs076).
- Sannikov, Yykiy (2008). “A continuous-time version of the principal–agent problem.” *Review of Economic Studies*, 75(3), pp. 957–984. doi:[10.1111/j.1467-937x.2008.00486.x](https://doi.org/10.1111/j.1467-937x.2008.00486.x).
- Scheuer, Florian (2014). “Entrepreneurial taxation with endogenous entry.” *American Economic Journal: Economic Policy*, 6(2), pp. 126–163. doi:[10.1257/pol.6.2.126](https://doi.org/10.1257/pol.6.2.126).
- Scheuer, Florian and Iván Werning (2016). “The taxation of superstars.” *The Quarterly Journal of Economics*, 132(1), pp. 211–270. doi:[10.1093/qje/qjw036](https://doi.org/10.1093/qje/qjw036).
- Shourideh, Ali (2013). “Optimal taxation of wealthy individuals.” 2013 Meeting Papers 261, Society for Economic Dynamics. URL <https://ideas.repec.org/p/red/sed013/261.html>.
- Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick (2019). “Capitalists in the twenty-first century.” Working Paper 25442, National Bureau of Economic Research. doi:[10.3386/w25442](https://doi.org/10.3386/w25442).

Song, Jae, David Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter (2014). “Firming up inequality.” Working Paper 21199, National Bureau of Economic Research. doi:[10.3386/w21199](https://doi.org/10.3386/w21199).

Stantcheva, Stefanie (2015). “Optimal taxation and human capital policies over the life cycle.” Working Paper 21207, National Bureau of Economic Research. doi:[10.3386/w21207](https://doi.org/10.3386/w21207).

## A Recursive analysis

In this section I will elaborate on the characterization of incentive compatibility given in the main text. I will then derive the Hamilton-Jacobi-Bellman equations for both the true value function and restricted value functions, show that the latter have closed-form solutions, and derive the Kolmogorov forward equation for the summary measure. Throughout this section I will take the underlying filtered probability space to be  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, P)$ , where  $\Omega = C([0, \infty))$ ,  $P$  is Wiener measure and  $\mathcal{F}$  is the  $\sigma$ -algebra generated by the evaluation maps.

### A.1 Incentive compatibility

**Definition A.1** (Shifted process). For each effort strategy  $(l_t)_{t \geq 0}$  (equivalently, choice of deterministic functions  $(\tilde{l}_t)_{t \geq 0}$  defined on  $C([0, t])$ ) the process  $Z^l := (Z_t^l)_{t \geq 0}$  is defined pointwise

$$Z_t^l(\omega) := \omega(t) - \sigma^{-1} \int_0^t \mu(\tilde{l}_s((\omega(s'))_{0 \leq s' \leq s})) ds$$

for each  $\omega \in \Omega := C([0, \infty))$ . Note that  $dZ_t^l := dZ_t - \sigma^{-1} \mu(\tilde{l}_t) dt$ .

In the main text I noted that a change in an agent's strategy may be viewed as a change in the measure she uses to evaluate the probability of the occurrence of certain paths of output. This is made formal in the following.

**Definition A.2** (Effort process and induced measure). An effort process for the agent is a progressively measurable process  $e = (l_t)_{t \geq 0}$  on  $(\Omega, \mathcal{F}, P)$ . This is equivalent to the existence of functions  $(\tilde{l}_t)_{t \geq 0}$  such that  $\tilde{l}_t : C([0, t]) \rightarrow \mathbb{R}$  for each  $t \geq 0$  and  $l_t = \tilde{l}_t((Z_s)_{0 \leq s \leq t})$  almost surely, for all  $t \geq 0$ , where the  $(\tilde{l}_t)_{t \geq 0}$  are measurable with respect to the filtration on  $\Omega$  generated by the evaluation maps. The measure on  $\Omega := C([0, t])$  implied by the process  $e$  is then defined by declaring

$$P^l(\omega \in \Omega \mid \omega(t_i) \in B_i, i = 1, \dots, N) = P\left(\omega \in \Omega \mid \int_0^{t_i} \tilde{l}_s((\omega(s'))_{0 \leq s' \leq s}) ds + \sigma \omega(t_i) \in B_i, i = 1, \dots, N\right)$$

for any points  $t_1, \dots, t_N \in [0, t]$  and Borel measurable sets  $B_1, \dots, B_N \subseteq \mathbb{R}$ .

Definition A.2 appears complicated but simply asserts that  $P^l$  is the probability measure used by the agent to evaluate the probability of occurrences of different output levels. Note that by the Girsanov theorem,  $Z^l$  is a Brownian motion on  $(\Omega, \mathcal{F}, P^l)$ .

**Definition A.3** (Time zero utility). For any allocation  $(c, l^P)$  chosen by the planner and agent strategy  $e$ , define a stochastic process  $V^{c, l} := (V_t^{c, l})_{t \geq 0}$  by

$$V_t := \rho \int_0^\infty \mathbb{E}^l[u(c_t, l_t) \mid \mathcal{F}_t] dt. \quad (32)$$

One may interpret  $V_t^{c, l}$  as the time zero utility based on time  $t$  information.

The law of iterated expectations ensures that for any allocation  $(c, l^P)$  and strategy  $e$  the process  $V^{c,l}$  is a martingale on  $(\Omega, \mathcal{F}, P^l)$ . Now, for the given underlying Brownian motion  $Z := (Z_t)_{t \geq 0}$  define  $\theta$  to be a strong solution to the SDE  $d\theta_t = \sigma\theta_t dZ_t$  and note that if  $(l_t)_{t \geq 0}$  is of the form  $l_t = \tilde{l}_t((Z_s)_{0 \leq s \leq t})$  almost surely then we have

$$d\theta_t = \sigma\theta_t dZ_t = \sigma\theta_t \left( \sigma^{-1} \mu(\tilde{l}(\theta.)) dt + dZ_t - \sigma^{-1} \mu(\tilde{l}(\theta.)) dt \right) = \mu(\tilde{l}(\theta.)) \theta_t dt + \sigma\theta_t dZ_t^l$$

where the process  $Z^l$  is defined in Definition A.1. It follows that  $(\theta, Z^l, P^l)$  is *always* a *weak* solution to

$$dX_t = \mu(\tilde{l}(X.)) X_t dt + \sigma X_t dW_t \quad (33)$$

and that choosing a strategy  $\tilde{l}$  amounts to choosing the distribution of the process given in (33). The Brownian motion  $(W_t)_{t \geq 0}$  is the noise process associated with the strategy  $e$ .

*Proof of Proposition 2.1.* The basic idea of the proof is that for *any* strategy  $e := (l_t)_{t \geq 0}$ , if we define a random variable  $X$  by

$$X := \rho \int_0^\infty e^{-\rho t} \frac{(c_t^{1-\alpha} l_t^\alpha)^{1-\gamma}}{1-\gamma} dt \quad (34)$$

then the process  $V^l := (V_t^l)_{t \geq 0}$  defined in (32) is a martingale on  $(\Omega, \mathcal{F}, P^l)$  by the law of iterated expectations. The strengthening of the martingale representation theorem given in Lemma 3.1 of Cvitanić et al. (2008) then asserts existence of  $Y = Y^{e,P,C}$  such that  $V_t^e = \rho \sigma \int_0^t Y_s dZ_s^l$  for all  $t \geq 0$  almost surely, which gives the result upon rearrangement.  $\square$

**Remark A.1.** The strengthening of the martingale representation proved in Cvitanić et al. (2008) and invoked above is necessary because  $\mathcal{F}$  is the filtration generated by the evaluation maps  $\omega$  and so is not necessarily equal to the natural filtration associated with  $Z^l$ , which is the assumption of the usual martingale representation theorem.

*Proof of Proposition 2.2.* For any strategy  $l$  and  $t \geq 0$ , define a random variable  $\hat{V}_t^l := \rho \int_0^t e^{-\rho s} u(c_s, l_s) ds + e^{-\rho t} W_t$  where  $W_t$  represents the continuation utility if the agent adheres to  $l^P$  after  $t$ . From the definition of  $dZ^l$  given in Proposition 2.1 we have  $dZ_t^{l^P} = dZ_t^l - \sigma^{-1} [\mu(l_t^P) - \mu(l_t)]$  and so by the expression (7),

$$\begin{aligned} d\hat{V}_t^l &= \rho e^{-\rho t} u(c_t, l_t) dt + d(e^{-\rho t} W_t) \\ &= \rho u(c_t, l_t) dt - \rho W_t dt + \left[ \rho (W_t - u(c_t, l_t^P)) dt + S_t dZ_t^{l^P} \right] \\ &= e^{-\rho t} (\rho [u(c_t, l_t) - u(c_t, l_t^P)] + \sigma^{-1} S_t [\mu(l_t) - \mu(l_t^P)]) dt + e^{-\rho t} S_t dZ_t^l. \end{aligned}$$

Using  $\mathbb{E}^l \left[ \int_0^t e^{-\rho s} S_s dZ_s^l \right] = 0$  for all  $t \geq 0$  and effort process  $e$ , it follows that

$$\mathbb{E}^l \left[ \hat{V}_t^l \right] = \hat{V}_0^l + \rho \mathbb{E}^l \left[ \int_0^t e^{-\rho s} (\sigma^{-1} S_s \mu(l_s) + \rho u(c_s, l_s) - [\sigma^{-1} S_s \mu(l_s^P) + \rho u(c_s, l_s^P)]) ds \right]. \quad (35)$$

Since the expected utility of the agent is exactly  $\mathbb{E}^l \left[ \lim_{t \rightarrow \infty} \hat{V}_t^l \right]$ , an effort process  $e$  is incentive compatible if and only if it maximizes the integral in (35) almost surely for all  $t \geq 0$ , which gives the result.  $\square$

## A.2 Hamilton-Jacobi-Bellman equation

Recall some notation from the main text. The utility function is given by  $u(c, l) = (c^{1-\alpha}l^\alpha)^{1-\gamma}/(1-\gamma)$  for some  $\alpha \in (0, 1)$  and  $\gamma > 0$ , and I will define  $\bar{\gamma}$  by  $(1-\gamma)(1-\alpha) = 1-\bar{\gamma}$  and assume  $\bar{\gamma}, \gamma > 1$ . Note that

$$\begin{aligned} u(c, l) &= \frac{(c^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\gamma} & u_2(c, l) &= \alpha l^{-1}(c^{1-\alpha}l^\alpha)^{1-\gamma} \\ \mu(l) &= \bar{\mu} - [\bar{\mu} - \underline{\mu}]l & \mu'(l) &= -[\bar{\mu} - \underline{\mu}] \end{aligned}$$

so the elasticity of utility with respect to output when  $l < 1$  is  $S_t = \sigma E(l_t)(1-\alpha)(c^{1-\alpha}l^\alpha)^{1-\gamma}$ , where

$$E(l) := \frac{\rho\alpha}{(1-\alpha)(\bar{\mu} - \underline{\mu})l}.$$

The joint law of promised utility and productivity is then

$$\begin{aligned} d\theta_t &= (\bar{\mu} - [\bar{\mu} - \underline{\mu}]l)\theta_t dt + \sigma\theta_t dZ_t \\ dU_t &= \rho\left(1 - (\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}\right)U_t dt + \sigma E(l)(1-\bar{\gamma})(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}U_t dZ_t \end{aligned} \quad (36)$$

where I have changed variables to  $\bar{c} = c[(1-\gamma)U_t]^{1-\bar{\gamma}}$ . Given the joint law of motion (36), the associated Hamilton-Jacobi-Bellman equation is

$$\begin{aligned} \rho V(U, \theta) &= \max_{\substack{C \geq 0 \\ l \geq 0}} \theta - C + \rho\left(1 - (\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}\right)U \frac{\partial V}{\partial U} + \frac{\sigma^2 E(l)^2}{2}(1-\bar{\gamma})^2(\bar{c}^{1-\alpha}l^\alpha)^{2-2\gamma}U_t^2 \frac{\partial^2 V}{\partial U^2} \\ &+ (\bar{\mu} - [\bar{\mu} - \underline{\mu}]l)\theta \frac{\partial V}{\partial \theta} + \frac{\sigma^2 \theta^2}{2} \frac{\partial^2 V}{\partial \theta^2} + E(l)(1-\bar{\gamma})(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma} \sigma^2 U \theta \frac{\partial^2 V}{\partial U \partial \theta}. \end{aligned} \quad (37)$$

Using the change of variables in Definition 2.4.

**Proposition A.1.** *The solution to the Hamilton-Jacobi-Bellman equation is of the form  $V(U, \theta) = v(u)\theta$  for some function  $v$  solving*

$$\begin{aligned} \rho v(u) &= \max_{\substack{\bar{c} \geq 0 \\ l \geq 0}} 1 - \bar{c}u + \rho\left(\frac{1 - (\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\bar{\gamma}}\right)uv'(u) + \frac{\sigma^2 E(l)^2}{2}(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}\bar{\gamma}uv'(u) \\ &+ (\bar{\mu} - [\bar{\mu} - \underline{\mu}]l)[v(u) - v'(u)u] + \frac{\sigma^2}{2}\left(1 - E(l)(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}\right)^2 u^2 v''(u). \end{aligned}$$

*Proof.* Under the change-of-variables  $u := [(1-\gamma)U]^{1-\bar{\gamma}}\theta^{-1}$ , the relevant algebra is

$$\begin{aligned} u &= [(1-\gamma)U]^{1-\bar{\gamma}}\theta^{-1} & U &= \frac{(u\theta)^{1-\bar{\gamma}}}{1-\gamma} \\ \frac{\partial u}{\partial U} &= \frac{1}{1-\alpha}[(1-\gamma)U]^{1-\bar{\gamma}}\theta^{-1} = \frac{u^{\bar{\gamma}}\theta^{\bar{\gamma}-1}}{1-\alpha} & \frac{\partial u}{\partial \theta} &= -\frac{u}{\theta}. \end{aligned}$$

Again defining  $v(u)\theta = V(U, \theta)$  gives

$$\begin{aligned} \frac{\partial V}{\partial U} &= \frac{\partial u}{\partial U}v'(u)\theta = \frac{\theta^{\bar{\gamma}}u^{\bar{\gamma}}}{1-\alpha}v'(u) & \frac{\partial V}{\partial \theta} &= v(u) - v'(u)u \end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 V}{\partial \theta^2} &= \frac{\partial}{\partial \theta} [v(u) - v'(u)u] = \frac{\partial u}{\partial \theta} [v'(u) - v'(u) - v''(u)u] = \frac{v''(u)u^2}{\theta} \\ \frac{\partial^2 V}{\partial U^2} &= \frac{\theta^{\bar{\gamma}}}{1-\alpha} \frac{\partial}{\partial U} [u^{\bar{\gamma}} v'(u)] = \frac{\theta^{\bar{\gamma}}}{1-\alpha} \frac{\partial u}{\partial U} [\bar{\gamma} u^{\bar{\gamma}-1} v'(u) + u^{\bar{\gamma}} v''(u)] = \frac{(u\theta)^{2\bar{\gamma}-2}}{(1-\alpha)^2} [\bar{\gamma} u v'(u) + u^2 v''(u)] \theta \\ \frac{\partial^2 V}{\partial U \partial \theta} &= \frac{\partial}{\partial U} [v(u) - v'(u)u] = \frac{\partial u}{\partial U} [-v''(u)u] = -\frac{v''(u)u^{1+\bar{\gamma}}}{1-\alpha} \theta^{\bar{\gamma}-1}.\end{aligned}$$

Since  $c(u) = \bar{c}(u)u$  (and  $C(U, \theta) = c(u)\theta$ ) for all  $u \geq 0$  we have

$$U - \frac{(C^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} = (1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}) U \quad (38)$$

for the drift, and

$$\frac{\rho \sigma \alpha}{[\bar{\mu} - \underline{\mu}] l} (C^{1-\alpha} l^\alpha)^{1-\gamma} = \frac{\rho \sigma \alpha}{[\bar{\mu} - \underline{\mu}] l} (1-\gamma) (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} U$$

for the diffusion. Substitution then gives

$$\begin{aligned}\rho v(u)\theta &= \max_{\substack{\bar{c} \geq 0 \\ l \geq 0}} (1 - \bar{c}u)\theta + \rho \left(1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}\right) U \frac{\theta^{\bar{\gamma}} u^{\bar{\gamma}}}{1-\alpha} v'(u) \\ &\quad + \frac{(\rho \sigma \alpha)^2}{[\bar{\mu} - \underline{\mu}]^2 l^2} (1-\gamma)^2 (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} U^2 \frac{(u\theta)^{2\bar{\gamma}-2}}{(1-\alpha)^2} [\bar{\gamma} u v'(u) + u^2 v''(u)] \theta \\ &\quad + (\bar{\mu} - [\bar{\mu} - \underline{\mu}] l) \theta [v(u) - v'(u)u] + \frac{\sigma^2 \theta}{2} v''(u) u^2 - \frac{\rho \alpha (1-\gamma)}{[\bar{\mu} - \underline{\mu}] l} (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} \sigma^2 U \theta \frac{v''(u) u^{1+\bar{\gamma}}}{1-\alpha} \theta^{\bar{\gamma}-1} \quad (39) \\ \rho v(u) &= \max_{\substack{\bar{c} \geq 0 \\ l \geq 0}} (1 - \bar{c}u) + \rho \left( \frac{1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\bar{\gamma}} \right) u v'(u) + \frac{(\rho \sigma \alpha)^2 (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma}}{(1-\alpha)^2 [\bar{\mu} - \underline{\mu}]^2 l^2} [\bar{\gamma} u v'(u) + u^2 v''(u)] \\ &\quad + (\bar{\mu} - [\bar{\mu} - \underline{\mu}] l) [v(u) - v'(u)u] + \frac{\sigma^2}{2} v''(u) u^2 - \frac{\rho \alpha}{(1-\alpha) [\bar{\mu} - \underline{\mu}] l} (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} \sigma^2 u^2 v''(u)\end{aligned}$$

as claimed upon factorization.  $\square$

### A.3 Restricted value and policy functions

Proceeding from Proposition A.1 gives the following, which is a slightly more general statement than Proposition 2.4.

**Proposition A.2.** *The restricted-action value and policy functions are given by*

$$v(u; l) = \frac{1}{\rho - \mu(l)} - \frac{(2\bar{\gamma} - 1)}{\rho(2\bar{\gamma} - x)} c(u; l) \quad c(u; l) = x^{\frac{1}{1-\bar{\gamma}}} l^{-\frac{\alpha}{1-\bar{\alpha}}} u$$

where  $x = x(l)$  solves the quadratic  $0 = E(l)^2(\bar{\gamma} - 1)(\bar{\gamma} - 1/2)\sigma^2 x^2 + \rho x - \rho$ . Further,  $\bar{v}(l)$  is decreasing in  $l$ .

*Proof.* Using Proposition A.1 and substituting the form  $v \equiv \bar{v}(l)u + \underline{v}(l)$  gives

$$\begin{aligned}\rho \bar{v}(l)u + \rho \underline{v}(l) &= \max_{\bar{c} \geq 0} 1 - \bar{c}u + \rho \left( \frac{1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\bar{\gamma}} \right) \bar{v}(l)u + \frac{\bar{\gamma} \sigma^2}{2} E(l)^2 (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \bar{v}(l)u \\ &\quad + (\bar{\mu} - [\bar{\mu} - \underline{\mu}] l) \underline{v}(l).\end{aligned}$$

Equating linear and constant parts gives  $v(l) = 1/(\rho - \bar{\mu} + [\bar{\mu} - \mu]l)$  and  $\bar{v}(l)$  solves

$$\rho\bar{v}(l) = \max_{\bar{c} \geq 0} -\bar{c} + \rho \left( \frac{1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} \right) \bar{v}(l) + \frac{\sigma^2 E(l)^2}{2} (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \bar{v}(l).$$

Combining this with the associated first-order condition gives

$$\begin{aligned} \rho\bar{v}(l) &= -\bar{c} + \rho \left( \frac{1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} \right) \bar{v}(l) + \frac{\sigma^2 D^2}{2(1-\alpha)^2 l^2} (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \bar{v}(l) \\ 0 &= -\bar{c} - \rho (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} \bar{v}(l) + (2-2\gamma) \frac{\sigma^2 E(l)^2}{2} (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \bar{v}(l). \end{aligned}$$

Combining, eliminating  $\bar{c}$  and dividing by  $\bar{v}(l)$  gives the following quadratic in  $x := (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}$

$$-\rho x + (2-2\gamma) \frac{\sigma^2 E(l)^2}{2} \bar{\gamma} x^2 = -\rho + \rho \left( \frac{1-x}{1-\bar{\gamma}} \right) + \frac{\sigma^2 E(l)^2}{2} \bar{\gamma} x^2.$$

Rearranging gives

$$(1-2\bar{\gamma}) \frac{\sigma^2 E(l)^2}{2} \bar{\gamma} x^2 = \frac{\rho \bar{\gamma}}{1-\bar{\gamma}} - \frac{\rho \bar{\gamma} x}{1-\bar{\gamma}}$$

multiplying by  $1-\bar{\gamma}$  simplifies to the claimed quadratic. The implied expression for consumption is then  $\bar{c} = x^{\frac{1}{1-\bar{\gamma}}} l^{-\frac{\alpha}{1-\bar{\gamma}}}$  and the coefficient of normalized utility solves (using the quadratic for  $x$  again)

$$\begin{aligned} \rho\bar{v}(l) &= -\bar{c} + \rho \left( \frac{1-x}{1-\bar{\gamma}} \right) \bar{v}(l) + \frac{\sigma^2 E(l)^2 x^2}{2} \bar{\gamma} \bar{v}(l) \\ \bar{c}/\bar{v}(l) &= \rho \left( \frac{1-x}{1-\bar{\gamma}} \right) - \rho + \frac{\bar{\gamma}}{2} \frac{\rho(1-x)}{(\bar{\gamma}-1)(\bar{\gamma}-1/2)} \\ \bar{c}/\bar{v}(l) &= -\frac{\rho(1-x)}{2\bar{\gamma}-1} - \rho \end{aligned}$$

which rearranges to the desired expression. To show that  $\bar{v}(l)$  is increasing in  $l$ , note that it is the fixed point of the equation  $\rho\bar{v} = T(\bar{v})$  where

$$T(\bar{v}) = \max_{\bar{c} \geq 0} -\bar{c} + \rho \left( \frac{(\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} - 1}{\bar{\gamma} - 1} \right) \bar{v} + \frac{\sigma^2 E(l)^2}{2} (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \bar{v}.$$

The right-hand side is clearly increasing in  $l$  whenever  $\bar{v}$  is negative since it is the pointwise maxima of functions increasing in  $l$ . The fixed point will then be increasing in  $l$  provided the right-hand side is convex in  $\bar{v}$ . This is indeed the case, since it is the pointwise maximum of linear functions.  $\square$

## A.4 Laws of motion

The proof of Proposition A.1 shows that normalized utility follows a diffusion process of the form  $du_t = \mu_u(u; \bar{c}, l)u_t dt + \sigma_u(u_t; \bar{c}, l)u_t dZ_t$

$$\begin{aligned} \mu_u(u; \bar{c}, l) &= \rho \left( \frac{1 - (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma}}{1-\bar{\gamma}} \right) + \frac{\sigma^2 E(l(u))^2}{2} (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma} \bar{\gamma} - \mu(l(u)) \\ \sigma_u(u; \bar{c}, l_t) &= \sigma \left( E(l(u)) (\bar{c}(u)^{1-\alpha} l(u)^\alpha)^{1-\gamma} - 1 \right) \end{aligned}$$

which is exactly the first claim of Proposition 2.6. Now note that if  $dU_t = \mu_U U_t dt + \sigma_U U_t dZ_t$  then

$$U_t = U_0 \exp\left([\mu_U - \sigma_U^2/2]t + \sigma_U dZ_t\right).$$

It follows that  $[(1 - \bar{\gamma})U_t]^{\frac{1}{1-\bar{\gamma}}} = [(1 - \bar{\gamma})U_0]^{\frac{1}{1-\bar{\gamma}}} \exp\left((1 - \bar{\gamma})^{-1}[\mu_U - \sigma_U^2/2]t + \frac{\sigma_U}{(1-\bar{\gamma})}dZ_t\right)$  and so

$$d[(1 - \bar{\gamma})U_t]^{\frac{1}{1-\bar{\gamma}}} = \left(\frac{\mu_U}{1 - \bar{\gamma}} + \frac{\bar{\gamma}\sigma_U^2/2}{(1 - \bar{\gamma})^2}\right)[(1 - \bar{\gamma})U_t]^{\frac{1}{1-\bar{\gamma}}} dt + \frac{\sigma_U}{1 - \bar{\gamma}}[(1 - \bar{\gamma})U_t]^{\frac{1}{1-\bar{\gamma}}} dZ_t.$$

Therefore, The law of motion for  $z_t := [(1 - \bar{\gamma})U_t]^{\frac{1}{1-\bar{\gamma}}}$  is  $dz_t = \mu_z z_t dt + \sigma_z z_t dZ_t$ , where

$$\begin{aligned} \mu_z &= \frac{\rho}{1 - \bar{\gamma}} \left(1 - (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}\right) + \frac{\bar{\gamma}}{2} \left(\frac{\rho\alpha\sigma}{(1 - \alpha)[\bar{\mu} - \underline{\mu}]l}\right)^2 (\bar{c}^{1-\alpha} l^\alpha)^{2-2\gamma} \\ \sigma_z &= \frac{\rho\alpha\sigma}{(1 - \alpha)[\bar{\mu} - \underline{\mu}]l} (\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma} \end{aligned}$$

which gives the second claim of Proposition 2.6. The volatility of normalized utility is then  $(E(l)x(l) - 1)\sigma$ , where  $x = x(l) = (\bar{c}(l)^{1-\alpha} l^\alpha)^{1-\gamma}$  solves the quadratic given in Proposition A.2. High shocks will imply a fall in normalized promised utility if and only if  $x(l) < 1/E(l)$ . This will be true if and only if

$$\rho < \frac{\sigma^2}{2}(2\bar{\gamma} - 1)(\bar{\gamma} - 1) + (1/\alpha - 1)[\bar{\mu} - \underline{\mu}]l$$

which is the final claim of Proposition 2.6.

## A.5 Stationary distribution proofs

In general the Kolmogorov forward equation will inform us of the evolution over time of the two-dimensional density of both promised utility and productivity. However, as shown in Ai ?, the homogeneity of the policy functions and the log-linear law of motion of productivity ensure that we need only solve for the density of a single variable, referred to as the *summary measure*.

*Proof of Lemma 3.4.* The process  $(\theta_t, u_t)_{t \geq 0}$  is a multi-dimensional diffusion process driven by the same Brownian motion, and so away from the initial point  $u_0$ , the joint density  $\Phi$  satisfies the Kolmogorov forward equation

$$\begin{aligned} \frac{\partial}{\partial t}[\Phi(\theta, u, t)] &= -\rho_D \Phi(\theta, u, t) - \mu_u(u) \frac{\partial}{\partial \theta}[\theta \Phi(\theta, u, t)] - \frac{\partial}{\partial u}[\mu_u(u) \Phi(\theta, u, t)] \\ &+ \frac{\sigma_\theta(u)^2}{2} \frac{\partial^2}{\partial \theta^2}[\theta^2 \Phi(\theta, u, t)] + \frac{\partial^2}{\partial \theta \partial u}[\theta \sigma_\theta(u) \sigma_u(u) \Phi(\theta, u, t)] + \frac{1}{2} \frac{\partial^2}{\partial u^2}[\sigma_u^2(u) \Phi(\theta, u, t)]. \end{aligned}$$

Note that for any smooth function  $f$  vanishing at zero with sufficiently rapid decay at  $\infty$ , integration by parts gives  $\int_0^\infty \theta f'(\theta) d\theta = [\theta f(\theta)]_{\theta=0}^\infty - \int_0^\infty f(\theta) d\theta = -\int_0^\infty f(\theta) d\theta$  and  $\int_0^\infty \theta f''(\theta) d\theta = -\int_0^\infty f'(\theta) d\theta = 0$ . Recalling the definition  $m(u, t) := \int_0^\infty \theta \Phi(\theta, u, t) d\theta$  and interchanging orders of integration, it follows that

for all  $(u, t)$  we have the following simplifications

$$\begin{aligned}
\mu(u) \int_0^\infty \theta \frac{\partial}{\partial \theta} [\theta \Phi(\theta, u, t)] d\theta &= -\mu(u)m(u, t) \\
- \int_0^\infty \frac{\partial}{\partial u} [\mu_u(u)\theta \Phi(\theta, u, t)] d\theta &= -\frac{\partial}{\partial u} [\mu_u(u)m(u, t)] \\
\frac{\sigma_\theta^2(u)}{2} \int_0^\infty \theta \frac{\partial^2}{\partial \theta^2} [\theta^2 \Phi(\theta, u, t)] d\theta &= 0 \\
\int_0^\infty \theta \frac{\partial^2}{\partial \theta \partial u} [\theta \sigma_u(u)\Phi(\theta, u, t)] d\theta &= -\frac{\partial}{\partial u} [\sigma_u(u)m(u, t)] \\
\int_0^\infty \theta \frac{\partial^2}{\partial u^2} [\sigma_u^2(u)\Phi(\theta, u, t)] d\theta &= \frac{\partial^2}{\partial u^2} [\sigma_u^2(u)m(u, t)].
\end{aligned}$$

Interchanging the order of integration, the multi-dimensional forward equation implies

$$\begin{aligned}
\frac{\partial}{\partial t} \int_0^\infty \theta \Phi(\theta, u, t) d\theta &= -\rho_D \int_0^\infty \theta \Phi(\theta, u, t) d\theta - \mu_u(u) \int_0^\infty \theta \frac{\partial}{\partial \theta} [\theta \Phi(\theta, u, t)] d\theta \\
&- \frac{\partial}{\partial u} \int_0^\infty \theta \mu_u(u)\Phi(\theta, u, t) d\theta + \frac{\sigma_\theta(u)^2}{2} \int_0^\infty \theta \frac{\partial^2}{\partial \theta^2} [\theta^2 \Phi(\theta, u, t)] d\theta \\
&+ \sigma_\theta(u) \int_0^\infty \theta \frac{\partial^2}{\partial \theta \partial u} [\theta \sigma_u(u)\Phi(\theta, u, t)] d\theta + \frac{1}{2} \int_0^\infty \theta \frac{\partial^2}{\partial u^2} [\sigma_u(u)^2 \Phi(\theta, u, t)] d\theta
\end{aligned}$$

which is equivalent to

$$\frac{\partial m}{\partial t} = -(\rho_D - \mu_u(u))m(u, t) - \frac{\partial}{\partial u} [(\mu_u(u) + \sigma_\theta(u)\sigma_u(u))m(u, t)] + \frac{1}{2} \frac{\partial^2}{\partial u^2} [\sigma_u^2(u)m(u, t)].$$

Setting the partial derivative with respect to time to zero then gives (15).  $\square$

The boundary conditions for the ODE in (15) are not obvious *a priori*, but we do not need to know them to find the stationary distribution. I will employ the finite-state Markov chain method of [Kushner and Dupuis \(1992\)](#) and find the stationary distribution of a finite-state Markov chain that is ‘close’ in distribution with the underlying continuous-time process.

For convenience I record the following lemma that shows that killed diffusion processes are distributed according to a double-Pareto distribution. The proof is standard and therefore omitted. I then specialize to the restricted-action versions of consumption and productivity.

**Lemma A.3.** *The stationary distribution of a stochastic process evolving according to  $dX_t = \mu_X X_t dt + \sigma_X X_t dZ_t$  that dies at rate  $\delta_X$  and is injected at some point  $\bar{X} > 0$  is given by*

$$f(x) = \begin{cases} Ax^{\alpha_+ - 1} & \text{if } x \leq \bar{X} \\ Bx^{\alpha_- - 1} & \text{if } x \geq \bar{X} \end{cases} \quad (40)$$

where

$$\alpha_\pm = \bar{\mu}_X / \sigma_X^2 \pm \sqrt{(\bar{\mu}_X / \sigma_X)^2 + 2\delta_X / \sigma_X^2}, \quad (41)$$

the constants  $A$  and  $B$  are chosen such that the density is continuous and integrates to unity, and I have written  $\bar{\mu}_X = \mu_X - \sigma_X^2/2$  for brevity.

Proposition 3.8 shows that by combining Lemma A.3 with the explicit policy functions for consumption given in Lemma 2.4 we get closed-form representations for the stationary distributions associated with the restricted-action allocations.

*Proof of Proposition 3.8.* From Lemma 2.6 the law of consumption is  $dc_t = \mu_c c_t dt + \sigma_c c_t dZ_t$ , where

$$\mu_c = \frac{\rho(1-x)}{1-\bar{\gamma}} + \frac{\bar{\gamma}E^2\sigma^2x^2}{2} \quad \sigma_c = \sigma Ex$$

where  $(\bar{\gamma}-1)(\bar{\gamma}-1/2)E^2\sigma^2x^2 = \rho(1-x)$ . It follows that

$$\mu_c - \frac{\sigma_c^2}{2} = \frac{\rho(1-x)}{1-\bar{\gamma}} + \frac{(\bar{\gamma}-1)E^2\sigma^2x^2}{2} = \frac{\rho(1-x)}{1-\bar{\gamma}} + \frac{\rho(1-x)}{2\bar{\gamma}-1} = \frac{\rho(1-x)\bar{\gamma}}{(1-\bar{\gamma})(2\bar{\gamma}-1)} = -\frac{E^2\sigma^2x^2\bar{\gamma}}{2}.$$

Using Lemma A.3 the tails are then

$$\alpha_{\pm} = \frac{\bar{\mu}_X}{\sigma_X^2} \pm \sqrt{\left(\frac{\bar{\mu}_X}{\sigma_X^2}\right)^2 + \frac{2\delta_X}{\sigma_X^2}} = -\frac{\bar{\gamma}}{2} \pm \sqrt{\frac{\bar{\gamma}^2}{4} + \frac{2\rho_D}{\sigma^2 E^2 x^2}}$$

which simplifies as claimed using the defining quadratic for  $x$ .  $\square$

Proposition 3.8 implies the following uniform bounds on consumption inequality since  $x \rightarrow 0$  as  $l \rightarrow 0$ .

**Corollary A.4** (Consumption bounds). *For all  $\gamma \geq 1$  and  $l \in (0, 1)$  we have*

$$\alpha_-(l) \leq -\frac{\bar{\gamma}}{2} \left( 1 + \sqrt{1 + \frac{4\rho_D}{\rho} (2 - 1/\bar{\gamma})(1 - 1/\bar{\gamma})} \right)$$

with equality in the limit as  $l \rightarrow 0$ .

## A.6 Planner preferences

Defining  $\Omega := V \times \Theta$ , aggregate consumption and output at any date  $t \geq 0$  are then,

$$\begin{aligned} C_t &:= e^{-\rho_D t} \underline{C}_t + \int_0^t e^{-\rho_D [t-T]} C_t^T dT \\ \underline{C}_t &:= \int_{\Omega} \mathbb{E}[c_t^{v,\theta}] \Phi(d\omega), \quad C_t^T := \int_{\Psi} \mathbb{E}[c_t^{T,\psi}] G(d\psi) \\ Y_t &:= e^{-\rho_D t} \underline{Y}_t + \int_0^t e^{-\rho_D [t-T]} Y_t^T dT \\ \underline{Y}_t &:= \int_{\Omega} \mathbb{E}[F(\theta_t^{v,\theta}, L_t^{v,\theta})] \Phi(d\omega), \quad Y_t^T := \int_{\Psi} \mathbb{E}[F(\theta_t^{T,\psi}, L_t^{T,\psi})] G(d\psi) \end{aligned}$$

while aggregate labor assigned to entrepreneurs and labor supplied by workers is

$$\begin{aligned} L_t^E &:= e^{-\rho_D t} \underline{L}_t^l + \int_0^t e^{-\rho_D [t-T]} L_t^{E,T} dT \\ \underline{L}_t^E &:= \int_{\Omega} \mathbb{E}[L_t^{v,\theta}] \Phi(d\omega), \quad L_t^{E,T} := \int_{\Psi} e^{-\rho_D [t-T]} \mathbb{E}[L_t^{T,\psi}] G(d\psi) \\ L_t^W &:= e^{-\rho_D t} \underline{L}_t^W + \int_0^t e^{-\rho_D [t-T]} L_t^{W,T} dT \\ \underline{L}_t^W &:= \int_{\Omega} \mathbb{E}[e_{W,t}^{v,\theta}] \Phi(d\omega), \quad L_t^{W,T} := \int_{\Psi} e^{-\rho_D [t-T]} \mathbb{E}[e_{W,t}^{T,\psi}] G(d\psi). \end{aligned}$$

I will also use the notation

$$\underline{U}_t = \int_{\Omega} \mathbb{E} \left[ u \left( c_t^{v,\theta}, e_t^{v,\theta} \right) \right] \Phi(d\omega) \quad U_t^T = \int_{\Psi} \mathbb{E} \left[ u \left( c_t^{T,\psi}, e_t^{T,\psi} \right) \right] G(d\psi)$$

for the total flow utility experienced by the first and subsequent generations. Note that each of the above aggregate quantities is written as the sum of two terms: the first is the contribution to the aggregate level of the initial generation, and the second is the contribution from all subsequent generations. For instance, since agents die at rate  $\rho_D$  and death is independent of the noise process, the aggregate consumption of surviving members of the initial generation at date  $t$  is  $e^{-\rho_D t}$  multiplied by the expected consumption of a single member of this generation.

**Lemma A.5.** *The preferences of the planner over allocations  $A$  are represented by the function*

$$U^P(A) = \int_0^{\infty} \left( e^{-\rho t} \underline{U}_t + \int_0^t e^{-\rho[t-T]} e^{-\rho_S T} U_t^T dT \right) dt.$$

*Proof.* From first principles, the planner's preferences may be represented

$$\begin{aligned} U^P &= \int_{\Omega} \int_0^{\infty} e^{-\rho t} \mathbb{E} [u(c_t^{v,\theta}, e_t^{v,\theta})] dt \Phi(d\omega) + \int_0^{\infty} e^{-\rho T} \int_{\Psi} \int_T^{\infty} e^{-\rho[t-T]} \mathbb{E} [u(c_t^{T,\psi}, e_t^{T,\psi})] dt G(d\psi) dT \\ &= \int_0^{\infty} e^{-\rho_S t} \left( \int_{\Omega} e^{-\rho_D t} \mathbb{E} [u(c_t^{v,\theta}, e_t^{v,\theta})] dt \Phi(d\omega) + \int_{\Psi} \int_0^t e^{-\rho_D[t-T]} \mathbb{E} [u(c_t^{T,\psi}, e_t^{T,\psi})] dT G(d\psi) \right) dt \end{aligned}$$

where I interchanged the order of integration and used the equalities

$$e^{-\rho T} e^{-(\rho+\rho_D)[t-T]} = e^{-\rho T} e^{-\rho[t-T]} e^{-\rho_D[t-T]} = e^{-\rho t} e^{-\rho_D[t-T]},$$

which gives the result upon simplification. □

## B Implementation

### B.1 Discrete-time analogues

In this section I will outline a discrete-time version of implementation with stock market equilibrium for clarity, spelling out the order of moves by the agent step-by-step. The agent problem in the main text ought to be viewed as a continuous-time version of the following.

- At  $t$  the agent has wealth  $a_t$  and chooses consumption  $\Delta c_t$ .
- She places  $b_t$  units in a risk-free bond earning after-tax return  $(1 - \tau_s)r$  and uses the remaining wealth  $\iota_t := a_t - b_t$  to purchase shares at price  $p_t$ . If  $x_t$  is the fraction of the business owned, then her budget constraint is  $a_t = b_t + p_t x_t$ .
- Between  $t$  and  $t + \Delta$  the firm yields output  $\Delta \theta_t$ . By  $t + \Delta$ , it has grown to  $\theta_{t+\Delta}$  and the price is  $p_{t+\Delta}$ .

- Wealth at  $t + \Delta$  comprises holdings of bonds and stocks, plus flow dividends minus consumption,

$$\begin{aligned}
a_{t+\Delta} &= \text{savings} + \text{interest} - \text{consumption} + \text{dividends} + \text{stocks} + \text{capital gains} \\
&= b_t + (1 - \tau_s)r\Delta b_t - \Delta c_t + \Delta(1 - \tau_d)\theta_t x_t + p_t x_t + (1 - \tau_{cg})(p_{t+\Delta} - p_t)x_t \\
&= a_t + \Delta[(1 - \tau_s)ra_t - c_t + [(1 - \tau_d)\theta_t/p_t - (1 - \tau_s)r]l_t] + (p_{t+\Delta}/p_t - 1)(1 - \tau_{cg})l_t
\end{aligned} \tag{42}$$

where I substituted  $x_t = l_t/p_t$ .

Now note that the approximate law of motion for the price is  $p_{t+\Delta}/p_t = 1 + \mu(\hat{l}(u))\Delta + \sigma\sqrt{\Delta}dX$  where  $dX$  is mean zero i.i.d. across time and assumes values  $\pm 1$ . Substituting into (42) gives

$$a_{t+\Delta} = a_t + \Delta \left[ (1 - \tau_s)ra_t - c_t + [(1 - \tau_d)\theta_t/p_t + (1 - \tau_{cg})\mu(\hat{l}(u)) - (1 - \tau_s)r]l_t \right] + \sigma\sqrt{\Delta}(1 - \tau_{cg})l_t dX.$$

Taking limits as  $\Delta \rightarrow 0$  gives the law of motion in the case of linear prices

$$da_t = [(1 - \tau_s)ra_t - c_t + l_t((1 - \tau_d)\theta_t/p_t + (1 - \tau_{cg})\mu(l_t) - (1 - \tau_s)r)]dt + \sigma(1 - \tau_{cg})l_t dZ_t. \tag{43}$$

Furthermore, if  $\hat{l}(u) \equiv l$  then by the Gordon growth formula

$$p_t = \frac{(1 - \tau_d)\theta_t}{r - \mu(\hat{l})} \tag{44}$$

and so the law of motion of wealth simplifies to

$$da_t = \left[ (1 - \tau_s)ra_t - c_t + \left( \tau_s r + (1 - \tau_{cg})\mu(l) - \mu(\hat{l}) \right) l_t \right] dt + \sigma(1 - \tau_{cg})l_t dZ_t. \tag{45}$$

## B.2 Agent problems and equilibrium characterization

This section contains proofs of claims pertaining to the implementation of the efficient allocations. Recall that Lemma 4.2 solves the problem of the individual agent and Proposition 4.3 characterizes the optimal linear taxes.

*Proof of Lemma 4.2.* When facing constant taxes and expectations of constant effort  $\hat{l}$ , the value function of the agent solves the Hamilton-Jacobi-Bellman equation

$$\begin{aligned}
\rho V(a) &= \max_{\bar{c}, l, \iota} \rho \frac{((\bar{c}a)^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} + [-\tau_a - \bar{c} + (1 - \tau_s)r]aV'(a) \\
&+ \left[ (1 - \tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r \right] \iota aV'(a) + \frac{\sigma^2 a^2}{2} \iota^2 (1 - \tau_{cg})^2 V''(a).
\end{aligned} \tag{46}$$

Assume a solution to this equation of the form  $V(a) = \bar{V}(\hat{l})a^{1-\bar{\gamma}}/(1-\gamma)$ , so that

$$aV'(a) = (1 - \alpha)\bar{V}(\hat{l})a^{1-\bar{\gamma}} \quad a^2V''(a) = -\bar{\gamma}(1 - \alpha)\bar{V}(\hat{l})a^{1-\bar{\gamma}}.$$

Substitution gives

$$\begin{aligned}
\frac{\rho\bar{V}(\hat{l})}{1-\gamma} &= \max_{\bar{c}, l, \iota} \frac{\rho(\bar{c}^{1-\alpha} l^\alpha)^{1-\gamma}}{1-\gamma} + [-\tau_a - \bar{c} + (1 - \tau_s)r](1 - \alpha)\bar{V}(\hat{l}) \\
&+ \left( \left[ (1 - \tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r \right] \iota - \frac{\sigma^2(1 - \tau_{cg})^2}{2} \iota^2 \bar{\gamma} \right) (1 - \alpha)\bar{V}(\hat{l}).
\end{aligned}$$

First-order conditions for consumption, leisure, and investment are then

$$\begin{aligned}(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma} &= \frac{\bar{c}}{\rho}\bar{V}(\hat{l}) \\(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma} &= \frac{l}{\rho\alpha}[\bar{\mu} - \underline{\mu}]l(1-\alpha)(1-\tau_{cg})\bar{V}(\hat{l}) = \frac{1}{E(l)}(1-\tau_{cg})\bar{V}(\hat{l})\iota \\ \iota &= \frac{(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r}{\sigma^2(1-\tau_{cg})^2\bar{\gamma}}.\end{aligned}$$

Substituting the third into the second gives the pair

$$\begin{aligned}\frac{(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\bar{V}(\hat{l})} &= \frac{\bar{c}}{\rho} \\ \frac{(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\bar{V}(\hat{l})} &= \frac{(1-\tau_{cg})}{E(l)} \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]}{\sigma^2(1-\tau_{cg})^2\bar{\gamma}},\end{aligned}$$

so that

$$\frac{\bar{c}}{\rho} = \frac{1}{E(l)} \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]}{\sigma^2(1-\tau_{cg})\bar{\gamma}}$$

We then use the HJB to obtain  $\bar{V}(\hat{l})$ ,

$$\begin{aligned}\frac{\rho\bar{V}(\hat{l})}{1-\gamma} &= \frac{\rho(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{1-\gamma} + [(1-\tau_s)r - \tau_a - \bar{c}](1-\alpha)\bar{V}(\hat{l}) + \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]^2}{2\sigma^2(1-\tau_{cg})^2\bar{\gamma}}(1-\alpha)\bar{V}(\hat{l}) \\ 1 &= \frac{(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\bar{V}(\hat{l})} + \left( -\tau_a - \bar{c} + (1-\tau_s)r + \frac{1}{2} \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]^2}{\sigma^2(1-\tau_{cg})^2\bar{\gamma}} \right) \frac{(1-\bar{\gamma})}{\rho}.\end{aligned}$$

The solution to the consumer problem reduces to the following three equations in three unknowns (coefficient of value function, and consumption and effort choices)

$$\begin{aligned}\frac{(\bar{c}^{1-\alpha}l^\alpha)^{1-\gamma}}{\bar{V}(\hat{l})} &= \frac{\bar{c}}{\rho} \\ \frac{\bar{c}}{\rho} &= \frac{1}{E(l)} \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]}{\sigma^2(1-\tau_{cg})\bar{\gamma}} \\ \rho &= \bar{c} + \left( -\tau_a - \bar{c} + (1-\tau_s)r + \frac{1}{2} \frac{[(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r]^2}{\sigma^2(1-\tau_{cg})^2\bar{\gamma}} \right) (1-\bar{\gamma}).\end{aligned}\tag{47}$$

The second rearranges to

$$\frac{\bar{c}}{\rho}\sigma^2(1-\tau_{cg})\bar{\gamma}E(l) = [(1-\tau_{cg})\mu(l) - \mu(\hat{l}) + \tau_s r].$$

The first equation in (47) serves only to determine  $\bar{V}(\hat{l})$ , while the second and third become the system in the statement of the lemma.  $\square$

*Proof of Proposition 4.3.* If we impose the equilibrium requirement  $\hat{l} = l$  the pair of equations in Lemma 4.2 becomes

$$\begin{aligned}1 &= \frac{\tau_s r - \tau_{cg}\mu(l)}{\sigma^2(1-\tau_{cg})E(l)} + \left( -\tau_a + (1-\tau_s)r + \frac{[\tau_s r - \tau_{cg}\mu(l)]^2}{2\sigma^2(1-\tau_{cg})^2\bar{\gamma}} \right) \frac{(1-\bar{\gamma})}{\rho} \\ \frac{\bar{\gamma}\bar{c}}{\rho} &= \frac{\tau_s r - \tau_{cg}\mu(l)}{\sigma^2(1-\tau_{cg})E(l)}.\end{aligned}\tag{48}$$

Recall that the drift and diffusion of consumption growth in the restricted-action allocations are given by

$$\mu_c = \sigma^2 E(l)^2 x(l)^2 \frac{(1 - \bar{\gamma})}{2} \quad \sigma_c = \sigma E(l)x(l).$$

We want the drift and diffusion in the above decentralized environment to match these expressions. The drift and volatility in the competitive equilibrium are given above by

$$\begin{aligned} \mu_a &= -\tau_a - \bar{c} + (1 - \tau_s)\rho + [\tau_s\rho - \tau_{cg}\mu(l)]\iota = -\tau_a - \bar{c} + (1 - \tau_s)\rho + \frac{[\tau_s\rho - \tau_{cg}\mu(l)]^2}{\sigma^2(1 - \tau_{cg})^2\bar{\gamma}} \\ \sigma_a &= \sigma\iota(1 - \tau_{cg}) = \frac{\tau_s\rho - \tau_{cg}\mu(l)}{\sigma(1 - \tau_{cg})\bar{\gamma}}. \end{aligned} \quad (49)$$

So we have five equations (characterizations of  $c$  and  $l$  in terms of taxes, drift and diffusion requirements for efficient allocations, and the definition of  $x(l)$ ) in six unknowns  $(c, l, \tau_s, \tau_{cg}, \tau_a, x)$ ,

$$\begin{aligned} 1 &= \frac{\tau_s r - \tau_{cg}\mu(l)}{\sigma^2(1 - \tau_{cg})E(l)} + \left( -\tau_a + (1 - \tau_s)r + \frac{1}{2} \frac{[\tau_s r - \tau_{cg}\mu(l)]^2}{\sigma^2(1 - \tau_{cg})^2\bar{\gamma}} \right) \frac{(1 - \bar{\gamma})}{\rho} \\ \frac{\bar{\gamma}\bar{c}}{\rho} &= \frac{\tau_s r - \tau_{cg}\mu(l)}{\sigma^2(1 - \tau_{cg})E(l)}. \\ \frac{\sigma^2}{2} E(l)^2 x^2 (1 - \bar{\gamma}) &= -\tau_a - \bar{c} + (1 - \tau_s)\rho + \frac{[\tau_s\rho - \tau_{cg}\mu(l)]^2}{\sigma^2(1 - \tau_{cg})^2\bar{\gamma}} \\ E(l)x &= \frac{\tau_s\rho - \tau_{cg}\mu(l)}{\sigma^2(1 - \tau_{cg})\bar{\gamma}} \\ E(l)^2 x^2 \sigma^2 (\bar{\gamma} - 1)(\bar{\gamma} - 1/2) &= \rho(1 - x). \end{aligned}$$

Now impose the requirement  $\tau_{cg} = \tau_s =: \tau$ . Then we have five equations in five unknowns:

$$\begin{aligned} 1 &= \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)} + \left( -\tau_a + (1 - \tau)\rho + \frac{1}{2} \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \right) \frac{(1 - \bar{\gamma})}{\rho} \\ \frac{\bar{\gamma}\bar{c}}{\rho} &= \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)}. \\ \sigma^2 E(l)^2 x^2 (1 - \bar{\gamma})/2 &= -\tau_a - \bar{c} + (1 - \tau)\rho + \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \\ E(l)x &= \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)\bar{\gamma}} \\ E(l)^2 x^2 \sigma^2 (\bar{\gamma} - 1)(\bar{\gamma} - 1/2) &= \rho(1 - x). \end{aligned} \quad (50)$$

Comparing the second with the fourth implies

$$\frac{\bar{\gamma}\bar{c}}{\rho} = \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)} = \bar{\gamma}x$$

and so  $x = \bar{c}/\rho$ . Replace  $x$  everywhere with this and eliminate the fourth equation.

$$\begin{aligned} 1 &= \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)} + \left( -\tau_a + (1 - \tau)\rho + \frac{1}{2} \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \right) \frac{(1 - \bar{\gamma})}{\rho} \\ \frac{\bar{\gamma}\bar{c}}{\rho} &= \frac{\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)}. \\ \sigma^2 E(l)^2 (\bar{c}/\rho)^2 (1 - \bar{\gamma})/2 &= -\tau_a - \bar{c} + (1 - \tau)\rho + \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \\ E(l)^2 \bar{c}^2 (\sigma/\rho)^2 (\bar{\gamma} - 1)(\bar{\gamma} - 1/2) &= \rho - \bar{c}. \end{aligned} \quad (51)$$

The system (51) then simplifies to

$$\begin{aligned}
\frac{\rho - \bar{\gamma}\bar{c}}{1 - \bar{\gamma}} &= -\tau_a + (1 - \tau)\rho + \frac{1}{2} \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \\
\bar{\gamma}\bar{c} &= \frac{\rho\tau(\rho - \mu(l))}{\sigma^2(1 - \tau)E(l)} \\
\sigma^2 E(l)^2 (\bar{c}/\rho)^2 \frac{(1 - \bar{\gamma})}{2} &= -\tau_a - \bar{c} + (1 - \tau)\rho + \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} \\
E(l)^2 \bar{c}^2 (\sigma/\rho)^2 \frac{(1 - \bar{\gamma})}{2} &= \frac{\bar{c} - \rho}{2\bar{\gamma} - 1}.
\end{aligned} \tag{52}$$

The third and fourth combine to give

$$\frac{2\bar{\gamma}\bar{c}}{2\bar{\gamma} - 1} - \frac{\rho}{2\bar{\gamma} - 1} - \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}} = -\tau_a + (1 - \tau)\rho.$$

The system is now

$$\begin{aligned}
\frac{\rho - \bar{\gamma}\bar{c}}{1 - \bar{\gamma}} + \frac{1}{2\bar{\gamma}} \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2} &= \frac{2\bar{\gamma}\bar{c}}{2\bar{\gamma} - 1} - \frac{\rho}{2\bar{\gamma} - 1} \\
\bar{\gamma}\bar{c}\sigma E(l)/\rho &= \frac{\tau(\rho - \mu(l))}{\sigma(1 - \tau)} \\
E(l)^2 \bar{c}^2 (\sigma/\rho)^2 \frac{(1 - \bar{\gamma})}{2} &= \frac{\bar{c} - \rho}{2\bar{\gamma} - 1}.
\end{aligned} \tag{53}$$

(The other equation gives  $\tau_a$ :  $\sigma^2 E(l)^2 (\bar{c}/\rho)^2 \frac{(1 - \bar{\gamma})}{2} = -\tau_a - \bar{c} + (1 - \tau)\rho + \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2\bar{\gamma}}$ ). One may show that the first two equations in (53) imply the third, and so we are left with the pair of equations

$$\begin{aligned}
\left(\frac{2\bar{\gamma} - 1}{1 - \bar{\gamma}}\right)(\rho - \bar{\gamma}\bar{c}) + \frac{(2\bar{\gamma} - 1)}{2\bar{\gamma}} \frac{\tau^2(\rho - \mu(l))^2}{\sigma^2(1 - \tau)^2} &= 2\bar{\gamma}\bar{c} - \rho \\
\bar{\gamma}\bar{c}\sigma E(l)/\rho &= \frac{\tau(\rho - \mu(l))}{\sigma(1 - \tau)}.
\end{aligned} \tag{54}$$

We can eliminate  $\bar{c}$  using the last equation, and finally end up with a single equation for  $\tau$  in terms of  $l$ . The first equation in (54) rearranges to

$$\begin{aligned}
(\rho - \bar{\gamma}\bar{c})(2\bar{\gamma} - 1)\bar{\gamma} + (2\bar{\gamma} - 1)(1 - \bar{\gamma}) \frac{\tau^2(\rho - \mu(l))^2}{2\sigma^2(1 - \tau)^2} &= \bar{\gamma}(2\bar{\gamma}\bar{c} - \rho)(1 - \bar{\gamma}) \\
\frac{1}{\rho}(\bar{\gamma} - 1/2)(1 - \bar{\gamma}) \frac{\tau^2(\rho - \mu(l))^2}{(1 - \tau)^2} &= \bar{\gamma}^2 \sigma^2 (\bar{c}/\rho - 1).
\end{aligned}$$

Substitution then gives the single equation for  $\tau$  in terms of  $l$ :

$$\frac{1}{\rho}(\bar{\gamma} - 1/2)(1 - \bar{\gamma}) \frac{\tau^2(\rho - \mu(l))^2}{(1 - \tau)^2} + \bar{\gamma}^2 \sigma^2 = \bar{\gamma} \frac{\tau(\rho - \mu(l))}{E(l)(1 - \tau)}.$$

This leads to a quadratic in  $y := \tau/(1 - \tau)$  (so that  $\tau = y/(1 + y)$  and  $1 - \tau = 1/(1 + y)$ ):

$$0 = \frac{1}{\rho}(\bar{\gamma} - 1/2)(1 - \bar{\gamma})(\rho - \mu(l))^2 y^2 - \bar{\gamma} \frac{(\rho - \mu(l))}{E(l)} y + \bar{\gamma}^2 \sigma^2 \tag{55}$$

which clearly has a unique positive solution for  $y$  (and hence  $\tau$ ). The expressions in the statement of the Proposition then follow upon simplification.  $\square$

## C Numerical schemes

This section outlines how I numerically solve the ordinary differential equations derived in the main text. Both the value function of the principal and the stationary distribution will be found using the method of [Kushner and Dupuis \(1992\)](#).

It is well known that the distribution of a variable evolving according to a diffusion process is given by the Kolmogorov forward (or Fokker-Planck) equation. I refer the reader to [Fleming and Soner \(2006\)](#) for details. As noted in the main text, I will find the summary measure of normalized promised utility by using the finite-state Markov chain method described in detail in [Kushner and Dupuis \(1992\)](#). If  $X = (X_t)_{t \geq 0}$  follows a diffusion process of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dZ_t \quad (56)$$

then  $X$  may be approximated arbitrarily well by a Markov chain  $X^\Delta$  indexed by a pair  $\Delta := (\Delta_T, \Delta_X)$  assuming values in the finite set  $\Sigma := \{\underline{x}, \underline{x} + \Delta_X, \dots, \bar{x} - \Delta_X, \bar{x}\}$  and subject to the transition probabilities

$$\begin{aligned} p(x, x + \Delta_X) &= \frac{\Delta_T}{\Delta_X^2} \left( \frac{\sigma^2(X)}{2} + \Delta_X \mu(X)^+ \right) \\ p(x, x - \Delta_X) &= \frac{\Delta_T}{\Delta_X^2} \left( \frac{\sigma^2(X)}{2} + \Delta_X \mu(X)^- \right) \\ p(x, x) &= 1 - p(x, x + \Delta_X) - p(x, x - \Delta_X) \end{aligned}$$

where  $x^+ := \max\{0, x\}$  and  $x^- := \max\{0, -x\}$ . It is shown on page 325 of [Fleming and Soner \(2006\)](#) and also in [Kushner and Dupuis \(1992\)](#) that  $X^\Delta$  converges weakly to  $X$  as  $\Delta \rightarrow 0$ . Now the trick is to interpret the summary measure as the stationary distribution of a diffusion process  $(M_t)_{t \geq 0}$  with drift  $\mu_M$  and diffusion  $\sigma_M$  given by

$$\mu_M(u) = \mu_u(u) + \sigma_\theta(u)\sigma_u(u) \quad \sigma_M(u) = \sigma_u(u), \quad (57)$$

which dies at the rate  $\rho_D - \mu(u)$  and is injected at the point  $\bar{u}$  at rate  $\rho_D$ . Solving for the stationary distribution of  $X^\Delta$  simply amounts to solving a particular linear system  $Ax = b$ , where  $A$  is constructed from the probabilities specified above and  $b$  depends upon the initial guess for normalized promised utility. The operation therefore has negligible computation time.